Transcending Patchiness in the Comparative Analysis of Paleocommunities: A Test Case from the Upper Cretaceous of New Jersey

J BRET BENNINGTON
Department of Geology, 114 Hofstra University, Hempstead, NY 11549, Email: geojbb@hofstra.edu

PALAIOS, 2003, V. 18, p. 22–33

Estimates of species abundance used to quantitatively describe paleocommunities are more precise and more reliable when sampling effort is distributed among many small replicate samples, rather than concentrated in the collecting of one or a few large samples. This is because sampling error is introduced by the patchy distribution of individuals within a fossil deposit. This study applies a dispersed sampling protocol to compare the fossil assemblages preserved within a marine shell bed at two different localities of the Upper Cretaceous Navesink Formation in east-central New Jersey. A spatial hierarchy of small bulk samples (replicate samples collected along an outcrop, samples from different outcrops within a locality, and samples collected from two different localities) reveals the magnitude and scale of patchiness in the distribution of macrofauna in the Navesink shell bed. Species abundance is highly variable between replicate samples and moderately variable between outcrops due to small scale patchiness. Nevertheless, estimates of species abundance generated by collecting across the patches at each locality reveal that the overall species abundance distribution for the Navesink shell bed is nearly identical between the two localities. When collecting effort is dispersed among many widely distributed samples, different patches of fossil remains are sampled and contribute to the overall estimate of average composition obtained for a locality. Comparisons of fossil assemblages between localities or horizons are rendered more reliable by decreasing the probability that compositionally different patches have been sampled within otherwise identical paleocommunities. The significance of differences detected between local paleocommunities can be assessed more confidently when replicate samples provide a measure of local variability arising from patchiness. Paleontologists sampling to describe the species abundance composition of a discrete region of a stratigraphic horizon (e.g., outcrop, locality) should define the spatial scale of the region they are describing and disperse their sampling effort within that region.

INTRODUCTION

The previous decade has seen growing interest among paleontologists in quantifying and comparing the composition of ecological communities throughout the pre-Quaternary history of Earth. Studies analyzing the ecological context for evolutionary change and the response of species and communities to environmental perturbations increasingly are attracting the attention of paleoecologists (Bambach and Bennington, 1996; Jackson et al., 1996; Patzkowsky, 1999). The hypothesis of ecological stability punctuated by rapid change through geologic time, referred to as coordinated stasis (Brett and Baird, 1995), is just beginning to be analyzed using high-quality species abundance data, both within its original area of study (the Siluro-Devonian of New York) and elsewhere (Ivany, 1999). Examples of recent studies attempting to quantify paleocommunity stasis and change in the rock record include Bennington and Bambach (1996), Holterhoff (1996), Jackson et al., (1996), Pandolfi (1996), Patzkowsky and Holland (1997, 1999), and Olszewski and Patzkowsky (2001). Other paleocommunity studies have examined the effect of environmental gradients on paleocommunity composition (Miller, 1998) or used paleocommunity data to detect environmental gradients (Lafferty et al., 1994). Also, there is new interest in using fossil assemblages to detect subtle environmental changes not recorded by lithological changes in stratigraphic sequences (Holland et al., 2001; Weber, 2001). All of these research questions are approached by making quantitative comparisons of the fossilized remnants of communities (paleocommunities) across space and time.

Fortunately, recent reviews of taphonomic studies addressing the fidelity with which the skeletonized members of living communities are represented in the fossil record show that fossil assemblages accurately represent the skeletonized (preservable) component of the original community, both in species composition (Kidwell and Bosence, 1991) and in relative abundance (Kidwell, 2001). However, to use fossil assemblages to make paleocommunity comparisons, data must be collected that accurately describe the species composition of paleocommunities within discrete regions and time intervals. These data are derived by sampling fossil assemblages in the field. The ultimate success of these research efforts hinges on how effectively their sampling protocols generate accurate representations of the paleocommunities being sampled.

Patchiness and Paleoecological Sampling

The Patchiness Problem: To obtain unbiased samples of the species proportions in fossil assemblages, paleoecologists usually collect and count all of the individuals present within a bulk sample or bedding plane quadrat. Using either method, the species abundances in the sample are taken as an estimate of the species abundances in the larger (statistical or target) population within the entire bedding plane or fossiliferous horizon within or beyond the locality being sampled. However, a single bulk sample

0883-1351/03/0018-0022/$3.00

constitutes a biased (unreliable) estimate of the target population, with the degree of bias increasing as the target population becomes decreasingly contagious, increasingly clumped, or less homogeneous in distribution (Hayek and Buzas, 1997). Benthic organisms in modern communities have been shown to be heterogeneous or patchy in their distribution across the sea floor (e.g., Hairston, 1959; Sanders, 1960; Buzas, 1968; Levinton and Bambach, 1975; Cummins et al., 1986). Furthermore, patches form at a variety of short time scales and exist at a variety of spatial scales governed by the heterogeneity of the substrate, previous history of colonization, subtle environmental gradients, and habitat perturbations (Kidwell and Bosence, 1991). Time-averaging has the potential to homogenize patchiness as different patches of living organisms shift position and sum over time to form a fossil assemblage (Kidwell and Bosence, 1991), but whether the amount of time-averaging present in a fossil assemblage was sufficient to homogenize a patchy community is not known. Within-habitat transport of shells also can homogenize a patchy fossil assemblage by causing species with similar hydrodynamic properties to covary (Cummins et al., 1986). At least three studies have tested explicitly for and detected significant spatial heterogeneity in fossil assemblages: two in the Paleozoic (Lafferty et al., 1994; Bennington and Bambach, 1996) and one in the Cenozoic (Cobabe and Allmon, 1994). Also, Cummins et al. (1986) detected patchy (contagious) species distributions in modern death assemblages. All of these studies found significant patchiness in spite of time-averaging. If the distribution of individuals in a fossil assemblage is patchy, then any single bulk sample likely will fail to estimate the target population accurately. Although the examples of patchiness cited above are all studies of marine invertebrates, the problem of patchiness in sampling also has been discussed for microfossils (Buzas, 1968) and should apply as well to plant fossil assemblages, which may be significantly less time averaged than shelly fossil assemblages (Kowalewski, 1996).

The Replicate Sampling Solution: If, as is likely the case, most fossil assemblages are patchy in their spatial distribution, then successful sampling must attempt to incorporate information from different patches into the estimate of the target population. This requires that multiple, replicate samples be collected to quantify accurately the species abundance composition of a bedding plane or fossiliferous horizon at a locality (Hayek and Buzas, 1997). Furthermore, because patchiness can occur over unknown spatial scales, scale itself becomes a critical consideration when making paleocommunity comparisons. For example, if the paleocommunity found in a single fossiliferous horizon is to be compared between two localities, the difference in the species abundance distributions at the larger scale, between localities, must be assessed relative to the variability present at the smaller spatial scale, within each locality (Lafferty et al., 1994). Likewise, a comparison between two fossiliferous horizons deposited at different times would be meaningless unless the differences observed were judged relative to the regional variation within each fossiliferous horizon (Bennington and Bambach, 1996; Ivany, 1999). Replicate samples are the only way to generate unbiased estimates of species abundance in the target population. Replicate samples also allow variability at one scale (local, due to patchiness) to form the baseline or null expectation for variability measured at a larger scale (regional, due to habitat shift, depth gradient, etc. or temporal, between re-establishments of similar habitat).

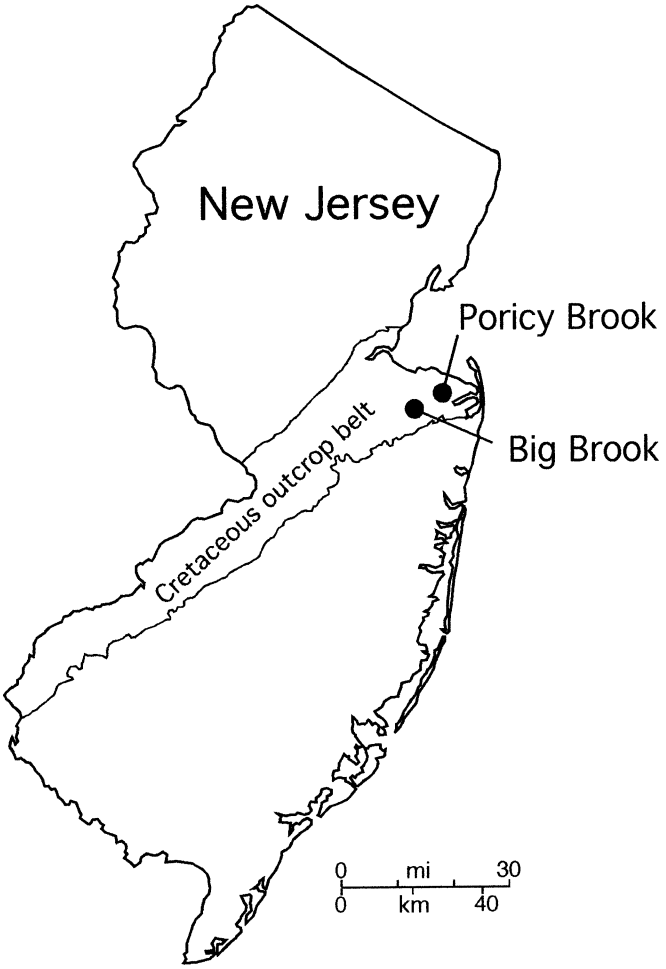

FIGURE 1—Map showing Big Brook and Poricy Brook sampling localities along the outer edge of the Cretaceous outcrop belt in eastern New Jersey.

A Test Case

To examine patchiness at different scales in a time averaged marine fossil assemblage and to demonstrate the importance of dispersed sampling for quantifying paleocommunity composition, a shell bed was sampled in the Upper Cretaceous Navesink Formation at two localities–Big Brook and Poricy Brook, Monmouth County, New Jersey (Fig. 1). The Navesink Formation in eastern New Jersey is an approximately 8-meter-thick transgressive sequence consisting of bioturbated, quartz-dominated sands in the lower part, changing upsection to bioturbated, muddy glauconite sands (Martino and Curran, 1990; Bonelli and Bennington, 2000). The upper, glauconite sand interval contains sporadic concentrations of shelly macrofauna and at least one laterally continuous shell bed less than 30 cm in thickness (Fig. 2). This shell bed hosts a fossil assemblage numerically dominated by the small ostreid oyster *Agerostrea mesenterica*, and includes the larger gry-

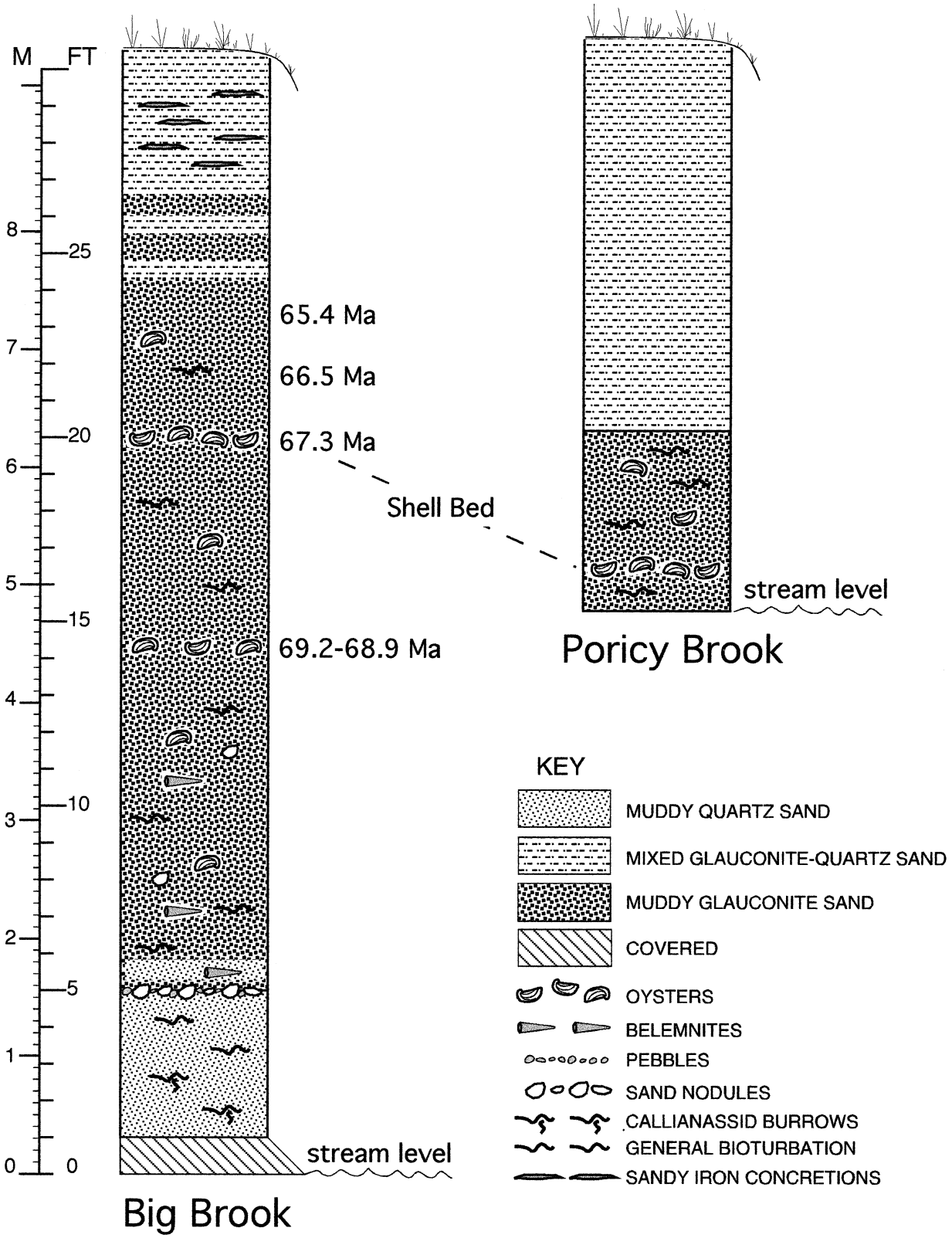

FIGURE 2—Stratigraphic diagram showing measured sections of the Navesink Formation at Big Brook and Poricy Brook. Age estimates from Sugarman et al. (1995).

phaeid oysters *Pycnodonte convexa* and *Exogyra costata*. Also present are the brachiopod *Choristothyris plicata*, and a fourth oyster species, *Gryphaeostrea vomer*. Other macrofossils (belemnite guards, and fragments of *Spondylus echinata* and small pectens) are rare. Based on the abundance of glauconite, the scarcity of quartz sand, the highly bioturbated fabric of the sediment, and the benthic foraminiferal assemblage, various authors have interpreted the upper interval of the Navesink Formation as having been deposited in a mid- to outer-shelf environment by gradual sediment accumulation (Olsson, 1963; Owens and Sohl, 1969; Martino and Curran, 1990).

The Navesink Formation at Poricy Brook is exposed incompletely and the shell bed cannot be correlated directly across the 10.6 km separating the Big Brook and Poricy Brook localities. Cut banks along the streams at both localities expose the shell beds at sporadic outcrops separated by tens of meters of cover. This pattern of exposure creates three spatial scales for sampling and two for comparison. Replicate samples can be collected at each outcrop, multiple outcrops can be sampled and compared at each locality, and the two localities can be compared using the samples collected across their respective localities. Because replicate samples are collected at each outcrop, the degree of patchiness can be measured between outcrops and the overall identity of the shell bed between localities can be assessed accurately, taking into account the patchiness inherent at each locality.

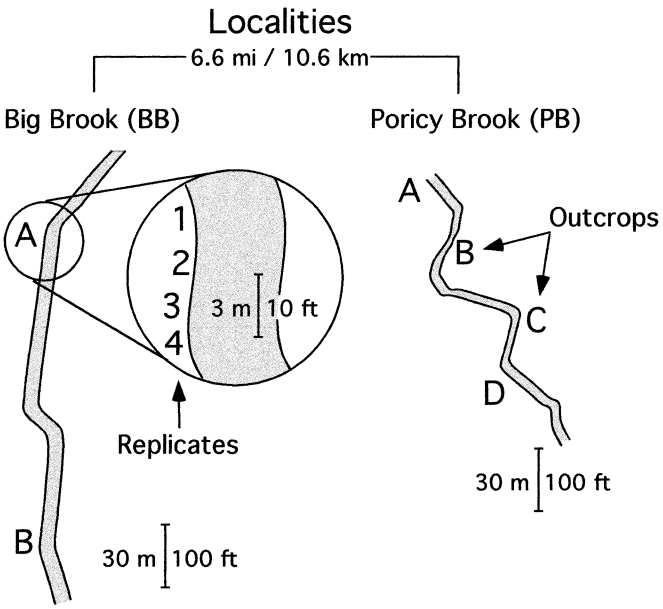

FIGURE 3—Diagram showing hierarchy of sampling and spatial relationships of samples collected in this study. Four replicate samples collected (1, 2, 3, 4) at each outcrop (A, B, C, D) at each locality (PB and BB).

METHODS AND DATA

Sampling

The Navesink shell bed was sampled at two localities in Monmouth County, New Jersey (Fig. 1), along the banks of Big Brook south of the Boundary Road Bridge (Marlboro, NJ 7.5‘ Quadrangle, UTM $^{5}66^{000}$m E, $^{44}63^{300}$m N) and along the banks of Poricy Brook at Poricy Park on the Middletown-Linecroft Road (Long Branch, NJ 7.5‘ Quadrangle, UTM $^{5}75^{000}$m E, $^{44}68^{900}$m N). Samples were dug from the midlevel of the shell bed and collected as bulk matrix and fossils. Each sample collected was approximately two liters in volume (filling a two-quart, plastic storage bag). Multiple outcrops were sampled at each locality and four replicate samples were taken from each outcrop, with replicate samples spaced from one to three meters apart (Fig. 3). Outcrops consisted of freshly exposed cut banks along meander bends in the streams. Four outcrops were sampled at Poricy Brook, whereas at Big Brook only two were sampled due to lack of good exposure. At both localities the total lateral extent of sampling was approximately similar at about 120 m. Separation between localities is approximately 10.6 km.

Replicate samples were processed in the laboratory by disaggregating the matrix by hand in water and detergent, and by washing the matrix through a 0.5 cm wire-mesh screen. This mesh size was sufficiently small to retain all of the identifiable shell fragments while allowing the sediment to pass through. Furthermore, Kidwell (2001) reports that molluscan macrobenthic surveys incorporating individuals greater than 2 mm in size produce the most reliable comparisons between the living community and the death assemblage. All shelly material retained on the screen was cleaned, dried, and sorted by species. The five most abundant species (*Agerostrea mesenterica*, *Pycnodonte convexa*, *Exogyra costata*, *Choristothyris plicata*, and *Gryphaeostrea vomer*), accounting for over 98% of the specimens recovered in each sample, were counted. Valve fragments were counted if they included more than half of the hingeline and beak, which form an unique element on each valve. Sample volume is equal for all replicate samples; however, sample size (total counts of individuals in each replicate sample) is highly variable (Table 1).

Because all of these species are bivalved organisms (four oysters and one brachiopod), the MNI method was used for counting total individuals (see Gilinsky and Bennington, 1994), which takes the larger number of disarticulated valves (left or right, brachial or pedicle) and adds it to the number of whole, articulated specimens in each sample. There are two reasons for using the MNI method is this study. First, it is not the number of unique individuals represented by the skeletal elements in the sample that is of interest in a paleoecological study, rather it is the density of individuals in the volume of the sample that is being measured to determine the relative contribution of each species to the total fossil assemblage (Hayek and Buzas, 1997). Pervasive within-assemblage mixing of paired valves could produce a sample wherein twenty valves are derived from twenty unique individuals, but to count each valve as a single individual would inflate the estimate of species density by a factor of two because twenty valves represents an average density of ten bivalved individuals. Second, it is impossible to know with any degree of certainty the size of the sampling domain (the total population of individuals from which valves were derived in the Navesink samples). In situations with little to no transport of skeletal elements within the environment of deposition, it is possible that the sample and

TABLE 1—Species-abundance data used in this study, tabulated by sample. Key to Sample ID: Localities: BB (Big Brook), PB (Poricy Brook), outcrops: A, B, C, D, replicate samples: 1, 2, 3, 4 (see Fig. 3.).

Samp. ID	*Agerostrea*	*Pycnodonte*	*Exogyra*	*Choristothyris*	*Gryphaeostrea*	Total
BBA1	35	7	4	9	1	56
BBA2	5	6	0	1	0	12
BBA3	3	7	0	0	0	10
BBA4	27	5	5	2	0	39
BBB1	46	4	1	8	0	59
BBB2	77	3	1	6	0	87
BBB3	65	1	0	11	0	77
BBB4	75	2	0	12	0	89
PBA1	46	5	0	7	1	59
PBA2	22	13	0	0	0	35
PBA3	21	5	0	1	0	27
PBA4	59	14	0	10	0	83
PBB1	42	10	1	0	0	53
PBB2	69	10	1	11	0	91
PBB3	124	6	1	10	0	141
PBB4	25	6	1	13	0	45
PBC1	57	20	1	24	0	102
PBC2	45	6	0	9	0	60
PBC3	121	4	0	6	0	131
PBC4	106	5	0	6	0	117
PBD1	15	9	0	0	0	24
PBD2	63	6	0	0	0	69
PBD3	21	9	0	0	0	30
PBD4	22	12	0	2	18	54
Total	1191	175	16	148	20	1550

the sampling domain are equivalent. When sampling is exhaustive, as in this study, it is not appropriate to count each valve as a unique individual and, therefore, the MNI approach is preferred (Gilinsky and Bennington, 1994). Given the muddy, low-energy depositional environment of the upper Navesink Formation, and the common occurrence of articulated and matching valves in samples, it is likely that the sampling domain is not very large relative to the volume of the exhaustively collected bulk samples. For this study, the MNI method is appropriately conservative and yields estimates of the density of individuals in samples that do not arbitrarily increase the power of the statistical comparisons used by inflating the number of individuals counted.

Analysis

Using the replicate samples collected at each outcrop, mean abundances for each species were calculated with 95% cluster confidence intervals for outcrops and localities. Cluster confidence intervals were calculated using the formula described in Bennington and Rutherford (1999) and explained in Hayek and Buzas (1997). Calculations were performed using the computer program SpeciesCI v.2.0 (Fortran code compiled for the Macintosh OS and available by request from the author) that calculates both binomial and cluster confidence intervals from species abundance data. Confidence intervals provide an easily calculated and intuitive way to assess the statistical significance of differences between mean species abundances. Tests of significance on linear regressions were performed at the 5% level ($a = 0.05$) using the Analysis Toolpack available as part of the Microsoft Excel 2000 software application.

RESULTS

Comparisons Between Replicate Samples

In spite of the equal volume of matrix collected for each replicate sample, total abundance varies by as much as an order of magnitude among all replicate samples and by as much as a factor of four between adjacent replicate samples (Table 1). The abundances of individual species also are highly variable from replicate sample to replicate sample (Table 1). Fidelity of species is variable, with two species (*Agerostrea* and *Pycnodonte*) present in all replicate samples, and three taxa (*Exogyra*, *Choristothyris*, and *Gryphaeostrea*) of the five counted absent from at least 25% of all replicate samples (Table 1). These differences between replicate samples are not caused by sampling error because the bulk volume of each sample was collected exhaustively. In other words, each replicate sample represents a small volume of the Navesink shell bed for which the total fossil assemblage is known. Therefore, the replicate samples directly show that there is a high degree of small-scale variation in the total density and individual abundance of macrofossil species preserved within the Navesink shell bed.

Comparisons at the Outcrop Scale

Plots by outcrop of mean relative abundance with 95% cluster confidence intervals for the four most abundant species are shown in Figures 4 and 5. *Agerostrea* varies in measured mean abundance from approximately 50% to 85%, depending on the outcrop sampled (Fig. 4). Non-overlapping confidence intervals indicate that this variability is statistically significant (not an artifact of sampling) be-

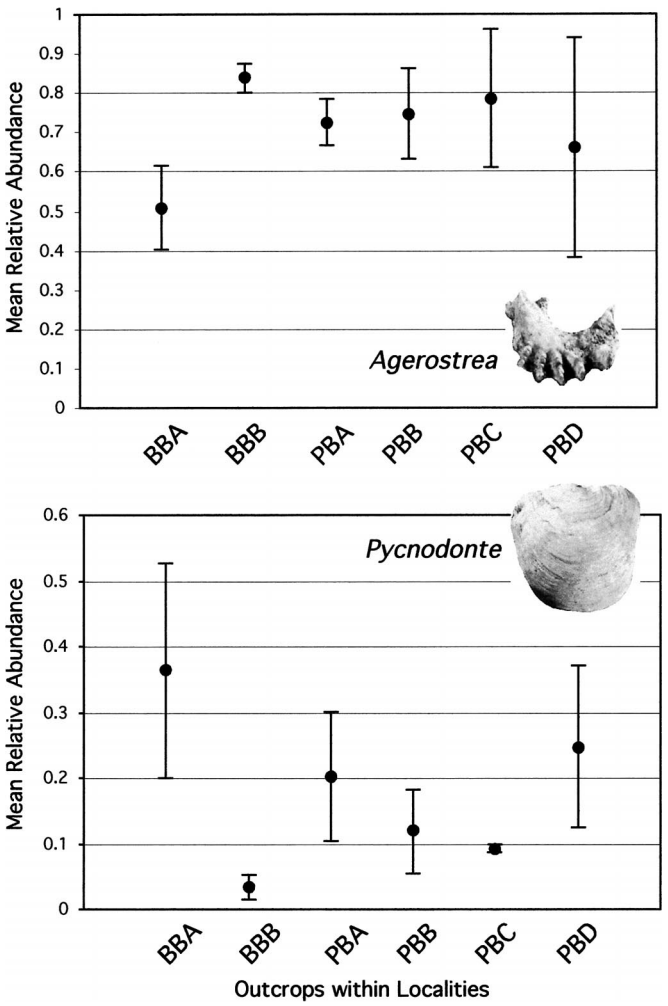

FIGURE 4—Plots of mean relative abundance and 95% cluster confidence intervals by outcrop for *Agerostrea mesenterica* and *Pycnodonte convexa*.

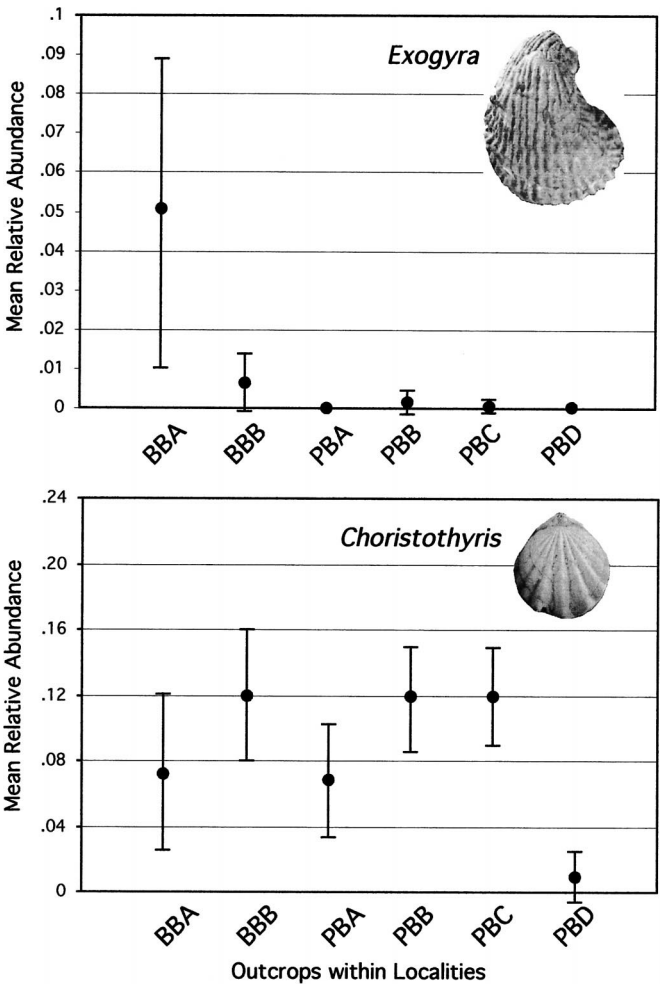

FIGURE 5—Plots of mean relative abundance and 95% cluster confidence intervals by outcrop for *Exogyra costata* and *Choristothyris plicata*.

tween outcrops BBA and BBB, and between PBA and both Big Brook outcrops. The large variation in confidence interval size from outcrop to outcrop indicates that individual replicate samples were highly variable at some outcrops (i.e., PBD) and very similar at others (i.e., BBB). *Pycnodonte*, which is even more variable in its mean abundance from outcrop to outcrop at both localities, also shows a wide range of confidence interval values (Fig. 4). *Exogyra* was found in very low abundance at all outcrops (Fig. 5). *Exogyra* was most abundant at outcrop BBA, contributing 5% to the total number of individuals. The wide confidence bands and examination of the count data (Table 1) show that this is due to an elevated number of individuals in a single replicate sample. *Choristothyris* was present at similar mean abundance in all outcrops except PBD (Fig. 5) where its low abundance is statistically significant. The large overlap in confidence intervals at the other five outcrops shows that the differences in mean abundance in *Choristothyris* at most outcrops (all except PBD) are not statistically significant (i.e., they are artifacts of sampling).

Comparison at the Locality Scale

To compare paleocommunity composition between Big Brook and Poricy Brook, mean relative abundances and 95% cluster confidence intervals were calculated for the five most abundant species using all replicate samples from each locality (Fig. 6). For all species, mean relative abundances are very similar and cluster confidence intervals show almost complete overlap between the localities. At this level of sampling effort, there are no significant statistical differences in species abundance composition between the Big Brook and Poricy Brook localities. The same result is obtained using a reduced data set in four trials culling two replicate samples at random from each outcrop at the Poricy Brook locality (Fig. 7). The increased range of the cluster confidence intervals for Poricy Brook in the reduced data set illustrates the loss of precision caused by the reduction in overall sampling effort. However, because sampling effort is still dispersed among all four localities, the means abundances are still comparable.

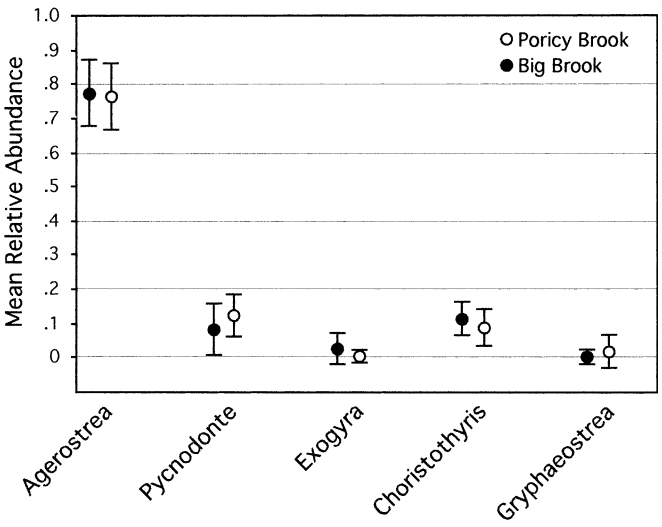

FIGURE 6—Plots of mean relative abundance and 95% cluster confidence intervals by locality for five most abundant species in the Navesink shell bed.

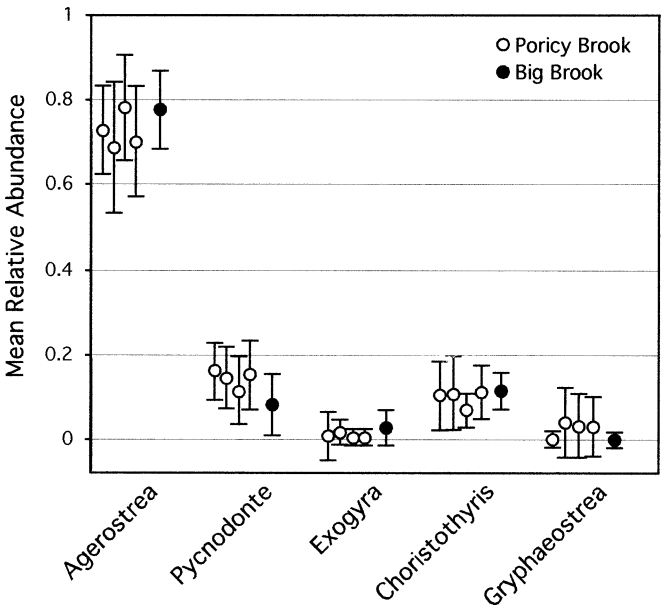

FIGURE 7—Plots of mean relative abundance and 95% cluster confidence intervals by locality for five most abundant species in the Navesink shell bed showing four random trials using reduced Poricy Brook data set (two replicate samples per outcrop).

DISCUSSION

Time Averaging, Transport, and Patchiness in the Navesink Shell Bed

Time Averaging: The degree to which a fossil assemblage has been time averaged is related directly to its taphonomic history and such variables as the rate of sediment accumulation, rate of shell destruction, and the amount of mixing induced by reworking and bioturbation (Kidwell and Bosence, 1991; Kowalewski, 1996; Olszewski and West, 1997). Patchiness is short-term variability that potentially can be homogenized over the long-term by time averaging. Shell beds deposited by sudden burial, such as census assemblages, should retain most of their original community patchiness, whereas shell beds that accumulate over long intervals of time are more likely to have lost the patchiness inherent in the living community. Likewise, replicate samples from shell beds that have been greatly time averaged should be more similar than samples from shell beds produced by less time averaging (Olszewski and West, 1997).

The taphonomic and sedimentologic characteristics of the Navesink shell bed, as well as its position near the maximum flooding surface of a transgressive systems tract (Sugarman et al., 1995), identify it as an example of a backlap shellbed (Kidwell, 1991; Kondo et al., 1998), developed within a condensed section deposited under conditions of sediment starvation (Bonelli and Bennington, 2000). A variety of observations demonstrate that the Navesink shell bed was deposited in a benthic environment characterized by low rates of sediment input, where shells remained exposed on the sediment surface for long periods of time and were mixed by bioturbation. Of the large oyster specimens in the Navesink shell bed, almost 100% show evidence of bioerosion, primarily in the form of clionid borings, but there are also borings attributable to lithophagid bivalves, acrothoracican barnacles, and polychaete annelids. Encrusting organisms are common and include several species of bryozoa, serpulid annelids, and small oysters. Many large oyster valves are biodegraded almost completely and some show evidence of having remained partially buried for extended periods of time. Most bivalved specimens are disarticulated, although it is not uncommon to find articulated specimens or matching valves in close proximity. Age dates based on strontium isotope stratigraphy from shelly material in the Navesink Formation at Big Brook (Sugarman et al., 1995) suggest an average rate of deposition of 1 meter per million years for the glauconite sand of the upper Navesink. Slow rates of sediment accumulation also are shown by the near absence of detrital quartz, intensely bioturbated sediment fabric, and accumulation of fecal pellets and their subsequent alteration to glauconite (Martino and Curran, 1990). The length of time over which the Navesink shell bed accumulated and the degree to which the fossil assemblage has been time averaged are impossible to measure precisely. Within-habitat, time averaged fossil assemblages that accumulate in outer shelf environments can encompass from hundreds to tens of thousands of years (Kidwell and Bosence, 1991). Even the lower boundary of this range would provide time for hundreds of generations of fauna to contribute hard parts to the shell bed and, over time, to homogenize the patchiness present in the distribution of living individuals at any one time.

Trends in Species Covariance: Each replicate sample consists of an equal volume of sediment collected in bulk. Therefore, changes in specimen abundance from replicate sample to replicate sample reflect changes in the local density of each species within the shell layer, which can create a patchy mosaic of relative abundance if all species do not covary in density together. Cummins et al. (1986) argued that covariance of species within a fossil assemblage from a single habitat does not usually result from biological interactions, because of the patchiness inherent in the distribution of the species in the living community. In-

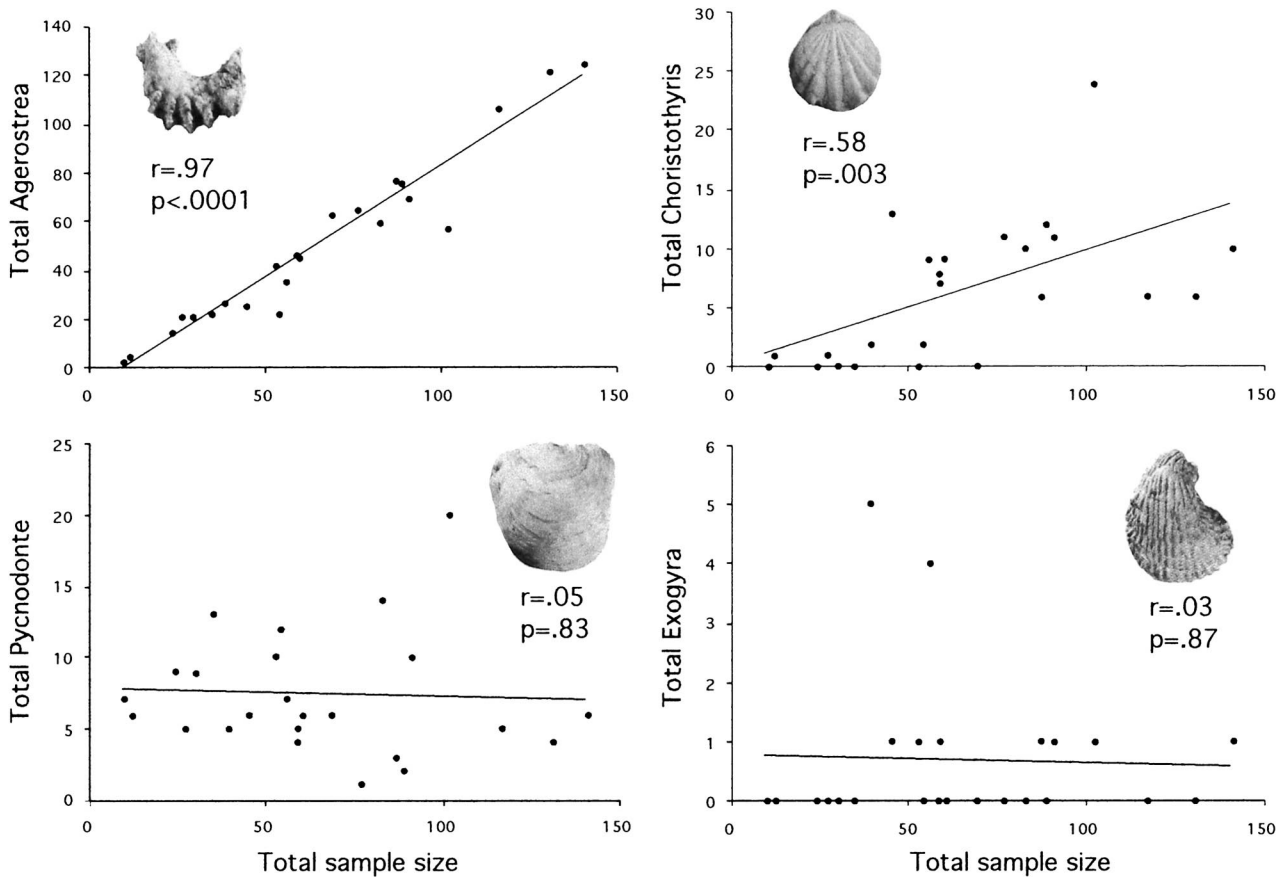

FIGURE 8—Density plots with best-fit linear regression lines showing number of individuals of each species counted versus total number of individuals in bulk replicate samples.

stead, they showed that the covariance of species in a death assemblage is usually a product of post-mortem shell transport, which mixes species from different patches according to their hydrodynamic characteristics (Cummins et al., 1986).

A graph of total individuals (density of a single species) versus replicate sample size (density of all species combined) for the most abundant species, *Agerostrea mesenterica* (Fig. 8), reveals a highly significant positive trend (r=0.97). Samples that contain more total specimens tend to have a high density of *Agerostrea*, which is the most abundant species in the shell bed. Hence, total sample density is controlled primarily by the density of *Agerostrea*, which varies considerably between adjacent replicate samples (Table 1), showing the strong tendency for *Agerostrea* to occur in small-scale patches. *Choristothyris*, which occurs in low abundance overall, also shows a trend of increasing species density with increasing total density (Fig. 8). Although this trend is not as robust (r=0.58) as that seen in *Agerostrea*, it is significant (p=0.003) and shows a tendency for *Choristothyris* to co-vary in density in patches along with *Agerostrea*. Neither *Pycnodonte* nor *Exogyra* show a trend of individual species density relative to total replicate sample density (Fig. 8), indicating that these two species, although patchy in their distribution among replicate samples (Table 1), do not co-vary with *Agerostrea*. This lack of covariance will cause the relative abundance of these species to vary greatly from sample to sample, depending on the density of *Agerostrea*, accentuating their already patchy small-scale distribution. Tests for correlation of species abundances in replicate samples found a weak but significant positive correlation between *Agerostrea* and *Choristothyris*, but no other species pair showed any significant correlation in abundance (Table 2). This supports the conclusion that there is little covariance of species in the patches that compose the Navesink shell bed.

Spatial Trends in Patchiness: In spite of the potential for time-averaging to homogenize patchiness in the Navesink shell bed, bulk replicate samples show a high degree of patchiness in species density and relative abundance on a spatial scale of meters (Fig. 3). This is shown by the changes in species density discussed above as well as by the large cluster confidence intervals around mean estimates of relative abundance seen at some outcrops (Figs. 4, 5). At the outcrop scale (tens of meters to hundreds of meters, Fig. 3), species are still somewhat patchy in their distribution. Although most localities had relatively similar mean species abundances, at least one locality did have a highly different mean species abundance with non-overlapping cluster confidence intervals in three of the four species analyzed (Figs. 4, 5). At the locality scale (kilometers, Fig. 3) all species show very similar mean abundances and all have mostly overlapping 95% cluster confidence intervals (Fig. 6), revealing no significant difference in the relative abundance of species between localities.

TABLE 2—Significance values (a = 0.05) for ANOVA test of regression on correlations of raw abundance of species pairs in all replicate samples. Statistically significant correlations are shown in bold.

	Agerostrea	*Pycnodonte*	*Exogyra*	*Choristothyris*
Agerostrea				
Pycnodonte	0.27			
Exogyra	0.71	0.87		
Choristothyris	**0.04**	0.32	0.61	
Gryphaeostrea	0.40	0.28	0.69	0.51

There are various reasons why large amounts of small-scale (within-outcrop and within-locality) patchiness in the Navesink shell bed persist in spite of the overall time averaged nature of the Navesink Formation. Shells that accumulated on the seafloor during the formation of the Navesink shell bed show evidence of having spent a long time in the taphonomically active zone (TAZ), perhaps due to the combination of slow sediment accumulation and lack of physical disturbance in the outer shelf environment. Bioerosion of shells in the TAZ could have had the effect of greatly reducing the number of older shells in the assemblage, leaving the Navesink shell bed time-incomplete (sensu Kowalewski, 1996) and dominated by species patches that existed late in the accumulation of the deposit (Olszewski, 1999). The lack of covariance in species abundance between replicate samples argues that within-habitat transport was not a significant factor in the formation of the Navesink shell bed. Cummins et al. (1986) suggested that within-habitat transport of shells is more effective than time averaging at homogenizing patchiness. Finally, in the muddy depositional environment of the Navesink Formation, the only available hard substrate would have been dead shells. This creates the potential for taphonomic feedback (Kidwell and Jablonski, 1983), whereby previously existing dead shells create suitable substrate for successive generations of attaching organisms. The extreme variability in the density of the most abundant species, *Agerostrea*, might have resulted from this species concentrating in patches where high numbers of dead shells from previous generations allowed greater numbers of larval oysters to attach successfully and grow to maturity. Adult *Agerostrea* in the Navesink Formation frequently are found cemented to other individuals. *Choristothyris*, a punduculate brachiopod that also requires a hard substrate for attachment, was the only species seen to covary with *Agerostrea* (Table 2), suggesting that its abundance also may be related to the density of shells in the substrate. The larger oysters *Exogyra* and *Pycnodonte* are gryphaeids, which were free-lying as adults (Seilacher, 1994) and thus may have been much less dependent on the availability of shell-rich patches, explaining their lack of covariance with *Agerostrea*. It is not possible to assess the degree to which *Gryphaeostrea* was influenced by taphonomic feedback because it only occurs in three replicate samples (Table 1).

Implications for Paleoecological Sampling

The results of this study demonstrate how important it is to consider spatial scale when sampling fossil assemblages to determine species abundance distributions or any other parameters used to characterize paleocommunities. Paleontologists need to sample in a way that makes it practical to collect data, but also provides data that are both representative of the entity being characterized and amenable to statistical analysis. One problem that perhaps has prevented achieving this goal in the past is confusion of terminology between statisticians and researchers in the field. The mass of sediment and fossils forming the *sample* collected by a paleontologist is an *observation* to a statistician (Hayek and Buzas, 1997). Alone, an observation is not very useful because it is a single attempt to characterize the larger population and is biased to an unknown degree. The only way to assess the bias of an observation is to make additional observations and compare them [". . . any one estimate can still be quite deviant from the true relative abundance value. . . replication is essential in natural population fieldwork." (Hayek and Buzas, 1997, p. 200)]. Multiple observations form a *statistical sample*, which is the statistician's characterization of the larger population that incorporates some measure of bias or uncertainty (e.g., mean species abundance with 95% error bars). In this study, the replicate samples are observations and they are combined to generate statistical samples at two different spatial scales–the outcrop scale and the locality scale (Fig. 3).

The Navesink shell bed provides an ideal test case to demonstrate why successful statistical sampling must incorporate spatially dispersed replicate samples (observations) if conclusions drawn from the data are to be reliable. The fossil assemblage preserved in the upper Navesink is a low-diversity, highly time averaged assemblage developed in what was likely a spatially homogeneous outer shelf habitat of muddy, glauconite sands. The conclusion of this study that the local paleocommunities present at the Poricy Brook and Big Brook localities are statistically indistinguishable is not surprising given these characteristics. What is instructive is the amount of patchiness preserved in the Navesink Shell bed at small spatial scales. If the paleocommunity comparison between localities had been based on comparing a single, large bulk sample from each locality, then there would have been a high probability of collecting two samples from dissimilar patches, meaning the comparison would not have been reliable because the samples would have been biased. No matter how large a single bulk sample becomes, if it is sampling the same patch it will remain biased. Subdividing a large bulk sample into several smaller samples or collecting replicate samples from the same place results in pseudoreplication; the resulting statistical sample is just as biased as a single large sample. Collecting closely-spaced replicate samples improves the reliability of the statistical samples used to make the comparison, but if the data for comparing the Navesink localities had been derived from replicate sam-

ples at only two outcrops, there would still be a reasonable chance that two statistically dissimilar outcrops would have been chosen for comparison. For example, if only the replicate sample data from localities BBA and PBA had been collected, the conclusion reached would have been that both *Agerosterea* and *Exogyra* have significantly different relative abundances at the two localities (Figs. 4 and 5). However, by collecting replicate samples from different outcrops within each locality, the patchiness present at the outcrop scale is incorporated into the statistical sample for each locality and no differences are found in species abundance between the two localities (Fig. 6). The lesson here is that replicate samples (observations) should be dispersed throughout the area being characterized by the statistical sample.

Bennington and Rutherford (1999) argued that most of the effort and cost in sampling comes in traveling to localities and in extracting and identifying the individual specimens. Once on site, collecting many small replicate samples is not significantly more arduous than collecting one or a few large samples. Furthermore, the parameter that determines the power of a comparison is the size of the statistical sample. There is no lower limit to the size of individual replicate samples that combine to form the statistical sample. Collecting a greater number of smaller, randomly dispersed replicate samples always will improve the precision and reliability of a comparison by decreasing the bias of the statistical sample (Bennington and Rutherford, 1999).

It cannot be known *a priori* what is an appropriate spatial scale across which to collect replicate samples. Ideally, one would collect many randomly distributed replicates across the region occupied by the paleocommunity being characterized. However, paleocommunities, like the living communities they record, often represent ecological systems with poorly defined boundaries (Miller, 2001). If paleocommunities intergrade because they formed along ancient environmental gradients, then trying to identify their boundaries is probably both impracticable and pointless. Hence, where does one stop collecting replicate samples when trying to define a paleocommunity quantitatively? One reasonable course of action is to collect replicate samples across several spatial scales, as illustrated in this study. This is particularly important if comparisons are to be made, because the variation in species abundance, diversity, or other quantitative parameters at a smaller scale can be employed to assess the significance of variation observed at a larger scale. In most cases for paleontologists, the available spatial scales for sampling will be constrained by the scale at which the fossil-bearing strata crop out. In the Navesink Formation, exposures are limited to localities where the shell bed is exposed by stream erosion at an accessible level. Locality exposures are not continuous along the stream, but are confined to outcrops along cut banks, and stretches of the stream between outcrops contain no exposures because of slumping and topography. Even if exposure is continuous for great distances, one can design a sampling protocol to incorporate one or more arbitrarily defined scales for the statistical samples. What is important is that the scale of the statistical sample be made explicit and that it be similar for all localities or horizons between which comparisons are made. Replicate samples then should be collected randomly within the bounds of the statistical sample. It is also important not to overextend the reach of conclusions drawn from the statistical samples. For example, to conclude that two shell bed horizons contain different paleocommunities because significant differences were detected in comparisons made of single localities is not very reasonable. It would be more informative to have a measure of the variability in each shell bed horizon among several localities before assessing the significance of differences observed between the beds (e.g., Bennington and Bambach, 1996). Likewise, comparisons made between local paleocommunities in a single shell bed at two different localities must incorporate replicate samples within each locality to sample small-scale patchiness. If replicates are collected in very close proximity, they may fail to sample large-scale patchiness adequately, leading one to over-emphasize the differences in species abundance and diversity that exist between localities (e.g., Cobabe and Allmon, 1994).

CONCLUSIONS

This study is not designed to be an exact model of sampling protocol for all studies incorporating paleocommunity comparisons. The sampling effort employed in this study to compare a single stratigraphic horizon between two localities is probably excessive. In most cases, it would not be necessary to analyze the scale of patchiness within each locality to make a reliable comparison between localities. What this study does provide is a test case to illustrate three very important, but under-appreciated, aspects of paleoecological sampling: (1) Species covary to different degrees at different spatial scales. At the larger scale of the habitat and community, species covary because they share similar habitat preferences or because their distributions are shaped by ecological interactions (Cummins et al., 1986). However, at the smaller, within-habitat scale, species tend to be distributed in patches and show little consistent covariance. Time averaging and within-habitat transport may, or may not, be effective at homogenizing different patches. The Navesink shell bed, in spite of its potential to host a highly time averaged fossil assemblage, contains species that are patchy in their distribution at spatial scales of meters and tens of meters (the replicate sample and outcrop scales), although its species abundance composition is consistent at the kilometer (locality) scale. (2) When fossil assemblages are patchy, sampling to quantify species abundance will be unreliable (biased by sampling error caused by patchiness) unless sampling effort is dispersed among replicate samples. Collecting additional replicate samples does not require that overall sampling effort (the total number of specimens collected) be increased at a locality, only that sampling effort be dispersed so that many small samples are collected rather than one or a few large samples. Dispersed sampling effort should be employed in any paleontological study attempting to quantify any aspect of a fossil assemblage distributed across a spatial scale larger than the size of the patches formed by the contiguous distributions of the component species or discontinuous nature of the habitat. An analogous argument for dispersing replicate samples also has been made for morphometric studies of colonial organisms. Hageman (2001) reported a strong patch effect in the variation detected among individuals in living

bryozoan colonies, and cautioned against collecting replicate samples of morphometric data from a single small fossil bryozoan fragment. (3) Paleontologists always should make explicit the scale encompassed by the statistical samples they are collecting. If a comparison is being made between the local paleocommunities at different localities (i.e., for a gradient analysis), then a consistent spatial scale for each statistical sample should be defined and replicates should be collected randomly within that scale at each locality. If a temporal comparison is to be made between paleocommunities preserved in different stratigraphic horizons, then the appropriate statistical samples must encompass replicate samples collected from the stratigraphic horizons across a consistent and sufficiently large spatial scale to encompass within-habitat patchiness. In either case, it may be of interest to collect replicate samples at two or more spatial scales to create a hierarchy of statistical samples such as in this study. In this way, differences observed between statistical samples at the smaller scale may be informative for interpreting the significance of differences observed at the larger scale.

ACKNOWLEDGEMENTS

I would like to thank all of the members of my fall 1998 paleontology class who did the bulk sampling and initial sample processing for this study, including Jeanette Greenidge, Lauren Kosmij, Mike Leone, and Jessica Miller, all of whom are now public-school science teachers. Nancy Forsberg coordinated mapping of the outcrop locations. James Bonelli, Mark Sels, and Jen Chandler helped greatly with additional sample processing. I also would like to thank Jason Williams and Russ Burke of the Hofstra Biology Department for their insightful editorial comments. Michal Kowalewski and Tom Olszewski reviewed the original manuscript and contributed many useful ideas for improving its substance. Hofstra University supported this research through travel grants, Faculty Development Grants, and Presidential Research Awards.

REFERENCES

BAMBACH, R.K., and BENNINGTON, J B., 1996, Do communities evolve? A major question in evolutionary paleoecology: *in* Jablonski, D., Erwin, D.H., and Lipps, J.H., eds., Evolutionary Paleobiology: University of Chicago Press, Chicago, p. 123–160.

BENNINGTON, J B., and BAMBACH, R.K., 1996, Statistical testing for paleocommunity recurrence: are similar fossil assemblages ever the same?: Palaeogeography Palaeoclimatology Palaeoecology, v. 127, p. 107–133.

BENNINGTON, J B., and RUTHERFORD, S.D., 1999, Precision and reliability in paleocommunity comparisons based on cluster-confidence intervals: how to get more statistical bang for your sampling buck: PALAIOS, v. 14, p. 506–515.

BONELLI, J., and BENNINGTON, J B., 2000, Sequence stratigraphic interpretation of the Upper Cretaceous Navesink formation, central New Jersey: evidence from macrofossil distribution and taphonomy: Geological Society of America, Abstracts with Programs, v. 32,No. 1, p. A-6.

BRETT, C.E., and BAIRD, G.C., 1995, Coordinated stasis and evolutionary ecology of Silurian to Middle Devonian faunas in the Appalachian Basin: *in* Erwin, D.H., and Anstey, R.L., eds., New Approaches to Speciation in the Fossil Record: Columbia University Press, New York, p. 285–315.

BUZAS, M.A., 1968, On the spatial distribution of foraminifera: Contributions from the Cushman Foundation for Foraminiferal Research, v. 19, p. 1–11.

COBABE, E.A., and ALLMON, W.D., 1994, Effects of sampling on paleoecologic and taphonomic analysis in high-diversity fossil accumulations: an example from the Eocene Gosport Sand, Alabama: Lethaia, v. 27, p. 167–178.

CUMMINS, H., POWELL, E.N., NEWTON, H.J., STANTON, R.J., JR, and STAFF, G., 1986, Assessing transportation by the covariance of species with comments on contagious and random distributions: Lethaia, v. 19, p. 1–22.

GILINSKY, N.L., and BENNINGTON, J B., 1994, Estimating numbers of whole individuals from collections of body parts: a taphonomic limitation of the paleontological record: Paleobiology, v. 20, p. 245–258.

HAGEMAN, S.J., 2001, Partitioning phenotypic variation into genotype, environment and GXE interaction and within-colony factors: implications for morphometric analyses (Bryozoa): Geological Society of America, Abstracts with Programs, v. 33, No.6, p. A-249.

HAIRSTON, N.G., 1959, Species-abundance and community structure: Ecology, v. 40, p. 404–415.

HAYEK, L.C., and BUZAS, M.A., 1997, Surveying Natural Populations: Columbia University Press, New York, 563 p.

HOLLAND, S.M., MILLER, A.I., MEYER, D.L., and DATTILO, B.F., 2001, The detection and importance of subtle biofacies within a single lithofacies: the Upper Ordovician Kope Formation of the Cincinnati, Ohio region: PALAIOS, v. 16, p. 205–217.

HOLTERHOFF, P.F., 1996, Crinoid biofacies in Upper Carboniferous cyclothems, midcontinent North America: faunal tracking and the role of regional processes in biofacies recurrence: Palaeogeography Palaeoclimatology Palaeoecology, v. 127, p. 47–81.

IVANY, L.C., 1999, So. . . now what? Thoughts and ruminations about coordinated stasis: PALAIOS, v. 14, p. 297–300.

JACKSON, J.B.C., BUDD, A.F., and PANDOLFI, J.M., 1996, The shifting balance of natural communities?: *in* Jablonski, D., Erwin, D.H., and Lipps, J.H., eds., Evolutionary Paleobiology: University of Chicago Press, Chicago, p. 89–122.

KIDWELL, S.M., 1991, Condensed deposits in siliciclastic sequences: expected and observed features: *in* Einsele, G., Ricken, W., and Seilacher, A., eds., Cycles and Events in Stratigraphy: Springer-Verlag, Berlin, p. 682–695.

KIDWELL, S.M., 2001, Preservation of species abundance in marine death assemblages: Science, v. 294, p. 1091–1094.

KIDWELL, S.M., and BOSENCE, D.W.J., 1991, Taphonomy and time-averaging of marine shelly faunas: *in* Briggs, D.E.G., and Allison, P.A., eds., Taphonomy, Releasing Information From the Fossil Record: Plenum Press, New York, p. 116–211.

KIDWELL, S.M., and JABLONSKI, D., 1983, Taphonomic feedback: ecological consequences of shell accumulation: *in* Tevesz, J.S., and McCall, P.L., eds., Biotic Interactions in Recent and Fossil Benthic Communities: Plenum Press, New York, p. 195–248.

KONDO, Y., ABBOTT, S.T., KITAMURA, A., KAMP, P.J.J., NAISH, T.R., KAMATAKI, T., and SAUL, G.S., 1998, The relationship between shellbed type and sequence architecture: examples from Japan and New Zealand: Sedimentary Geology, v. 122, p. 109–127.

KOWALEWSKI, M., 1996, Time-averaging, overcompleteness, and the geological record: Journal of Geology, v. 104, p. 317–326.

LAFFERTY, A.G., MILLER, A.I., and BRETT, C.E., 1994, Comparative spatial variability in faunal composition along two Middle Devonian paleoenvironmental gradients: PALAIOS, v. 9, p. 224–236.

LEVINTON, J.S., and BAMBACH, R.K., 1975, A comparative study of Silurian and recent deposit-feeding bivalve communities: Paleobiology, v. 1, p. 97–124.

MARTINO, R.L., and CURRAN, H.A., 1990, Sedimentology, ichnology, and paleoenvironments of the Upper Cretaceous Wenonah and Mt. Laurel Formations, New Jersey: Journal of Sedimentary Petrology, v. 60, p. 125–144.

MILLER, A.I., 1988, Spatial resolution in subfossil molluscan remains: implications for paleobiological analyses: Paleobiology, v. 14, p. 91–103.

MILLER III, W., 2001, What's in a name? Ecologic entities and the marine paleoecologic record: *in* Allmon, W.D., and Bottjer, D.J., eds., Evolutionary Paleoecology: Columbia University Press, New York, p. 15–34.

OLSSON, R.K., 1963, Latest Cretaceous and earliest Tertiary stratigraphy of the New Jersey Coastal Plain: Bulletin of the American Association of Petroleum Geologists, v. 47, p. 643–665.

OLSZEWSKI, T.D., 1999, Taking advantage of time-averaging: Paleobiology, v. 25, p. 226–238.

OLSZEWSKI, T.D., and PATZKOWSKY, M. E., 2001, Measuring recurrence of marine biotic gradients: a case study from the Pennsylvanian-Permian midcontinent: PALAIOS, v. 16, p. 444–460.

OLSZEWSKI, T.D., and WEST, R.R., 1997, Influence of transportation and time-averaging in fossil assemblages from the Pennsylvanian of Oklahoma: Lethaia, v. 30, p. 315–329.

OWENS, J.P., and SOHL, N.F., 1969, Shelf and deltaic paleo-environments in the Cretaceous and Tertiary formations of the New Jersey Coastal Plain: *in* Subitzky, S., ed., Geology of Selected Areas in New Jersey and Eastern Pennsylvania and Guidebook of Excursions: Rutgers University Press, New Brunswick, NJ, p. 235–278.

PANDOLFI, J.M., 1996, Limited membership in Pleistocene reef coral assemblages from the Huan Peninsula, Papua New Guinea: constancy during global change: Paleobiology, v. 22, p. 152–176.

PATZKOWSKY, M.E., 1999, A new agenda for evolutionary paleoecology—or would you in the background please step forward: PALAIOS, v. 14, p. 195–197.

PATZKOWSKY, M.E., and HOLLAND, S.M., 1997, Patterns of turnover in Middle and Upper Ordovician brachiopods of the eastern United States: Paleobiology, v. 23, p. 420–443.

PATZKOWSKY, M.E., and HOLLAND, S.M., 1999, Biofacies replacement in a sequence stratigraphic framework: Middle and Upper Ordovician of the Nashville Dome, Tennessee USA: PALAIOS, v. 14, p. 301–323.

SANDERS, H.L., 1960, Benthic studies in Buzzards Bay III. The structure of the soft-bottom community: Limnology and Oceanography, v. 5, p. 138–153.

SEILACHER, A., 1984, Constructional morphology of bivalves: evolutionary pathways in primary versus secondary soft-bottom dwellers: Palaeontology, v. 27, p. 207–237.

SUGARMAN, P.J., MILLER, K.G., BUKRY, D., and FEIGENSON, M.D., 1995, Uppermost Campanian-Maestrichtian strontium isotopic, biostratigraphic, and sequence stratigraphic framework of the New Jersey Coastal Plain: Geological Society of America Bulletin, v. 107, p. 19–37.

WEBER, A.J., 2001, Use of high-resolution fossil data to assess the causes of meter-scale cyclicity in the type Cincinnatian series (Upper Ordovician): Geological Society of America, Abstracts with Programs, v. 33, p. A-31.

ACCEPTED MAY 26, 2002