The effects of taxonomic standardization on sampling-standardized estimates of historical diversity

Peter J. Wagner¹,*, Martin Aberhan², Austin Hendy³ and Wolfgang Kiessling²

¹Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, IL 60615, USA
²Invalidenstraße 43, Museum für Naturkunde, Humboldt University, 10115 Berlin, Germany
³Department of Geology, University of Cincinnati, 500 Geology/Physics Building, Cincinnati, OH 45221, USA

Occurrence-based databases such as the Palaeobiology database (PBDB) provide means of accommodating the heterogeneities of the fossil record when evaluating historical diversity patterns. Although palaeontologists have given ample attention to the effects of taxonomic practice on diversity patterns derived from synoptic databases (those using first and last appearances of taxa), workers have not examined the effects of taxonomic error on occurrence-based diversity studies. Here, we contrast diversity patterns and diversity dynamics between raw data and taxonomically vetted data in the PBDB to evaluate the effects of taxonomic errors. We examine three groups: Palaeozoic gastropods, Jurassic bivalves and Cenozoic bivalves. We contrast genus-level diversity patterns based on: (i) all occurrences assigned to a genus (i.e. both species records and records identifying only the genus), (ii) only occurrences for which a species is identified, and (iii) only occurrences for which a species is identified, but after vetting the genus to which the species is assigned.

Extensive generic reassignments elevate origination and extinction rates within Palaeozoic gastropods and origination rates within Cenozoic bivalves. However, vetting increases generic richness markedly only for Cenozoic bivalves, and even then the increase is less than 10%. Moreover, the patterns of standing generic richness are highly similar under all three data treatments. Unless our results are unusual, taxonomic standardization can elevate diversity dynamics in some cases, but it will not greatly change inferred richness over time.

Keywords: global biodiversity; taxonomy; sampling-standardization; gastropods; bivalves

1. INTRODUCTION

Many studies examine how flawed taxonomy might compromise perceived historical biodiversity patterns using both empirical (Smith & Patterson 1988; Wagner 1995; Adrain & Westrop 2000; Ausich & Peters 2005) and simulation approaches (e.g. Sepkoski & Kendrick 1993; Robeck et al. 2000). These studies focus on synoptic databases of first and last appearances (e.g. Sepkoski 1982, 2001). Recently, workers have begun to re-evaluate historical biodiversity patterns using occurrence-based databases that attempt to record all finds for fossil taxa. Such databases allow sampling standardization over time and/or space in ways that are not feasible for synoptic databases (e.g. Miller & Foote 1996; Alroy 1996; Alroy et al. 2001). However, occurrence-based databases frequently rely on published fossil lists that are many years old. These lists often have outdated generic taxonomy, owing to either subsequent generic revision or inexpert systematic knowledge by the lists’ compilers (Smith 2003; Forey et al. 2004). The same species might be assigned to different genera on different lists, which introduces source of possible taxonomic error not present in synoptic studies.

We assess the effects of taxonomic standardization on analyses of marine invertebrate data from the Paleobiology Database (PBDB; http://paleodb.org; see Alroy et al. 2001). The PBDB is particularly interesting because: (i) it is by far the most comprehensive palaeontological database available to the community with over 600 000 generic occurrences, (ii) the vast majority of its occurrences use generic assignments from the original literature, (iii) it includes many faunal lists over 20 years old and many identifications by non-specialists, (iv) although it uses taxonomy tables to dynamically update generic assignments, those tables have information for only a small proportion of species with occurrence records, (v) it assigns a large percentage of occurrences to a small percentage of genera, which is consistent with many assignments to ‘wastebasket’ taxa (Plotnick & Wagner 2006), and (vi) initial analyses of this database suggest historical diversity patterns very different from those suggested by synoptic databases, with the PBDB implying fairly constant generic richness over time (Alroy et al. 2001) and synoptic data implying a twofold increase in generic richness in the Cenozoic (Sepkoski 1997). Since synoptic ranges typically come from taxonomic specialists (Sepkoski 2002), outdated generic taxonomy might be
partly responsible for the different historical patterns implied by synoptic databases and the PBDB. We use Middle Ordovician–Middle Carboniferous gastropods, Jurassic bivalves and Cenozoic bivalves to assess whether and (if so) how taxonomic revision alters sampling-standardized diversity patterns and to find possible general effects of taxonomic standardization.

2. MATERIAL AND METHODS
Since most analyses of historical diversity focus on numbers of genera and subgenera through time, we focus on the generic/subgeneric assignments of species. (Following studies such as Sepkoski (1997), we treat genera and subgenera as being of equal rank). We impose some species-level synonymizations, which occasionally reduce two or more generic occurrences to one generic occurrence when a species is listed twice under two names on the same list. In part, we preview what the PBDB will show in the future, at least for the records used here. Since the PBDB reflects especially the easily obtained published literature, we also evaluate literal readings of the prominent marine macroinvertebrate palaeontology literature.

We examine Ordovician–Carboniferous gastropods, Jurassic bivalves and Cenozoic bivalves owing to the authors’ familiarity with these taxa. We have prior reason to think that the taxonomic quality of PBDB data varies among these taxa. Less than 20% of Palaeozoic gastropod and Cenozoic bivalve records were published after 1994 (see Fig. S1 of the electronic supplementary material), and there has been substantial taxonomic revision for both groups in that time. In contrast, 54% of Jurassic bivalve records are from post-1994 papers and 39% reflect the work of two current researchers (F. Fürsich & M. Aberhan), whose taxonomic expertise we followed in the revisions.

(a) Taxonomic assignment and reassignment
True ‘taxonomic standardization’ is impossible because there are no universally accepted criteria for delimiting genera. Here, we assign species to ‘correct’ genera based on recently published opinions and also based on our own expertise and opinions (see Appendix of the electronic supplementary material). Like the taxonomic opinions of any worker, our assignments are also imperfect. However, they impose uniform and current standards on data compiled over several decades and almost certainly improve many obvious imperfections.

We distinguish between two types of occurrences: species records, where a species is identified (e.g. *Bellerophon leda*), and genus-only records (e.g. *Bellerophon sp*.). All records include both types. Species records account for 4974 of 7659 Palaeozoic gastropod occurrences, 10 814 of 13 813 Jurassic bivalve records and 9966 of 13 060 Cenozoic bivalve records. Most analyses of PBDB data (e.g. Alroy et al. 2001) use all records. However, taxonomy tables modify the genus assignment only of species records unless an entire genus is synonymized with another. Therefore, we contrast three different treatments of the data: (i) all records, (ii) unvetted species records, which use the generic assignment as entered into the PBDB, and (iii) vetted species records, which use the generic assignment considered to be ‘correct’ by one of us (e.g. *Retispira* for *B. leda*). In some cases, we acknowledge that a genus assignment is actually indeterminate for a species, or we separate species into informal ‘new’ genera.

Our standardization applies only to generic assignments and species synonymies, and thus assumes that initial species identifications are correct. Examining the effects of specimen reidentifications on similar subsets of data is beyond the scope of this study, and a separate study assessing the effects of specimen reidentifications should be conducted.

(b) Contrasting historical diversity patterns
We assess the effects of taxonomic vetting on sampling-standardized diversity patterns by subsampling fossil occurrences (=records). Subsampling of occurrences rather than localities can misrepresent relative diversity among intervals if alpha diversity changes markedly among these intervals (Bush et al. 2004). Also, restricting subsampling to occurrences of a particular taxonomic group will dampen changes in richness because occurrences are units of richness at each locality and different numbers of localities must be sampled to obtain the same number of occurrences if the diversity of a taxon relative to other taxa changes markedly (Miller et al. 1998). However, our purpose is not to produce definitive statements about diversity patterns within these taxa. Instead, we want to assess how taxonomy affects the implications of the PBDB (and literal readings of the literature) given the methods used in PBDB data.

We use the average number of genera and subgenera subsampled in each interval (sampled-in-bin or SIB) to estimate relative generic richness from interval to interval. However, estimates of origination and extinction rates require stratigraphic ranges within the subsampled data. Therefore, every replication standardizes sampling in each interval and then reconstructs stratigraphic ranges for each genus. We then estimate origination rate as the average number of genera first appearing divided by the average number of genera surviving from the prior interval. We estimate extinction rate as the average number of genera last appearing divided by the number of genera extant in the interval. To ameliorate ‘monographic’ effects (Raup & Boyajian 1988), we exclude taxa known from only a single interval from both rate metrics (Foote 2000). This precludes origination and extinction rates for the first and last intervals.

For Palaeozoic gastropods, we use temporal bins of ca 10 million years each (see Fig. S2A of the electronic supplementary material). For Jurassic and Cenozoic bivalves, we use a finer scale (stages; see Fig. S2B–C of the electronic supplementary material) because the power of our tests increases with the number of time-intervals and the data are rich enough to permit this. For the sake of larger subsample sizes, we lump the Devonian 4 and Devonian 5 bins (Frasnian and Famennian) for Palaeozoic gastropods and the Aalenian and Bajocian stages for Jurassic bivalves.

We use Spearman’s rank correlation among first differences (changes in diversity from one interval to the next, e.g. Gould & Callaway 1980) to assess whether taxonomy affects the shape of historical diversity patterns. We then use the same test to examine correlations in the origination and extinction rates among the data treatments.

A second issue is whether vetting alters the scale of the diversity patterns. This is particularly important if we are to assess whether taxonomic vetting could account for the differences between synoptic and sampling-standardized depictions of Cenozoic diversity relative to pre-Cenozoic diversity. We use t-tests to assess whether subsampled generic richness and subsampled rates differ significantly among different data treatments. We do not use a non-parametric analogue of the t-test because the scales of the differences are important, not just the rank orders of the differences.
Table 1. Initial PBDB data and effects of vetting. (‘PBDB genera’ is the number of genera with occurrences in the raw data. ‘Genera with no unvetted or vetted species’ and ‘genera with unvetted species but no vetted species’ both are effectively removed from the final analysis. ‘Genera’ is the number of genera with occurrences in the vetted data.)

<table>
<thead>
<tr>
<th></th>
<th>PBDB genera</th>
<th>genera with no unvetted or vetted species</th>
<th>genera with unvetted species but no vetted species</th>
<th>genera added by vetting</th>
<th>net change in genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordovician–Carboniferous gastropods</td>
<td>521</td>
<td>34</td>
<td>88</td>
<td>59</td>
<td>-63</td>
</tr>
<tr>
<td>Jurassic bivalves</td>
<td>275</td>
<td>33</td>
<td>40</td>
<td>38</td>
<td>-35</td>
</tr>
<tr>
<td>Cenozoic bivalves</td>
<td>557</td>
<td>79</td>
<td>47</td>
<td>176</td>
<td>50</td>
</tr>
</tbody>
</table>

Figure 1. Standing generic richness given the raw data for: (a) Middle Ordovician–Middle Carboniferous gastropods, (b) Jurassic bivalves, and (c) Cenozoic bivalves. Note that this counts only genera actually sampled within each bin rather than all of the taxa implied to be present from first/last appearance data. Time-scales here and elsewhere based on Gradstein et al. (2005).

We plot richness on a log 2 scale here and elsewhere to emphasize whether vetting can double apparent generic richness in some cases, as is necessary if flawed taxonomy is responsible for the failure of occurrence-based studies to show a doubling in generic richness during the Meso-Cenozoic.

3. RESULTS

(a) Effects on data structure

Taxonomic vetting reassigned 34.9% (1803 of 4974) of Palaeozoic gastropod species records, 38.2% (3822 of 9996) of Cenozoic bivalve species records and 19.1% (2068 of 10 814) of Jurassic bivalve species records (see table 1 and also Fig. S3 of the electronic supplementary material). Reassignments result in Palaeozoic gastropods and Jurassic bivalves ‘losing’ genera and Cenozoic bivalves ‘gaining’ genera (table 1). All three groups have numerous genera solely known from genus-only records both before and after vetting (‘genera with no unvetted or vetted species’ in table 1). Overall, vetting reduces the total generic richness for Palaeozoic gastropods and Jurassic bivalves (‘net change’ in table 1), but increases the total generic richness for Cenozoic bivalves.

Prior to sampling standardization, we must sample at least as many genera from all records (open symbols in figure 1) as we do from unvetted species records (grey symbols) because the latter is a subset of the former. Only for Cenozoic bivalves do vetted species records (black symbols) introduce enough new genera to yield higher richness than do all records.

Unsurprisingly, the genera that lose the most records are the commonly occurring genera (see Fig. S4 of the electronic supplementary material). However, the genera that gain occurrences are not always the rare genera, as moderately common genera also ‘profit’ from vetting. Regardless, the redistribution of records typically increases the evenness of generic occurrences distributions (see Fig. S5 of the electronic supplementary material). Finally, generic stratigraphic ranges given vetted species records tend to be longer than the ranges given unvetted species records, but shorter than ranges given all records (see Fig. S6 of the electronic supplementary material).

(b) Effects on implied historical diversity

Subsampling all records (open symbols in figure 2) yields greater SIB generic richness than does subsampling unvetted species records (grey symbols) for all intervals for Jurassic and Cenozoic bivalves, but only for 7 out of 12 intervals (with two ties) for Palaeozoic gastropods. Subsampled vetted species records (black symbols) yield greater SIB richness than do subsampled unvetted species records in 8 out of 9 intervals for Jurassic bivalves and 11 out of 11 intervals for Cenozoic bivalves, but in only 6 out of 13 intervals for Palaeozoic gastropods (with three ties). Finally, subsampled vetted species records consistently yield higher SIB richness than do all records in 8 out of 12 intervals for Cenozoic bivalves and 7 out of 13 intervals for Palaeozoic gastropods, but in zero intervals for Jurassic bivalves.

Subsampled SIB curves show significant correlations for all three examples (table S1 of the electronic supplementary material). Palaeozoic gastropods show no
significant differences in the scale of subsampled SIB richness. Jurassic bivalves show significantly greater subsampled SIB richness given vetted species than given unvetted species records, and significantly greater SIB richness given all records than given vetted species records. Finally, Cenozoic bivalves show significantly greater SIB richness given all records than given unvetted species records, but significantly greater SIB richness given vetted species records than given all records.

Subsampled origination rates ($\lambda$; figure 3 and table S2 of the electronic supplementary material) and extinction rates ($\mu$; figure 4 and table S3 of the electronic supplementary material) show significant correlations over time among all three data treatments in all three datasets with two exceptions: $\lambda$ among all records and unvetted species records for Jurassic bivalves, and $\mu$ among unvetted and vetted Palaeozoic gastropod species records. However, taxonomic vetting raises both origination and extinction rates significantly for Palaeozoic gastropods relative to either all records or species records. Vetting raises origination rates for Cenozoic bivalves but has no significant effect on turnover rates for Jurassic bivalves.

4. DISCUSSION

(a) Implications for historical diversity patterns

Extensive taxonomic revision alters subsampled richness in PBDB data notably for only one of the three datasets examined here, Cenozoic bivalves. Superficially, this appears to corroborate the idea that taxonomy might account for the differences between synoptic and sampling-standardized diversity patterns (e.g. Sepkoski 1997 versus Alroy et al. 2001). However, this actually requires evidence that taxonomic revision can double subsampled richness in some cases. The extensive revision of Cenozoic genus and subgenus assignments falls far short of this, and typically increases subsampled richness by an average of 2.3 times that implied by current assignments. Unless prominent Cenozoic taxa are very different from any of the three taxa analysed here, a ‘Cenozoic wastebasket’ hypothesis is not a probable explanation for the absence of a major Cenozoic diversification after sampling standardization.
Although taxonomic standardization does not affect apparent turnover rates for Jurassic bivalves, it does affect turnover rates for both Palaeozoic gastropods and Cenozoic bivalves. For Palaeozoic gastropods, the difference reflects the redefinition of polyphyletic taxa, which both shortens apparent ranges and adds originations and extinctions (see also Nützel 2005). Vetted rates now emphasize events the redefinition of polyphyletic taxa, which both shortens apparent turnover rates for Jurassic bivalves, it does affect the species richness of genera such as Palaeozoic gastropods and Cenozoic bivalves. For Palaeozoic gastropods, the difference reflects the introduction of enough ‘new’ taxa (especially subgenera for Neogene species) to elevate net richness. This necessitates elevated origination rates without requiring elevated extinction rates. Anderson’s (1996) extensive synonymizations of the highly speciose genera Bellerophon and Murchisonia by one of us (P.J.W.) show that phylogenetic structure is fairly weak within these genera relative to contemporaneous clades with comparable numbers of species. Sophisticated phylogenetic methods might tease out phylogenetic signal among these species, but any new genera stemming from such analyses will not be easy for non-experts to identify. We expect similar results in any future phylogenetic analyses that include species-rich, long-lived bivalve genera such as Anadara, Chlamys, Nuculana, Ostrea or Tellina.

Systematic revisions also can alter the numbers of species and records within genera by synonymizing species. For example, Johnson’s (1984) revision of European Jurassicpectinid bivalves markedly reduced the species richness of genera such as Camptonectes, Chlamys, Entolium and Eocene through synonymization rather than reassignment. Anderson’s (1996) extensive revision of Cenozoic corbulids greatly reduced the species richness of Corbula in the same way. A comparable revision is not available for other species-rich groups such as Jurassic limids and some anomalodesmatans. As a result, Jurassic genera such as Plagiostoma, Pholadomya and Pleuronyx, and Cenozoic genera such as Anadara, Barbatia and Crassatella retain high species numbers after vetting. Although we are confident that our generic assignments are appropriate, we expect richness within these genera to drop with future species-level revisions that lump species that we have kept separate.

(c) Generic diversity: whether to use species records or all genus records

Unless one synonymizes a genus, taxonomy tables can alter generic assignments only for species records. This alone is a reason to infer generic richness patterns from species records rather than from all genus records. Palaeozoic gastropods also suggest that using all generic records instead of only species records can dampen turnover rates. For example, the end-Ordovician and Late Devonian mass extinctions are not obvious when using all genus records. However, both are observable when generic diversity is derived from species characters reliably diagnosing genera simply by chance (Plotnick & Wagner 2006). Preliminary phylogenetic analyses of the highly speciose genera Bellerophon and Murchisonia by one of us (P.J.W.) show that phylogenetic structure is fairly weak within these genera relative to contemporaneous clades with comparable numbers of species. Sophisticated phylogenetic methods might tease out phylogenetic signal among these species, but any new genera stemming from such analyses will not be easy for non-experts to identify. We expect similar results in any future phylogenetic analyses that include species-rich, long-lived bivalve genera such as Anadara, Chlamys, Nuculana, Ostrea or Tellina.

The nature of speciose genera

Vetting greatly reduces both the number of highly speciose genera and the richness of these genera (see Appendix and Fig. 57 of the electronic supplementary material). Nevertheless, all three datasets retain a similar rank order of species-rich genera after vetting, and familiar names such as Bellerophon, Plagiostoma and Anadara remain speciose after vetting. Since character change is probabilistic, it is unavoidable that parts of a phylogeny will conserve characters reliably diagnosing genera simply by chance (Plotnick & Wagner 2006). Preliminary phylogenetic analyses of the highly speciose genera Bellerophon and Murchisonia by one of us (P.J.W.) show that phylogenetic structure is fairly weak within these genera relative to contemporaneous clades with comparable numbers of species. Sophisticated phylogenetic methods might tease out phylogenetic signal among these species, but any new genera stemming from such analyses will not be easy for non-experts to identify. We expect similar results in any future phylogenetic analyses that include species-rich, long-lived bivalve genera such as Anadara, Chlamys, Nuculana, Ostrea or Tellina.

Systematic revisions also can alter the numbers of species and records within genera by synonymizing species. For example, Johnson’s (1984) revision of European Jurassicpectinid bivalves markedly reduced the species richness of genera such as Camptonectes, Chlamys, Entolium and Eocene through synonymization rather than reassignment. Anderson’s (1996) extensive revision of Cenozoic corbulids greatly reduced the species richness of Corbula in the same way. A comparable revision is not available for other species-rich groups such as Jurassic limids and some anomalodesmatans. As a result, Jurassic genera such as Plagiostoma, Pholadomya and Pleuronyx, and Cenozoic genera such as Anadara, Barbatia and Crassatella retain high species numbers after vetting. Although we are confident that our generic assignments are appropriate, we expect richness within these genera to drop with future species-level revisions that lump species that we have kept separate.
records even prior to taxonomic vetting. This strongly implies that specimens assigned only (and incorrectly) to a genus distort the subsampled ranges of genera. Obviously, assessing the generality of this pattern requires studies of additional taxonomically neglected clades. Moreover, restricting generic diversity patterns to species records is seemingly unnecessary for well-studied groups such as Jurassic bivalves. Nevertheless, palaeontologists might be erring on the side of caution to use only records that identify species, if they are interested in turnover rates instead of standing richness. This might be especially desirable for analyses at fine temporal and/or geographical scales.

5. CONCLUSIONS
Extensive vetting of generic assignments elevates apparent turnover rates for Palaeozoic gastropods and elevates origination rates for Cenozoic bivalves, but it has little effect on turnover rates for Jurassic bivalves. This supports the idea that extensive vetting will elevate apparent turnover rates (especially among taxonomically neglected groups) rather than dampen them. However, extensive vetting of generic assignments has little effect on sampling-standardized patterns of standing generic richness for Palaeozoic gastropods, Jurassic bivalves and Cenozoic bivalves. Even though vetting does elevate subsampled generic richness among Cenozoic bivalves, it does not come close to reconciling synoptic and sampling-standardized results. Unless these results are anomalous, differences between synoptic results and sampling-standardized results probably are owing to other causes such as variable sampling intensity over time and the Pull of the Recent (Raup 1979).

NSF grants EAR-990328 and EAR-0207874 funded P.J.W.’s contributions. A.J.W.H. acknowledges funding support from the NASA Exobiology grant NAG5-13426 to Arnold I. Miller. W.K. was supported by the Volkswagen-Stiftung. This work was conducted as part of the Marine Invertebrate Working Group of the Paleobiology Database, supported by the National Center for Ecological Analysis and Synthesis, and NSF grant DEB-0083983 to C. Marshall, J. Alroy & A. Miller. This is PBDB publication no. 90.

REFERENCES
Supplementary Information for “The Effects of Taxonomic Standardization on Sampling-Standardized Estimates of Historical Diversity”

1. Statistical tables. 
(a) Subsampled richness among data treatments. Table S1 provides the correlations and contrasts among the subsampled richnesses. The correlations are evaluated using Spearman’s rank correlation coefficient, which tests a null hypothesis of random association between ranks. In this case, the ranked variables are the changes in subsampled richness from comparable time units among the three data sets. Thus, significant p-values indicate strong similarity between two data treatments. The contrasts are evaluated using a paired t-test, where the null hypothesis is that the absolute values are drawn from the same normal distribution. Here, those values are the subsampled richness given the different data treatments. Here, significant results indicate that one data treatment persistently has greater subsampled richness than does another.

(b) Subsampled origination rates among data treatments. Table S2 provides the correlations and contrasts among origination rates implied by the three data treatments. The main text describes the calculation of these rates. Here, Spearman’s ρ tests a null hypothesis of random association between ranked origination rates, and significant p-values again denote very similar results among different data treatments. The t-tests assess a null hypothesis of rates being drawn
from the same normal distribution; therefore, significant results indicate that origination rates are significantly higher in one data treatment than in another.

**Table S2.** Contrasts of subsampled estimates of origination rates given all records, species records, and vetted species records. Spearman’s ρ values give correlations (accommodating ties) between rates. The t-statistic summarizes the distribution around zero of the differences in estimated rates. “d.f.” gives the degrees of freedom for evaluating the t-statistic (i.e., number of appropriate intervals-1). See Fig. 3.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>d.f.</th>
<th>Unv. Species</th>
<th>All vs.</th>
<th>Vett. Species</th>
<th>All vs.</th>
<th>Unv. vs. Vett.</th>
<th>Assigned Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ord.-Carb. Gastropods</td>
<td>10</td>
<td>0.825 0.009</td>
<td>-1.118 0.290</td>
<td>0.719 0.021</td>
<td>-3.911 0.003</td>
<td>0.918 0.004</td>
<td>-5.000 0.001</td>
</tr>
<tr>
<td>Jurassic Bivalves</td>
<td>7</td>
<td>0.690 0.068</td>
<td>-0.050 0.961</td>
<td>0.905 0.017</td>
<td>-1.060 0.324</td>
<td>0.786 0.038</td>
<td>-1.185 0.275</td>
</tr>
<tr>
<td>Cenozoic Bivalves</td>
<td>8</td>
<td>0.967 0.006</td>
<td>-0.732 0.485</td>
<td>0.967 0.006</td>
<td>-3.130 0.014</td>
<td>0.950 0.007</td>
<td>-4.250 0.003</td>
</tr>
</tbody>
</table>

(c) **Subsampled extinction rates among data treatments.** Table S3 provides the correlations and contrasts among extinction rates implied by the three data treatments. The main text describes the calculation of these rates. Here, Spearman’s ρ tests a null hypothesis of random association between ranked extinction rates, and significant p-values again denote very similar results among different data treatments. The t-tests assess a null hypothesis of rates being drawn from the same normal distribution; therefore, significant results indicate that origination rates are significantly higher in one data treatment than in another.

**Table S3.** Contrasts of subsampled estimates of extinction rates given all records, species records, and vetted species records. Spearman’s ρ values give correlations (accommodating ties) between rates. The t-statistic summarizes the distribution around zero of the differences in estimated rates. “d.f.” gives the degrees of freedom for evaluating the t-statistic (i.e., number of appropriate intervals-1). See Fig. 4.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>d.f.</th>
<th>Unv. Species</th>
<th>All vs.</th>
<th>Vett. Species</th>
<th>All vs.</th>
<th>Unv. vs. Vett.</th>
<th>Assigned Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ord.-Carb. Gastropods</td>
<td>10</td>
<td>0.882 0.005</td>
<td>-1.593 0.142</td>
<td>0.627 0.047</td>
<td>-3.987 0.003</td>
<td>0.591 0.062</td>
<td>-3.246 0.009</td>
</tr>
<tr>
<td>Jurassic Bivalves</td>
<td>7</td>
<td>0.905 0.017</td>
<td>-1.329 0.225</td>
<td>0.929 0.014</td>
<td>-0.796 0.452</td>
<td>0.976 0.010</td>
<td>-0.798 0.451</td>
</tr>
<tr>
<td>Cenozoic Bivalves</td>
<td>8</td>
<td>0.900 0.011</td>
<td>-0.605 0.562</td>
<td>0.950 0.007</td>
<td>-1.868 0.099</td>
<td>0.967 0.006</td>
<td>-1.273 0.239</td>
</tr>
</tbody>
</table>
2. General Data Summaries.

(a) *Age of published data.* A large proportion of both records (Fig. S1A-C) and collections (S1D-F) are from the last 25 years for all three data sets. However, Ordovician-Carboniferous gastropods and Cenozoic bivalves show a much greater proportion of both records and collections from papers published prior to 1980. Correspondingly, both Ordovician gastropods and Cenozoic bivalves show a much large proportion of records that are reassigned to different genera (see below).

![Fig. S1.](#) Age of PBDB records (A-C) and collections (D-F) for Ordovician-Carboniferous gastropods (A, D), Jurassic bivalves (B, E), and Cenozoic bivalves (C, F). Records represent the number of occurrences in the collections. White bars for records indicate all records, whereas grey bars reflect only those that identify a species.
(b) Sampling over time. Sampling within each group varies substantially over time for all three studies (Fig. S2). The variance is greatest for Palaeozoic gastropods (Fig. S2A), but both bivalve groups show nearly a ten-fold range in variation for records (Fig. S2B-C). Moreover, “species records” (i.e., occurrences in which a particular species is identified, and thus where the generic assignment can be changed without examining the original material) represent a smaller proportion of “all records” for Palaeozoic gastropods than for either Jurassic or Cenozoic bivalves (see also Table S4).

---

**Fig. S2.** Number of occurrences for: A) Ordovician-Carboniferous gastropods; B) Jurassic bivalves; and, C) Cenozoic bivalves. Open symbols give all occurrences, regardless of whether a species is identified. Grey symbols give only occurrences of identified species.

**Table S4.** PBDB occurrences. “Total records” count all occurrences of genera in the PBDB as of May 2005. “Species records” counts occurrences in which a species is identified. “Reassigned records” counts species records where we assign the species to a different genus than does the PBDB record.

<table>
<thead>
<tr>
<th>Group</th>
<th>Total Records</th>
<th>Species Records</th>
<th>Reassigned Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ord.- Carb. Gastropods</td>
<td>7659</td>
<td>4974</td>
<td>1803</td>
</tr>
<tr>
<td>Jurassic Bivalves</td>
<td>13813</td>
<td>10814</td>
<td>2068</td>
</tr>
<tr>
<td>Cenozoic Bivalves</td>
<td>13060</td>
<td>9996</td>
<td>3822</td>
</tr>
</tbody>
</table>
(c) **Effects of vetting over time.** The proportion of species records that we re-assign to different genera varies substantially over time within each group (Fig. S3). However, whereas only two Jurassic intervals have over 25% of the species records assigned to a different genus (Fig. S3B), only one Palaeozoic (Fig. S3A) and one Cenozoic interval (Fig. S3C) have fewer that 25% of species records assigned to a different genus.

2. **Effects of Vetting on General Distributions.**

(a) **Numbers of occurrences.** Vetting preferentially removes records from commonly occurring genera in all three groups (Fig. S4). This highly significant correlation is almost necessary: only genera with numerous occurrences can lose numerous occurrences. The converse is not necessarily true, as both commonly occurring and rarely occurring genera can gain occurrences. If we restrict the analysis to only genera that gain records, the only Cenozoic bivalves show a negative (and significant) correlation between the original number of records and the number of records gained, indicating that records are non-randomly reassigned to genera with few records. This likely reflects the large number of genera introduced by vetting. The correlation is positive albeit insignificant for Palaeozoic gastropods and Jurassic bivalves.
(b) Evenness of generic occurrences. Ecologists typically use evenness, i.e., some measure of how close a distribution is to uniform, to describe abundances within a community. However, the evenness of records (be they specimens or occurrences) affects the expected sampled richness at a given sample size: that is, at a given richness and sample size, we sample more taxa from an even distribution than from an uneven distribution (Hurlbert 1971). Thus, the effect of vetting on the evenness of generic occurrence distributions is important because it is possible to change the expected sampled richness without introducing or deleting any genera.

There are many evenness metrics, all of which are highly correlated with each other (Peters 2004). We use Pielou’s (1966) $J$, calculated as:

$$J = \sum_{i=1}^{S} f_i \ln(f_i) / \ln(S)$$

where $S$ is the sampled generic richness and $f_i$ is the proportion of total occurrences belonging to genus $i$. 

Fig. S4. Association between the number of species records assigned to genera in the PBDB and the change in assigned records after re-assigning species for Palaeozoic gastropods (A), Jurassic bivalves (B); and, Cenozoic bivalves (C). $\tau$ gives Kendall’s rank correlation coefficient. If one restricts the contrast to genera that gain records, then the correlations are very different: for Palaeozoic gastropods, $\tau=0.030, p=0.534$; for Jurassic bivalves, $\tau=0.067, p=0.290$; for Cenozoic bivalves, $\tau=-0.071, p=0.011$. 

Wagner et al. S6 Effects of Taxonomy Standardization
In general, generic occurrences given species records are more even than are generic occurrences given all records (Fig. S5). Generic occurrences given vetted species record tend to be more even than generic occurrences given unvetted species records. Because increasing evenness increases expected sampled richness (Hurlbert 1971), vetting generic richness can increase the expected number of sampled genera even without introducing new genera.

(c) Stratigraphic ranges of genera. The original “synoptic” stratigraphic ranges of genera necessarily limit the subsampled ranges. As we use these to estimate origination and extinction rates in our analyses, altering these ranges alters our ideas of turnover. Notably, generic ranges given vetted species are significantly lower than are generic ranges given all records for both Ordovician-Carboniferous gastropods (Fig. S6A) and Jurassic bivalves (Fig. S6B). This applies to both numbers of decreases and the average amount of increase. Generic ranges given vetted species are significantly shorter than are generic ranges given unvetted species records for Ordovician-Carboniferous gastropods, although the number of decreases does not differ significantly from the number of increases (Fig. S6D). In contrast, generic ranges given vetted species records show no significant change under either criterion for Jurassic bivalves (Fig.
Cenozoic bivalves show a very different pattern from either, as generic ranges given vetted species are nearly indistinguishable from those given all records (Fig. S6C), and significantly longer than those given unvetted species records (Fig. S6F).

(d) Species richnesses of genera. Just as vetting preferentially removes records from commonly occurring genera, it also removes species from speciose genera in all three groups (Fig. S7). Again, this correlation is almost a necessary one, as the only genera that can lose numerous
species are those that have numerous genera. As with records, the converse is not true, as even speciose genera can gain new species. If the correlation is restricted only to those genera that gain species, then there is a weak but significant and positive correlation for Cenozoic bivalves. To correlation is also positive for Palaeozoic gastropods and Jurassic bivalves, but not significant.
REFERENCES


Ecology 52, 577 - 586.


Pielou, E. C. 1966 Species diversity and pattern diversity in the study of ecological succession. 