

# Diversity biases in terrestrial mammalian assemblages and quantifying the differences between museum collections and published accounts: A case study from the Miocene of Nevada

Edward Byrd Davis <sup>a,\*</sup>, Nicholas D. Pyenson <sup>b</sup>

<sup>a</sup> *University of California Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, Berkeley, CA 94720, USA*

<sup>b</sup> *Department of Integrative Biology and Museum of Paleontology, 1101 Valley Life Sciences Building, University of California, Berkeley, California 94720, USA*

Received 11 January 2007; received in revised form 7 March 2007; accepted 12 March 2007

## Abstract

Understanding diversity through time in the fossil record has primarily relied on the raw count of species within a given time interval, or species richness. These estimates are often derived from published fossil data, and standardized for sample size or geographic area. However, most methods that standardize richness by sample size are sensitive to changes in evenness, which introduces a potential problem with relying on published records: published accounts could be more even than the museum collections from which they are drawn. We address this bias in the context of mammalian paleodiversity, comparing published and museum collections of the Hemphillian Thousand Creek fauna to those of the Barstovian Virgin Valley fauna. We rarified specimen data, both number of identified specimens (NISP) and minimum number of individuals (MNI), and presence/absence data to compare published and museum data within and between faunas. Within faunas, published numbers of specimens are more even than museum samples, but the difference for localities in Virgin Valley is not significant. Neither published nor museum numbers of specimens indicate a significant difference between faunas, but the diversity pattern is reversed between the two data sets. Presence/absence rarefactions show no differences between sources; here, published data adequately sample the underlying museum records. Specimen-based evenness is not accurate in the published sample, and therefore we suggest that future studies of diversity in terrestrial mammalian assemblages must assess unpublished collections. Additionally, NISP data for Thousand Creek are more even than MNI data, suggesting that relying solely on NISP for assessing species diversity can also be misleading. Because publication bias alters richness and evenness, diversity estimates using published data must be circumspect about data sources.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Species diversity; Miocene; Mammalia; Collections; Publications; Biodiversity

## 1. Introduction

Measuring diversity in the fossil record has traditionally relied on species richness, or the raw count of species

for a given time and geographic unit. For regional faunal assemblages in both the terrestrial and marine realm, measuring species richness provides an easy means to assess diversity at that scale, although species richness values are sensitive to sampling intensity, and so larger sample sizes lead to greater richness (Raup, 1975). In comparison, evenness measures diversity by considering the relative abundances of individuals within the taxa of

\* Corresponding author. Fax: +1 510 642 1822.  
E-mail addresses: [daviseb@berkeley.edu](mailto:daviseb@berkeley.edu) (E.B. Davis),  
[pyenson@berkeley.edu](mailto:pyenson@berkeley.edu) (N.D. Pyenson).

an assemblage. Thus, if all taxa have the same number of individuals, the assemblage is considered even, but if one taxon dominates, the assemblage is uneven. Compared with relative or absolute sums of species counts for species richness, evenness is usually calculated by one of a number of complex indices (e.g. [Simpson's \(1949\)](#) index or [Hurlbert's \(1971\)](#) index), and comparisons can be difficult to interpret. However, many indices of evenness are robust with respect to sample size.

To account for differences in sampling intensities between sites, species counts are often rarified, which jackknives the assemblage data ([Raup, 1975](#); [Tipper, 1979](#)). This method creates average richness estimates for a known smaller sample size, making different sample sizes comparable. Ideally, the data for these rarefaction analyses would be drawn from original collections of the assemblages in question.

In the past decade, an array of accessible paleobiological databases has emerged (e.g., the Paleobiology Database, FAUNMAP, MIOMAP, NOW, and many others), and these online sources of data have provided a rich source for large-scale studies of diversity through time. Much of the specimen data in these databases comes from the published literature, which allowed enterers of these databases to quickly create robust or vetted data sources. The specimen data published in the peer-reviewed literature, however, is a biased subsample of the original collection made in the field. Because of logistical and economic constraints, the published literature usually contains only trophy or voucher specimens for specific taxa known from an assemblage; it is unusual for published accounts to make an exhaustive account of all the specimens found in an assemblage. In fact, [Koch \(1978\)](#) found that overall molluscan species richness was underestimated by a factor of three to four in publications, with bivalves and gastropods underestimated by a factor of five.

In this study, we expand on the work of [Koch \(1978\)](#) through quantification of the discrepancy between published accounts and museum collections in paleomammalian diversity estimates. Additionally, we consider the effect of publication bias on evenness, and its consequences for sample standardization. This comparison will provide a basis for researchers to judge the relative effect of using published data over fossil assemblage data carefully culled from museum collections. Although collecting biases in the field can also have an impact on relative abundances, we do not address this problem here because our study sample is a result of exhaustive collection efforts, not solely trophy hunting. We emphasize that the purpose and method of collecting are additional variables that should be considered by future studies.

Rarefaction analyses that produce curves of richness against sample size are affected by the evenness of the sample from which they are created ([Raup, 1975](#); [Olszewski, 2004](#)). At one extreme, exceptionally even samples tend to have relatively steep rarefaction curves, racing towards saturation and then leveling out, indicating that the sampled assemblage has similar proportions of all taxa. Exceptionally uneven samples, dominated by a single taxon, will have much shallower rarefaction curves, slowly building in richness as many new individuals must be sampled to increase the sampled richness. Because the initial slope of a rarefaction curve is equal to a measure of evenness, [Hurlbert's \(1971\)](#) probability of interspecific encounter ([Olszewski, 2004](#)), and because rarefaction curves are dependent on evenness for their shape, any bias introduced into the evenness of a sample can have an effect on richness comparisons.

As a case study, we focus on the Thousand Creek and Virgin Valley mammalian paleofaunas from the middle and late Miocene of northwestern Nevada ([Fig. 1A–C](#)). These faunas are found in geographic proximity in neighboring basins but are separated by 8 Myr of geologic time. Both faunas have been the subject of extensive field work by parties from the University of California Museum of Paleontology (UCMP) and these faunas have also been extensively published ([Gidley, 1908](#); [Merriam, 1909](#); [Furlong, 1910](#); [Kellogg, 1910](#); [Merriam, 1910](#); [Merriam, 1911](#); [Butterworth, 1916](#); [Merriam and Stock, 1928](#); [Hall, 1930](#); [Stock, 1936](#); [Wood, 1936](#); [Wilson, 1937](#)). By comparing published faunal data to data derived from the UCMP collections, we assess whether the published data reliably capture the evenness and richness patterns of the more complete set of unpublished museum data. While the results are most applicable to mammalian paleodiversity analysis, the approach and underlying concepts are relevant to future paleodiversity research in both the terrestrial and marine realms.

## 2. Materials and methods

### 2.1. Thousand Creek and Virgin Valley local faunas

We examined published accounts and museum collections of two fossil mammal faunas in northwestern Nevada ([Fig. 1C](#)): the Barstovian (middle Miocene) Virgin Valley and the Hemphillian (late Miocene) Thousand Creek localities, 15–16 Ma and 7–8 Ma, respectively ([Perkins et al., 1998](#); [Castor and Henry, 2000](#)). We extracted published faunal data for these sites from the MIOMAP database, an archival, online database that summarizes published taxonomic information in a uniform way ([Carrasco et al., 2005](#)). Detailed

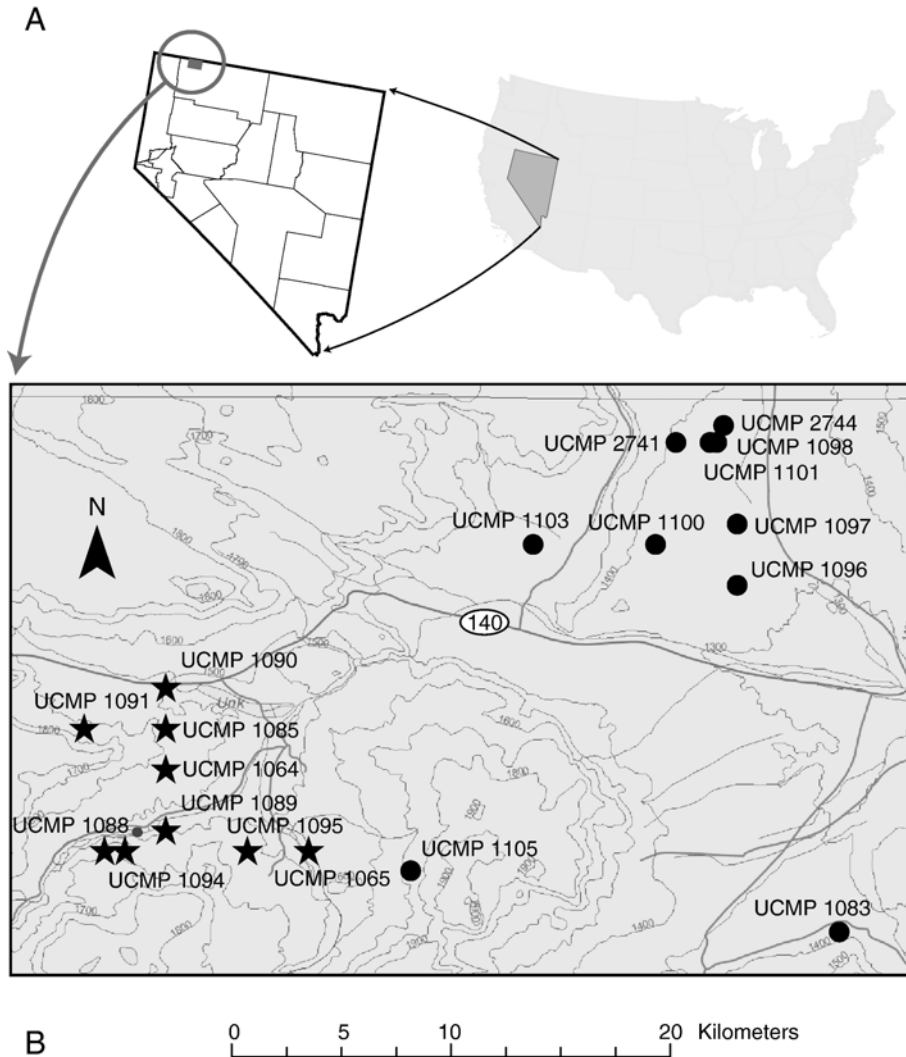


Fig. 1. A) Map of U.S.A., indicating location of Nevada, with inset map of Nevada; highlighted area in circle indicates extent of B. B) Map of northwestern most Humboldt County, with Thousand Creek Formation UCMP localities (circles) and Virgin Valley UCMP localities (stars). Note that the sites from the Thousand Creek Formation cover a larger geographic area, which may affect the relative diversity sampled.

information about the MIOMAP database can be found online at <http://www.ucmp.berkeley.edu/miomap/> and has been summarized in Barnosky (2001) and Barnosky and Carrasco (2002). We included only curated museum collections in this analysis. The processes of collection and curation can also introduce important biases in representation, so we were fortunate in this respect that the materials from both Thousand Creek and Virgin Valley were collected and curated by the same workers. The geographic proximity of the sites in both faunas makes a collecting trip at any one locality worth a trip to the other. The samples from both Thousand Creek and Virgin Valley also reflect the early twentieth century UCMP tradition of exhaustively

collecting all identifiable material. Additionally, >90% of the material collected at the sites included in our analysis has been curated, and we broke down batch-cataloged specimens into actual numbers of specimens. We also personally inspected every specimen in the collections to fully account for each specimen in our catalog (see Appendix 1). The combination of the aforementioned factors affects the quality of the collections. Consequently, these collections do a better job of representing the taphocoenosis than other collections might. Certainly, if there is any suspicion that the collections may be biased in proportions, workers should return to the field for additional sampling, if possible. Current collections should not simply be discarded for

paleoecological research (Allmon, 2005); field notes and other documentation can give an insight into the attitude of the original collectors.

The Virgin Valley and Thousand Creek faunas are preserved in very similar depositional environments and appear to have been affected by similar taphonomic filters. The initial publication of the faunas had difficulty distinguishing the Virgin Valley Formation from the overlying Thousand Creek Formation (Merriam, 1910). Despite an approximately 8 Myr unconformity between the Virgin Valley and Thousand Creek formations, it remains difficult to distinguish the strata without reference to faunal composition (Green, 1984). Both formations consist of tuffaceous claystones with local exposures of ash beds and poorly sorted conglomerates (Green, 1984). The vertebrate fossil assemblages consist primarily of mammalian postcrania, especially podial elements. When combined with sedimentologic and tectonic contexts (Ach and Swisher, 1990), the taphonomic data suggest that these fossils were deposited near the shores of a single large lake or several smaller lakes, where the bones were scavenged before burial, judging from criteria listed in Behrensmeyer et al. (1992). The presence of deposits of fossil wood in Virgin Valley (Crabtree, 1983) substantiates this assessment.

## 2.2. Faunal analysis

Our study focuses on collections from 19 out of the 30 total published localities from Virgin Valley and Thousand Creek, all housed at the UCMP. We chose the 10 richest localities of the 21 available from Thousand Creek in order to have a sample comparable to the nine available from Virgin Valley. To create the database for our analysis, we tabulated the number of identified specimens (NISP), minimum number of individuals (MNI) and simple presence/absence (P/A) data for all taxa at each locality. NISP is a very poor proxy for MNI (Shotwell, 1955; Grayson, 1973), as it counts as more abundant those animals that leave more specimens, whether or not they were truly more abundant. The P/A data, on the other hand, counts all taxa equally, completely removing all evenness data from an analysis. There are several methods for calculating MNI from a set of NISP data: Shotwell's (1955) method normalizes by preservability; and Grayson (1973) reviewed the more standard method of counting only the largest number of left or right elements for a species, thereby determining the minimum number of individual animals that could have created the observed assemblage. We used the latter method to calculate the MNI for each UCMP locality, and then summed MNI values from all localities to derive an overall MNI for each

taxon in the fauna. This approach tacitly assumes that individuals preserved at each locality on the landscape were independent of one another. Our method for calculating MNI is similar to Grayson's (1973) maximum distinction method, because we use arbitrary geographic divisions (i.e., collecting localities) to inform our MNI. Moreover, we do not have the problem he discusses of MNI inflation because Grayson (1973) referred to small, often contiguous sampling grids from archaeological excavations, while our geographic units are from fossiliferous exposures, hundreds to thousands of meters apart (Fig. 1B).

The NISP, MNI, and P/A data cover the spectrum of interpreting fossil abundance data. For the NISP and MNI catalog of published literature, we counted only UCMP specimens that were given with their specimen numbers in text, tables, or figures. Some papers mentioned occurrences of taxa from certain localities, but they did not provide voucher specimen numbers. In those cases, the taxa were counted present in the P/A data, but were not counted in the NISP or MNI data. We then counted NISP and MNI in those same localities in the collections of the UCMP, using the UCMP electronic catalog (available online at <<http://elib.cs.berkeley.edu/ucmp/>>) as a template for our identifications. We examined all 2066 specimens from these localities, updating catalog identifications where necessary either because of misidentification or subsequent systematic revision. From these total tallies, we compiled faunal lists following standard methodology (Alroy, 1992), omitting “?”, “cf.,” and “aff.” tags, and counting specifically and generically indeterminate specimens only when no specimens in those genera or families were identifiable to lower levels. The list of specimens ( $n=790$ ) that contributed to these faunal lists can be found in the Appendix table.

We compared the published and museum samples using rarefaction to standardize sample sizes, a commonly accepted procedure (Raup, 1975; Alroy, 1996; Bush et al., 2004; Olszewski, 2004). Although sampling issues are always a concern in paleobiological data, the similarities in taphonomy, taxonomic composition, and collection history between Virgin Valley and Thousand Creek localities indicate that the rarefaction approach is appropriate for this study (Tipper, 1979). The within-locality comparisons that are the basis of our analysis isolate the effects of publication bias on these statistics.

## 2.3. Hypothesis testing

For a given published assemblage, we hypothesize the published fauna will have a similar number of

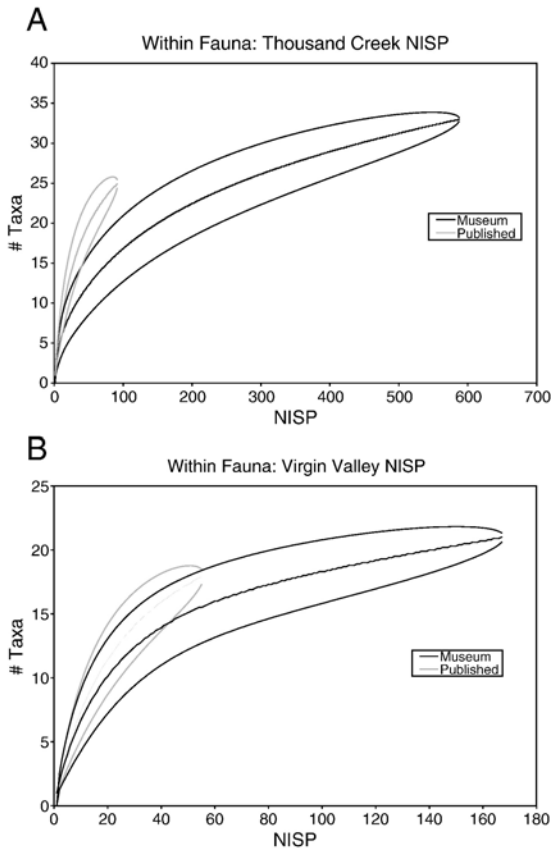


Fig. 2. Within-fauna comparisons of NISP-based rarefaction. A) Thousand Creek. B) Virgin Valley. The distance between the museum-based curve (black) and publication-based curve (gray) indicates the amount richness would be overestimated by simply referring to published data. The outside curves bracketing each rarefaction curve indicate 95% confidence intervals and can be used to judge significance; see Section 4 for additional discussion of significance tests.

representative specimens for each taxon, thereby artificially depressing the importance of numerically dominant species. Consequently, we predict that 1) published collections would be significantly more even than museum collections, generating NISP and MNI rarefaction curves that are steeper than those generated from museum collections. Additionally, we predict that 2) P/A data from published data and museum collections will be significantly different, as museum collections will capture occurrences that were not published because of redundancy: i.e., identifiable specimens from localities discovered at a later time would only be published if they represented taxa new to the fauna, or unusual individuals of previously published taxa. Finally, we predict that 3) comparisons between Virgin Valley and Thousand Creek rarefied data will be affected by publication bias, with the published data exhibiting artificially similar evenness.

These potential biases would result in published data showing no significant diversity differences between the two faunas even if museum data do so. Our null hypothesis is that the published data do not significantly differ from the museum data because the published data adequately capture the diversity (richness and evenness) of the museum data.

To test these hypotheses, we rarified our NISP, MNI, and P/A data using Holland's Analytical Rarefaction program for Windows (<http://www.uga.edu/~strata/software/>). This program is based on the equations presented by Raup (1975), originally derived by Hurlbert (1971) and Heck et al. (1975). The graphs we present illustrate the estimated richness for each integer number of specimens (NISP), minimum number of individuals (MNI) or number of occurrences (P/A). In addition, the graphs present the upper and lower 95% confidence bounds around each of those richness estimates. We judge curves to be significantly different when both curves move outside of the other's confidence limits as they reach their maximal values. This visual distinction approximates a two-sample *t*-test on the rarified richness

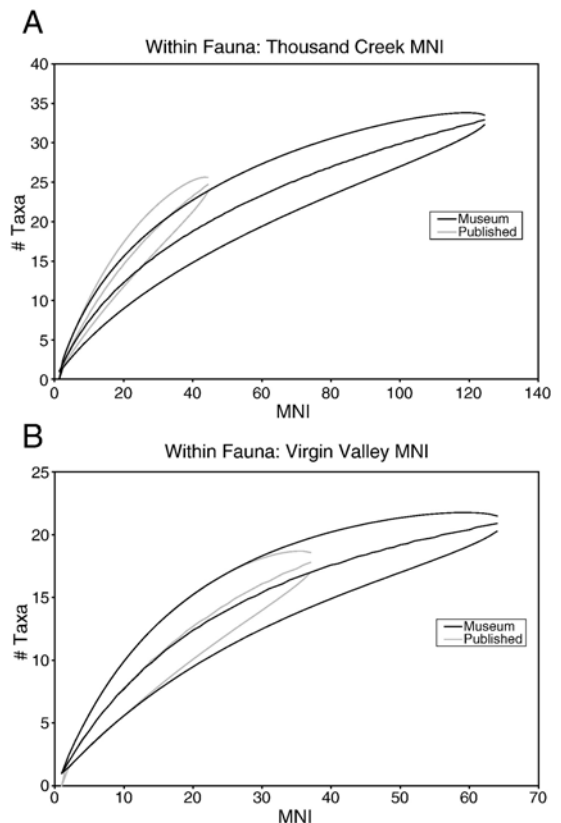


Fig. 3. Within-fauna comparisons of MNI-based rarefaction. A) Thousand Creek. B) Virgin Valley. Formatting of curves as in Fig. 2.

values because it incorporates information about the variance of both samples.

### 3. Results

#### 3.1. Within-fauna comparisons

NISP rarefactions differ as predicted for both Virgin Valley and Thousand Creek, with the curves from published material steeper than those from the museum collections. This result means that rarefactions based on published collections will be biased towards higher richness for a given number of specimens. The Thousand Creek NISP curves diverge early, and the curve from published data is significantly higher than the museum-based curve for sample sizes higher than only 14 specimens (Fig. 2A). Although the Virgin Valley NISP curves show the predicted pattern, they never diverge enough to be considered statistically distinct (Fig. 2B).

The MNI curves (Fig. 3A, B) show qualitatively identical results to the NISP curves (Fig. 2A, B). Both of the publication-based curves are steeper than their corresponding museum-based curves. The Thousand Creek curves diverge significantly above 34 individuals (Fig. 3A), and the Virgin Valley MNI curves diverge

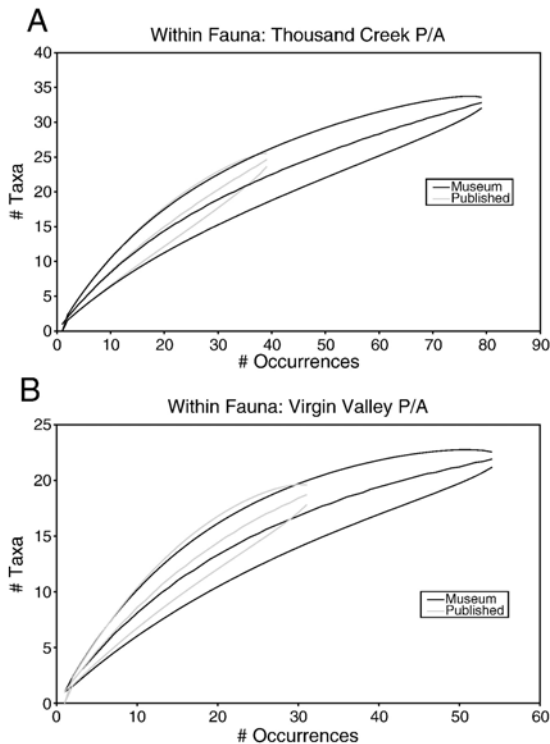


Fig. 4. Within-fauna comparisons of occurrences-based rarefaction. A) Thousand Creek. B) Virgin Valley. Formatting of curves as in Fig. 2.

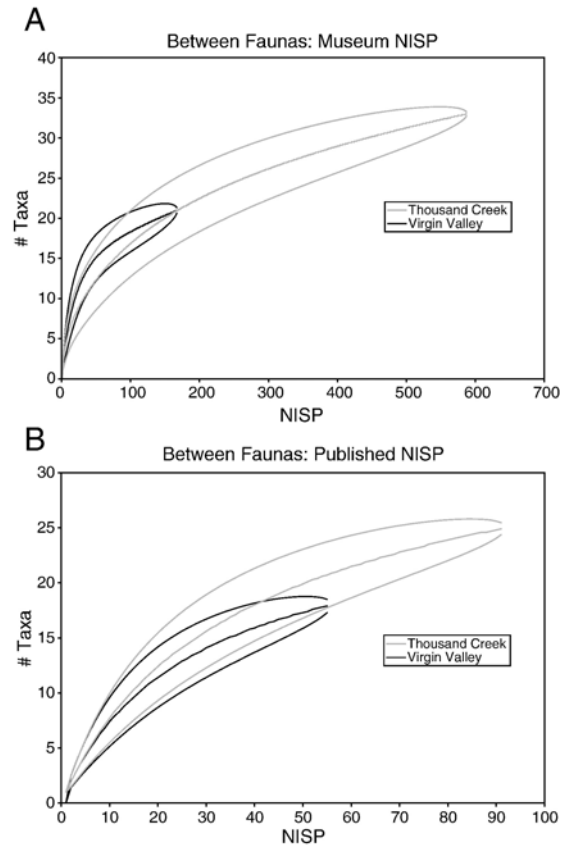


Fig. 5. Between-faunal comparisons of NISP-based rarefaction. A) Museum collections-based curves. B) Publication-based diversity curves. The distance between the Thousand Creek (gray) and Virgin Valley (black) curves indicates the relative difference in richness for the faunas at comparable sampling intensities. Formatting of 95% confidence curves as in Fig. 2.

very little and are, again, not statistically significantly different (Fig. 3B).

Presence/absence rarefactions for both faunas also fit our predictions but never diverge to the point of statistical significance (Fig. 4A, B).

#### 3.2. Between-fauna comparisons

NISP rarefactions of museum data indicate Virgin Valley is more diverse than Thousand Creek at any rarefied sample size (Fig. 5A), but the curves never diverge significantly. Conversely, NISP rarefactions of published data indicate Thousand Creek is more diverse than Virgin Valley at any rarefied sample size (Fig. 5B). Again, the difference between the published data sets for both faunas is not statistically significant. The Thousand Creek curve lies above the upper 95% confidence limit of the Virgin Valley curve above 44 specimens, but the Virgin Valley curve is never lower than the lower 95%

confidence limit of the Thousand Creek curve (Fig. 5B). This result signifies that the museum and published NISP data are not statistically inconsistent, even though they have reversed patterns of relative richness.

The MNI comparisons are very different from the NISP results for between-fauna comparisons. Both the museum-based and publication-based curves indicate that the Thousand Creek fauna is richer than the Virgin Valley fauna (Fig. 6A, B). The museum-based results are not statistically significant (Fig. 6A). The publication-based results are significant (Fig. 6B), suggesting that if one were to make this comparison based only on the published data, one would reasonably conclude that the Thousand Creek fauna is richer than the Virgin Valley fauna.

In contrast with the NISP results and in agreement with the MNI results, the P/A rarefactions are very similar for both museum and published data (Fig. 7A, B). For both data sources the Thousand Creek curve lies above the Virgin Valley curve. Unlike the MNI results, the museum data have a significant divergence of curves near the upper limit of Virgin Valley sampling (Fig. 7A). For published

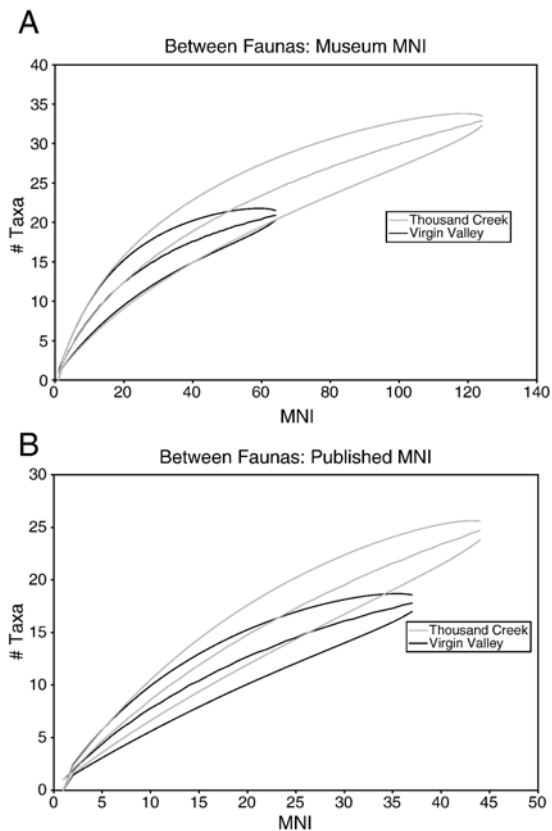


Fig. 6. Between-faunal comparisons of MNI-based rarefaction. A) Museum collections-based curves. B) Publication-based diversity curves. Formatting of curves as in Fig. 5.

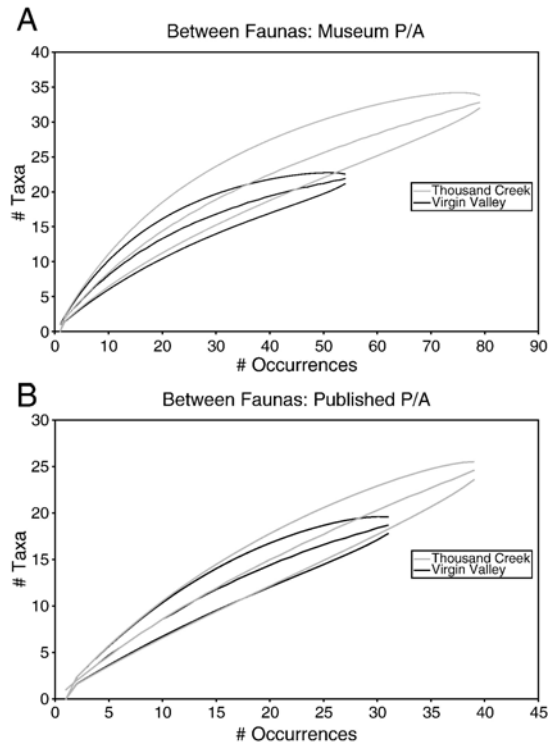


Fig. 7. Between-faunal comparisons of occurrences-based rarefaction. A) Museum collections-based curves. B) Publication-based diversity curves. Formatting of curves as in Fig. 5.

data, although the curves never statistically diverge, the Thousand Creek curve lies above the Virgin Valley upper 95% confidence limit above 29 occurrences (Fig. 7B). So, the P/A data agree with the MNI data that Thousand Creek is richer than Virgin Valley, but reverse which data set, museum or published, shows a significant difference.

#### 4. Discussion

The rarefaction curves are consistently steeper for the publication-derived NISP and MNI data (Figs. 2 and 3), suggesting they are biased towards higher evenness, as we predicted. Although these differences are not consistently statistically significant, the divergence between the rarefied Thousand Creek published richness and museum richness is quite striking for NISP data (Fig. 2A), and the data are still significant for MNI data, even if they are not as equally striking (Fig. 3A). The Virgin Valley data, on the other hand, do not show a significant difference in richness between museum and published data for either NISP or MNI (Figs. 2B and 3B). The consistency between NISP and MNI data is not surprising, since MNI is based on NISP. NISP and MNI are both biased estimators of the number of individuals,

with NISP biased towards too high a count and MNI biased towards too low a count. MNI has reasonably been assumed to be closer to the true count of individuals sampled, but that true count must lie somewhere between the two. There is, in the end, no clear signal that using only published data significantly affects comparisons between faunas.

The significant difference between the museum and publication Thousand Creek NISP curves effectively reverses the between-fauna NISP comparison (Fig. 5). The differences are not so extreme that they differ statistically: Both the UCMP and publication comparisons are not statistically significant, emphasizing the importance of accounting for uncertainty when making diversity comparisons.

These results would change if we were to use a less strict criterion for statistical difference, either through relaxing the  $p$  value or counting significance when only one curve is outside the confidence limits of the other. Keeping the 95% confidence limits as in Fig. 5 and asking whether the Thousand Creek fauna is more diverse than (above the confidence limit of) the Virgin Valley fauna at 50 specimens illustrate this point. For the museum data (Fig. 5A), the answer is no, but for the published data (Fig. 5B), the answer is yes.

As explained in Section 2, we estimated significance using the criterion of both curves beyond both 95% confidence limits. This is a quick visual way to approximate a  $t$ -test using the pooled variances of the two estimates of richness. Applying a one-sided  $t$ -test with pooled variance to the same question of whether Thousand Creek is more diverse than Virgin Valley at 50 specimens results in no significant differences for either museum data or published data, because the  $t$ -test includes the larger uncertainty in the Thousand Creek diversity at 50 specimens. Uncertainty in both parameter estimates must be included in an analysis of this sort.

Unsurprisingly, the MNI results are more consistent than the NISP results, since MNI values will be less affected by unpublished specimens of published taxa. The publication-based MNI data show a significant difference between Thousand Creek and Virgin Valley faunas, but the museum collections-based data do not. The excessive evenness of the published Thousand Creek data is again responsible for this difference, as the publication bias pulls the Thousand Creek curve up and outside of the 95% confidence intervals. The abundance issues associated with using NISP instead of MNI create the situation seen in the between-fauna comparison for museum data (Figs. 5A and 6A). The NISP data make Thousand Creek look more even than it really is because additional specimens that add no new MNI information are counted

in the NISP data. These results provide more evidence that paleoecological workers should always go through the effort to calculate MNI values for their faunas.

Although the two paleomammalian faunas in this study share similar taphonomic and collecting histories, there are some differences, both geological and publication-based between Thousand Creek and Virgin Valley that might account for the differences in their NISP and MNI rarefaction results. The geological reasons for these differences likely relate to the greater area of outcrop available for Thousand Creek localities, compared to those from Virgin Valley (Fig. 1). The Thousand Creek Formation crops out over a larger area than the Virgin Valley Formation, and the localities included in this analysis are from a slightly larger geographic area. Area of sampling has an important effect on paleodiversity (Barnosky et al., 2005) and also has an effect on Beta (between locality) diversity (Mac Nally et al., 2004), which also influences regional rarefaction curves (Olszewski, 2004). This difference in area is a consequence of the depositional history of the two formations, with the Virgin Valley Formation filling a single caldera and the Thousand Creek Formation deposited over a more regional scale (Green, 1984; Ach and Swisher, 1990).

The difference in publication histories between the two faunas is more striking. Although both faunas were treated in the earliest publications (Gidley, 1908; Furlong, 1910; Kellogg, 1910; Merriam, 1910; Merriam, 1911), all of the follow-up publications treated new occurrences from the Thousand Creek fauna, often published as short notes (Butterworth, 1916; Merriam and Stock, 1928; Hall, 1930; Stock, 1936; Wood, 1936; Wilson, 1937). This variation in publication means that a much larger proportion of the taxa in the Virgin Valley fauna were published in Merriam's (1911) faunal analysis, a venue that allowed a reasonably complete accounting of the relative abundances.

Thus, the greatest risk these results present for large-scale analyses concerns comparing data from faunas published under differing conditions. Some faunas, such as those exhaustively worked for theses and dissertations (e.g., Shotwell, 1953; Vanderhill, 1980; Morea, 1981) or as government-funded remediation projects (Voorhies, 1990) are not subject to page limits, and so might be published with much more complete NISP data, which could lead to better MNI data. More completely published faunas will tend to be more faithful to the underlying evenness of the fossil sample, and may not work well with faunas published in venues with page limitations. In the end, publication venue might be treated as another taphonomic factor to be considered when compiling data for analysis.

The NISP and MNI results are also pertinent to estimating relative abundance in paleobiological analyses.



The differences in relative abundances we found between publication and museum data made subtle but important differences to the overall faunal evenness and richness estimates, which certainly have implications for large-scale studies, but studies on individual taxa could be more impacted by publication bias, particularly if those taxa have been over or underpublished before. In the case of Virgin Valley and Thousand Creek, there are several taxa which appear to be less important in the publication data than they are in the museum data. The camelids are the prime example, with no specimens published in Virgin Valley, and only one specimen published in Thousand Creek (see Appendix). The camelids are, however, an important part of these faunas, contributing 12% of the MNI to Virgin Valley and 18% to the Thousand Creek fauna. Of course, the camelid condition is so extreme that a camelid worker would be required to return to the museum; other abundant taxa show a similar pattern, with *Hypolagus vetus* contributing 4% to the published fauna and 13% to the museum collections of Thousand Creek. Contrastingly, the well-published mylagaulids of Virgin Valley appear to be 21% of the fauna, based on publications, but only represent 15% of the museum data. In this case, the underrepresentation of many other taxa in the fauna makes this well-represented taxon appear more abundant than it really is. This phenomenon is taken to an extreme for the truly rare taxa, such as *Diprionomys minimus* from Thousand Creek, which appears to be only 2% of the fauna through the filter of publication, but represents even less, only 0.8% of the fauna in the museum collections. Our results make apparent the need to carefully consult museum collections before performing paleobiological analyses that rely on relative abundance information.

The results of our P/A analysis (Figs. 4 and 7) were more straightforward. The within-fauna comparisons do not show any significant divergence, and the between-fauna comparisons are consistent in their patterns. The museum-based between-fauna comparison does show a significant divergence and the publication-based comparison does not (Fig. 7A, B). The lack of significance in the latter might indicate a loss in statistical power because of the additional random variance introduced by the publication filter. Our results suggest that analyses of published faunal data that rely on occurrence data will not be subject to biases introduced by the filter of publication, only to an increase in variance. This conclusion makes intuitive sense, as one of the goals of all paleontological publications is to produce an accurate accounting of where and when certain taxa are found. Our results do disagree with those of Koch (1978), who found a major bias in the published record

of mollusk richness. Perhaps the lesser overall richness and representation in the terrestrial record have allowed terrestrial workers to publish a greater proportion of all taxa than marine workers. For terrestrial systems, the biggest problems arise when contextual analyses begin to consider the relative numbers of specimens representing each taxon in an assemblage.

## 5. Summary

Because of the limitations of publication, some published relative abundance data do not meet all of the assumptions that underlie paleoecological studies. Although much work has been done towards accounting for sampling biases (Raup, 1972; Raup, 1975; Alroy, 2000; Foote, 2001) and taxonomic biases (Adrain and Westrop, 2000; Alroy, 2002; Alroy, 2003; Isaac et al., 2004), little work has assessed the effect of biases in the specimens chosen for publication (e.g. Koch, 1978).

Through comparisons of published accounts and the collections of the UCMP for two faunas from northwestern Nevada, we have been able to explore the effects of the filter of publication on estimates of richness. Richness rarified by NISP or MNI is affected by publication, as rarefaction is very sensitive to evenness in a sample (Raup, 1975; Olszewski, 2004). Although publication does make assemblages appear overly even, the filtering process does not result in a consistent significant difference in richness within faunas for our data. Comparing the Hemphillian Thousand Creek fauna to the Barstovian Virgin Valley reveals the effect of publication bias in comparing diversity between faunas. Rarefaction curves based on published NISP and both published and museum MNI indicate the Thousand Creek fauna is more diverse than that of Virgin Valley for any sample size, but not significantly so for NISP or museum MNI. Museum-based rarefaction curves of NISP indicate Virgin Valley is more diverse, but not significantly so. The non-significance of the NISP comparisons makes the results statistically equivalent and emphasizes the importance of properly accounting for variance in paleobiological analyses. The significance of the published MNI results emphasizes the effect that accounting for the minimum number of individuals can have on a faunal analysis. All of these results make it clear that simple qualitative comparisons of rarefaction results may be misleading. Workers must be explicit about the nature of statistical comparisons they make between faunas.

Results from presence/absence rarefaction indicate that occurrences are properly transmitted from museum collections through the filter of publication. This result is

hardly surprising, as all paleontologists focus on accurately communicating occurrences, but disagrees with Koch's (1978) findings for marine invertebrates, suggesting different precautions be taken by marine and terrestrial workers.

Taphonomy will always make fossil assemblages different from the original biological assemblage they represent (Behrensmeyer et al., 1992). Publication bias can be seen as an additional taphonomic filter placed upon fossils before they reach the published paleontological record (see Fig. 1 of Crampton et al., 2003), with its most important effect on the relative evenness of taxa in the sample. Our results agree with Allmon (2005), that additional museum and field work is necessary to properly evaluate a fossil assemblage; paleobiology cannot proceed without careful maintenance and study of museum collections.

### Acknowledgements

We would like to thank A.D. Barnosky, D.R. Lindberg, J.H. Lipps, P.A. Holroyd, the members of the Barnosky and Lindberg labs, and the University of California Museum of Paleontology community. We also extend a sincere thanks to all the members of the Nevada expedition during the 2003 Integrative Biology Vertebrate Paleontology Field Methods course. Four anonymous reviewers contributed greatly to the improvement of this paper. EBD thanks S.S.B. Hopkins for forbearance, emotional support, and always picking holes in my ideas. NDP would like to thank G.P. Wilson and R.B. Irmis for comments. This research was completed while both authors were Graduate Research Fellows of the National Science Foundation. Map produced using ArcGIS software funded by the MIOMAP project (National Science Foundation grant EAR-0310221) and provided by the University of California, Berkeley, GIS Center. This paper is University of California Museum of Paleontology Contribution No. 1936.

### Appendix A

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.palaeo.2007.03.006](https://doi.org/10.1016/j.palaeo.2007.03.006).

### References

- Ach, J.A., Swisher, C.C., 1990. The High Rock caldera complex; nested "failed" calderas in northwestern Nevada. *Eos* 71 (43), 1614.
- Adrain, J.M., Westrop, S.R., 2000. An empirical assessment of taxic paleobiology. *Science* 289, 110–112.
- Allmon, W.D., 2005. The importance of museum collections in paleobiology. *Paleobiology* 31 (1), 1–5.
- Alroy, J., 1992. Conjunction among taxonomic distributions and the Miocene mammalian biochronology of the Great Plains. *Paleobiology* 18 (3), 326–343.
- Alroy, J., 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology* 127 (1–4), 285–311.
- Alroy, J., 2000. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* 26 (4), 707–733.
- Alroy, J., 2002. How many named species are valid? *Proceedings of the National Academy of Sciences* 99 (6), 3706–3711.
- Alroy, J., 2003. Taxonomic inflation and body mass distributions in North American fossil mammals. *Journal of Mammalogy* 84 (2), 431–443.
- Barnosky, A.D., 2001. Distinguishing the effects of the Red Queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *Journal of Vertebrate Paleontology* 21 (1), 172–185.
- Barnosky, A.D., Carrasco, M.A., 2002. Effects of Oligo-Miocene global climate changes on mammalian species richness in the northwestern quarter of the USA. *Evolutionary Ecology Research* 4 (6), 811–841.
- Barnosky, A.D., Carrasco, M.A., Davis, E.B., 2005. The impact of the species–area relationship on estimates of paleodiversity. *Public Library of Science: Biology* 3 (8), 1356–1361.
- Behrensmeyer, B.K., et al., 1992. Paleoenvironmental contexts and taphonomic modes. In: Behrensmeyer, A.K., et al. (Eds.), *Terrestrial Ecosystems Through Time: Evolutionary Paleocology of Terrestrial Plants and Animals*. The University of Chicago Press, Chicago, pp. 15–136.
- Bush, A.M., Markey, M.J., Marshall, C.R., 2004. Removing bias from diversity curves: the effects of spatially organized biodiversity on sampling standardization. *Paleobiology* 30 (4), 666–686.
- Butterworth, E.M., 1916. A new mustelid from the Thousand Creek Pliocene of Nevada. *University of California Publications in Geological Sciences* 10 (2), 21–24.
- Carrasco, M.A., Kraatz, B.P., Davis, E.B., Barnosky, A.D., 2005. Miocene Mammal Mapping Project (MIOMAP). <http://www.ucmp.berkeley.edu/miomap/>.
- Castor, S.B., Henry, C.D., 2000. Geology, geochemistry, and origin of volcanic rock-hosted uranium deposits in northwestern Nevada and southeastern Oregon, USA. *Ore Geology Reviews* 16 (1–2), 1–40.
- Crabtree, D.R., 1983. *Picea wolfei*, a new species of petrified cone from the Miocene of northwestern Nevada. *American Journal of Botany* 70 (9), 1356–1364.
- Crampton, J.S., et al., 2003. Estimating the rock volume bias in paleobiodiversity studies. *Science* 301, 358–360.
- Foote, M., 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology* 27 (4), 602–630.
- Furlong, E.L., 1910. An apodont rodent from the Tertiary of Nevada. *University of California Publications in Geological Sciences* 5 (25), 397–403.
- Gidley, J.W., 1908. Notes on a collection of fossil mammals from Virgin Valley, Nevada. *University of California Publications in Geological Sciences* 5 (15), 235–242.
- Grayson, D.K., 1973. On the methodology of faunal analysis. *American Antiquity* 39 (4), 432–439.
- Green, R.C., 1984. Geologic appraisal of the Charles Sheldon Wilderness Study Area, Nevada and Oregon. *Geological Survey Bulletin* 1538, 13–34.

- Hall, E.R., 1930. Three new genera of Mustelidae from the later Tertiary of North America. *Journal of Mammalogy* 11, 146–154.
- Heck, K.L.J., Van Belle, G., Simberloff, D., 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56 (6), 1459–1461.
- Hurlbert, S.H., 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52 (4), 577–586.
- Isaac, N.J.B., Mallet, J., Mace, G.M., 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution* 19 (9), 464–469.
- Kellogg, L., 1910. Rodent fauna of the late Tertiary beds at Virgin Valley and Thousand Creek, Nevada. University of California Publications in Geological Sciences 5, 421–437.
- Koch, C.F., 1978. Bias in the published fossil record. *Paleobiology* 4 (3), 367–372.
- Mac Nally, R., Fleishman, E., Bulluck, L.P., Betrus, C.J., 2004. Comparative influence of spatial scale on beta diversity within regional assemblages of birds and butterflies. *Journal of Biogeography* 31, 917–929.
- Merriam, J.C., 1909. The occurrence of strepsicerine antelopes in the Tertiary of northwestern Nevada. University of California Publications in Geological Sciences 5 (22), 319–330.
- Merriam, J.C., 1910. Tertiary mammal beds of Virgin Valley and Thousand Creek in northwestern Nevada. Part I — geologic history. University of California Publications in Geological Sciences 6 (2), 21–53.
- Merriam, J.C., 1911. Tertiary mammal beds of Virgin Valley and Thousand Creek in northwestern Nevada. Part II — vertebrate faunas. University of California Publications in Geological Sciences 6 (11), 199–304.
- Merriam, J.C., Stock, C., 1928. A further contribution to the mammalian fauna of the Thousand Creek Pliocene, northwestern Nevada. Contributions to Palaeontology from the Carnegie Institution of Washington 393, 5–21.
- Morea, M.F. 1981. The Massacre Lake local fauna (Mammalia, Hemingfordian) from northwestern Washoe County, Nevada. Ph.D Thesis, University of California, Riverside, 262 pp.
- Olszewski, T.D., 2004. A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos* 104, 377–387.
- Perkins, M.E., Brown, F.H., Nash, W.P., McIntosh, W., Williams, S.K., 1998. Sequence, age, and source of silicic fallout tuffs in middle to late Miocene basins of the northern Basin and Range province. *Geological Society of America Bulletin* 110 (3), 344–360.
- Raup, D.M., 1972. Taxonomic diversity during the Phanerozoic. *Science* 177, 1065–1071.
- Raup, D.M., 1975. Taxonomic diversity estimation using rarefaction. *Paleobiology* 1 (4), 333–342.
- Shotwell, J.A. 1953. A Hemphillian mammalian fauna from McKay Reservoir, Oregon. Ph.D Thesis, University of California, Berkeley, 150 pp.
- Shotwell, J.A., 1955. An approach to the paleoecology of mammals. *Ecology* 36 (2), 327–337.
- Simpson, E.H., 1949. Measurement of diversity. *Nature* 163, 688.
- Stock, C., 1936. A Pliomastodon skull from the Thousand Creek beds, northwestern Nevada. Contributions to Palaeontology from the Carnegie Institution of Washington 473 (3), 35–39.
- Tipper, J.C., 1979. Rarefaction and rarefaction; the use and abuse of a method in paleoecology. *Paleobiology* 5 (4), 423–434.
- Vanderhill, J.B. 1980. Geology and paleontology of the Patrick Buttes, Sioux County, Nebraska, and Goshen County, Wyoming. M.S. Thesis, University of Nebraska, Lincoln, 98 pp.
- Voorhies, M.R., 1990. Vertebrate paleontology of the proposed Norden Reservoir Area, Brown, Cherry, and Keya Paha Counties, Nebraska. Division of Archeological Research, University of Nebraska at Lincoln, Technical Report, 82-09: 1-A593.
- Wilson, R.W., 1937. Pliocene Rodents of Western North 21–73.
- Wood, A.E., 1936. Fossil heteromyid rodents in the collections of the University of California. *American Journal of Science* 32 (188), 112–119.