



ORIGINAL
ARTICLE

The botanist effect: counties with maximal species richness tend to be home to universities and botanists

Daniel E. Moerman^{1*} and George F. Estabrook²

¹Behavioral Sciences, University of Michigan-Dearborn, Dearborn, MI, ²Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Aim To investigate the distribution of local flowering plant species richness in areas surrounding American universities.

Methods Species richness in university counties was compared with neighbouring counties. Data were derived from *Synthesis of the North American Flora* (<http://www.phylosystems.com/prepublication>). Probabilities of the resultant distribution were calculated, and the results were also simulated.

Results In almost every case there were more species reported in the university county than in its neighbours. Several possible explanations were considered. We conclude that the key element is the presence of botanists in these counties who have, apparently, paid more attention to plants near at hand, and found more species of them there.

Main conclusions Ecologists must be aware that numerical data that appear very solid, collected over many decades, may represent not only the qualities of 'nature' but also something of the collectors of the data.

Keywords

Collecting bias, comparative analysis, distributional data, inventory data, North American flora, plant diversity, plant richness, sampling bias, scientific behaviour, simulation.

*Correspondence: Daniel Moerman, 6515 Cherry Hill Rd, Ypsilanti, MI 48198, USA. E-mail: dmoerman@umich.edu

INTRODUCTION

In a classic paper in this journal, C. Barry Cox argued that the apparent taxonomic diversity of the Holarctic Kingdom had been 'artificially inflated' because, he noted, quoting Good (1974), botanists have given 'exaggerated importance' to that part of the world (Cox, 2001). Cox gives no explicit reason why this may be the case. Conversely, Lomolino (2004, p. 296) has expressed serious concern over the 'Wallacean shortfall', where 'the paucity of information on the geography of nature' seriously inhibits our ability to protect endangered species.

It has been argued that in the still largely unknown Brazilian Amazon, botanical collection 'concentrated in widely scattered centres' has led to a bias that has distorted arguments about the evolutionary history of the area (Nelson *et al.*, 1990, p. 714). On a much smaller geographical scale, it has been shown that the number of botanical collecting trips to each of the Galapagos islands was a better predictor of species number than any physical characteristics of the respective islands (Connor & Simberloff, 1978). In this paper we provide

evidence over a substantial area of the North American continent, suggesting that botanical diversity increases with the presence of botanists, a phenomenon we call the 'botanist effect.' We believe that the principles underlying this rather different case of collection bias may have widespread applicability, and should serve as a cautionary tale for those doing broad-scale comparative research in a number of areas in science, particularly ecology.

We noted that Washtenaw County, where the University of Michigan was established in the mid-19th century, has more species of flowering plants reported to occur within it than any of the counties with which it shares boundaries, according to the best available data sources (Table 1). We learned this, and most of what follows, from the compilation of data available from Version 2.0 of the *Synthesis of the North American Flora* (Kartesz & Meacham, 2005); most of the data for Michigan in the *Synthesis* are derived from *Michigan Flora* (Voss, 1972–1996), which we also consulted. This interesting, but by itself perhaps fortuitous, coincidence of local maximal species richness with the location of a major state university prompted

us to investigate whether this was an isolated coincidence or a widespread phenomenon.

METHODS

First we located the major university campuses in the state of Michigan, and for each noted the county in which it was located. Using data from the *Synthesis* we noted the number of flowering plant species reported from this county, and from each county that shared a common boundary. The Michigan data led us to consider additional university counties, first across the mid-west, then further afield. The process involved a number of factors, as follows:

1 We selected major, generally well known, public research universities (with a few exceptions for private institutions contiguous to public ones), usually no more than one or two per state, for which county-level data were available (they were not available for Ohio).

2 We allowed a maximum of 11 surrounding counties, selecting universities that were relatively isolated; having made an exception for the University of Michigan (our first case), we made subsequent exceptions for North Carolina, Colorado–Wyoming, Utah and Virginia; note that before selecting this sample we were not necessarily aware of the contiguity of some of these universities (we were aware of the North Carolina ‘research triangle’ but not the Colorado–Wyoming triad).

3 Reasoning that if a county were 10 times the size of a neighbour, it might have more plants than its neighbours for this reason alone, we excluded states with dramatically variable county sizes (Nevada, California), or if all the counties were very large, hence more than ordinarily likely to include a university (Arizona).

4 We excluded New England, with its very erratic county sizes and boundaries, and many contiguous universities (public and private).

5 Once we selected a university, we never subsequently excluded it from the analysis unless, by then, it failed to meet one of our criteria (e.g. more than 11 surrounding counties).

We analysed these data statistically under the null hypothesis that the number of flowering plant species (richness) in a county is not related to whether or not a state university has been founded there. We denote with $m[i]$ the number of counties that share a common boundary with the county in which university i is located, plus one to include the university county itself. We can now state the null hypothesis more precisely as follows: the $m[i]$ richness values for the $m[i]$ counties in the case of university i could have been associated with these counties in any possible way with equal probability. Under this hypothesis, the probability $P[i]$ that the largest richness value is associated with the university county is simply $P[i] = 1/m[i]$.

In the case of Washtenaw County with the University of Michigan, there are also universities in two of the contiguous counties, Wayne and Oakland, so we included the three counties contiguous with these as well (Table 1, case 1). Under our null hypothesis, among these 10 counties the probability that the three with universities would also be the three richest

in species is the number of ways to choose three from 10 (in no particular order), which is:

$$P[i] = [1/(10 \times 9 \times 8)]/[1/(1 \times 2 \times 3)] = (1/720)/(1/6) = 1/120.$$

To test several universities simultaneously in this way, we extended the null hypothesis to state that one university occurs in a county of specified species richness independently of the species richness of the county in which another university might occur. Thus the significance of the observed pattern for four universities in Michigan (Table 1, cases 1–4) would be:

$$P = \prod_i^4 P(i) \text{ in which } \prod \text{ means product.}$$

In this case, P is 1 in 14 400, that is, $1/(120 \times 5 \times 4 \times 6)$. Because the results for Michigan were very significant, we assembled similar data for other large universities throughout the USA, selected as described earlier, using the national data base compiled by Kartesz & Meacham (2005) from state floras and other floristic surveys. These data were analysed numerically in the same way as described above.

The data for many central US universities showed a clear association between the county where a university was established, and locally maximal species diversity. We considered several propositions that might account for the pattern.

To calculate the realized significance of the observed data reported in Table 1, we cannot simply multiply together the $P[i]$ for all 37 cases, because not all are instances of the university counties containing the most species locally. Instead, we wrote a computer program to simulate the null hypothesis to create tables like Table 1. For each of the 37 cases in Table 1, this program replaced the $m[i]$ richness values with an equiprobably chosen permutation of them over the $m[i]$ counties. Then it counted the number of cases in which the universities were in the most species-rich counties, and used this number as our test statistic. We simulated 1 million tables like Table 1 to estimate a probability distribution for this test statistic under our null hypothesis.

RESULTS

Our primary data are displayed in Table 1, in which we have grouped together neighbouring counties containing universities, as in south-eastern Michigan with neighbouring Washtenaw, Oakland and Wayne counties, respectively the homes of the University of Michigan and Eastern Michigan University, Oakland University (and the Cranbrook Institute of Science, which has a herbarium), and Wayne State University. All the counties neighbouring these three are listed; none has more species than any of the three. We located two other cases of three neighbouring counties with universities, one in North Carolina, and one with three crossing the state line between Colorado and Wyoming. Of 1 million tables like Table 1 simulated under the null hypothesis, none of them had more than 17 of the 37 cases in which the universities were in the most species-rich counties. Thus the realized significance of our observed 31 cases is much less than one in a million.

Table 1 Species richness in US counties with universities, and neighbouring counties

#	University	County	1	2	3	4	5	6	7	8	9	10	11	12	Notes
1	U of Michigan	Washtenaw	1467	933	530	578	796	850	843	860					
	Oakland U, Cranbrook	Oakland	1313												
	Wayne State U	Wayne	1301												
2	Western Michigan U	Kalamazoo	1482	638	721	746	949								1
3	Northern Michigan U	Marquette	847	626	410	645									2
4	Michigan State U	Ingham	1124	431	881	816	416	623							
5	U of Mich Biol Sta.	Cheboygan	1073	692	485	436	845	1031							3
6	Indiana U	Monroe	991	644	680	810	1095	708	682						
7	U of Illinois	Champaign	1132	443	936	662	817								
8	U of Arkansas	Washington	1247	1075	688	571	539								
9	U of Missouri	Boone	1250	683	486	763	648	510	554						
10	U of Minnesota	Hennepin	1100	906	881	882	706	493	802	717					
11	U of Iowa	Johnson	981	817	275	1019	564	309	318						
12	Iowa State U	Story	1057	259	556	590	619	190							
13	U of New Mexico	Bernalillo	1016	1290	1004	692	489	891							4
14	U of Texas	Travis	1270	485	279	820	328	619	554						
15	Penn State	Centre	1415	914	740	717	1322	1270	922	642					5
16	U of New Hampshire	Strafford	1226	1295	1145	737	1034								6
17	U of South Carolina	Richland	1256	979	989	844	1116	693							
18	North Carolina State U	Wake	1249	590	866	648	914	980	856	657	553				7
	U of North Carolina	Orange	1233												
	Duke U	Durham	1059												
19	U of Georgia	Clarke	1046	534	156	93	250	237							
20	U of Alabama	Tuscaloosa	1306	403	710	851	719	572	348	313					
21	U of Mississippi	Lafayette	707	390	295	433	196	178	255	131					
22	U of Louisiana	Lafayette	862	713	543	819	801								8
23	U of Utah	Utah	1362	604	414	1017	886	738	931	1024	896				
	Brigham Young U	Salt Lake	1302												
24	Colorado State	Larimer, CO	1574	811	766	888	690	197	697	638	928	767	1333		
	U of Colorado	Boulder, CO	1535												
	U of Wyoming	Albany, WY	1416												
25	U of North Dakota	Grand Forks	536	329	789	182	383	287							9
26	U of South Dakota	Clay	528	368	479	464	378	255							10
27	U of Nebraska	Landcaster	183	159	155	144	112	107	137						
28	U of Kansas	Douglas	1154	771	855	821	818	715	841						
29	U of Oregon	Lane	1666	692	1076	1276	1053	1687							
30	U of Kentucky	Fayette	964	329	713	1047	714	531	183						
31	U of Tennessee	Knox	1340	523	624	334	1163	1272	523	841					
32	U of Oklahoma	Cleveland	1131	856	280	496	467	447							
33	Washington State U	Whitman	1075	1055	426	911	763	689	486	676	313	412			

Table 1 *continued*

#	University	County	1	2	3	4	5	6	7	8	9	10	11	12	Notes
34	U of Washington	King	1201	829	1317	1270	1106	396							
35	U of Florida	Alachua	1503	438	508	967	1203	1198	582	920					
36	U of West Virginia	Monongalia	1442	771	1094	1168	609	750	804						
37	James Madison U	Augusta	1369	1215	862	831	845	933	924	1079	1233	1014	942	1045	
	U of Virginia	Albemarle	1284												

Maximal richness indicated in bold type. Numbered counties ordinarily begin to the north or north-east of the university counties, and circle around in a clockwise manner (Kartesz & Meacham, 2005).

¹County 5 (Van Buren) borders Lake Michigan.

²Marquette County is bordered on the north by Lake Superior.

³The UM Biological Station is a field station for biological research and education in northern Michigan on the shore of Douglas Lake. The Station is just a few miles east of the boundary between Cheboygan and Emmett Co (no. 6 in the table). The lake continues west into Emmett County which is, therefore, accessible by boat from the Station.

⁴Bernalillo County is much smaller than all the surrounding counties.

⁵These counties vary significantly in size.

⁶County 2 (York County) is in Maine.

⁷Three major North Carolina universities (UNC, NC State, Duke) are in neighbouring counties. The remaining eight counties surround those three.

⁸The universities of Mississippi and Louisiana are each in Lafayette County. The four counties surrounding Lafayette County, LS, are each about twice its size; county 4 borders the Gulf of Mexico.

⁹County 3 (Polk County) is in Minnesota.

¹⁰Counties 4 and 5 (Dixon, Cedar) are in Nebraska.

The counties home to James Madison University and the University of Virginia are surrounded by 11 neighbouring counties, largely because of the very irregular shapes of Virginia's counties (Table 1, case 37). None of the 11 neighbours has more reported plant species than the two counties with universities. Assuming our null hypothesis of randomness, the probability of this particular constellation of species richness in 13 counties is approximately 1/78.

Overall, the pattern we saw for Washtenaw County is true for most counties with universities. There are only five exceptions among the 37 instances we compiled. Indiana University, the University of Iowa, the University of Kentucky and the University of Oregon are in counties second-richest in diversity among neighbours. One, the University of Washington, is third in a set of six neighbouring counties.

DISCUSSION

How can we account for these surprising findings? A first possibility is that, somehow, universities were situated in counties particularly well endowed with botanical resources. However, these universities were established over a period of 200 years under dramatically different cultural, historical and educational constraints. Many of the institutions listed in Table 1 are 'land-grant colleges', created under the terms of the Morrill Act of 1862 designed to provide instruction in agriculture, the mechanic arts and military tactics, as well as the liberal arts and sciences more generally. In the 19th century, the study of agriculture often entailed the general study of plants, and many land-grant colleges to this day have 'weed science' programs and the like. A Google search for 'weeds' on the Michigan State University website (MSU is Michigan's land-grant college with a college of agriculture) produced 30 800 hits; the same search at the University of Michigan website yielded 341 hits. There are two major reasons that lead us to reject this proposition, however. These universities were sited before the botanical research on which our data are based was accumulated; no-one knew the species-richness data until very recently. Second, a review of the record shows that many other factors were clearly important in locating these institutions; they were primarily political and economic (Burke, 1906; Hatch, 1967). The University of Mississippi (Table 1, case 21) was apparently placed in Oxford because it was deemed a healthier environment than the surrounding, lower-lying counties (Butterworth, 1980; Sansing, 1999). Moreover, it seems plausible that, if this effect were due to the appearance of species deliberately introduced by agricultural scientists, there would, in counties with land-grant and/or agricultural schools, be an excess of exotic (non-native) species. We searched our data for states with two universities listed, one a land-grant college, the other not, and determined the proportions of exotic plants in eight counties for four such pairs of colleges. In one case (University of Michigan, Michigan State University), the proportion was the same in both counties (24% and 23%, respectively). In three other

states (NC, CO, WA), there was actually a lower percentage of exotics in the counties with the land-grant universities.

We considered the possibility that, even given our precautions mentioned above, university counties were systematically larger than neighbouring counties, hence having more plants. We used a null hypothesis that a university county and its neighbours all independently sample the same continuous size distribution, specific to that particular cluster of counties, that is, the distributions may vary independently from cluster to cluster around the country, but not within a cluster. Under this hypothesis, the probability would be 0.5 that an arbitrarily selected county next to a university would have a smaller area.

Using a random number generator, we randomly chose 10 university counties from Table 1 that had the locally highest species richness; we did not consider multi-university groups. Then we compared its area (from US census data) with that of the county listed immediately next to it in the table (usually the county to the north or north-east). The areas of these university counties and their neighbours are given in Table 2, in which the five cases where the university county was larger are marked with an asterisk. Were the null hypothesis false, the number of asterisks would be improbably small. However, the realized significance (the probability of five or more asterisks) is 0.5, which is entirely consistent with the null hypothesis that there is no systematic difference in the areas of these counties.

Another possible explanation is that universities are places where there is a lot of traffic in and out, often of an international nature. Students come and go every year; faculty are often quite cosmopolitan. Such movement may lead to the inadvertent (or even deliberate) parallel movement of plants into the local environment, as seeds in the toilet kit, or perhaps in a trouser turnup. Such cosmopolitan activity might also be generally true of cities to which many people travel for business or pleasure; and, of course, many great universities are located in cities. Many are not; but since the vast majority of cities also have universities (as well as symphony orchestras, museums, etc.) it is hard to differentiate the two factors. If a substantial number of cities lacking universities could be

shown to have high species diversity, this might be considered a factor. We can't find enough to make the case.

Another possibility, and our preference, is that university counties have high species diversity because they have a disproportionate share of botanists, and are, as a result, subject to more intense botanical attention. Such attention may lead to increased collection, and it has been argued that the number of species in a sample is a function of the number of individual plants collected (Gotelli & Colwell, 2001). A good example of this comes from case 5 of Table 1 showing data for Cheboygan County in Northern Michigan. Cheboygan County does not have a university, but it is the site of the University of Michigan Biological Station, where students have taken field botany courses, and collected plants, for a century. The Station is situated on the shore of Douglas Lake, which extends westward into neighbouring Emmett County which, in turn, shows the second highest species diversity in the set of related counties. An intriguing case is seen in the example of *Symphytum asperum* Lepechin (Boraginaceae), a native of the Caucasus, which is scattered across the northern states and Canada; it has been reported in two counties in Michigan: Washtenaw and Cheboygan. Note that another famous international student attraction in northern Michigan is the Interlochen Center for the Arts in Grand Traverse County, only about 50 miles from the Biological Station. Grand Traverse has seven neighbouring counties; of the eight, Grand Traverse is third in species diversity. While it has many violinists, singers and dancers, it is not a particular attraction for botanists.

Our favoured explanation is also supported by at least one of the cases in Table 1 where the university county comes in second. Indiana University is in Monroe County with 991 flowering plant species, which is exceeded by neighbouring Lawrence County with 1095. According to Webster (1961), Ralph Kriebel, who died in 1946 and lived in Bedford (Lawrence County), Indiana, probably made the most thorough sampling of the flora of any Indiana county by collecting in Lawrence County. He also collected extensively throughout the state, which resulted in a large private herbarium. In the *Flora of Indiana* (Deam, 1940), the letter K in the county maps associated with each species refers to specimens Deam examined in the (then private) herbarium of Ralph Kriebel. The Kriebel herbarium is now housed at Purdue University in West Lafayette, Indiana.

Somewhat more speculative is a second case where the university county comes second in species richness. The University of Iowa in Iowa City is in Johnson County, with 981 plant species; five neighbouring counties average 457 species (range 275–817). However, Muscatine County to the east (and bordering the Mississippi River) has 1019 species. Checking with a botanist at the University of Iowa we learned this: 'Why are there more species reported for Muscatine than Johnson County? First of all note that, according to your figures, Muscatine has only 38 more species. That relatively small increase can be explained to a large extent by comparing the habitat diversity of those two

Table 2 Area (km²) of 10 university counties (randomly selected from Table 1) and the neighbouring county listed as number 2 in that table

University	University county area	Neighbouring county area
U South Carolina	1999	1917*
Western Michigan U	1502	1494*
U Alabama	909	2085
U Tennessee	1362	640*
U Louisiana	699	2432
Michigan State U	1453	1401*
U West Virginia	948	1497
U Georgia	313	1093
U Kansas	1228	1494
U Missouri	1790	1264*

*Neighbouring county is smaller than university county.

counties; virtually all the habitats that occur in Johnson Co. also occur in Muscatine. In addition, Muscatine county has an especially large sandy flood plain area known locally as Muscatine Island, which provides unique habitats for plants and animals not found elsewhere in the state. There are also extensive oak-savanna areas along the Cedar River in Muscatine County, as well as northward range extensions of the Ozark flora along the Mississippi River. Although Muscatine did not have a college or university, it was an older and wealthier city than Iowa City, and probably had interested amateur botanists who passed on their collections to the University of Iowa.' (J. Schabillon, pers. comm.)

County occurrence records are based on plant collections, which are mostly housed in the herbaria of the state universities, and mostly collected by trained botanists, most of whom were employed by the same state universities. Thus, except for unusual circumstances in which competent and extensive botanical activity centred in an adjacent county before a state university was founded, botanical activity would occur disproportionately in the vicinity of the state university where trained botanists were employed. This effect is clearly reflected in the phenomenon we report here: plant species richness is, at least to some degree, a function of the location of botanists.

Ecology is often described as the study of factors that influence the abundance and distribution of species in space (Andrewartha & Birch, 1954; Krebs, 1985). Ecological research over the past half century has clearly shown that many natural factors influence these outcomes. Here we have shown that an additional significant factor determining the data on the distribution of plants in space is the distribution of botanists in space. We suggest that this situation may exist elsewhere in science, and that all ecologists should be aware of the issue and attempt to control for it in analysis.

REFERENCES

- Andrewartha, H.G. & Birch, L.C. (1954) *The distribution and abundance of animals*. University of Chicago Press, Chicago, IL.
- Burke, H.A. (1906) *History of the University of Michigan*. University of Michigan Press, Ann Arbor, MI.
- Butterworth, J.K. (1980) *A people's university*. University Press of Mississippi, Jackson, MS.
- Connor, E.F. & Simberloff, D. (1978) Species number and compositional similarity of the Galapagos flora and avifauna. *Ecological Monographs*, **48**, 219–248.
- Cox, C.B. (2001) The biogeographic regions reconsidered. *Journal of Biogeography*, **28**, 511–523.
- Deam, C.C. (1940) *Flora of Indiana*. Wm. B. Burford, Indianapolis, IN.
- Good, R. (1974) *The geography of the flowering plants*. Longman, London.
- Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

- Hatch, R.A. (ed.) (1967) *Convention of friends of agricultural education*. University of Illinois Press, Urbana, IL (Originally published in 1872 as the Fourth Annual Report of the Trustees of the Illinois Industrial University for the year 1870–1, Springfield, IL, USA).
- Kartesz, J.T. & Meacham, C.A. (2005) *Synthesis of the North American flora*, prepublication version 2.0. Missouri Botanical Garden, St Louis, MO. <http://www.phylosystems.com/prepublication>.
- Krebs, C. (1985) *Ecology: the experimental analysis of distribution and abundance*. Harper & Row, New York.
- Lomolino, M.V. (2004) Conservation biogeography. *Frontiers of biogeography: new directions in the geography of nature* (ed. by M.V. Lomolino and L.R. Heaney), pp. 293–296. Sinauer Associates, Sunderland, MA.
- Nelson, B.W., Ferreira, C.A.C., da Silva, M.F. & Kawasaki, M.L. (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature*, **345**, 714–716.
- Sansing, D.G. (1999) *The University of Mississippi*. University Press of Mississippi, Jackson, MI.
- Voss, E.G. (1972–1996) *Michigan flora*. Cranbrook Institute of Science and University of Michigan Herbarium, Ann Arbor, MI.
- Webster, G.L. (1961) The Kriebel Herbarium at Purdue University. *Proceedings of the Indiana Academy of Science*, **71**, 233–234.

BIOSKETCHES

Daniel E. Moerman, until his recent retirement, was William E. Stirton Professor of Anthropology at the University of Michigan-Dearborn, MI, USA. He has carried out research primarily in two areas – medicinal plants (primarily of Native American peoples), and the impact on health of people's knowledge and understanding of it. His book *Native American Ethnobotany* (Timber Press, 1998) received the Annual Literature Award of the Council on Botanical and Horticultural Libraries for 2000. In 2002, his book *Meaning, Medicine and the 'Placebo Effect'*, on the role of meaning in the healing process, was published by Cambridge University Press. He is Editor-in-Chief of the journal *Economic Botany*.

George Frederick Estabrook is Professor of Botany in the University of Michigan, Ann Arbor, and Research Scientist in the University of Michigan Herbarium. His research interests are in the application of ecological methods in ethnobotanical studies, and in the development of computational methods for application more generally to studies of ecology and evolution. For the past several years he has applied ecology to elucidate the relations between the culture and technology of subsistence peasants in the interior of Portugal.

Editor: Nicholas Gotelli