Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influences

M.H. MacRoberts and B.R. MacRoberts
Herbarium, Museum of Life Sciences, Biology Department, Louisiana State University, Shreveport, LA 71115, and Bog Research, 740 Columbia, Shreveport, LA 71104. E-mail: mmacrobe@lsus.edu

To determine influences on the production of a scientific article, the content of the article must be studied. We examined articles in biogeography and found that most of the influence is not cited, specific types of articles that are influential are cited while other types of that also are influential are not cited, and work that is "uncited" and "seldom cited" is used extensively. As a result, evaluative citation analysis should take uncited work into account.

Introduction

In a previous series of studies, we examined articles and compared influences evident in the text with those referenced in the bibliography. Regardless of subject examined, we found that authors cite only a fraction of their influence (M.H. MacRoberts, 1997; M.H. MacRoberts & MacRoberts, 1986, 1987a, 1987b, 1988, 1989, 1996, 1997a). While it is undoubtedly impossible to detect all influence on a article, a large percentage can be detected. Thus, instead of one population of influences there are two: those that are cited and those that are not cited. These two populations can be identified by studying articles, and the articles that are cited and not cited can be compared.

In the course of doing research on Southeastern United States flora, we encountered a discipline that might be of interest to citation analysts. This is biogeography, a field that traces its roots to such luminaries as Alexander von Humboldt, ardent explorer, and Alfred Russel Wallace, codiscoverer with Charles Darwin of the theory of evolution by natural selection (Darwin & Wallace, 1858). Von Humboldt's *De Distributione Geographica Plantarum*, published in 1817, is considered to be the founding text of phytogeography (the science of plant distribution and numbers), and Wallace's *The Geographical Distribution of Animals*, published in 1876, is considered to be the beginning for zoogeography (the science of animal distribution and numbers). Together, the field is called biogeography (for a modern textbook coverage of the field, see Lomolino, Riddle, & Brown, 2006). The main question asked by biogeographers is why are plants and animals distributed as they are, and thus "why are there this many species in any particular locality?" Major sources of information for plants are herbarium specimens and floras; main sources for zoogeography are museum collections of animals and "faunas" (The word is not currently in general use, but since there is no general term today for these works, we use it here.). Authors of biogeographical articles have not personally collected the data they use but rely heavily on extremely large databases, compiled by thousands of individuals over centuries. However, what is particularly interesting about articles in this field from the standpoint of citation analysis is that there is a generally accepted protocol by which authors provide substantial information about the databases they use, *but they do not cite them*. This is part of what Szava-Kovats (2008) called the nonindexed reference stock brought up by the author(s), but not cited.

Using Google Scholar and *Web of Science* (ISI *Web of Knowledge*) and suggestions from colleagues, we reviewed the citation-analysis literature to determine the current state of research on the subject of noncitation of influence. Fortunately, several recent reviews (e.g., Bornmann & Daniel, 2008; Camacho-Minano & Nunez-Nickel, 2009; Cronin, 2005; Nicolaisen, 2007; Van Raan, 2004) brought us up to date. From our reading, it would appear that there are only a few who have studied scientific articles to determine whether influences are cited (Kostoff, 2005; Kostoff, Morse, & Oncu, 2007). Hoerman and Nowicke (1995) studied secondary and tertiary citing. Szava-Kovats (2002) studied physics articles and found that there is an "over-abundance of relevant literature" (p. 3). Szava-Kovats (2008) also found that the vast majority of influence on a article is not cited: "The oppressive weight of the vast literature that should be cited to give precise and full documentation . . . is so vast that it is . . . impossible to cite . . ." (p. 32). Greenberg (2009) studied how citation distortions create unfounded authority. Cronin (2005) recognized that not all influences are cited:

> . . . we are invariably challenged to cite the most precise and most relevant work on a given subject, for the simple reason that few, if any, of us are wholly and authoritatively

Received May 8, 2009; revised July 16, 2009; accepted August 18, 2009

 • Published online 12 October 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21228

> familiar with the scattered literature of our specialties, let alone the wider scientific literature. Even if we were familiar with the entire corpus of relevant literature, we will still have to make difficult choices. (p. 1506)

Camacho-Minano and Nunez-Nickle (2009) recognized much the same thing: “It seems clear that authors cannot cite all references when large numbers exist . . . or when space is limited. Thus, they select specific papers depending on personal preferences” (p. 756). Consequently, we know that there is a lot of citable (i.e., influential) work of which only a portion is cited. The question becomes “What is cited and what is not cited?”

There is no way to precisely define influence, but we digress briefly here to consider it. In a previous article, we (MacRoberts & MacRoberts, 1986) stated that

> we are obviously not defining “influence” in the most general sense imaginable but are using it precisely in the manner used by citation analysts. For example, we do not consider it “influence” if a physicist used geometrical analogues to solve mathematical problems because twenty-five years earlier in high school he had an excellent geometry teacher, while we do consider Mossbauer to have influenced his colleagues and hence, the course of physics, if the Mossbauer Effect is used by other physicists. (p. 170)

When it is evident in the text that an author makes use of another’s work, either directly or through secondary sources, he or she has been influenced by that work. We will briefly discuss “obliterated” work (i.e., work that has become so accepted that it is no longer cited) in the discussion, but it should be understood at the outset that “obliterated” work in the Mertonian sense (Merton, 1973) is not what we are discussing here.

This article is divided into three parts. First, we document the uncited works in biogeographical articles. Second, we consider some of the characteristics of cited and uncited works that are used. Third, we document how some of these uncited works are actually cited and used.

How Thoroughly Do Biogeographical Articles Capture Influence?

We describe in some detail two examples. Twenty additional examples are given in Appendix A.

Example 1

Hong Qian, Jason Fridley, and Michael Palmer in “The latitudinal gradient of species-area relationships for vascular plants of North America,” published in 2007 in *American Naturalist*, described the decrease in species richness with increase in latitude across North America. In their methods section, they stated that “A total of 1,742 floras of North America north of Mexico were included in our study, derived from a compilation of state and provincial data, the Floras of North America database, and additional floristic literature” (Qian et al., 2007, p. 272). What is a flora? Withers, Palmer, Wade, White, and Neal (1998) described a flora as

> a list of plant species that are known to occur within a region of interest. Often, the list is accompanied by a scientific description of each species that would permit it to be identified. Botanists have written floras for every type of region imaginable: city, state, and national parks, counties and states, a single pond or rock outcrop, and even for an entire country. Thus, floras span a wide range of land areas, from 1 ha or less to many millions of hectares. (p. 24)

Floras range from government documents, theses, unpublished reports, online Web sites, short notes and lists in botany journals to 2,000-page books with thousands of references (Palmer, Wade, & Neal, 1995). McLaughlin (2007), another phytogeographer, adds that most floras usually comprise between 200 and 1,000 species and are focused on “local intact remnants of the regional flora in areas of intense agricultural and urban development” (p. 2). To further investigate the Qian et al. (2007) database, we contacted Michael Palmer, who is in charge of the “Floras of North America Project,” which constitutes part of the database for the Qian et al. (2007) article. He supplied us with the online database mentioned but not included in the print version of their article. Of the 1,742 floras used for the database of the article, one turned out to be our “Annotated checklist of the vascular flora of the Hickory Creek Unit of the Big Thicket National Preserve, Tyler County, Texas” published in *Sida* (B.R. MacRoberts, MacRoberts, & Brown, 2002). Thus, we are familiar with the type of material that Qian et al. (2007) used because we are the authors of some of it.

One question remains about floras: Are they of any value to phytogeographers? According to Kreft and Jetz (2007), both phytogeographers whose work we discuss in Appendix A, “to date, our quantitative understanding of [species] diversity gradients . . . such as plants has been hampered by the paucity of distribution data” (p. 5925). If the meaning of that statement should be missed, they repeated: “Progress toward more general and, importantly, global models of gradients of species richness to date have been hampered by the many species that remain only poorly documented in their geographic occurrence or altogether unknown” (p. 5925). Further, on the “Floras of North America Project” Web site (http://www.okstate.edu/artsci/botany/floras/), we read:

> Botanists and ecologists in North America have a rich tradition of writing Floras, or lists of vascular plants. The potential of using these floras as baselines for understanding patterns of, and threats to, modern biodiversity cannot be overstated. Nevertheless, floristic work is often published in obscure sources and is very difficult for biodiversity stakeholders to obtain. Although information technology and data transfer amongst scientists are both growing by leaps and bounds, floras are largely confined to dusty shelves. Imagine how biodiversity research could be facilitated if there was a georeferenced, easily accessible database that allows scientists to access raw floristic data from thousands of locations. . . .

Returning to Qian et al. (2007), they used 1,742 floras, *but cite none of them*. Thus, our work was used but not cited;

DOI: 10.1002/asi

neither were the other 1,741 floras. Instead, Qian et al. (2007) referenced 85 other works. What are these? They are largely articles just like their own and published in the same type of journals (mainly Thomson Reuters-monitored journals) that also attempt to explain species richness across latitudinal gradients. We will discuss the citation characteristics of these two groups of publications later.

Example 2

In 2007, Stephen McLaughlin published "Tundra to Tropics: The Floristic Plant Geography of North America" in *Sida, Bototanical Miscellany*. McLaughlin is one of the few authors who included data sources in his work. He stated that "The 245 local floras selected for this study are listed in Appendix A" (p. 3). *Although listed in an appendix, they are not included in the bibliography*. Instead, his bibliography consists of 28 other publications, mostly books and articles in books, but also articles in Thomson Reuters-monitored journals.

McLaughlin (1986) presents an interesting study of citing. In his earliest phytogeography article on the flora of the southwest United States, he used 50 floras, but he cites only 17 of them. He stated that "complete bibliographic citations for 33 of the 50 local floras can be found in Bowers' (1982) annotated bibliography" (p. 48). But when he wrote a second article expanding the geographical coverage to the entire Western United States, although he included the data from all 50 previously used floras, he cited only the 50 new ones he added to the list, and refers to the uncited 50 by citing his earlier article (McLaughlin, 1989). Thus, in this second article, *he cited only half of the floras used*. When McLaughlin later wrote a article about a natural-areas park in Arizona in 2006 and compared his study area phytogeographically to the flora of all of North America, he noted that "Geographic affinities of the floras of [the study areas] were examined by mapping the percentage of native taxa found in 245 local floras from Mexico, the United States, and Canada, using the database compiled by the author" (p. 669). He did *not cite any of these floras*. It is only in his 2007 article described earlier that he lists the sources in an appendix. In all, he used 640 sources, but cited only 67 of them.

Are Qian et al. (2007) or McLaughlin (1986, 1989, 2006, 2007) exceptional in their citing practices? Not in the field of biogeography. The reader can examine 20 additional examples in Appendix A. The number could be greatly increased. The point, which need not be belabored, is that while the work of a *very* large number of individuals (both published and unpublished) is used in the production of the databases used by biogeographers, and while they are acknowledged in a very general way, *they are not cited*.

Citation Patterns of Cited and Uncited Biogeographic Influences

In the previous section (including examples in Appendix A), we showed that only a fraction of influence on biogeographical articles is cited. What was evident to us as investigators of these articles was that they were all theoretical/analytical and that they cited predominately theoretical/analytical articles. In this section, we compare the citations in these articles and citations in the uncited works.

TABLE 1. Characteristics of references in theoretical/analytical articles and data articles. Both samples consist of 530 citations.

	Characteristics					
Type of papers	1	2	3	4	5	6
Theoretical/analytical articles	42	314	159	7	0	8
Data articles (sources)	239	95	63	65	50	18

We selected 10 theoretical/analytical biogeographical articles that were published in journals that are monitored by Thomson Reuters and examined their bibliographies. These 10 articles are from the 22 described in this study. We recorded whether the cited item was (a) a article from a journal not monitored by Thomson Reuters, (b) an article from a Thomson Reuters-monitored journal, (c) a book or chapter in a book, (d) an unpublished report, (e) a thesis or dissertation, or (f) an online Web source or CD-ROM. There were 530 citations in these 10 articles. We selected five data articles or sources used for data in the 22 theoretical/analytical articles, such as those listed in McLaughlin (2007). We recorded the same information that we recorded for the other sample. We stopped when we reached 530 items so that the two samples would be equal. Table 1 shows the data for the two groups.

Statistical comparison of the two groups shows that there is no possibility of them being derived from the same population. Theoretical/analytical biogeographical articles predominantly cite theoretical/analytical articles from Thomson Reuters-monitored journals; data articles do not cite many Thomson Reuters-monitored articles but instead cite work from journals not monitored by Thomson Reuters, unpublished reports, theses/dissertations, and the grey literature.

Other disciplines apparently follow the same pattern. Years ago, McGervey (1974) wrote:

> Consider a 1968 article by Gell-Mann, Oakes, and Renner . . . . It cited 26 papers; all were by theorists. A check of the cited papers shows that almost all of the papers cited in them were also by theorists. The only references to experimental work were "second generation" citations of books or review articles. (p. 30)

Cole and Cole (1972) found that highly cited physics papers also cited highly cited papers, and so on. If we did not know that biogeographic papers cited very few of their influences, we might come to the same conclusion as the Coles: that only a small number of individuals contribute to scientific progress. But knowing that the vast majority of influence is not cited and that that influence is not to be found in the Thomson Reuters-monitored journals leads to a different conclusion: that many—not a few—contribute to scientific progress.

DOI: 10.1002/asi

Are "Uncited" and "Seldom-Cited" Articles Really Not Cited or Seldom-Cited and Not Used or Seldom Used?

There is a widely held belief that the majority of scientific articles are either never cited or seldom-cited and not used (Hamilton, 1990, 1991; Pendlebury, 1991). Aksnes (2003) stated that "Citation distributions are extremely skewed. The large majority of the scientific articles are never or seldom cited in the subsequent scientific literature" (p. 159). Opthof (1997) noted that "A considerable amount of published scientific work is *never* cited" (p. 2). Meho (2007) wrote: "It is a sobering fact that some 90% of articles that have been published in academic journals are never cited. Indeed, as many as 50% of papers are never read by anyone other than their authors, referees, and journal editors. We know this thanks to citation analysis . . ." (p. 32). These statements are incorrect (e.g., Larivière, Gingras, & Archambault, 2009), but they underline the point that we wish to make in this section: (a) When citation analysts speak of "uncited" or "seldom-cited," they are usually referring to uncited or seldom-cited in the journals monitored by Thomson Reuters and other similar databases, not to all journals, books, and reports; and (b) "uncited" or "seldom-cited" is *not* a synonym for "not used."

Next, we give one example each of an uncited and a seldom-cited article. Several more examples are given in Appendix B.

Example 1

In 1995, we published the following two-page article: "*Palhinhaea cernua* (L.) Vasconcellos & Franco (Lycopodiaceae) new to Texas" in *Phytologia*, a journal not monitored by Thomson Reuters. This article (It could better be called a "Note.") describes our discovery of nodding club-moss in Texas. Nodding club-moss had been found in Louisiana and farther east, but never in Texas (Correll & Johnston, 1970; Flora of North America, 1993). Our article has not been cited in Thomson Reuters-monitored journals or Google Scholar. It, therefore, classifies as an uncited (i.e., not used) article. But this is far from the case. The information contained in the article has been incorporated into the phytogeographical literature of both Texas and North America. The information is used in Turner, Nichols, Denny, and Doron's (2003) *Atlas of the Vascular Plants of Texas*, Diggs, Lipscomb, Reed, and O'Kennon's (2006) *Illustrated Flora of East Texas* (It is cited in that work.), the USDA Plants database (www.plants.usda.gov/), and the NatureServe database (www.natureserve.org). In other words, it has been incorporated into the phytogeographical literature in both print and electronic form. The purpose of the article has been achieved. Those who use these databases will use that information, but will not cite the primary source or, for that matter, even know from whence the information came.

There are thousands of publications by botanists and zoologists identical to the one illustrated earlier, and the information in them is quickly incorporated into the general biological literature. It is not odd that they receive few citations, and none in Thomson Reuters-monitored journals; it would be odd if they were heavily cited and/or cited in such journals. The only report of this type that has received great attention in recent years is the rediscovery of the ivory-billed woodpecker published in *Science* (Fitzpatrick et al., 2005), and the only reason it was published there is because the ivory-billed woodpecker is "the Holy Grail of birders—the one sighting every birder fantasizes about" (Jackson, 2004, p. 1). If the rediscovery had been of *Agalinis caddoensis* (Caddo purple false-foxglove), a plant discovered and collected in 1913 in Caddo Parish, Louisiana, by Francis Whittier Pennell, Curator of Botany, Academy of Natural Sciences in Philadelphia, and never found again, it would never have found its way into *Science* or any other "top" journal (D.T. MacRoberts, 1978).

One final point: Turner et al.'s (2003) *Atlas of the Vascular Plants of Texas*, Diggs et al.'s (2006) *Illustrated Flora of East Texas*, the USDA Plants database (www.plants.usda.gov/), the NatureServe database (www.natureserve.org), and similar publications are databases used extensively by botanists, phytogeographers, and other biologists who do publish in Thomson Reuters-monitored journals. The information in those so-called *uncited* articles is used; it is just not being cited.

Example 2

In 1985, J.A. Matos and D.C. Rudolph published "The vegetation of the Roy E. Larsen Sandylands Sanctuary in the Big Thicket of Texas" in *Castanea*. An examination of Thomson Reuter (*Web of Knowledge*) shows one journal-article citation, and Google Scholar shows citations in one book, five journal articles, three unpublished reports, and one self-citation. A brief survey of journals and books in our possession relating to the Big Thicket region adds citations in two more books, 12 more journals, one master's thesis, and one more self-citation. In other words, counting only the published material and dropping the self-citations, Thomson Reuters had one citation, and Google Scholar had six citations, for a total of six citations (Google duplicated the one *Web of Knowledge* citation.) We were able to more than triple the citations, adding 14 for a total of 20. If theses, self-citations, and unpublished reports were added, the count rises to 27. This is a substantially different number than those given by Thomson Reuters and Google Scholar. But, further, the Matos and Rudolph article is one of the 1,742 floras used in the Qian et al. (2007) database discussed earlier. Consequently, it is used over and over again in any biogeographical publication by those authors, but it is not cited.

The point, of course, as a number of researchers have made, is that the various citation databases give widely different results (Bornmann, Mutz, Neuhaus, & Daniel, 2008; Meho, 2007). We would therefore suggest that before making statements like those of Aksnes (2003, p. 159), Opthof (1997, p. 2), and Meho (2007, p. 32) (whether the percentage uncited in Thomson Reuters or Google Scholar is 90 or 50 or 25 or 5), authors check a few of those uncited or seldom-cited publications to confirm that they are actually not cited

DOI: 10.1002/asi

or seldom cited; more important, that they are not or seldom used. *Not cited does not mean not used*.

Discussion and Conclusions

To summarize our findings in biogeography: (a) Most of the work used is not cited; (b) theoretical/analytical articles that are predominantly published in Thomson Reuters-monitored journals are cited, and data articles that provide the basis of the theoretical/analytical work and that are seldom published in Thomson Reuters-monitored journals are not cited; and (c) so-called *uncited* and *seldom-cited* articles are not only cited but used.

Our approach, unlike traditional citation analysis, does not begin and end with lists of citations but goes to the text and beyond to determine what the influences on scientific work actually are. This approach produces a very different dataset (and understanding) than that produced by examining only citations. Thus, we find that the articles (i.e., works) used for biogeographical databases are not only read but are redacted and their data *repeatedly* used. *But they are not cited*. Let us make this point clear: We are not talking here about articles in the journals not monitored by the Thomson Reuters but instead about the articles in the journals monitored by Thomson Reuters: the so-called "top 10% of journals," journals such as those from which our examples largely come: *Science*, *Nature*, *American Naturalist*, *Ecology*, *Journal of Biogeography*, *Annals of Botany*, *Systematic Zoology*, *Ecology Letters*, *Applied Vegetation Science*, *Castanea*, *Great Basin Naturalist*, and so on. We are talking about the thousands of floras and faunas, atlases, millions of herbarium and zoological specimens, thousands of unnamed fieldworkers, and unnamed persons consulted, the information synthesized into massive databases, the data collected by the Natural Heritage Programs, the data amassed by the USDA Plants database, and so on, that are used *but not cited*. We are talking about the literature not monitored by Thomson Reuters: the nonprestigious literature, the grey literature, the "notes" published by botanists and zoologists that describe range extensions, master's theses, and birdwatchers' distribution reports. This is information used (*but not cited*) in articles published in the journals scanned by Thomson Reuters. These workers and works are invisible to citation analysts who rely on standard citation databases.

There is a difference between influences that biogeographers cite and those they do not cite. The difference is obvious: The articles described in Appendix A are almost all theoretical/analytical. The author is trying to explain the pattern of plant and animal distribution, often using statistical methods, and comparing and contrasting his results with those of earlier biogeographers. The data sources are not contentious but are simply the thousands of observations that make up the data the author needs to make the analysis. Biogeographers are aware of the distinction. The implication for citation analysis is that if one is interested in "influence," then there is no distinction between articles manipulating the data and articles that provide the data. In science, the theoretical/analytical level is no more important than the observational level; in some ways, it is often less important because data can be and often are used by many individuals for different purposes. All science begins with observation: Alfred Russel Wallace in the Malay Archipelago and Alexander von Humboldt in South America made observations, and they and others have built on them. Their 200-year-old plant and animal collections are preserved in the Instituto de Ecologia y Sistematica in Havana, the Museum National d'Histoire Naturalle in Paris, and the British Museum of Natural History in London, and these collections continue to be used.

Biogeographers are not unique in their citing practices. In previous articles, we and others have shown that authors cite only a fraction of their influences, that many citations go to secondary, not primary, sources, and that an informal level of communication is not captured (e.g., Hicks & Potter, 1991; M.H. MacRoberts & MacRoberts, 1996 and references therein). Additionally, there are areas of research that are simply not cited. Protein crystallography is apparently one (Tainer, 1991), and botany and ecology show this pattern as well (see Leimu & Koricheva, 2005; Taborsky, 2009). In biochemistry, methods are cited, but not reagents (Seglen, 1996). As we noted in a previous article (M.H. MacRoberts & MacRoberts, 1996), floras are seldom cited even though they are indispensable to all levels of botanical research and are continuously used, as attested to by the tattered copies that litter workbenches in herbaria.

Evaluative citation analysis is based on the assumption that authors cite their influences, and even if not all are cited, those that are cited are more important (i.e., better, higher quality) than those not cited (Borgman & Furner, 2002; Cronin, 2005; Meho, 2007; Nicolaisen, 2007; Smith, 1981; Van Raan, 2004). Thus, "the number of citations an author attracts is a reliable measure of the attention the author receives from the scientific community, or, in other words, of the scientific impact of the author" (Krell, 2009, p. 6; also see Moed 2002). But as we have seen, without the vast uncited majority of articles there would be no biogeography at all: sine qua non. Why, for example, did Qian et al. (2007) cite fewer than 5% of their influences? Was it because they were citing only the best or because they were constructing a article and following the protocol of a discipline that requires only a general description of the database? Imagine what their article would look like if they cited all influences: not 12 pages of which 2 are citations, but 53 pages of which 43 are citations. Qian et al. undoubtedly judged some articles to have more merit than others, but they certainly did not judge all floras to have less merit than the 85 publications they cited. We certainly do not judge the works we used (*but did not cite*) in our articles described in Appendix A to have less (or more) merit than the works we cited. In none of the biogeographical articles we examined in the preparation of this article is there any implication of a quality gradient, but rather authors clearly acknowledge their dependence on works they do not cite. What is detectable in virtually all of these works is the awareness of the vast amount of data that are being drawn upon and the frustration of the author(s) in having little or no means to

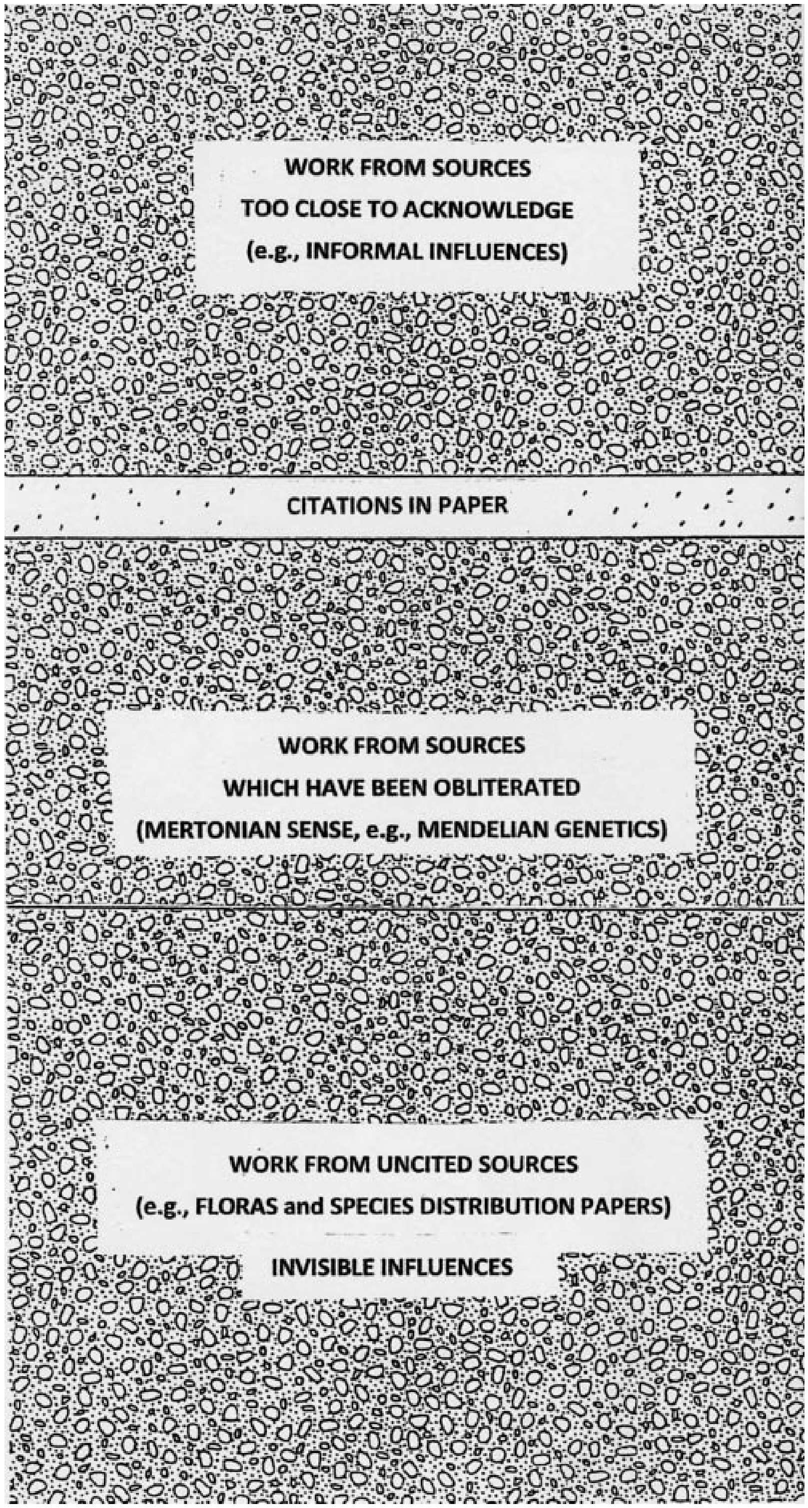

FIG. 1. The geology of citations (after Hicks & Potter, 1991).

describe, much less acknowledge it, which, as Szava-Kovats (2008) noted, is not new to the era of "Big Science." Nonetheless, there are a few institutional means to accomplish this: the online lists of sources mentioned by Qian et al. and the hard copy that is "available on request" (see Conant & Collins, 1998; Sorrie & Weakley, 2006, in Appendix A). In this manner, the "invisible" become "visible."

Using geology as a metaphor, Hicks and Potter (1991) presented a visual summary of influence and citing that we have modified slightly by making it consist of four strata instead of three (Figure 1) (Edge, 1979, used the iceberg metaphor much in the same way.) The upper stratum is the informal communication discussed so extensively by Edge, Gilbert, Mulkay, and others in sociology of scientific knowledge (as cited in

DOI: 10.1002/asi

Hicks & Potter, 1991; M.H. MacRoberts & MacRoberts, 1996), which as Hicks and Potter said "suggests that the most important influence and impact on researchers often come from those with whom they have a close informal contact—the colleagues with whom one has lunch, and the person on the next lab bench . . ." (p. 483). These influences are not included in the citations. Below this stratum are the citations.

> Citations are just a thin . . . band, sandwiched between the rock of eons. And it is this highly limited, highly unrepresentative, yet alluringly available band of rock that the ISI has fetishized and turned into a highly desirable and marketable commodity (Hicks & Potter, 1991, p. 483).

Below this layer is the work that was once cited but is now "obliterated;" work so accepted that originators are no longer cited (Merton, 1973): The "assumptions and historical residue of ideas and techniques incorporated into even the simplest article . . . " (Hicks & Potter, 1991, p. 481). But a stratum not mentioned by Hicks and Potter is that of works that are used, but not cited. This work has not gone through the "obliteration" process described by Merton (1973), but is immediately "obliterated" (i.e., absorbed) into other work. It is invisible except to the specialist who knows the discipline.

This four-layered geological metaphor also represents the probable extent of influence with citations, as Hicks and Potter (1991) noted, being only a "thin band" amounting to perhaps less than 10% of the total influence. One further point: Hicks and Potter made explicit that the citation layer is "unrepresentative:" It is, as in our observations, neither a random sample of all influences nor the best of the influences, but a very specific subset. In the case of biogeography, it consists largely of theoretical/analytical articles.

It seems to some that we are setting up a straw man, but this is not the case. Garfield (1997) believed that the question is "Do scientific articles cite most of the relevant articles that led up to the current work?" Our answer, of course, is a resounding *no* (Greenberg, 2009; M.H. MacRoberts, 1997). In the present case, without the data articles, sine qua non. But if all one wants to know is who is cited and how often in the journals monitored by Thomson Reuters, then turn to Thomson Reuters. But if one wants to know who contributes to science and how information is used and moves through the system, then another course is necessary. This will not involve "grand narratives and large-scale number crunching" (Cronin, 2005, p. 1505), but instead research on what goes on at the lab bench, what scientists do as they work and interact with colleagues, what they read, how they develop their data, and how they construct their articles within the culture of their disciplines (for references, see Hicks & Potter, 1991; M.H. MacRoberts & MacRoberts, 1996; also see Greenberg, 2009). As Nobel laureate Peter Medawar (1969) stated, "it is no use looking to scientific 'papers,' for they not merely conceal but actively misrepresent the reasoning that goes into the work they describe. Only unstudied evidence will do—and that means listening at the keyhole" (p. 32).

Acknowledgments

Michael Palmer made available unpublished information regarding the *Floras of North America Project* and the database for the 2007 article. The comments of three anonymous reviewers are appreciated, as is the effort of Blaise Cronin.

References

Aksnes, D.W. (2003). Characteristics of highly cited papers. Research Evaluation, 12, 159–170.

Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), Annual Review of Information Science and Technology, 36, 3–72.

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 64, 45–80.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H. (2008). Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics in Science and Environmental Politics, 8, 93–102.

Camacho-Minano, M., & Nunez-Nickel, M. (2009). The multilayered nature of reference selection. Journal of the American Society for Information Science and Technology, 60, 754–777.

Carr, W.R. (2005). An annotated list of G3/T3 and rarer plant taxa of Texas. Unpublished report. The Nature Conservancy of Texas, Austin, TX.

Cole, J.R., & Cole, S. (1972). The Ortega hypothesis. Science, 178, 368–375.

Conant, R. (1975). A field guide to reptiles and amphibians of eastern and central North America. Boston: Houghton Mifflin.

Conant, R., & Collins, J.T. (1998). A field guide to the reptiles and amphibians: Eastern and central North America. New York: Houghton Mifflin.

Cook, R.E. (1969). Variation in species density of North American birds. Systematic Zoology, 18, 63–84.

Correll, D.S., & Johnston, M.C. (1970). Manual of the vascular plants of Texas. Texas Research Foundation, Renner, TX.

Cronin, B. (2005). A hundred million acts of whimsy? Current Science, 89, 1505–1509.

Currie, D.J. (1991). Energy and large-scale patterns of animal- and plant-species richness. American Naturalist, 137, 27–49.

Currie, D.J., & Paquin, V. (1989). Large-scale biogeographical patterns of species richness in trees. Nature, 329, 326–327.

Darwin, C., & Wallace, A.R. (1858). On the tendency of species to form varieties; and on the perpetuation of the varieties and species by natural means of selection. Proceeding of the Linnean Society, 3, 45–62.

Diggs, G.M., Lipscomb, B.L., Reed, M.D., & O'Kennon, R.J. (2006). Illustrated flora of east Texas. Sida, Botanical Miscellany, 26, 1–1594.

Dobson, A.P., Rodriguez, J.P., Roberts, W.M., & Wilcove, D.S. (1997). Geographic distribution of endangered species in the United States. Science, 275, 550–553.

Edge, D. (1979). Quantitative measures of communication in science: A critical review. History of Science, 17, 102–134.

Elias, T.S. (1980). The complete trees of North America. New York: Reinholt.

Estill, J.C., & Curzan, M.B. (2001). Phytogeography of rare plant species endemic to the southeastern United States. Castanea, 66, 2–23.

Fitzpatrick, J.W., Lammertink, M., Luneau, M.D., Gallager, T.W., Harrison, B.R., Sparling, G.M., et al. (2005). Ivory-billed woodpecker (Campephilus principalis) persists in continental North America. Science, 308, 1460–1462.

Floras of North America Project. (1993). Retrieved March 12, 2009, from http://www.okstate.edu/artsci/botany/floras/

Flora of North America Editorial Committee. (2003). Flora of North America. New York: Oxford University Press.

Francis, A.P., & Currie, D.J. (2003). A globally consistent richness-climate relationship of angiosperms. American Naturalist, 161, 523–536.

Fridley, J.D., Peet, R.K., Wentworth, T.R., & White, P.S. (2005). Connecting fine- and broad-scale species-area relationships of southeastern U.S. flora. Ecology, 86, 1172–1177.

Garfield, E. (1997). Validation of citation analysis. Journal of the American Society of Information Science, 48, 962.

Greenberg, S.A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. British Medical Journal, 339. Retreived September 27, 2009, from http://www.bmj.com/cgi/content/full/339/jul20_3/b2680

Hall, E.R., & Kelson, K.R. (1959). The mammals of North America. New York: Ronald Press.

Hamilton, D.P. (1990). Publishing by—and for?—the numbers. Science, 250, 1331–1332.

Hamilton, D.P. (1991). Research papers: Who's uncited now? Science, 251, 25.

Harrison, J.A., Underhill, L.G., & Barnard, P. (2008). The seminal legacy of the Southern African bird atlas project. South African Journal of Science, 104, 82–84.

Heywood, V.H. (Ed.). (1993). Flowering plants of the world. New York: Oxford University Press.

Hicks, D., & Potter, J. (1991). Sociology of scientific knowledge: A reflexive citation analysis of science disciplines and disciplining science. Social Studies of Science, 21, 459–501.

Hoerman, H.L., & Nowicke, C.E. (1995). Secondary and tertiary citing: A study of referencing behavior in the literature of citation analysis deriving from the Ortega hypothesis of Cole and Cole. Library Quarterly, 65, 415–434.

Hosie, R.C. (1980). Native trees of Canada. Ottawa: Supply and Services Canada.

Jackson, J.A. (2004). In search of the ivory-billed woodpecker. Washington, DC: Smithsonian Books.

Kostoff, R.N. (2005). Science and technology transition metrics. Unpublished report online. Office of Naval Research, Arlington, VA. Retrieved September 29. 2009, from http://www.onr.Navy.mil/sci_tech/33/332/docs/metrics_book_review1.doc

Kostoff, R.N., Morse, S.A., & Oneu, S. (2007). Seminal literature of anthrax research. Critical Reviews in Microbiology, 33, 171–181.

Kreft, H., & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. Proceedings of the National Academy of Sciences, USA, 104, 5925–5930.

Krell, F.-T. (2009). The poverty of citation databases: Data mining is crucial for fair metrical evaluation of research performance. BioScience, 59, 6–7.

Larivière, V., Gingris, Y., & Archambault, E. (2009). The decline in the concentration of citations, 1900–2007. Journal of the American Society for Information Science and Technology, 60, 858–862.

Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? Trends in Ecology and Evolution, 20, 28–32.

Little, E.L. (1971). Atlas of United States trees. Washington, DC: USDA Forest Service.

Lomolino, M.V., Riddle, B.R., & Brown, J.H. (2006). Biogeography. Sunderland, MA: Sinaver.

Lonsdale, W.M. (1999). Global patterns of plant invasions and the concept of invisibility. Ecology, 80, 1522–1536.

Louisiana Natural Heritage Program. (2008). Rare plant list. Louisiana Department of Wildlife and Fisheries, Baton Rouge.

MacRoberts, B.R., & MacRoberts, M.H. (1995). Palhinhea cernua (L.) Vasconcellos & Franco (Lycopodiaceae) new to Texas. Phytologia, 78, 402–403.

MacRoberts, B.R., MacRoberts, M.H., & Brown, L.E. (2002). Annotated checklist of the vascular flora of the Hickory Creek Unit of the Big Thicket National Preserve, Tyler County, Texas. Sida, 20, 781–795.

MacRoberts, D.T. (1978). The status of Agalinis caddoensis Pennell. Phytologia, 40, 1–6.

MacRoberts, M.H. (1997). Rejoinder. Journal of the American Society for Information Science, 48, 963.

MacRoberts, M.H., & MacRoberts, B.R. (1986). Quantitative measures of communication in science: A study of the formal level. Social Studies of Science, 16, 151–172.

MacRoberts, M.H., & MacRoberts, B.R. (1987a). Testing the Ortega hypothesis: Facts and artifacts. Scientometrics, 12, 293–295.

MacRoberts, M.H., & MacRoberts, B.R. (1987b). Another test of the normative theory of citing. Journal of the American Society for Information Science, 38, 305–306.

MacRoberts, M.H., & MacRoberts, B.R. (1988). Author motivation for not citing influences: A methodological note. Journal of the American Society for Information Science, 39, 432–433.

MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. Journal of the American Society for Information Science, 40, 342–349.

MacRoberts, M.H., & MacRoberts, B.R. (1996). Problems of citation analysis. Scientometrics, 36, 435–444.

MacRoberts, M.H., & MacRoberts, B.R. (1997a). Citation content analysis of a botany journal. Journal of the American Society for Information Science, 48, 274–275.

MacRoberts, M.H., & MacRoberts, B.R. (1997b). Talinum rugospermum Holsinger new to Louisiana with notes on terete-leaved Talinum in Louisiana. Phytologia, 82, 86–93.

MacRoberts, M.H., & MacRoberts, B.R. (2008). Rare and endemic plants of the Big Thicket: A phytogeographical analysis. Journal of the Botanical Research Institute of Texas, 2, 1475–1479.

MacRoberts, M.H., MacRoberts, B.R., Sorrie, B.A., & Evans, R.E. (2002). Endemism in the West Gulf Coastal Plain: Importance of xeric habitat. Sida, 20, 767–780.

Matos, J.A., & Rudolph, D.C. (1985). The vegetation of the Roy E. Larsen Sandylands Sanctuary in the Big Thicket of Texas. Castanea, 50, 228–249.

McGervey, J.D. (1974). Citation analysis. Science, 183, 28–30.

McLaughlin, S.P. (1986). Floristic analysis of the southwestern United States. Great Basin Naturalist, 46, 46–65.

McLaughlin, S.P. (1989). Natural floristic areas of the western United States. Journal of Biogeography, 16, 239–248.

McLaughlin, S.P. (2006). Vascular floras of Sonoita Creek State Natural Area and San Rafael State Park: Arizona's first natural-area parks. Sida, 22, 661–704.

McLaughlin, S.P. (2007). Tundra to tropics: The floristic plant geography of North America. Sida, Botanical Miscellany, 30, 1–58.

Medawar, P. (1969). The art of the soluble. London: Harmondsworth.

Meho, L.I. (2007). The rise and rise of citation analysis. Physics World, January, 32–36.

Merton, R.K. (1973). The sociology of science. Chicago: University of Chicago Press.

Moed, H.F. (2002). The impact-factors debate: The ISI's uses and limits. Nature, 415, 731–732.

NatureServe. (2009). Retrieved September 27, 2009, from http://www.natureserve.org

Nicolaisen, J. (2007). Citation analysis. Annual Review of Information Science and Technology, 41, 609–641.

Nixon, E.S. Marietta, K.L., & McCray, M. (1980). Brachyelytrum erectum and Talinum rugospermum, new species to Texas and notes on Schoenolirion wrightii. Sida, 8, 355–356.

Opthof, T. (1997). Sense and nonsense about impact factor. Cardiovascular Research, 33, 1–7.

Palmer, M.W., Wade, G.L., & Neal, P. (1995). Standards for the writing of floras. BioScience, 45, 339–345.

Pendlebury, D.A. (1991). Science, citation, and funding. Science, 251, 1410–1411.

Poole, J.M., Carr, W.R., Price, D.M., & Singhurst, J.R. (2007). Rare plants of Texas. College Station: Texas A&M University Press.

Qian, H. (1999). Spatial pattern of vascular plant diversity in North America north of Mexico and its floristic relationship with Eurasia. Annals of Botany, 83, 271–283.

Qian, H. (2001). A comparison of generic endemism of vascular plants between east Asia and North America. International Journal of Plant Science, 162, 191–199.

Qian, H., Fridley, J.D., & Palmer, M.W. (2007). The latitudinal gradient of species-area relationships for vascular plants of North America. American Naturalist, 170, 690–701.

DOI: 10.1002/asi

Rahbek, C., & Graves, G.R. (2001). Multiscale assessment of patterns of avian species richness. Proceedings of the National Academy of Sciences, USA, 98, 4534–4539.

Seglen, P.O. (1996). Bruk av siteringer og tidsskriftimpaktfaktor til forskningsevaluering. Biblioteksarbejde, 48, 27–34.

Simpson, G.G. (1964). Species density of North American recent mammals. Systematic Zoology, 13, 57–73.

Smith, L.C. (1981). Citation analysis. Library Trends, 30, 83–106.

Sorrie, B.A., MacRoberts, M.H., MacRoberts, B.R., & Walker, S.B. (2003). Oxypolis ternata (Apiaceae) deleted from the Texas flora. Sida, 20, 1323–1324.

Sorrie, B.A., & Weakley, A.S. (2006). Conservation of endangered Pinus palustris ecosystem based on Coastal Plain centers of plant endemism. Applied Vegetation Science, 9, 59–66.

Stebbens, R.G. (1966). A field guide to the western reptiles and amphibians. Boston: Houghton Mifflin.

Stein, B.A. (2001). A fragile cornucopia: Assessing the status of U.S. biodiversity. Environment, 43, 11–22.

Stein, B.A., Kutner, L.S., & Adams, J.S. (Eds.). (2000). Precious heritage: The status of biodiversity in the United States. New York: Oxford University Press.

Storch, D., Evans, K.L., & Gaston, K.J. (2005). The species–area–energy relationship. Ecology Letters, 8, 487–492.

Szava-Kovats, E. (2002). Indirect-collective referencing (ICR) in the elite journal literature of physics: II. A literature science study on the level of communications. Journal of the American Society for Information Science and Technology, 53, 47–56.

Szava-Kovats, E. (2008). Phenomenon and manifestation of the "Author's effect of showcasing" (AES): A literature science study, I. Emergence, causes and traces of the phenomenon in the literature, perception and notion of the effect. Journal of Information Science, 34, 30–44.

Taborsky, M. (2009). Biased citation practice and taxonomic parochialism. Ethology, 115, 105–111.

Tainer, J.A. (1991). Science, citation, and funding. Science, 251, 1408.

Thomas, R.D. (2002). Cynosurus echinatus (Poaceae) new to Texas. Sida, 20, 837.

Thomas, R.D., & Allen, C.M. (1993–1998). Atlas of the vascular flora of Louisiana. Baton Rouge: Louisiana Department of Wildlife and Fisheries.

Turner, B.L., Nichols, H., Denny, G., & Doron, O. (2003). Atlas of the vascular plants of Texas. Sida, Botanical Miscellany, 24, 1–888.

USDA NRCS. (2009). The PLANTS Database. Baton Rouge, LA: National Plant Database Center. Retrieved September 27, 2009, from http://www.plants.usda

Von Humboldt, A. (1817). De distributione geographica plantarum. Paris: Lutetiae Parisiorum.

Van Raan, A.F.J. (2004). Measuring science. In M.F. Moed, W. Glanzel, & U. Schmoch (Eds.), Handbook of quantitative science and technology research (pp. 19–50). Dordrecht, The Netherlands: Kluwer.

Wallace, A.R. (1876). The geographical distribution of animals. New York: Harper.

Wilson, D.E., & Reeder, D.M. (Eds.). (1993). Mammal species of the world: A taxonomic and geographic reference. Washington, DC: Smithsonian Institution Press.

Withers, M.A., Palmer, M.W., Wade, G.L., White, P.S., & Neal, P.R. (1998). Changing patterns in the number of species in North American floras. In T.D. Sisk (Ed.), Perspectives on the land use history of North America: A context for understanding our changing environment (pp. 23–32). Reston, VA: U.S. Geological Survey, Biological Resource Division.

Zollner, D., MacRoberts, M.H., MacRoberts, B.R., & Ladd, D. (2005). Endemic vascular plants of the Interior Highlands, U.S.A. Sida, 21, 1781–1791.

Appendix A

A sample of biogeography publications. This sample of 20 publications was selected because they happened to be in our possession. Hundreds more could be included.

Conant, R., & Collins, J.T. (1998). A field guide to the reptiles and amphibians: Eastern and central North America. New York: Houghton Mifflin. An earlier edition of this is one of the main data sources for the Currie (1991) article (listed later). In their six-page acknowledgments, the authors thanked over 250 individuals who helped in various ways, but admitted that "space limitations preclude the listing of all of their names. Their contributions, however, are carefully recorded in our workbooks, our files, and on our base maps, and their findings, properly accredited, are available on request" (p. ix). They continued: "Credit must also be given to the hundreds of herpetologists whose published works have been consulted, often repeatedly, and whose names would appear frequently if this were a documented, scientific publication. We are deeply grateful to them, but there is room to list only relatively few of their works on the pages devoted to references" (p. xiv). Even so, 141 references are made, virtually all of which are regional reviews, with such titles as *Amphibians and Reptiles of Kentucky*, *The Herpetology of Michigan*, and *Snakes of Canada*. The works that are synthesized by these works *are not cited* by Conant and Collins (1998).

Currie, D.J. (1991). Energy and large-scale patterns of animal- and plant-species richness. American Naturalist, 137, 27–49. Here, Currie used not only the same data from his previous study of trees but adds mammals, birds, amphibians, and reptiles. He used a number of secondary syntheses of the distribution of these animals based on the work of thousands of individuals over centuries (e.g., Conant, 1975; Cook, 1969; Hall & Kelson, 1959; Stebbens, 1966). In other words, he used thousands upon thousands of records, published and unpublished, *but cited only four syntheses.*

Currie, D.J., & Paquin, V. (1989). Large-scale biogeographical patterns of species richness in trees. Nature, 329, 326–327. These authors went to the secondary literature and used three sources (Elias, 1980; Hosie, 1980; Little, 1971) to determine tree-species distribution across North America. Little (1971) said: "The maps in this volume have been compiled from various sources .... Principal records on tree distribution include publications, herbarium specimens, field work, and review by botanists, foresters, and others. The more detailed publications consulted are listed under Selected References .... *The list of publications and persons consulted is too long for citation here*[italics added] .... State floras, manuals, and catalogs have been consulted" (p. 5). Little cited about 250 publications; these, in turn, cite thousands of publications. *The primary sources for the data are not cited.*

Dobson, A.P., Rodriguez, J.P., Roberts, W.M., & Wilcove, D.S. (1997). Geographic distribution of endangered species in the United States. Science, 275, 550–553. This publication is based on a very large database that is referred to in a single reference "(9)," which is explained in the "References

DOI: 10.1002/asi

and Notes" as: " 9. United States Environmental Protection Agency, *Endangered Species by County Database* (Office of Pesticide Programs, Washington, DC, 1995)." The "document" referred to does not exist as such but is a compilation of many documents from agencies such as the USDA, Natural Heritage programs, and the EPA. Its true nature is revealed in the "acknowledgments," which state: "We thank L. Turner and M. Hood at the Environmental Protection Agency for comments on the manuscript and for providing us with the raw data for the analysis . . ." (p. 553). *But no primary data source is cited.*

Estill, J.C., & Curzan, M.B. (2001). Phytogeography of rare plant species endemic to the southeastern United States. Castanea, 66, 2–23. Based heavily on secondary and tertiary sources and on herbarium specimens, in their acknowledgments, they said: "Most importantly, the data allowing for this review is founded on the fieldwork of hundreds of botanists working in the Southeast over the past two centuries, without them this work would not have been possible" (p. 21). *But they are not cited.*

Francis, A.P., & Currie, D.J. (2003). A globally consistent richness-climate relationship of angiosperms. American Naturalist, 161, 523–536. "Angiosperm family richness was calculated based on range maps taken from Heywood (1993)," referring to V.H. Heywood, Consulting Editor, *Flowering Plants of the World.* New York: Oxford University Press, a 336-page reference work written by 44 experts. The number of works used in the production of this compendium is not given, but the "Acknowledgments" state: "The Publishers acknowledge the following reference sources" (p. 336), and listed 57 works, among these such notable works as the entire *Transactions of the Linnean Society of London* (1791–present); and Engler, H.G.A. (Ed.). (1900–1953). *Das Pflanzenreich. Regni vegetabilis Conspectus* (pp. 1–107). Berlin. *The number of individual works used by these 44 experts is not indicated, but it would involve thousands spanning centuries.*

Fridley, J.D., Peet, R.K., Wentworth, T.R., & White, P.S. (2005). Connecting fine- and broad-scale species-area relationships of southeastern U.S. flora. Ecology, 86, 1172–1177. In this example, the authors use all recently collected data of a very particular format. In their unusually lengthy acknowledgments, they stated: "We gratefully acknowledge the >600 individuals who have participated in collection of the Carolina Vegetation Survey data set . . ." (p. 1176). While these are not individual reports, more than 600 individuals helped collect specific data from plots, and these individuals had to know sufficient botany to identify plants. *None is individually acknowledged nor are their names in any citation database.*

Kreft, H., & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. Proceedings of the National Academy of Sciences, USA, 104, 5925–5930. "We analyzed the species richness of vascular plants . . . across 1,032 geographic units worldwide . . . . Geographical units represent natural . . . or political units . . . and were derived from floras, checklists, and other literature sources. The original data set consists of >3,300 species-richness accounts referring to >1,800 geographic units" (p. 5929). *None is cited.*

Lonsdale, W.M. (1999). Global patterns of plant invasions and the concept of invisibility. Ecology, 80, 1522–1536. In the methods section, he noted: "Data on the number of exotic and native plant species from sites around the world came from various sources, mostly compilations (Table 2)" (p. 1525). Lonsdale cited 15 compilations, but *primary sources are not cited.*

MacRoberts, M.H., & MacRoberts, B.R. (2008). Rare and endemic plants of the Big Thicket: A phytogeographical analysis. *Journal of the Botanical Research Institute of Texas*, 2, 1475–1479. We wrote that "using the Carr (2005) and Poole et al. (2007) annotated lists of rare Texas species and several sources for state distribution (e.g., Carr, 2005; Diggs et al., 2006; Poole et al., 2007; Turner et al., 2003) we mapped species richness . . . across Texas" (p. 1475). The Poole, Carr, Price, and Singhurst (2007) publication has about 1,640 references; the Carr (2005) publication has about 400 references; the Turner, Nichols, Denny, and Doron (2003) publication is based on approximately 500,000 herbarium specimens collected by hundreds of botanists; the Diggs, Lipscomb, Reed, and O'Kennon (2006) publication has about 4,000 references. Another publication we used was the *Flora of North America*, a 26-volume work coauthored by hundreds of botanists. This 16,000 page work contains approximately 45,000 references and was edited not by an individual or individuals but by a "committee." Our 2008 publication, on the other hand, has only 24 references. We used the data from hundreds of individuals *without citing them and instead cited syntheses.*

MacRoberts, M.H., MacRoberts, B.R., Sorrie, B.A., & Evans, R.E. (2002). Endemism in the West Gulf Coastal Plain: Importance of xeric habitat. Sida, 20, 767–780. In the methods section, the authors stated: "To develop a list of endemics, we obtained species distributional data from a wide variety of sources, including regional, state and local floras, floristic atlases, published papers, and monographs . . . . Natural Heritage Program databases, our own field work, and an examination of herbarium specimens notably at ASTC, BRIT, Corpus Christi Museum of Science and History, GH, LSU, LSUS, NCU, NLU, SBSC, SHST, TAMU (on line), TEX, VDB, and WWF" (p. 768). As before, we used the data from hundreds of individuals *without citing them and instead cited only syntheses.*

Qian, H. (1999). Spatial pattern of vascular plant diversity in North America north of Mexico and its floristic relationship with Eurasia. Annals of Botany, 83, 271–283. In his methods section, Qian gave his data sources: "To record geographic distribution (presence/absence) information and determine the native/exotic status of each species in each of the twelve regions utilized in this study, over 200 reference books (including continental, regional, state/provincial, and local floras, checklists, atlases, monographs, and theses) and more than 1000 journal papers . . . were consulted" (p. 272). *None is cited.*

Qian, H. (2001). A comparison of generic endemism of vascular plants between east Asia and North America.

DOI: 10.1002/asi

International Journal of Plant Science, 162, 191–199. Qian stated that "The database for the endemic genera of East Asia and North America was assembled largely on a larger database, 'Phytogeographic Checklist of the Vascular Plant Genera of the Northern Hemisphere [VPGNH]' . . . which has been developed through an effort of 2 decades. The VPGNH was compiled based on an extensive review of the literature, including continental, regional, state/provincial, and local floras, checklists, atlases, monographs, theses, and journal papers pertinent to the floras of the two continents" (p. 192). *None is cited.*

Rahbek, C., & Graves, G.R. (2001). Multiscale assessment of patterns of avian species richness. Proceedings of the National Academy of Sciences, USA, 98, 4534–4539. This publication is particularly interesting because the authors skipped secondary and tertiary sources and went straight to the primary data: museum collections and sight records. In their acknowledgments, they noted: "Primary distributional data were derived from the collections of the Academy of Natural Sciences (Philadelphia); American Museum of Natural History (New York); Carnegie Museum of Natural History (Pittsburgh); Colleccion Ornithologica Phelps (Caracus, Venezuela) . . ." (p. 4538), and *29* additional museums. And this does not include the "documented sight records" (p. 4538). We do not know how many specimens this involves, but undoubtedly thousands collected by hundreds of individuals. But *no one is cited.*

Simpson, G.G. (1964). Species density of North American recent mammals. Systematic Zoology, 13, 57–73. We include this classic biogeographical article in our list because it deals with mammals, not plants. Simpson explained his database briefly, one that would be understood by all mammalogists: "The species lists and distributions were based on Hall and Kelson (1959) . . ." (p. 57). The Hall and Kelson work, entitled *The Mammals of North America*, consists of two volumes that synthesize the distributional information on all North American mammals and contain about 1,450 references by several hundred authors. Simpson's article has only 13 references, but *he does not cite any of the references in Hall and Kelson,* although they are the source of virtually all of the information he used.

Sorrie, B.A., & Weakley, A.S. (2006). Conservation of endangered Pinus palustris ecosystem based on Coastal Plain centers of plant endemism. Applied Vegetation Science, 9, 59–66. In their methods section, Sorrie and Weakley stated: "We have reviewed hundreds of taxonomic and floristic papers, too numerous to cite here, but a list is available on request. We have also conducted supplementary distributional studies involving more than twenty herbaria (A, AMES, AUA, DUKE, FLAS, FSU, GA, GH, IBE, LSU, MISSA, NCSC, NEBC, NCU, NLU, US, USA, USCH, SWSL, VSC, VDB at BRIT, and several private herbaria) to verify distributions . . ." (p. 60). In their acknowledgments, they said: "We wish to thank . . . Coastal Plain botanists and Southeastern herbaria for their contributions to the set of data on which this paper is based" (p. 65). There is clearly a lot of influence here, but *it is not cited.*

Stein, B.A. (2001). A fragile cornucopia: Assessing the status of U.S. biodiversity. Environment, 43, 11–22. This work deals with all sorts of animals and plants and draws almost all of its information from *Precious Heritage: The Status of Biodiversity in the United States* (Stein, Kutner, & Adams, 2000). This single source brings together more than a quarter-century of information from the network of Natural Heritage programs that "operate in all 50 U.S. states, across Canada, and in a dozen countries in Latin America and the Caribbean, with each center maintaining detailed maps and computer records about the species and ecosystems that are of greatest conservation concern within the state" (Stein, 2001, p. 15). This work is, however, not entirely confined to employees of Natural Heritage programs but also involves many cooperators under contract as well as volunteers. As a case in point, over the past 20 years, we have supplied both the Texas and Louisiana Natural Heritage programs with data on rare species on both a volunteer and a contractual basis. *Precious Heritage* relies on many major syntheses such as Wilson & Reeder (Eds.). (1993). *Mammal Species of the World: A Taxonomic and Geographic Reference. Precious Heritage* has about 400 references, Stein's (2001) "A fragile cornucopia . . ." has only 11. *Virtually none of this work is cited except indirectly through secondary sources.*

Storch, D., Evans, K.L., & Gaston K.J. (2005). The species-area-energy relationship. Ecology Letters, 8, 487–492. These biogeographers, whose work is on birds, used "two extensive data sets on avian distributions in different biogeographic regions" (p. 488). These are the South African Bird Atlas Project and the New Atlas of Breeding Birds in Britain and Ireland: 1988–1991. The former involved "more than 5000 'citizen scientists,'" and the latter required the help of "thousands of birdwatchers" (Harrison, Underhill, & Barnard, 2008, p. 82). As Harrison, Underhill, and Barnard (2008) noted: "By 2007, the [South African Bird Atlas Project] database had spawned 50 research publications and eight Ph.D.s and master's degrees" (p. 82). *But the individuals contributing to the database are not cited.*

Withers, M.A., Palmer, M.W., Wade, G.L., White, P.S., & Neal, P.R. (1998). Changing patterns in the number of species in North American floras. In T.D. Sisk (Ed.), Perspectives on the land use history of North America: A context for understanding our changing environment (pp. 23–32). Reston, VA: U.S. Geological Survey, Biological Resource Division. In their methods section, the authors stated that they ". . . rely on information collected by hundreds of different scientists over more than 200 years . . . . We used many different strategies to acquire the floras used in this study. We searched university and government libraries, used computerized index searches, scanned more than 30 key botanical journals, read the bibliography and reference list of botanical surveys and floras, used government document searches, corresponded with other botanists, and even posted e-mail requests on relevant bulletin boards. Still we are constantly discovering new floras" (p. 27). *None is cited.*

Zollner, D., MacRoberts, M.H., MacRoberts, B.R., & Ladd, D. (2005). Endemic vascular plants of the Interior

DOI: 10.1002/asi

Highlands, U.S.A. Sida, 21, 1781–1791. They noted: "We searched all available sources of information, including extensive consultation with knowledgeable experts, to determine global ranges of species in the vascular flora of Arkansas, Illinois, Kansas, Missouri, and Oklahoma. This included general references . . . and more specific papers . . . . Also included are various lists of species of concern kept by the Arkansas, Illinois, Kansas, Missouri, and Oklahoma Natural Heritage programs and the Ozark, Ouachita, Mark Twain, and Shawnee National Forests" (p. 1783). *But they are not cited.*

Appendix B

These three examples of uncited articles suffice to make the point. All of these articles read alike, so a small sample is sufficient to make the point.

We (M.H. MacRoberts & MacRoberts (1997b) published a short article: "Talinum rugospermum Holsinger new to Louisiana with notes on terete-leaved Talinum in Louisiana." This article described our discovery of rough-seeded flameflower in Louisiana. This species had not been previously reported for the state and had only been recently reported for Texas (Nixon, Marietta, & McCray 1980). Our article has not been cited in Thomson Reuters-monitored journals or Google Scholar. Like our previous example, this one also classifies as one of the uncited (i.e., not used) articles. But, the information contained in it has been incorporated into the phytogeographical literature of both Louisiana and North America. The information is used in Thomas and Allen's (1998), Atlas of the Vascular Flora of Louisiana, Flora of North America Vol. 4 (2003) (It is cited in that work.), the USDA Plants database (www.plants.usda.gov/), and the NatureServe database (www.natureserve.org). It also has been incorporated into the *Louisiana Natural Heritage Database* and is included in their *Louisiana Rare Plant List* (Louisiana Natural Heritage Program, 2008). Like our *Palhinhaea* article, this one has been incorporated into the phytogeographical literature in both print and electronic forms. And, like our *Palhinhaea* article, its purpose has been achieved: Its findings are imbedded in the knowledge base of plant distributions.

In 2002, R.D. Thomas published a note, *Cynosaurus echinatus (Poaceae) new to Texas.* While occurring in Louisiana, Arkansas, and Oklahoma, *Cynosaurus echinatus* (bristly dogtail) had not been reported in Texas. This article has not been cited in Thomson Reuters-monitored journals or Google Scholar. The Thomas report was incorporated into Turner, Nichols, Denny, and Doron (2003), *Atlas of the Vascular Plants of Texas*, included and cited in Diggs, Lipscomb, Reed, and O'Kennon (2006), *Illustrated Flora of East Texas*, and in both the USDA Plants (www.plants.usda.gov/) and NatureServe (www.natureserve.org) databases. The information is included in Vol. 24 of the *Flora of North America*, but cited only in the "geographic bibliography." But the *Flora of North America*, the premier synthesis of botany in our time, is not monitored by Thomson Reuters.

In 2003, Sorrie, MacRoberts, MacRoberts, & Walker published *Oxypolis ternata (Apiaceae) deleted from the Texas flora*. The article described the reexamination of the herbarium specimens that constituted the evidence for including *Oxypolis ternata* (Piedmont cowbane) in the Texas flora and the discovery that the specimens were misidentified *Oxypolis rigidior*, a common species in east Texas. Neither Google Scholar nor Thomson Reuters mentions this article. But, *Oxypolis ternata* is not included in Turner, Nichols, Denny, and Doron's (2003) *Atlas of the Vascular Plants of Texas*. Poole, Carr, Price, and Singhurst (2007) in *Rare Plants of Texas* cited the article and described that the species has been removed from the Texas flora. NatureServe (www.natureserve.org) removed it from their database, but USDA Plants (www.plants.usda.gov/) has not done so yet (as of this time). One other source has picked it up. A.S. Weakley in his online *Flora of the Carolinas, Virginia, Georgia, and Surrounding Areas* cited it. Thus, it is clear that this article, while not cited in the Thomson Reuters-monitored journals, has been used, and its purpose is being fulfilled.

DOI: 10.1002/asi