The Global Flora; Descriptive statistics with a commentary, and an ethnobotanical example

Daniel E. Moerman

Research

Abstract

This paper briefly describes the world's flora, based on the materials available at ThePlantList.org, a large web site built by a collaboration of botanists at Kew Gardens and the Missouri Botanical Garden. The paper details the number and distribution of families, genera, species, authors, publication dates and several other elements of the flora. The author notes several of the most notable features of this global scientific enterprise. Though it might seem arcane, the database is widely utilized, and as such seems worth examining. For example, a search of Wikipedia.com found 901 references to ThePlantList.org.

Introduction

This paper presents a statistical and analytic description of what is, currently, the most complete listing of scientifically identified plants in the world. "The Plant List, a working list of all plant species" is available online at www.ThePlantList.org (TPL), and is a collective work primarily by botanists and others at the Royal Botanic Gardens, Kew, and the Missouri Botanical Garden.

TPL characterizes 4 sorts of "plants": the flowering plants or Angiosperms; conifers, cycads and allies or Gymnosperms; ferns and fern allies or Pteridophytes, and mosses and liverworts or Bryophytes. In what follows, I will consider only the first two groups, Angiosperms and Gymnosperms, and will ignore the ferns and mosses.

Moving down the taxonomic classification, the major groups are divided into families, and the families into genera, the genera into species. At the level of the family, TPL offers a list of all accepted names in the family in comma separated file format (.csv), that is a listing of all the species in that family; .csv files are commonly readable by various database and spreadsheet programs. There are .csv files for families and genera, but not for the highest orders of classification in TPL (Angiosperms, Gymnosperms, Pteridophytes, and Bryophytes). In order to look at the entire database, I had individually to download all the family files (this is part of the reason that I decided to ignore the ferns and mosses about which I know very little anyway; 200 more downloads – 35 ferns, 165 mosses – would have driven me around the bend).

Why did I do this? Basically for two reasons: first, I plan to do some comparative analytics of medicinal floras from around the world, similar to analyses I have done in the past for the medicinal plants of North America (Moerman 1991) but using a global rather than regional lists of existing species; and second, I wanted to be able to update all these analyses to the usage of the now widely accepted system for classification of Angiosperms, APG III (Bremer *et al.* 2009).

A major motivation for this work is to have a new comparator, a global flora, to standardize comparisons of eth-

Correspondence

Daniel E. Moerman, William E Stirton Professor Emeritus of Anthropology, University of Michigan, Dearborn, Michigan, U.S.A. dmoerman@umich.edu

Ethnobotany Research & Applications 11:109-119 (2013)

Published: August 8, 2013

www.ethnobotanyjournal.org/vol11/i1547-3465-11-109.pdf

nofloras around the world. In earlier work (Moerman *et al*. 1999) we compared 5 different medicinal floras (from North America north of the Rio Grande, Chiapas, Korea, Kashmir, and Ecuador); in each case, the medicinal flora was compared with the existing local flora, and then those comparisons were compared. The four northern floras, and medicinal floras, were very similar, while the tropical flora (Ecuador) was quite different. I want to repeat that (with more floras) comparing them all to the same overall flora, TPL. In addition, TPL subscribes to APGIII which I wished to use. For those reasons, I had to study TPL and get to know it. At the end of this paper I will preview this analysis.

Note that this paper is not written by a botanist, but by an anthropologist, indeed, one with a lot of experience with the human use of plants, but an anthropologist nonetheless. The Plant List is, among other things, a human artifact, fitting somewhere on that continuum of objects from Australopithecine pebble tools to Silicon Valley cell phones. Indeed, one can query TPL from his smart phone should he wish to. It is in that context, as a human artifact, that I seek to describe and characterize TPL.

Methods

The major methodological issue was how to transform all the .csv files into one file which could be examined in toto. To do this was quite tedious; it took the better part of a week in November 2012. Each .csv file was opened in a page in my Chrome browser; each file was then selected and copied to the computer clipboard. A blank Notepad file was then opened, and the clipboard was copied there. The file was then briefly edited, removing the first line in the file which was the listing of the 20 database fields (ID, Major group, etc.). The file was then named and saved on my desktop computer in .csv format to the appropriate directory where it was then appended to a database (constructed earlier) using a command like “Append from Aca.csv csv” where Aca represents the Acanthaceae with its 2,955 species. This procedure was followed 420 times.

The database program utilized is Microsoft FoxPro Version 9.0 SP2 (Microsoft 2007a). I have been using it (or its many predecessor versions back to dBase) for 30 years, and am comfortable with it.

After entering all the family .csv files, I spent several weeks cleaning and fixing errors in the original (unsurprisingly, the online material contains errors) and in my own downloading (the unrelieved tedium led to a number of errors, which, I believe, were all ultimately found and repaired). Finally, a number of supplementary files of various sorts were created both in FoxPro and Excel (Microsoft 2007b), and various analyses were carried out.

Results

TPL divides the Angiosperms into 406 families, and the Gymnosperms into 14. The Angiosperms have 298,536 species (including subspecies and varieties) while the Gymnosperms have 1303, for a total of 299,839 considered here. (Note that whenever I report something like this without qualification, my count matches that of TPL, so far as I can tell.) I count 14,011 genera of Angiosperms while TPL reports 14,038; I can’t account for the difference. The Gymnosperms have 88 genera.

Families

Of the 420 families, all but two have the standard contemporary form with the “-aceae” ending. The two for which TPL have elected to use the retained names of long usage are Compositae and Leguminosae. They do not explain why they use these two of the ten permitted by the International Code of Botanical Nomenclature (article 18.5). As I plan in the future to match my version of TPL with other more limited listings (for ethnobotanical analysis), I elect to use the names Asteraceae and Fabaceae for these families.

The size of these families varies dramatically: 32 families have only 1 species; 32 families have only 2 species; 33 families have from 3 to 5 species. In aggregate, these 97 small families contain a total of 226 species.

By contrast, there are a relatively small number of families which contain most of the species. The three largest families are Asteraceae (32,006 species), Orchidaceae (27,950) and Fabaceae (26,423); these three are each roughly twice the size of the next largest family Rubiaceae (14,506); the largest three comprise 29% of the species. The largest 11 families comprise over half of all Angiosperms and Gymnosperms.

Genera

As noted, there are 14,011 genera in the two classes. These range dramatically in size as do the families. The largest three genera, with the most species, are *Hieracium* (Asteraceae, 2,992 species), *Astragalus* (Fabaceae, 2,765), and *Carex* (Cyperaceae, 2,420). Ten percent of the genera, the 141 largest genera, contain 31% of the species.

By contrast, there are many very small genera: 4,563 of them have only 1 species; 1,795 of them have 2; 2,307 have from 3 to 5.

Curiously, there are 12 genera which occur twice, one each in two different families (Table 1). For example, the genus *Benthamia* (named after botanist George Bentham) has 4 species in the Boraginaceae family, and 31 species in the Orchidaceae. If you search for Benthamia in TPL, only the orchids show up. But if you look at the listing for Boraginaceae, it includes the four Benthamia species. There is some indication that the proper generic name for the Boraginaceae items may be *Amsinckia*, but this is disputed by various authorities. However they might be resolved, these cases are all in violation of the International code.

The code also states "The specific epithet, with or without the addition of a transcribed symbol, may not exactly repeat the generic name (such repetition would result in a tautonym)." (McNeill & International Association for Plant Taxonomy 2012). TPL, however, includes 46 tautonyms, or epithets very close to being such. Among them are *Cajanus cajan* (L.) Millsp., *Inga inga* (Vell.) J. Moore, *Cubilia cubili* (Blanco) Adelb., *Apios apios* (L.) MacMill., *Meum meum* (L.) H. Karst., *Opopanax opopanax* (L.) H. Karst., and *Batatas batatas* (L.) H. Karst.

Table 1. The Plant List genera which are used in two different families.

Genus	Family	Species
Benthamia	Boraginaceae	4
	Orchidaceae	31
Bridgesia	Cactaceae	1
	Sapindaceae	1
Ceratocephalus	Asteraceae	3
	Ranunculaceae	1
Fagelia	Fabaceae	1
	Scrophulariaceae	1
Heeria	Anacardiaceae	1
	Melastomataceae	1
Huertea	Staphyleaceae	2
	Tapisciaceae	1
Laxmannia	Asparagaceae	16
	Rubiaceae	1
Lepidostemon	Brassicaceae	5
	Convolvulaceae	1
Malacocarpus	Cactaceae	1
	Zygophyllaceae	1
Matthiola	Brassicaceae	34
	Rubiaceae	3
Monniera	Rutaceae	1
	Scrophulariaceae	1
Oreocallis	Ericaceae	1
	Proteaceae	1

Species.

Generic names are not supposed to appear in more than one family. This is not the case for species epithets which are used repeatedly in different genera. In TPL, while there are nearly 300,000 species, there are only 82,705 specific epithets used for them; the average specific epithet is used 3.6 times.

The most commonly used specific eipthet is "*gracilis*" ("slender") used 534 times, while the variant "*gracile*" is used 141 times. All in all, the "root" "*gracil-*" with 47 different suffixes is used 818 times including 94 with "*gracillima*," 3 with "*graciliscapum*," 1 with "*gracilibracteatum*," and 1 with "*graciloides*."

The specific epithet "*angustifolia*" ("narrow leaf") occurs 479 times; the root "*angusti-*" occurs in 91 different forms including 59 species with "*angustifolius*," 8 with "*angustipetalum*," 2 with "*angustielliptica*," and 1 with "*angustivenosa*."

The specific epithet "*parviflora*" ("tiny flowers") occurs 420 times; the root "*parvi-*" occurs 102 times with a total of 1120 instances including 1 with "*parvizania*."

The most common "root" in specific epithets is"*foli-*" with 14,311 instances, "*folia-*" with 9711, "*foliu-*" with 4138, and "*folio-*" with 451. Including that root's occurrence in generic names (c.f., *Trifolium*), it shows up a total of 15,547 times.

There are 282 specific epithets used more than 100 times; these 282 account for 44,621 species epithets out of a total of 299,839; hence, 0.3% of names are used for 16% of all species. By contrast, 54,025 specific epithets are unique, ranging from "*aageododonta*" to "*zypaquirensis*." In addition, 4563 taxa are "monospecific genera," that is, genera with only one species. These do not seem to favor individual families. There are, of course, more of them in large families than in small ones, but the variation is very modest. The largest family, Asteraceae has the most monospecific genera, 617, but that is only 1.9% of the family's 32,006 species. Among larger families, the largest proportion of monospecific genera is in Aizoaceae, with 68 of 1130 species, or 6% of the family.

The shortest species names in TPL are 8 characters long; there are 11 of them, from *Aa erosa* (Rchb.f.) Schltr. to *Zea mays* L. Including the author, the shortest name is *Zea mays* L. with 11 characters.

The longest name in TPL is *Leucospermum hypophyllocarpodendron* (L.) Druce subsp. *hypophyllocarpodendron*, with 65 characters (excluding authors). Unfortunately, this is an autonym without a second subspecies, at least in TPL. However, Tropicos.com reports a taxon *Leucospermum hypophyllocarpodendron* Druce subsp.

canaliculatum Druce, but at the time of this writing it does not occur in TPL.

The longest name with author is *Nepeta mallophora* Webb & Heldr. ex Nyman subsp. *anticaria* (Ladero & Rivas Goday ex Ubera & Valdés) M.B.Crespo, Camuñas & J.C.Cristóbal with 110 characters.

Authors.

TPL includes 68,507 different entries for authors of botanical names. Unfortunately, they aren't necessarily distinct people, or sets of people. For example, there are 1,978 items carrying the name DC., the abbreviation for the great Swiss botanist Augustin Pyramus de Candolle (1778-1841). There are also 345 entries for A.DC., the abbreviation for his son, Alphonse Louis Pierre Pyrame de Candolle (1806-1893); but there are also 38 entries for A._DC. (with a space), likely the same person. And there are 116 entries for A.DC.'s son Anne Casimir Pyrame de Candolle (1836-1918) known as C.DC., but there are an additional 389 for C._DC. (with a space). Given the apparent difficulty of eliminating such errors, we will probably never know just how many botanical authors there are. The preponderance of authors is multiple, as in (L.) DC., or "Stübing, J.B. Peris, Olivares & J. Martín." Combining multiple names and errors, consider the difficulty in assessing this one -- (Tang & F.T. Wang ex S.C. Chen & G.H. Zhu) X.Qi Chen, S.W. Gale & P.J. Cribb, the author for an orchid with the otherwise agreeably brief name *Neottia fangii* -- and recognizing that there are some 68,507 of them to check.

Of course the most familiar author is "L." with 3,422 species, plus "L.f." with 202 (this refers to Linnaeus' son, Carl Linnaeus the younger, that is, "Linnaeus *filius*"), plus "(L.) L." with 80. There are 2365 with "(L.)SomeoneElse," the most common being "(L.) DC." with 137 which seems fitting.

Infraspecific ranks

TPL recognizes only 3 infraspecific ranks, form (f.), subspecies (subsp.) and variety (var.). Subspecies occur in 12,561 species, varieties in 12,532, and forms in 484. TPL states "We do not intend The Plant List to be complete for names of infraspecific rank. These are primarily included because names of species rank are synonyms of accepted infraspecific names." Indeed, there are many inconsistencies in how these ranks are handled. According to standard practice, there cannot be only one subspecies; a subspecies can be defined only in opposition to another part of the species which then becomes a subspecies of its own. So, if the species *Alpha beta* is determined to have a subspecies *gamma*, the proper designation of this set of taxa is:
Alpha beta,
Alpha beta subsp. *beta* (according to the International Code of Botanical Nomenclature (McNeill & IAPT 2012), this form, known as an "autonym" is created automatically when the first subspecies, *gamma* is created) as
Alpha beta subsp. *gamma*.

In this case, the international code cautions that *Alpha beta* refers to the species in its entirety including both subspecies, and using that epithet to refer only to the autonym subspecies "may lead to confusion." (McNeill & IAPT 2012) While it is very hard to count, there are many species with infraspecific ranks which do not have this format. One example is complete with one subspecies, but three listings:
Abarema alexandri (Urb.) Barneby & J.W. Grimes, the species,
Abarema alexandri var. *alexandri*, the autonym, and
Abarema alexandri var. *troyana* (Urb.) Barneby & J.W. Grimes the subspecies.
But many others lack this format, for example
Aaronsohnia pubescens (Desf.) K. Bremer & Humphries, and
Aarohsohnia pubescens subsp. *maroccana* Ball) Förther & Podlech, (no autonym) and alternately,
Abarema curvicarpa (H.S.Irwin) Barneby & J.W.Grime and
Abarema curvicarpa var. *curvicarpa* (an autonym, but no additional variety).

There are many cases of each of these types, but I can't find a straightforward way to count them. In any case, with a total of 25,577 items with infraspecific epithets in TPL, that is less than 10% of the total, and is not likely to make much difference in counts and percentages.

Infraspecific taxa are not randomly distributed among families. Of 420 families, 146 have none. Most of the families with many are, of course, large to begin with: Asteraceae with 4,233 (13% of the total in the family), Fabaceae with 2,888 (11%), Lamiaceae with 1,134 (13%). Among larger families, 24% of Rosaceae's 2,595 species have infraspecies ranks while Cactaceae has 22% of 2,610 species. In considering these numbers, recall that the data, as noted above, are imprecise.

Finally, since autonyms are automatically created when the first infraspecific taxon is defined, they do not need authors, dates, or publications, since that is "known information"; that is, it is part of the designation of the species. TPL however includes dates with 223 autonyms, and includes publications with 3104 of them. I don't imagine it does any harm.

Crosses

TPL contains 292 taxa classified as generic crosses, and 3099 as species crosses. Three taxa are both generic and specific crosses: ×*Cirsiocarduus* ×*jaubertianus*, ×*Conyzigeron* ×*huelsinii*, and ×*Crepidiastrixeris* ×*denticulato-lanceolata*. All three are in the Asteraceae family. Combining all the species which are either generic or species crosses yields 3,388 species. The Orchidaceae family has by far the most crosses; 21% of the crosses are orchids. The 6 largest families have over half the crosses (Orchidaceae, Poaceae, Lamiaceae, Cyperaceae, Asteraceae, and Fagaceae have 2000 crosses) while 296 families have none. The largest proportion of cross species in a family is in Potamogetonaceae, with 50 crosses of a total of 177.

Publication dates

Examining publication dates was difficult as the dates, ordinarily a simple year (ca “1850”), often were more elaborate. So one finds things like “(1859) pub 1861” or “1907-1908 publ. 1913,” or “Feb. 1886,” etc. There are also errors, or perhaps oddities like “2002)(2002” or “(1963)(” or “(18310”; the last one is understandable since the terminal zero is the unshifted right parenthesis. To resolve these problems, and to make a more useful field, I corrected all these in my version of the database by replacing the odd ones with a simple 4 digit year, from 1753 to 2010, using the first year listed in the online version, so “(1800)1801” became “1800”. I checked a number of the dates in IPNI (www.ipni.org, the International Plant Names Index) and found that many of these errors/oddities were in IPNI. So, without other information, I changed things like “200x” to “2000.” This means that my list of dates includes some errors (200x might actually be 2004, not 2000), but I can’t find any plausible, easy method for finding and correcting these errors. And most of them will not be off by more than 3 or 5 years.

The array of these (sometimes approximate) dates is very interesting. Figure 1 shows a graph of the number of species named per year from the publication of Linnaeus’ *Species Plantorum* (1753) up to the year 2010. The line represents a linear relation with an annual increase of approximately 1.3 percent. The spike at year 1753 represents the publication of Linnaeus’ Species Plantarum; the spike in 1891 probably represent the publication of Otto Kuntze’s *Revisio Generum Plantarum*. While the graph does appear to shows a substantial drop in plant naming during the Second World War, there is no such drop indicated for the American Civil War, the First World War, or any other particular historical circumstance. Should botanists continue to name 1.3% of plants per year (the long term rate), and if there are 352,000 species of An-

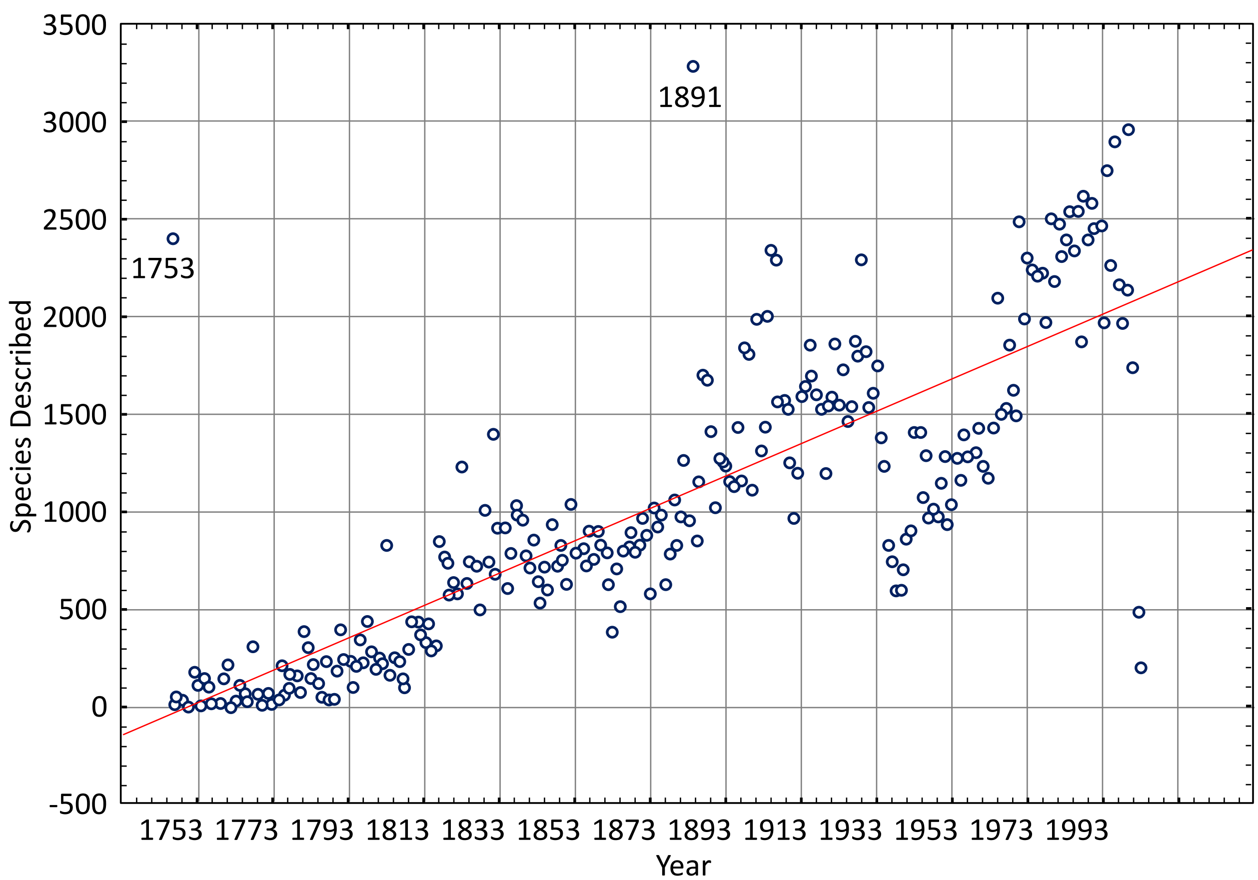

Figure 1. Number of species in ThePlantList.org described each year (with corrected and unified dates modified as noted in the text). Two outliers are well above the linear regression line. The earlier represent Linnaeus (1753) and the later probably represents Kuntze (1891).

giosperms (as TPL estimates), the job should be done in about 2025.

Publications

There are 16,380 different publications represented in TPL. 43,945 of the species do not have any publication given, and 3,317 are listed as "unknown." The most citations, 6,247, are to *Bot. Jahrb. Syst*., the *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie*, Leibzig, volumes 1-127, 1881-2009. IPNI notes that this title has been superseded by *Plant Diversity and Evolution*. Many of the major publications are serials like *Kew Bull*. with 5620 citations, *Phytologia* with 3418, and *Novon* with 3400. Linnaeus' *Sp. Pl*. is 12th in the list with 2436, in three editions in 1753, 1762 and 1764. Later works by subsequent authors are, confusingly, also listed as *Sp.Pl*. In addition, many of the TPL publications have more detailed information on publication date, edition, pages, etc. It is hard to know just how many of these items were listed in Linnaeus' several editions of *Sp. Pl*. Counting all the items whose publication includes the designation *Sp. Pl.* and whose author is *L*. yields 2353 species. He was a busy man, with pen, pencil, paper, and no computer, laptop, tablet or smartphone in sight.

Discussion

TPL, The Plant List

Several observations follow from this listing of genera, species, etc. First, although it has many flaws, TPL is an astonishingly valuable source. If 10% of its entries are not completely accurate, that leaves 90% -- that is something like 270,000 items -- which are just fine. It is frustrating that, even though it seems designed for download (since .csv files are provided), it takes such an arduous effort to get the whole thing (or even 2 of 4 major groups). It is also the case that the online version included many more hundreds of thousands of synonyms, and other illegitimate plant names. One can, with the online version, find the correct version for all or at least most of these millions of names; this is not possible with my downloaded version.

Botany

One can, after perusing some of the information in this collection express some opinions about the state of the botanical enterprise. There are several obvious issues. First, in recent years, largely as a result of the efforts of the Angiosperm Phylogeny Group (APG) (Bremer *et al*. 2009), there has been a rapid expansion of the number of families; and, most of the new families are quite small. For example, the important summary of botanical knowledge in the 1970s, Vernon Heywood's *Flowering Plants of the World* (Heywood 1978) described 306 families of plants globally. The subsequent book, Heywood's *Flowering Plant Families of the World* (Heywood *et al*. 2007), is described as "the successor" to the earlier one, includes 506 families. The authors explicitly credit APG II and several of its subsequent modifications (APG II 2003).
I have no problem with new tools in the kit of the systematist; rbcL is as interesting as stamens and pistils. But one does worry about the evolutionary implications of a system with so many families -- that is "groups" of plants – which seem to be largely evolutionarily distinct. This impression is increased by TPL in that it does not include any of the larger phylogenetic categories; there are no orders included, which presumably have some sort of evolutionary coherence.

Similarly for genera. The genus with the most species (*Hieracium*, Asteraceae, with 2992) is outnumbered by the genera with only 1 species (4,563 of them, one third of all genera). It seems unlikely that so many species are so marginally related to any of the rest.

The most remarkable part of this effort to me is what is shown in Figure 1; that the cumulative effort of thousands of botanists over centuries has continued to find more and more plants, and to name them, and to classify them is quite extraordinary. That the work of individuals living in a world so vastly different from ours can be seen to be participating in a scientific enterprise only modestly different from our own, warms my heart.

Ethnobotanical application

Years ago, I did an analysis of the medicinal plants of native American peoples, counting the total number of medicinal species per family, and comparing it with the total number of species per family. The null hypothesis, that medicinal plants were randomly distributed throughout the overall flora, was resoundingly rejected. The technique I used was simple linear regression analysis (Moerman 1979). I found this necessary because the obvious method, looking at percentages, didn't work because a number of very small families with one or two or three species sometimes had one or two or three medicinal species, giving them 100% medicinal status. By contrast large families like Rosaceae had hundreds of medicinals, but never more than about 30%. Hence, I elected to use regression analysis because I could control for family size in that way. Note that this method, although widely utilized, has become somewhat controversial in recent years. Bradley Bennett has recommended using a binomial analysis (Bennett & Husby 2008); Caroline Weckerle and colleagues have recommended using Bayesian analysis (Weckerle *et al.* 2011); Weckerly has also recommended an Imprecise Dirichlet model (IDM) (Weckerle *et al*. 2012). Other methods have also been suggested. A recent review of my residual method, plus binomial, Bayesian and IDM methods has been carried out (Turi & Murch 2013); the authors conclude that the results are very similar, some finding a few more, others a few less, "overutilized" and "underutilized" families. Given that, I elect to continue using the regression residual method because, as I have observed elsewhere (Moerman 2012), it is the simplest, the most easily explained, and it has a very nice graphic dimension which the others lack.

I made three analyses of ThePlantList data. In the first, I did a comparison of the north American flora (with some 26,000 species) with the global flora from ThePlantList (with some 299,000 species). See Figure 2.

Second, I repeated the analysis first done in 1979 comparing the north American medicinal flora to the total North American flora. Note that in 1991, the medicinal flora I had was what had appeared in the second edition of my work on Native American Medicinal plants, and compared 2095 medicinal species (Moerman 1986) with 16,270 species of North American plants derived from a preliminary checklist of species for the Flora North America (Shetler & Skog 1978). In the current analysis, I am operating with 2630 medicinal species, from the third edition of my native American plant work (Moerman 1998), and 26,598 species of north American plants derived from the work of John Kartesz at BONAP and the USDA plants database; the most useful single recent source for this listing is from Kartesz & Meacham (2005). (Note that USDA plants database may be more complete, but it continues to rely on the Cronquist organization of families.) The current analysis uses APGIII family categories; the change from Cronquist's classification (Cronquist 1988) has led to some modest changes in the outcome. See Figure 3 and Tables 2 and 3. Third, I did a comparison of the North American medicinal flora with the global flora. See Figure 4 and Table 4.

Results

The comparison of the north American flora and the global flora is very interesting. The former, with about 26,000 species is somewhat less than a tenth of TPL with some 299,000 species. Yet the two floras are very similar. A re-

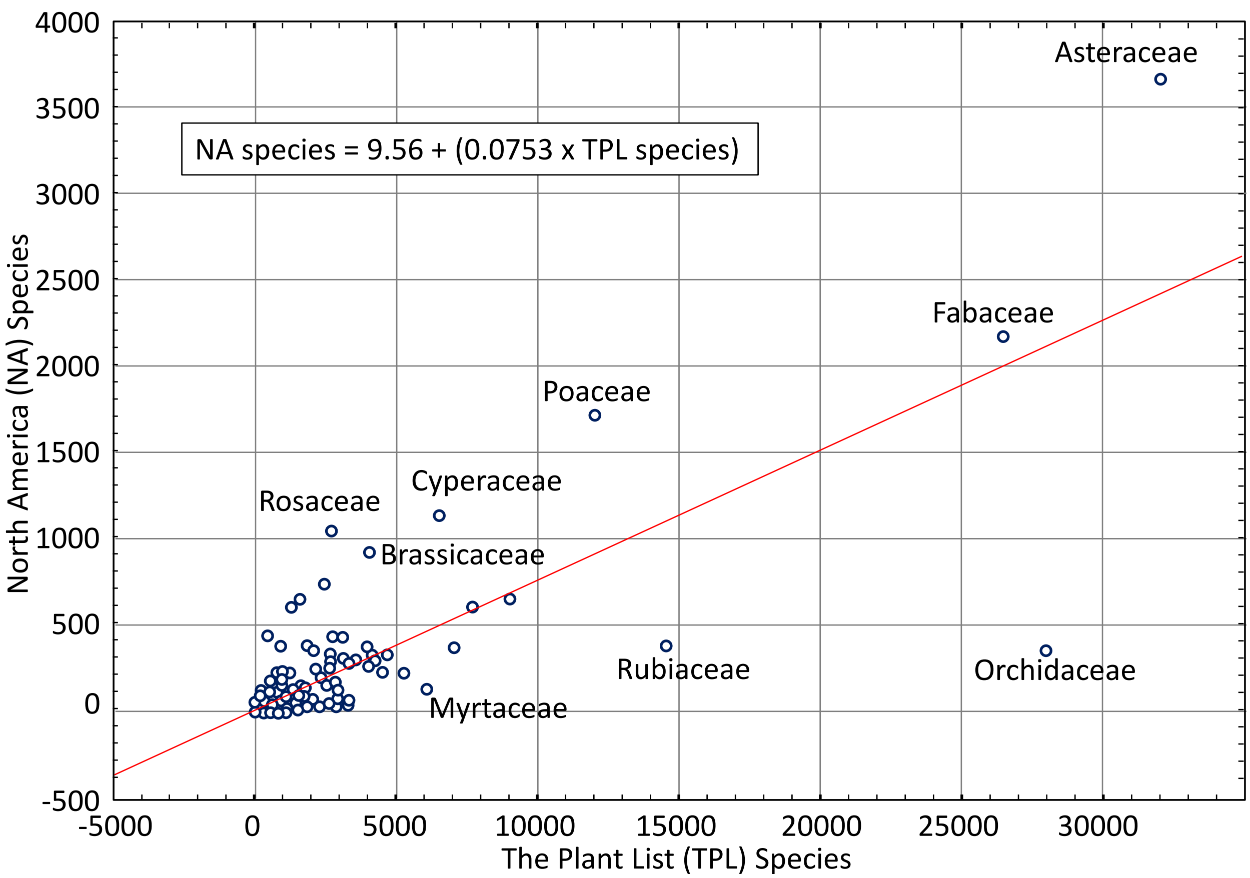

Figure 2. Scattergram associated with a regression of North American species per family on Global species per family from TPL. The regression line is the one line drawn most closely to all the data points. The correlation coefficient between the two series is r=0.8098. This indicates that the global flora and the continental flora, even though dramatically different in size (approximately 11 to 1), are proportionally of similar size families. The outliers like Asteraceae and Orchidaceae respectively indicate families relatively larger in the local flora (the former) and relative smaller in the local flora (the latter).

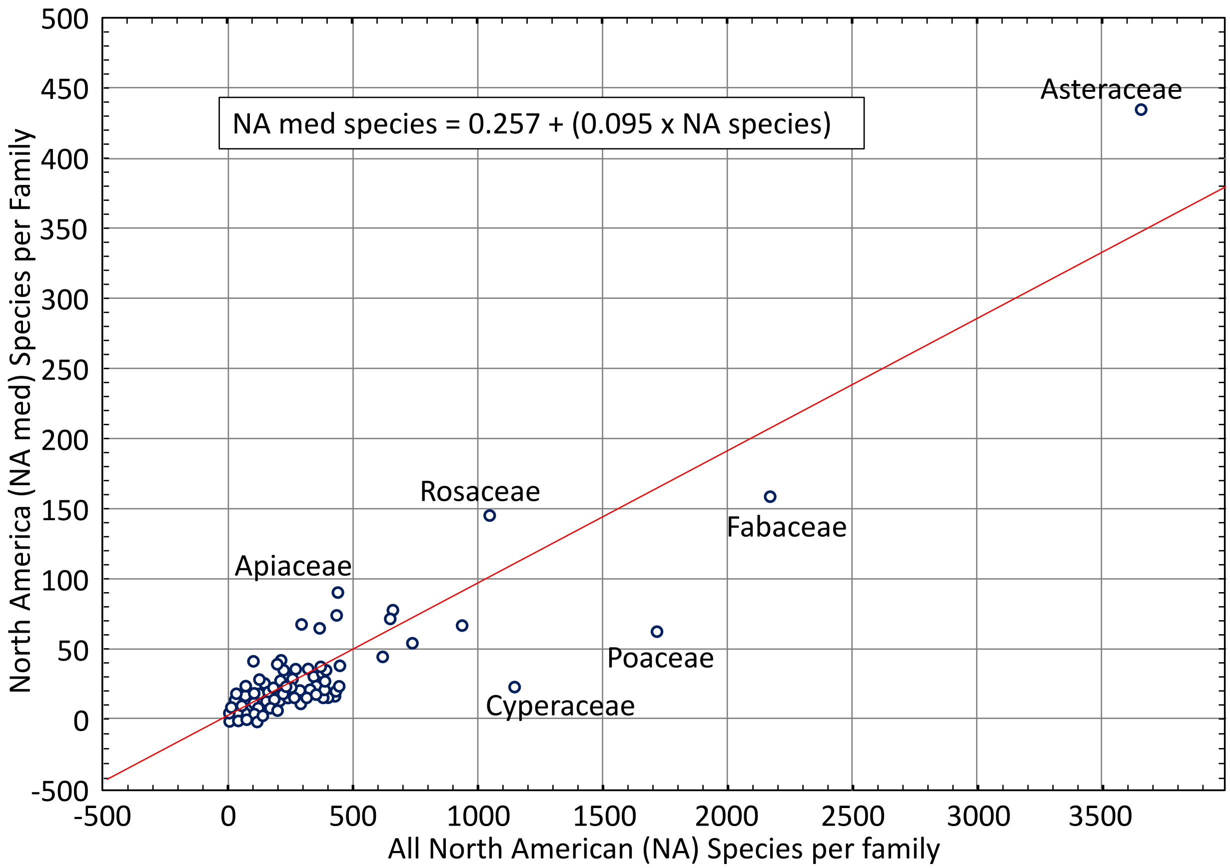

Figure 3. Scattergram associated with a regression of North American medicinal species per family on total species per family in North America. The residuals, that is, the vertical distances between the data point and the line indicate how the family is either over- or under- utilized by native American peoples. This regression replicates the analysis first presented 22 years ago (Moerman 1991).

Table 2. Comparison of the "top 10" families most utilized for medicinal species in the original analysis published in 1991, and in the current analysis in 2013. Most of the top 10 then are still top ten now. Exceptions are the Corylaceae family (*Alnus*, *Corylus*), switched to Betulaceae by APGIII, which now takes position 10. And Liliaceae, an important family in the past with nearly 400 species in 1991, now has had many switched by APGIII to Asparagaceae, Amaryllidaceae, or Nartheciaceae. Although the order changes a bit, the important medicinal families as classified in 1991 are still important now.

Rank 1991	Family	Rank 2013	Rank 2013	Family	Rank 1991
1	Asteraceae	1	1	Asteraceae	1
2	Rosaceae	3	2	Apiaceae	9
3	Lamiaceae	10	3	Rosaceae	2
4	Ranunculaceae	5	4	Ericaceae	12
5	Pinaceae	6	5	Ranunculaceae	4
6	Caprifoliaceae	9	6	Pinaceae	5
7	Salicaceae	8	7	Salicaceae	7
8	Liliaceae	34	8	Caprifoliaceae	6
9	Apiacae	2	9	Lamiaceae	3
10	Corylaceae		10	Betulaceae	

Table 3. Comparison of the "bottom 5" families least utilized for medicinal species (given the size of the family, Juncaceae and Caryophyllaceae slip up a bit; Rubiaceae drops substantially (globally much larger than in North America) but the bottom three are the same today as over 20 years ago.

Rank 1991	Family	Rank 2013	Rank 2013	Family	Rank 1991
228	Juncaceae	411	416	Rubiaceae	195
229	Caryophyllaceae	417	417	Brassicaceae	220
230	Fabaceae	418	418	Fabaceeae	230
231	Cyperaceae	419	419	Cyperaceae	231
232	Poaceae	420	420	Poaceae	231

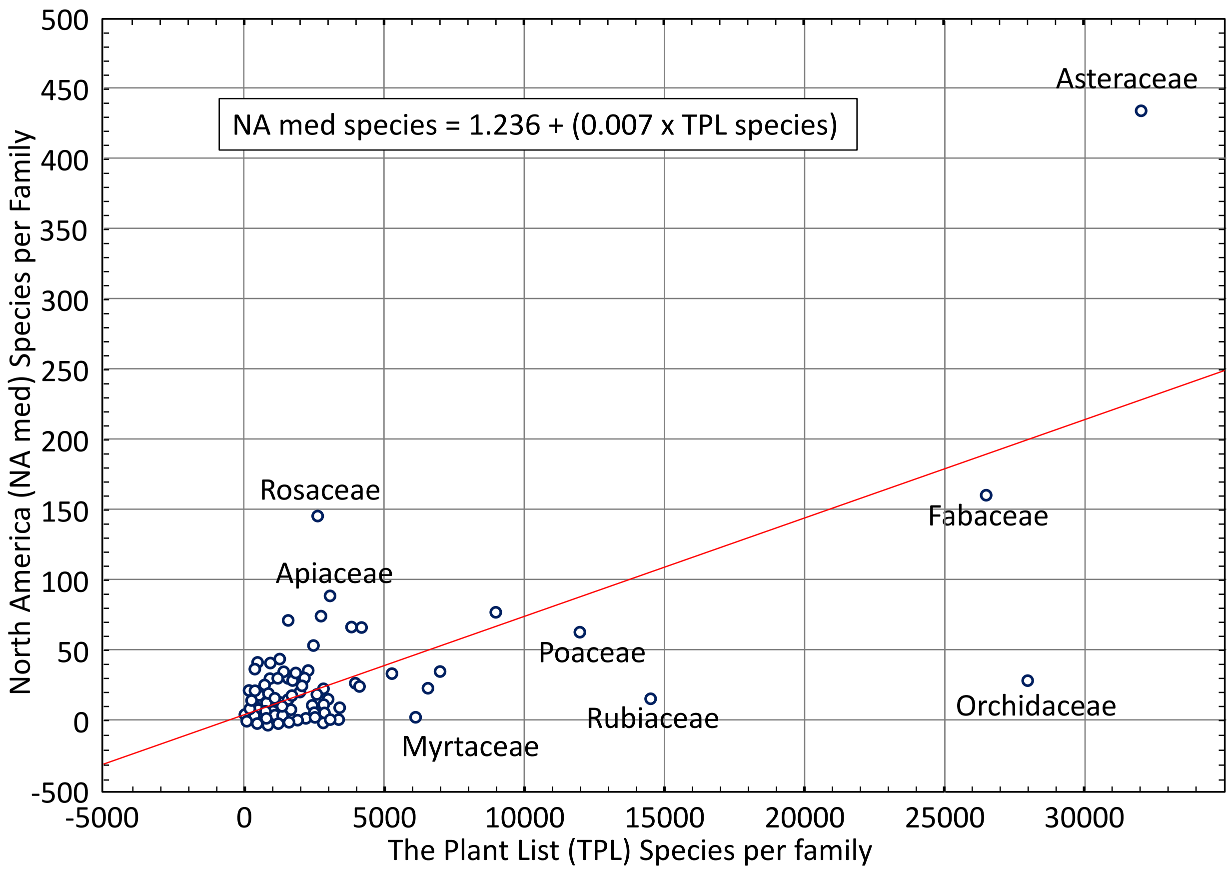

Figure 4. Scattergram associated with a regression of North American medicinal species per family and the total number of global species from ThePlantList.org. Outliers above the regression line represent global families with most of the medicinal species of North America. The ones below the line show the families either ignored in North America, or which don't grow there. There are, for example, 356 orchid species in North America, of which 28 are used medicinally; but globally, there are nearly 28,000 orchid species, which may be used medicinally elsewhere, but not in north America.

Table 4. The relationship of the medicinal species of North America to the global flora of The Plant List, as shown in Figure 4. These are the large global families from which native American people selected medicinal species.

Rank 2013	Family
1	Asteraceae
2	Rosaceae
3	Apiaceae
4	Polygonaceae

Rank 2013	Family
5	Ranunculaceae
6	Ericaceae
7	Brassicaceae
8	Pinaceae

Rank 2013	Family
9	Boraginaceae
10	Plantaginaceae

Table 5. The bottom 10 families in the relation between North American medicinal species and global species, by family. These are exceedingly large families, with very few medicinal species in north America. The ten families contain a total of 105,615 species, but only 298 North American medicinal species. The lowest families in Table 3, Poaceae, Cyperaceae and Fabaceae, show up here too.

Rank	Family	Rank	Family	Rank	Family
411	Acanthaceae	415	Cyperaceae	419	Rubiaceae
412	Gesneriaceae	416	Bromeliaceae	420	Orchidaceae
413	Melastomataceae	417	Fabaceae		
414	Poaceae	418	Myrtaceae		

gression analysis comparing family size finds a correlation of R=0.74 between the two floras, shown in Figure 2. Most of the points lie very close to the regression line indicating that they are proportionally very much the same size. The outliers above the line show North American families which have proportionately more species than occur in TPL (Asteraceae, Rosaceae, Poaceae, Cyperaceae and Brassicaceae, for example) while the outliers below are those where the north American flora has substantially fewer than the global flora (Orchidaceae, Rubiaceae, Myrtaceae, Bromeliaceae and Araceae, for example). Regardless, the two floras, the north American and the Global are remarkably alike, at least in terms of family sizes.

Second, the comparison of the north American medicinal flora with the overall north American with the much larger sample than was used in 1991, and one organized according to APGIII rather than Cronquist is remarkably unchanged. Table 2 compares the leading, highest residual, families in the two analyses, done 22 years apart, with 2 much larger samples now than then. Eight of the top 10 from 1991 are still in the top 12 in 2013. The Liliaceae family no longer appears near the top of the list since APGIII has parceled out species to 11 other families; the now much smaller Liliaceae doesn't make the top of the list. However, the Corylaceae family, which APGIII has combined with Betulaceae, lives on at the top of the list in its new family, 10th overall.

Likewise, the least utilized families (see Table 3) are still very much the same; the bottom three (Poaceae, Cyperaceae, and Fabaceae) are the same, and in the same order, now as 22 years ago. Juncaceae has moved from 5th to 11th from the bottom.

Third, the comparison of the north American medicinal flora with the global, TPL, flora (Tables 4 and 5) shows some substantial similarity to the comparison with the north American flora (predicted by the similarity between the American and the Global flora), but with some very interesting and significant differences. The "overselected" families are roughly the same, while the "underselected" ones include some of the originals, but add some too, notably the very large families Orchidaceae (27,950 species) and Rubiaceae (14,506) which have in North America a total of 44 medicinal species between them.

Accounting for families which are relatively small in North America, this new listing of underrepresented families is quite plausible. And one concludes that a multi-cultural comparison against this global flora will be an effective way to interpret local medicinal (and other: food, fiber, ritual, etc.) floras, and will in particular be an effective platform for comparisons among them.

Acknowledgements

Dan Austin tutored the author in some of the arcana of botanical nomenclature, and I thank him for it. Nearly 40 years ago, Bob Brill taught me to program the system which is today FoxPro; that I can still do it indicates an amazing bit of pedagogy. This paper is dedicated to the memory of my dear friend George Estabrook who died on Thanksgiving Day, 2011; I miss him every day.

Literature cited

APG II. 2003. An update of the Angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. B*otanical Journal of the Linnean Society* 141(4):399-436.

Bennett, B.C. & C.E. Husby. 2008. Patterns of medicinal plant use: An examination of the Ecuadorian Shuar medicinal flora using contingency table and binomial analyses. *Journal of ethnopharmacology* 116(3):422-430.

Bremer, B., K. Bremer, M.W. Chase, M.F. Fay, J.L. Reveal, D.E. Soltis, P.S. Soltis & P.E. Stevens. 2009. An update of the Angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161(2):105-121.

Cronquist, A. 1988. *The Evolution and Classification of Flowering Plants*. Second edition. New York Botanical Garden Press, Bronx, New York, New York.

Heywood, V.H. 1978. *Flowering Plants of the World*. First American edition. Mayflower Books, New York, New York.

Heywood, V.H., R.K. Brummit, A. Culham, O. Seberg. 2007. *Flowering Plant Families of the World*. Updated and revised edition. Firefly Books, Buffalo, New York.

Kartesz, J.T. & C.A. Meacham. 2005. *Synthesis of the North American Flora*. Prepublication Version 2.0. Missouri Botanical Garden, St. Louis Missouri. www.phylosystems.com/prepublication/, November 20, 2005.

Kuntze, O. 1891. *Revisio Generum Plantarum Vascularium Omnium et Cellularium Multarum Secundum Leges Nomenclaturae Internationales cum Enumeratione Plantarum in Itinere Mundi Collectarum-Mit Erlaeuterungen*. Volume 1. A. Felix, Leipzig, Germany.

Linnaeus, C. 1753. *Species Plantarum: Exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. Volumes 1 & 2. Impensis Laurentii Salvii, Stockholm.

McNeill, J. & International Association for Plant Taxonomy. 2012. *International Code of Nomenclature for Algae, Fungi and Plants (Melbourne code): Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011*. Koeltz Scientific Books, Konigstein, Germany. www.iapt-taxon.org/nomen/main.php?page=art6, July 29, 2013.

Microsoft. 2007a. *Microsoft FoxPro Version 9.0 SP2*. Redmond, Washington.

Microsoft. 2007b. *Microsoft Office Excel 2007, version 12*. Redmond, Washington.

Moerman, D.E. 1979. Symbols and Selectivity: A statistical analysis of Native American medical ethnobotany. *Journal of Ethnopharmacology* 1:111-119.

Moerman, D.E. 1986. *Medicinal Plants of Native America*. Museum of Anthropology, University of Michigan, Ann Arbor, Michigan.

Moerman, D.E. 1991. The Medicinal Flora of Native North America: An analysis. *Journal of Ethnopharmacology* 31:1-42.

Moerman, D.E. 1998. *Native American Ethnobotany*. Timber Press, Portland, Oregon.

Moerman, D.E. 2012. Commentary: Regression residual vs. Bayesian analysis of medicinal floras. *Journal of Ethnopharmacology* 139(3):693-694.

Moerman, D.E., R.W. Pemberton, D. Kiefer & B. Berlin. 1999. A comparative analysis of five medicinal floras. *Journal of Ethnobiology* 19(1):46-67.

Shetler, S. & L.E. Skog. 1978. *A Provisional Checklist of Species for Flora North America Monographs in Systematic Botany.* Volume 1. Missouri Botanical Gardens Press, St. Louis, Missouri.

Turi, C.E. & S.J. Murch. 2013. Spiritual and Ceremonial Plants in North America: An assessment of Moerman's ethnobotanical database comparing residual, binomial, Bayesian and imprecise Dirichlet model (IDM) analysis. *Journal of Ethnopharmacology* 148(2):386-394.

Weckerle, C.S., S. Cabras, M.E. Castellanos & M. Leonti. 2011. Quantitative methods in ethnobotany and ethnopharmacology: Considering the overall flora - Hypothesis testing for over- and underused plant families with the Bayesian approach. *Journal of Ethnopharmacology* 137(1):837-843.

Weckerle, C.S., S. Cabras, M.E. Castellanos & M. Leonti. 2012. An imprecise probability approach for the detection of over and underused taxonomic groups with the Campania (Italy) and the Sierra Popoluca (Mexico) medicinal flora. *Journal of ethnopharmacology* 142(1):259-264.