Assessing the fidelity of the fossil record by using marine bivalves

James W. Valentine*†, David Jablonski‡, Susan Kidwell‡, and Kaustuv Roy§

*Department of Integrative Biology, University of California, Berkeley, CA 94720; ‡Department of Geophysical Sciences, University of Chicago, Chicago, IL 60637; and §Section of Ecology, Behavior, and Evolution, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116

Contributed by James W. Valentine, March 6, 2006

Taxa that fail to become incorporated into the fossil record can reveal much about the biases of this record and provide the information needed to correct such biases in empirical analyses of the history of life. Yet little is known about the characteristics of taxa missing from the fossil record. For the marine Bivalvia, which have become a model system for macroevolutionary and macroecological analysis in the fossil record, 308 of the 1,292 living genera and subgenera (herein termed ''taxa'') are not recorded as fossils. These missing taxa are not a random sample of the clade, but instead tend to have small body size, reactive shell structures, commensal or parasitic habit, deep-sea distribution, narrow geographic range, restriction to regions exposing few Neogene marine sediments, or recent date of formal taxonomic description in the neontological literature. Most missing taxa show two or more of these features and tend to be concentrated in particular families. When we exclude the smallest taxa (<1 cm) and deep-sea endemics, date of published description and geographic range become the strongest predictors of the missing taxa; other factors are statistically insignificant or have relatively small effects. These biases might influence a variety of analyses including the use of fossil data in support of phylogenetic analyses, molecular clock calibrations, and analyses of spatial and temporal dynamics of clades and biotas. Clade inventories such as these can be used to develop protocols that minimize the biases imposed by sampling and preservation.

Bivalvia | taphonomy | unfossilized taxa

Evaluating the fidelity and completeness of the fossil record has been important to evolutionary studies beginning with Darwin's seminal work (1). Understanding the biases of the fossil record is also critical for interpreting the temporal dynamics underlying large-scale biodiversity patterns. In the sea, the modern evolutionary fauna (2) expanded dramatically after the end-Permian extinctions ≈250 million years ago, with bivalve and gastropod mollusks as dominant components. The living marine Bivalvia are particularly well studied, and their preservability is generally high, so that they provide an excellent model system for evolutionary and macroecological studies (3, 4). Nevertheless, 24% of the 1,292 living marine bivalve genera and subgenera are unknown as fossils. Previous local and regional studies have suggested that important factors negatively affecting preservation include small size, fragile shells, deeper-water habitats, and rarity (e.g., refs. 3–7). In this study we use a global survey of genera and subgenera of living marine bivalves to test whether those *not* found in the fossil record (hereafter *missing taxa*) represent a random sample with respect to their body sizes, life habits, habitats, distributions, and shell composition. Our results show that the missing taxa represent a nonrandom subset of the living bivalves and suggest that empirical analyses of fossil data need to take these biases into account.

Results

General Taxonomic Distribution of the Missing Genera. Almost a quarter (24%) of all genera and subgenera of living bivalves are missing from the fossil record. The proportion of missing taxa in each of the seven major living bivalve clades ranges from 6% to 45%, with three clades diverging significantly from the overall percentage (log-likelihood ratio tests, with Bonferroni correction) (Table 1; we exclude Trigonioida here because it contains only one living genus). This significant interclade variation suggests that missing taxa are not simply a random sample of the living biota.

Table 1. Major living clades of Bivalvia, showing the number of families and genera/subgenera assigned to each and the percentage of genera found fossil in each

Clade	No. families	No. genera	Missing genera, %
Protobranchs	11	74	24
Arcoids	7	74	15
Mytiloids	1	61	20
Pteriomorphs	16	143	15*
Trigonioids	1	1	0
Lucinoids	2	52	6**
Heteroconchs	52	774	25
Anomalodesmatans	16	113	45***
Total	106	1,292	26

Classification follows Crame (18); lucinoids and anomalodesmatans may be monophyletic groups nested within the heteroconchs (46). *, differ from overall average at $P < 0.05$; **, differ from overall average at $P < 0.01$; ***, differ from overall average at $P < 0.001$ (log-likelihood ratio tests with sequential Bonferroni correction; see ref. 47).

Of the 106 extant bivalve families, 4 completely lack a fossil record, and these 4 each contain 3 or fewer taxa (see Tables 3 and 4, which are published as supporting information on the PNAS web site). The observed number of missing families is within the random expectation; given the distribution of genera within bivalve families, in a random sample of 984 genera we would expect the fossil record to capture an average of 100 families (sampling without replacement; 95% confidence interval is 96–104 families based on 1,000 replicates). However, as discussed below, the missing families do not appear to be a random ecological sample of Bivalvia; for example, three of the four missing families are restricted to the deep sea. At the other end of the richness scale, all five families with 50 or more genera lie at or below the mean in their proportions of missing taxa. Families that are significantly above the mean in terms of missing taxa are of intermediate taxonomic richness, but otherwise no single factor characterizes them all.

Body Size. The body-size distribution of all living bivalve taxa is not currently known. Thus, to investigate whether the missing taxa, for which we do have comprehensive size data, are size-

Conflict of interest statement: No conflicts declared.

Abbreviations: SAV, surface area-to-volume; SFD, size–frequency distribution.

†To whom correspondence should be addressed. E-mail: jwvsossi@socrates.berkeley.edu.

© 2006 by The National Academy of Sciences of the USA

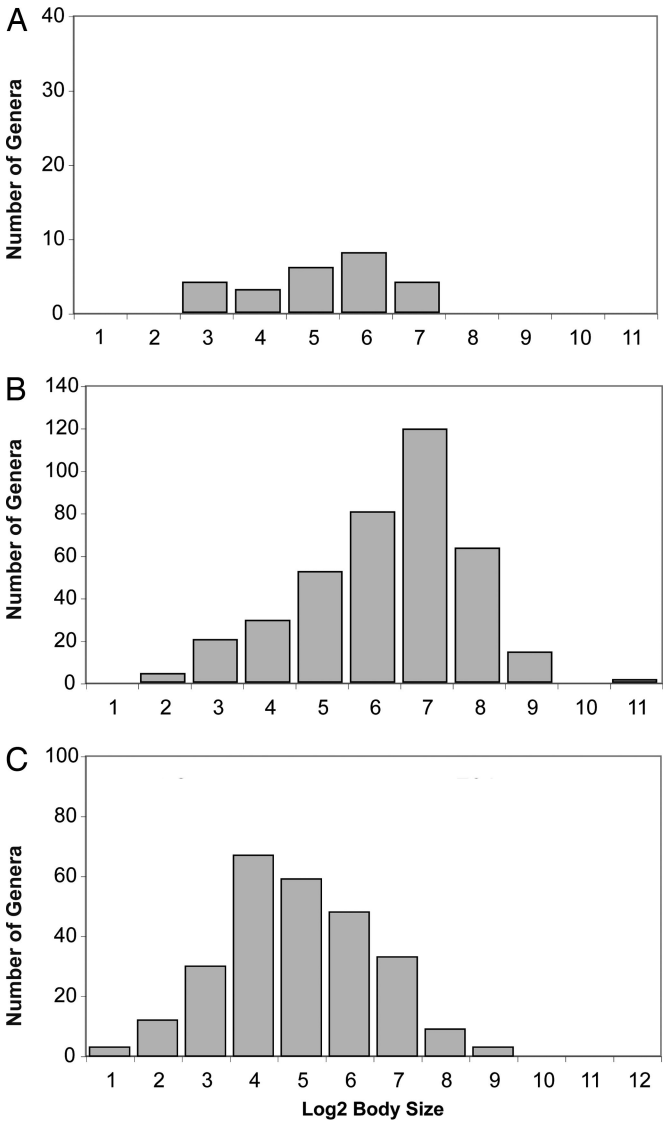

Fig. 1. Body sizes of marine bivalve genera and subgenera. The eastern Pacific bivalve genera and subgenera having a fossil record (*B*; $n = 382$, median = 5.03 $\log_2$ units = 32.7 mm) are significantly larger as a group than the eastern Pacific bivalve taxa missing from the fossil record (*A*; $n = 25$, median = 3.95 $\log_2$ units = 12.2 mm); $P = 0.0004$ by Mann–Whitney *U* test. The eastern Pacific genera having a fossil record are also significantly larger than the other missing bivalve taxa (*C*; $n = 268$, median = 3.31 $\log_2$ units = 10.0 mm).

biased, we used information for bivalve species from the northeastern Pacific continental shelf (Point Barrow, Alaska, 71°N, to the southern edge of the Panamic Province, 5°S), which is the best-known set of regional faunas (8). The close similarities in mode, median, and range of the size–frequency distributions (SFDs) from multiple biogeographic provinces along this shelf (8) suggest that the northeastern Pacific fauna, which includes 32% of all living genera and subgenera, is a reasonable first-order estimate for the global SFD. The bivalve species of the Antarctic are skewed to small body sizes as a group (9), but 12 of the 13 taxa (92%) restricted to that region lack a fossil record (see below), so that this anomalous SFD reinforces rather than undermines the comparisons reported here.

Data on body sizes of species within 407 shelf-depth genera in the northeastern Pacific show that the missing genera are significantly smaller than those with a fossil record (Fig. 1). The SFD of the missing northeastern Pacific genera ($n = 25$; median 1.2 cm) is evidently a reasonable sample of the global assemblage of missing taxa ($n = 268$; median 1.0 cm), because their SFDs do not differ significantly ($P = 0.22$ by Mann–Whitney *U* test). Moreover, the global SFD for the missing taxa differs significantly from the SFD of the northeastern Pacific bivalves having a fossil record, even when the missing eastern Pacific genera are excluded ($P < 0.000001$ by Mann–Whitney *U* test). More conservatively, using 1 cm (almost exactly the median size of all missing genera globally) as a threshold measure for "small" body size, 119 (48%) of the 247 missing shelf-depth taxa for which we have body sizes are small-bodied, whereas only 16% of the northeastern Pacific shelf-depth taxa are small-bodied by this criterion. Thus, there is a clear bias against the preservation of small bivalves as fossils, which may be exacerbated if small specimens are overlooked during collection or, as is commonly the case, are difficult to assign taxonomically.

Shell Composition. At least three aspects of shell composition have been hypothesized to affect the preservation of bivalve shells: (*i*) mineralogy, with aragonite being more vulnerable to chemical dissolution than calcite; (*ii*) organic content, with microstructures having high ratios of organic matrix to mineral phases being more prone to postmortem microbial disintegration; and (*iii*) surface area-to-volume (SAV) of mineral crystallites, with high-SAV microstructures being more reactive chemically (4, 7, 10–12). As shown in Table 2, only mineralogy and SAV show significant differences between missing and fossilized taxa, although the observed differences in organic content are also in the expected direction. SAV has a much stronger effect than mineralogy, but SAV data are less reliable (4) and so these results should be interpreted with caution. The influence of shell composition is not independent of body size, however. For example, 34 ± 11% of missing taxa that are ≤1 cm (with 95% binomial confidence limits) have high-SAV shells, whereas only 12 ± 6% of missing taxa >1 cm are high-SAV (see Table 4).

Depth. Although the deep sea is a rich environment in terms of biodiversity (13, 14), much of the bivalve fauna there involves genera and subgenera that also occur at shelf depths; only 61 of the 1,292 living taxa are restricted to depths >200 m. Of these exclusively deep-sea forms, which are dominated by protobranchs and carnivorous anomalodesmatans, 67% lack a fossil record. This strong bias is nonetheless not driven by a single group. For example, 50% of the deep-sea protobranchs and 75% of the deep-sea anomalodesmatans are missing.

The under-representation of deep-sea taxa is not unexpected given that the sampled Cenozoic fossil record of bivalves is overwhelmingly from shelf depths, and we are impressed that as many as one-third of the exclusively deep-sea bivalve taxa are recorded as fossils. We suspect that this is because many deep-sea bivalve taxa are widely distributed, so that only a few exposures of fossiliferous deep-sea sediments (as are present along the coasts of Japan, the western United States, and southern Italy for example) can register a surprisingly large proportion of deep-sea bivalve genera.

Life Habit. Bivalves have a broad range of life habits, including species that are parasitic, carnivorous, and chemosymbiotic, and some that are permanently cemented, bore in rocks, or even swim. However, the most common life-habit categories are suspension- vs. deposit-feeders, burrowers (infauna) vs. surface-dwellers (epifauna), and free-living forms vs. commensals.

Burrowing bivalves might be expected to have a richer fossil record than surface-dwellers, because the burrowers live embedded in sediments, but this is not the case in our data. The richest epifaunal and infaunal families have similar generic representation in the fossil record; e.g., compare infaunal Veneridae (9% of taxa missing) and Tellinidae (21% missing) to

Table 2. Tests for the effects of factors hypothesized to influence occurrence of bivalve taxa in the fossil record, using log-likelihood ratio (*G*) test, with Williams' correction (47)

Category	Number missing	Proportion missing	Number fossil	Adjusted *G*
Shell composition				
Entirely aragonite	275	0.25	831	8.56**
Calcitic or bimineralic	25	0.15	142	
High-organic	27	0.28	70	1.31
Moderate and low-organic	264	0.23	903	
High SAV	40	0.51	38	41.13***
Moderate and low SAV	170	0.17	802	
Depth				
Shelf	266	0.22	965	57.80***
Deep-sea	42	0.69	19	
Life habit				
Epifauna	40	0.15	235	18.20***
Infauna	269	0.26	749	
Suspension feeder	226	0.23	769	
Deposit feeder†	41	0.25	125	0.31
Chemosymbiotic†	12	0.15	66	2.41
Carnivore†	29	0.55	24	23.47***
Commensal	64	0.57	62	61.78***
Free-living	244	0.21	922	
Geographic range‡				
Confined to one province	128	0.50	72	118.87***
More than one province	127	0.50	468	
Date of description				
Pre-*Treatise*	220	0.19	921	95.91***
Post-*Treatise*	88	0.58	63	

Taxa living partially exposed above the sediment–water interface ("semi-infaunal") grouped with epifaunal taxa. Numbers in category groupings do not sum consistently because of missing data. **, $P < 0.01$; ***, $P < 0.001$ (after sequential Bonferroni correction).

†Compared with suspension feeders.

‡For the 540 genera in Flessa and Jablonski's (31) database having a fossil record.

the epifaunal Pectinidae (7% missing) and Mytilidae (20% missing) (see Table 3). In fact, a full comparison shows that epifaunal taxa are more fully represented as fossils than are infaunal ones (Table 2).

This difference may be linked to factors other than infaunal vs. epifaunal habits *per se*. For example, all 1,018 infaunal taxa are aragonitic whereas 66% of the 274 epifaunal taxa are calcitic or bimineralic (4, 7). Body size may also play a role here; only two of the predominantly epifaunal families (Dimyidae and Gaimardiidae) are dominated by small taxa, and 50–66% of their taxa are missing from the record. Furthermore, aside from the Dimyidae, only 5 of the 39 missing epifaunal taxa have sizes <1 cm (vs. 135 of the missing 259 infaunal taxa for which we have size data), and in our full northeast Pacific sample only 10 out of 147 living epifaunal genera (7%) contain species <1 cm, all suggesting that epifaunal, and correlatively calcitic, clades have largely abandoned the lower end of the size spectrum.

Among feeding categories, the proportion of missing taxa in the two dominant modes, suspension- and deposit-feeders, are indistinguishable (Table 2). This is surprising, given that Protobranchia, the largest group of bivalve deposit-feeders, tend toward small body size, comprise a significant proportion of the deep-sea bivalves (13, 14), and have more reactive shells (high-organic aragonitic microstructures), and yet are missing almost precisely the overall class average (24%; Table 1).

Commensal and parasitic taxa are underrepresented as fossils by a large margin (Table 2). Roughly 10% of living bivalve taxa fall into this category, despite the relatively small amount of study they have received. This relatively high diversity suggests that commensal habits may entail morphological specializations, possibly coadaptive to their hosts or host burrows, which are manifested taxonomically at generic and subgeneric levels (15–17). These genera comprise chiefly small-bodied species (i.e., their median size is <1 cm, and no species exceeds 16 mm) and tend to have thin shells, so we must again consider whether small body size accounts for the preservational bias. Additional factors may be involved: commensal individuals may not survive their hosts and, especially those living in burrows (many with stomatopods), may have their shells destroyed during host decomposition; and many host-specific commensals live in the tropics, which is undersampled for Cenozoic fossils (see below).

Region. In many respects the spatial distribution of missing taxa echoes the general diversity pattern in the sea, with a distinct latitudinal gradient in both Northern and Southern hemispheres, and a strong concentration in the Indo-West Pacific (which contains 35% of the missing shelf-depth taxa). This situation suggests that the first-order spatial distribution of missing taxa is numerically a random sample of the full biota, but more precise data on the full biota are needed. As noted above, commensals are concentrated in the tropics, but anomalodesmatans are relatively rich at high latitudes (18), and we do not yet know whether these biogeographic trends among groups rich in missing taxa will cancel out. An extratropical hotspot of missing taxa occurs in southeastern Australia, with 21 of the missing genera (= 8% of shelf taxa for which we have biogeographic data). This concentration reflects especially intensive local sampling and taxonomic description of the living fauna, and we suspect that many of these taxa will eventually prove to occur in the adjoining tropical regions today and in the geologic past.

This spatial distribution of missing taxa derives from at least two influences: (*i*) regional geology, which influences spatial variation in quality of the fossil record, as in tropical Australia, which supports a large array of living bivalve genera but has few fossiliferous marine Neogene deposits (19); and (*ii*) undersampling of the late Cenozoic tropical fossil record even when deposits of the appropriate age and environment exist (e.g., refs. 20–23). For example, our compilation of all published records of marine bivalves in the late Neogene of Indonesia, which Beu (24) notes as the richest and most extensive paleontological sample of the entire Indo-West Pacific core of marine molluscan diversity (25), found a total of only ≈1,600 late Miocene specimens and 4,500 Pliocene specimens in this vast area (D.J., unpublished work). These numbers are dwarfed by the sample sizes available from many individual and regional studies in temperate latitudes (e.g., refs. 26–30). Owing to these biases in sampling intensity, tropical Indo-Pacific taxa almost certainly have a significantly lower probability of being recorded as fossils than do taxa from higher latitudes (20–22).

Geographic distribution may also affect the capture of taxa as fossils, which may be a positive function of geographic range. This effect cannot be fully assessed until a complete genus-level biogeography is available for the fossilized taxa, but our preliminary analysis supports this possibility. Of the 255 missing shelf-depth taxa for which we have distributional data, 50%

occur in only one biogeographic province (using the scheme in ref. 25, and dividing the huge Indo-West Pacific region into a west Indian Ocean, an east Indian Ocean-West Pacific, and a central Pacific section). In contrast, only 13% of the 540 bivalve taxa having a fossil record inventoried by Flessa and Jablonski (31) are restricted to one province. The Flessa–Jablonski data set did not include the protobranchs, anomalodesmatans, and several heteroconch groups, so these results should be viewed cautiously, but the results strongly suggest a geographic-range effect.

Taxonomic Lag. Some bivalve genera are missing from fossil compendia simply because the systematics of fossil taxa have lagged behind new discoveries and taxonomic revisions in the living fauna. For example, Jablonski *et al.* (3) added 144 bivalve taxa (and deleted 27) to Sepkoski's (32) seminal compilation of genera with fossil records; 83% of the newly added taxa had been described before the publication of the treatise (33) that formed the starting point for Sepkoski's compendium. These taxa were added to the fossil lists simply by shifting fossil records of species into their appropriate, more narrowly defined, genera and subgenera.

Further taxonomic standardization is clearly needed. Of the 151 marine bivalve genera and subgenera named since the treatise (33) was published, 88 (58%) are not yet known as fossils. We expect that a large fraction of those missing taxa will prove to have fossil occurrences (and broader geographic ranges) once both older and newly described species are placed within a more refined taxonomy. However, 31% of the new taxa are known only in deep-sea settings, and some of these may never be recorded as fossils.

Interactions Among Variables. The biases described above do not influence the record independently. Instead, many of these biases are strongly correlated; for example, commensal and parasitic taxa tend to have small body sizes, thin shells, and tropical distributions, and their systematics is poorly known. All of these effects contribute to the poor fossil record of these taxa. Thus, a better understanding of these interactions using a multivariate analysis of the variables discussed here would clearly be valuable. Unfortunately, such an analysis is not feasible at this time because comprehensive data on one of the key variables, body size, is not yet available for all bivalve genera and subgenera.

Nevertheless, we can gain some insights into the effects of such interactions by partitioning the missing taxa to exclude small-bodied (<1 cm) and deep-sea taxa, thus emulating the great bulk of the macrofossil record (see Table 4). We then find no significant differences between missing and fossil taxa in life habits and mineralogy. Significant differences remain with respect to shell organic content and SAV, and the effects of organic content become slightly stronger (e.g., moderate and low-organic taxa are 85% of the missing taxa but 93% of the fossil taxa; viewed another way, 77% of the high-organic taxa and 89% of the moderate- and low-organic taxa are known as fossils). However, date of taxonomic description and geographic range have the strongest effect, indicating the pressing need for taxonomic standardization of living and fossil biotas.

Discussion

The fossil record, although rich, is obviously neither complete nor unbiased. Furthermore, secular changes in the global marine environment might favor or disfavor the preservation of taxa in certain categories (ref. 34, but see refs. 4 and 23), thereby imposing long-term shifts in the nature of the biases in the record. Our data are clearly most applicable to the modern evolutionary fauna and are likely best suited to quantitative evaluation only of Cenozoic mollusks. Nevertheless, our data can serve as a foundation, the details of which may be modified, for understanding preservational conditions in other eras and other taxa.

The fossil record is affected by both preservation and sampling. Some missing taxa may truly not have been preserved anywhere, whereas others may exist as fossils but have not yet been discovered and/or properly identified by paleontologists, owing to such factors as rarity within communities, rarity of habitats, spatial distribution relative to well studied fossil biotas, and taxonomic difficulties. Some particularly vulnerable groups, such as small-bodied forms, may be affected by both preservation and sampling problems.

Further data compilation and analysis may eventually permit us to partition the missing taxa of all phyla more precisely among the responsible factors. However, the main aim here is to identify key variables that affect our ability to analyze macroevolutionary and macroecological patterns in the fossil record and estimate their magnitudes within the modern evolutionary fauna. Thus, even the question of whether small taxa are missing owing to preservation failure or to sampling failure (that is, whether these are structural zeros or sampling zeros) is secondary to the fact that small-bodied clades will be less reliable even for large-scale analyses, although sampling zeros may he filled in once they are better understood (see ref. 30). Restriction to the deep sea and high-SAV microstructures are other factors that strongly weigh against discovery of taxa as fossils, although these factors covary with body size, and all effects are accentuated by taxonomic lag. On the other hand, a number of factors previously suspected to impose important biases (epifaunal habitat, aragonitic mineralogy, and high-organic microstructures) show minor or undetectable effects (Table 2).

The pervasive bias against small size probably derives from positive interactions among several factors: many small bivalves have thin shells and should thus be more easily fragmented, dissolved, or macerated after death (35); they may commonly elude collection, passing through screens or being easily destroyed during preparation; and many pose especially difficult identification problems [many minute bivalves are paedomorphic (36–38) and tend to resemble juveniles of their larger relatives, which are themselves difficult to identify from immature shells]. Thus, although small bivalves represent a significant fraction of taxa, they are likely to be poorly known in both living and fossil assemblages. This bias is probably true for most macrofossil groups: the number of small-bodied taxa, both past and present, is almost certainly underestimated. There is also a weak but significant inverse relation between body size and date of description ($P < 0.05$; Spearman's $r = -0.11$; $n = 294$), as has been noted for a few animal groups (39). Multiple regression indicates that neither size nor date is independent of geographic range, although again effects are small (size and range together explaining only 9% of the variation in date of description).

These results are broadly consistent with those of previous comparisons of living and fossil bivalve faunas. Valentine (6) analyzed Californian province bivalves and found effects of both body size and depth on the taxa missing from the Pleistocene, and emphasized among-province variation in preservation quality as an important determinant of the quality of the fossil record, as we have found here. Using different approaches, Harper (7) and Jablonski *et al.* (3) found that deep-sea and aragonitic taxa had poorer fossil representation, and Jablonski *et al.* (3) and Cooper *et al.* (30) found evidence for bias against small body sizes. Harper (7) also found, as we have, no difference between epifaunal and infaunal taxa in fossil representation. However, Kidwell's (4) analysis of the stratigraphic ranges of fossil bivalves found a far weaker effect of shell composition than we have. In particular, the geologic durations of high-SAV taxa are equal to or greater than those for low-SAV taxa. Because this is opposite to the preservational effects documented here, it is

likely to involve a biological signal, probably along phylogenetic lines, and invites further study.

Remedies. The analyses presented here are intended to promote the development of more rigorous protocols for bringing fossil data to bear on a range of paleobiological and evolutionary questions. At least three approaches can be used to minimize the effects of the biases identified here.

Data partitioning. Analyses such as those presented here can be used to identify the most robust subset(s) of the data, and to omit or down-weight the least robust and dependable subset(s). For example, in using fossils to infer ancestral character states, phylogenetic information could be weighted according to the clade-level and/or functional-group data presented here. Fortey *et al.* (40) suggested that phylogenies wherein 70% or more of the species are known as fossils should be regarded as representing good records. A similar approach might be used for higher taxonomic levels such as genera, with the cutoff value depending on the nature of the question and the temporal resolution required. Clearly, clades with the highest proportions of fossilized taxa are best suited for placing confidence intervals on evolutionary events used to calibrate molecular clocks. Thus, our analyses suggest that Peterson *et al.* (41) made a prudent choice when they used mytiloids and nuculoids to calibrate a protostome molecular clock.

Null models. The direction and magnitude of known biases could be explicitly integrated into null hypotheses for paleobiological analyses; that is, hypotheses could be framed to illuminate effects that are contrary in direction or magnitude to the observed bias. For example, our results suggest the null expectation that geologic ranges of high-SAV taxa should be truncated relative to low-SAV forms. However, high-SAV taxa have been shown to have durations equal to or greater than those for low-SAV taxa (4), indicating that the signal from the durational data is unlikely to be preservational but arises from some other, probably biological, source. Similarly, contrary to the expectations based on the preservational biases documented here, epifaunal suspension-feeding bivalve taxa had significantly shorter recorded durations than infaunal suspension feeders during the Mesozoic (42).

Standing richness estimates. Recorded fossil diversity is the major source of data when modeling ancient diversity for a given time, place and taxon, including global diversity. However, missing taxa form a nontrivial fraction of fossil faunas, about one-quarter in our data, which may represent a best case, because we deal here with living taxa and therefore the more recent parts of the record. Thus, to ignore them entirely would significantly misrepresent the true numbers of ancient taxa. There can be no simple formula for dealing with this problem, because many factors that cause taxa to be missing will vary in time and space, so that calculations of missing taxa would have to be tailored to some context. At the very least, as Cooper *et al.* (30) note, a concerted effort to retrieve small-bodied taxa from the fossil record would in itself significantly improve the accuracy of diversity and stratigraphic-range estimates.

Although our study focused exclusively on marine bivalve taxa, our results are likely to be relevant to benthic marine macroinvertebrates in general. The other dominant clade in Sepkoski's (2) modern fauna, Gastropoda, is at least twice as diverse as Bivalvia but likely suffers from similar biases, judging from the few cases where gastropods and bivalves have been studied in concert (e.g., refs. 6 and 30). Analyses such as this are needed for other major constituents of the fossil record to fully understand the biases and limitations of paleobiological data.

Methods

Our inventory of marine bivalve taxa with fossil records is an updated version of Sepkoski's (32) *Compendium of Marine Animal Genera* (posted at www.pbdb.org). Information about bivalve taxa lacking a fossil record were compiled from a few major sources (e.g., refs. 31, 33, 43, and 44) and the primary literature, chiefly through searches of the on-line version of the *Zoological Record* (http://scientific.thomson.com/products/zr; last accessed January 2006). Our use of the generic and subgeneric levels minimizes the effects of sampling and of taxonomic instability at the species level, while including subgenera maximizes the number and proportion of missing taxa and thus is a conservative protocol. For each missing taxon, we compiled data on shell size (geometric mean of length and height; following ref. 8 and references cited therein), bathymetric range, life habit (feeding type and living position), geographic distribution (in bins modified from provinces of ref. 25), and shell mineralogy, organic content, and microstructural surface area (following ref. 4). Geographic distributions were drawn from the taxonomically standardized global database of Flessa and Jablonski (31) augmented by published sources for taxa not yet included there.

Because we are dealing with taxa above the species level, we rely on the observation that bivalve life habits are closely reflected in shell morphology and are highly conserved at the generic and familial levels, as is shell composition (we followed ref. 4 in dealing with families having a range of shell compositions). In quantifying body size, we could not survey all living species (which will require a multiyear project by a large consortium) but instead used the first few species from each taxon that we encountered in our literature searches to characterize a taxon; in this sense, they constitute a random sample. Furthermore, because most of the missing taxa are species-poor, our data set includes a large fraction of their species, and thus our sample is unlikely to contain major biases. Because bathymetric distributions of many bivalve species are still poorly known, we simply partitioned the missing taxa into two categories, those that are represented on the continental shelf (taxa with any species found living above 200 m) and those that occur exclusively in the deep sea (i.e., below 200 m).

To restrict our analysis exclusively or primarily to marine taxa, we excluded the Corbiculidae, Dreissenidae, Pisidiidae, Unionidae, and a few other brackish-water and freshwater taxa scattered through predominantly marine families. The Teredinidae, wood-dwelling wormlike forms, were excluded from the size analyses because their body sizes are only weakly related to the dimensions of their small, highly modified shells, and from geographic analyses because their distributions have been drastically altered by human activities (45).

We thank K. Amano, L. C. Anderson, A. G. Beu, J. S. Crampton, E. V. Coan, T. A. Darragh, H. H. Dijkstra, the late K. Lamprell, P. A. Maxwell, P. M. Mikkelsen, P. Middelfart, N. J. Morris, G. Paulay, F. Scarabino, J. A. Schneider, J. D. Taylor, J. D. Todd, T. R. Waller, A. Warén, and T. E. Yancy for generously sharing their expertise on bivalve systematics, ecology, and shell structure; and P. S. Anderson, E. Hunt, R. M. Price, T. A. Rothfus, R. J. Rundell, and H. J. Sims for assistance in data entry and the occasional bibliographic emergency. The manuscript was kindly reviewed by E. M. Harper and J. S. Crampton. Early stages of this work were supported by the National Science Foundation, and later stages were supported by the National Aeronautics and Space Administration.

1. Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection* (Murray, London).
2. Sepkoski, J. J., Jr. (1981) *Paleobiology* **7,** 36–53.
3. Jablonski, D., Roy, K., Valentine, J. W., Price, R. M. & Anderson, P. S. (2003) *Science* **300,** 1133–1135.
4. Kidwell, S. M. (2005) *Science* **307,** 914–917.
5. Schopf, T. J. M. (1979) *Paleobiology* **4,** 261–270.
6. Valentine, J. W. (1989) *Paleobiology* **15,** 83–94.
7. Harper, E. M. (1998) in *The Adequacy of the Fossil Record*, eds. Donovan, S. K. & Paul, C. R. C. (Wiley, Chichester, U.K.), pp. 243–267.

8. Roy, K., Jablonski, D. & Martien, K. K. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13150–13155.
9. Dell, R. K. (1990) *Bull. R. Soc. N.Z.* **27,** 1–311.
10. Glover, C. P. & Kidwell, S. M. (1993) *J. Geol.* **101,** 729–747.
11. Simon, A., Poulicek, M., Velmirov, B. & MacKenzie, F. T. (1994) *Biogeochemistry* **25,** 167–195.
12. Sanders, D. (2003) *J. Afr. Earth Sci.* **36,** 99–134.
13. Zardus, J. D. (2002) *Adv. Mar. Biol.* **42,** 1–65.
14. Levin, L. A., Etter, R. J., Rex, M. A., Gooday, A. J., Smith, C. R., Pineda, J., Stuart, C. T., Hessler, R. R. & Pawson, D. (2001) *Annu. Rev. Ecol. Syst.* **32,** 51–93.
15. Boss, K. J. (1965) *Malacologia* **3,** 183–195.
16. Morton, B. & Scott, P. H. (1989) *Asian Mar. Biol.* **6,** 129–160.
17. Mikkelsen, P. M. & Bieler, R. (1992) *Malacologia* **34,** 1–24.
18. Crame, J. A. (2000) *Paleobiology* **26,** 188–214.
19. Johnson, D. (2004) *The Geology of Australia* (Cambridge Univ. Press, Cambridge, U.K.).
20. Jablonski, D. (1993) *Nature* **364,** 142–144.
21. Jackson, J. B. & Johnson, K. G. (2001) *Science* **293,** 2401–2404.
22. Johnson, K. G. (2001) *Paleobiology* **29,** 19–21.
23. Bush, A. M. & Bambach, R. K. (2004) *J. Geol.* **112,** 615–642.
24. Beu, A. G. (2005) *Scr. Geol.* **130,** 1–186.
25. Valentine, J. W. (1973) *Evolutionary Paleoecology of the Marine Biosphere* (Prentice–Hall, Englewood Cliffs, NJ).
26. Kidwell, S. M. (1986) *Palaios* **1,** 239–255.
27. Ferrero, E. & Merlino, B. (1992) *Boll. Malacol.* **28,** 101–138.
28. Kowalewski, M., Gürs, K., Nebelsick, J. H., Oschmann, W., Piller, W. E. & Hoffmeister, A. P. (2002) *Geol. Soc. Am. Bull.* **114,** 239–256.
29. Crampton, J. S., Beu, A. G., Cooper, R. A., Jones, C. M., Marshall, B. & Maxwell, P. A. (2003) *Science* **301,** 358–360.
30. Cooper, R. A., Maxwell, P. A., Crampton, J. S., Beu, A. G., Jones, C. M. & Marshall, B. A. (2006) *Geology* **34,** 241–244.
31. Flessa, K. W. & Jablonski, D. (1996) in *Evolutionary Paleobiology*, eds. Jablonski, D., Erwin, D. H. & Lipps, J. H. (Univ. of Chicago Press, Chicago), pp. 376–397.
32. Sepkoski, J. J., Jr. (2002) *Bull. Am. Paleontol.* **363,** 1–560.
33. Moore, R. C., ed. (1969–1971) *Treatise on Invertebrate Paleontology, Part N, Mollusca 6* (Geol. Soc. of Am. and Univ. of Kansas, Lawrence).
34. Stanley, S. M. & Hardie, L. A. (1998) *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **144,** 3–19.
35. Kidwell, S. M. & Bosence, D. W. J. (1991) in *Taphonomy: Releasing the Data Locked in the Fossil Record*, eds. Allison, P. A. & Briggs, D. E. G. (Plenum, New York), pp. 115–209.
36. Ockelmann, K. W. (1964) *Ophelia* **1,** 121–146.
37. Gofas, S. & Salas, C. (1996) *J. Conch. (London)* **35,** 427–435.
38. Middelfart, P. (2002) *Zootaxa* **112,** 1–124.
39. Collen, B., Purvis, A. & Gittleman, J. L. (2004) *Global Ecol. Biogeogr.* **13,** 459–467.
40. Fortey, R. A. & Jefferies, R. P. S. (1982) in *Problems of Phylogenetic Reconstruction*, eds. Joysey, K. A. & Friday, A. E. (Academic, London), pp. 197–234.
41. Peterson, K. J., Lyons, J. G., Nowak, K. S., Takacs, C. M., Wargo, M. J. & McPeek, M. A. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 6536–6541.
42. Jablonski, D. (2005) *Paleobiology* **32,** Suppl., 192–210.
43. Millard, V. (2001) *A Classification of Worldwide Mollusca* (Millard, 22 Rhine Road, Cape Town 8050, South Africa), 2nd Ed.
44. Taylor, J. D., Kennedy, W. J. & Hall, A. (1969) *Bull. Br. Mus. Nat. Hist. Zool. Suppl.* **3,** 1–125.
45. Turner, R. D. (1966) *A Survey and Illustrated Catalogue of the Teredinidae (Mollusca: Bivalvia)* (Harvard Univ. Press, Cambridge, MA).
46. Giribet, G. & Wheeler, W. (2002) *Invertebr. Biol.* **121,** 271–324.
47. Sokal, R. R. & Rohlf, F. J. (1995) *Biometry* (Freeman, San Francisco), 3rd Ed.