

Influence of biological and social-historical variables on the time taken to describe an angiosperm¹

Evelin K. S. Cavallin², Cássia B. R. Munhoz², Stephen A. Harris³, Daniel Villarroel², and Carolyn E. B. Proença^{2,4}

PREMISE OF THE STUDY: By convention, scientific naming of angiosperm species began in 1753; it is estimated that 10–20% of species remain undescribed. To complete this task before rare, undescribed species go extinct, a better understanding of the description process is needed. The South American Cerrado biodiversity hotspot was considered a suitable model due to a high diversity of plants, habitats, and social history of species description.

METHODS: A randomized sample of 214 species (2% of the angiosperm flora) and 22 variables were analyzed using multivariate analyses and analysis of variance.

KEY RESULTS: Plants with wide global distributions, recorded from many areas, and above 2.6 m were described significantly earlier than narrowly distributed, uncommon species of smaller stature. The beginning of the career of the botanist who first collected the species was highly significant, with an average delay between first collection and description of 29 yr, and between type collection and description 19 yr; standard deviations were high and rose over time. Over a third of first collections were not cited in descriptions. Trends such as scientific specialization and decline of undescribed species were highlighted. Descriptions that involved potential collaboration between collectors and authors were significantly slower than those that did not.

CONCLUSIONS: Results support four recommendations to hasten discovery of new species: (1) preferential collecting of plants below 2.6 m, at least in the Cerrado; (2) access to undetermined material in herbaria; (3) fieldwork in areas where narrow-endemic species occur; (4) fieldwork by knowledgeable botanists followed by descriptive activity by the same.

KEY WORDS botanical nomenclature; history of botany; new species; plant collecting; plant conservation; plant taxonomy

Publication of Linnaeus' (1753) *Species Plantarum* is the reference point for modern vascular plant species nomenclature (McNeill et al., 2012). Linnaeus' network of botanical correspondents and collaborators helped him catalogue the natural world and systematize the botanical knowledge of previous generations (Cruz, 2002; Jarvis, 2007), including plants described by early traveling naturalists (Reveal and Jarvis, 2009; Sequeira et al., 2010); approximately one third of the genera treated by Linnaeus were non-European (Walters, 1961). These factors, as well as his systematic application of binomial nomenclature, made Linnaeus' work the logical nomenclatural reference point for flowering plants at the rank of genus and below (Parkinson, 1975).

Since 1753, more than one million plant names (bryophytes, pteridophytes, gymnosperms, angiosperms) have been formally described, of which ca. 33% are accepted (Paton et al., 2008; The Plant List, 2013). Joppa et al. (2011) estimated that undescribed angiosperms species account for 10–20% of approximately 300,000–350,000 species (The Plant List, 2013). Some researchers have suggested the true number of undescribed species may be much higher (Schatz, 2002), while others have pointed out that methods such as interpolation from species discovery curves have high margins of error (Bebber et al., 2007b). However, independent of their numbers, authors tend to agree that undescribed species are likely to have similar characteristics: they are rare, very local, and/or occurring in areas with high levels of endemism, such as threatened biodiversity hotspots. If the task of species description is to be completed, knowledge of how biological, social, and historical variables influence a species age (years since 1753), and the lag between first known collection and description, is needed to optimize the process before rare, undescribed species go extinct.

¹ Manuscript received 15 March 2016; revision accepted 26 September 2016.

² Departamento de Botânica, Universidade de Brasília, Campus Universitário Darcy Ribeiro, Asa Norte, Postal Code 4457, CEP 70910-970, Brasília, Distrito Federal, Brazil; and

³ Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB UK

⁴ Author for correspondence (e-mail: cproenca@unb.br)
doi:10.3732/ajb.1600120

Little research has been undertaken on the factors affecting the process of plant species description (Bebber et al., 2010). The descriptive episode is a time series of four phases: (1) collection, (2) identification as an undescribed species, (3) description, and (4) publication. Our model is the Central South American Cerrado Biodiversity Hotspot (Myers et al., 2000), henceforth Cerrado. This area was chosen for four reasons: (1) it is a large, diverse area (ca. 2,390,000 km²) in which altitudes range from ca. 180 to 1900 m above sea level (SIEG, 2015), dominated by savannas but with diverse habitats such as grasslands, marshlands, isolated highland vegetation, dry forest, and seasonally inundated and noninundated gallery forests (de Oliveira Filho and Ratter, 2002; Batalha, 2011); (2) plant endemism (36.5%; Forzza et al., 2012), across many different families, is the highest of the world's savanna-dominated vegetation (Klink and Machado, 2005); (3) historical and cultural diversity of plant description is very high; (4) species description began in 1753 with continuous descriptive activity to the present; and (5) European, North American, and Latin American botanists, with different levels of access to literature, type material, and the field, were responsible for describing ca. 12,000 Cerrado species (Dean, 1991; Mendonça et al., 2008).

Botanical exploration of Brazil began with Georg Marcgraf (1648) who eventually produced *Historia Naturalis Brasiliae* (Hemming, 2007). The most famous botanists of 18th-century Brazil were Friar José Mariano da Conceição Vellozo, author of the aborted *Flora Fluminensis*, and Alexandre Rodrigues Ferreira, head of the *Viagem Filosófica* (Areia et al., 1991). By the middle of the 19th century, the country had become a botanical “Mecca” (von Martius et al., 1904; Leite, 1995; Kury, 2001; Giulietti et al., 2005; Felipe and Zaidan, 2008), and was “botanized” by naturalists from many European countries, including (listed by order of arrival in Brazil): Spix and von Martius (1938), Pohl (1976), Wied-Neuwied (1942), Saint Hilaire (1937), Gardner (1846), Warming (1908) and Freyreiss (1907), and by a few Brazilians such as Antônio Luiz Patrício da Silva Manso, and in the latter half of the century, Guilherme Schüch, Barão de Capanema, and João Barbosa Rodrigues (von Martius et al., 1904).

The start of the 20th century saw an increase in botanical collectors and authors who were born or living in Brazil, many of German origin, for example (by date of birth), Adolpho Ducke, Alexander Curt Brade, Frederico Carlos Hoehne, João Geraldo Kuhlmann, Henrique Mello Barreto, Ezequias Paulo Heringer, and Guido Pabst (Marx, 2004). Kuhlmann and Hoehne were particularly important during this period since they collected widely and collaborated on the description of nearly 1000 new taxa for the South American flora (Franco and Drummond, 2009; Heizer, 2010). Ezequias Paulo Heringer was a particularly important collector in the Cerrado region.

Bolivian botanical exploration, which began in the 16th century, was spurred by the commercial interests of the Spanish Crown in what was then Alto Peru, although until 1825 the focus was mostly the Andes and the northern provinces (de León, 1962). The first plant collections from the Bolivian Cerrados date from 1830 to 1833, when French botanist Alcide Charles Victor Dessalines D'Orbigny traveled through eastern Bolivia, Paraguay, Argentina and Brazil. During the 20th century, to the present day, the eastern Bolivian Cerrados have been botanized by staff from three herbaria: Museo de Historia Natural Noel Kempff Mercado, Santa Cruz, Missouri Botanical Garden (Timothy Killeen and collaborators), and University of Oxford (John Wood and collaborators).

The purpose of this study was to investigate the factors that delay the four phases of angiosperm species description. How do species' biological characteristics perform in contrast to the sociohistorical factors associated with the description episode? Ultimately, the questions we want to answer are: Why were some species described in 1753, while others took 260 yr to describe or remain undescribed? What can be done to hasten the description of undescribed species of flowering plants in time for them to be targeted by conservation efforts?

MATERIALS AND METHODS

Species selection—A representative sample of 1176 angiosperm species (ca. 9% of the Cerrado flora; Forzza et al., 2012) was selected for their ecological variability and taxonomic diversity (Proença et al., 2010; Villarroel, 2011). The taxa were from the following families or large genera: *Aspidosperma* (Apocynaceae); Araceae; *Vernonia* sensu lato (Asteraceae); *Handroanthus*, *Jacaranda*, and *Tabebuia* (Bignoniaceae); Bromeliaceae; *Mimosa* (Fabaceae); Loranthaceae; *Miconia* (Melastomataceae); *Psidium* (Myrtaceae); *Cyrtopodium* and *Habenaria* (Orchidaceae); *Paspalum* (Poaceae); Santalaceae; and *Solanum* (Solanaceae). This study took advantage of a published herbarium-based, geographic distribution survey, with additional information gathered on habitat, habit, and pollination and dispersal syndromes; taxonomic experts checked identifications of ca. 16,000 georeferenced specimens from Brazil (Proença et al., 2010) and 2200 specimens from Bolivia (Villarroel, 2011). However, despite these resources, to make data collection and identification checks feasible, it was necessary to further reduce the number of taxa sampled. For the 16 taxa (12 genera and 4 additional families, totalling 1176 species), batches of 20 species in alphabetical order were built; residual batches of fewer than 20 species were ignored, resulting in 43 batches. The computer program Sorteador (Uaise, Belo Horizonte, Brazil, 2014, <http://www.sorteador.com.br>) was used to select one species from each batch, resulting in the selection of the 43 species used for the pilot analysis. To choose the species for the final analysis, we repeated this process another four times; if a species that had already been included was drawn again, a new random selection was made until an unselected species was drawn. This process resulted in 215 species.

The taxa in the geographic study (Proença et al., 2010; Villarroel, 2011) were mostly from medium-sized to very large angiosperm families, so to avoid possible bias in family diversity, a second selection of species from small families was undertaken. These families were selected by building a global angiosperm species richness vs. family curve for the 178 Cerrado families (Fig. 1). The unsampled part of the curve was divided into six regular intervals of 16 families each. Two families were randomly selected from each interval, from which three species were selected at random based on their order of global richness using Sorteador. These additional families would have added another 36 species to the database; however, three of the families drawn had fewer than three species in the Cerrado: Hernandiaceae (two species), Laxmaniaceae (two species), and Magnoliaceae (one species). *Carica papaya* L. was eliminated because its widespread cultivation would distort the values for its geographic distribution and Cerrado occupancy. Thirty-one additional species were added from the following families: Achariaceae, Bixaceae, Calophyllaceae, Caricaceae, Hernandiaceae, Laxmaniaceae, Loganiaceae, Magnoliaceae, Nymphaeaceae, Pentaphragmataceae, Potamogetonaceae, and Smilacaceae.

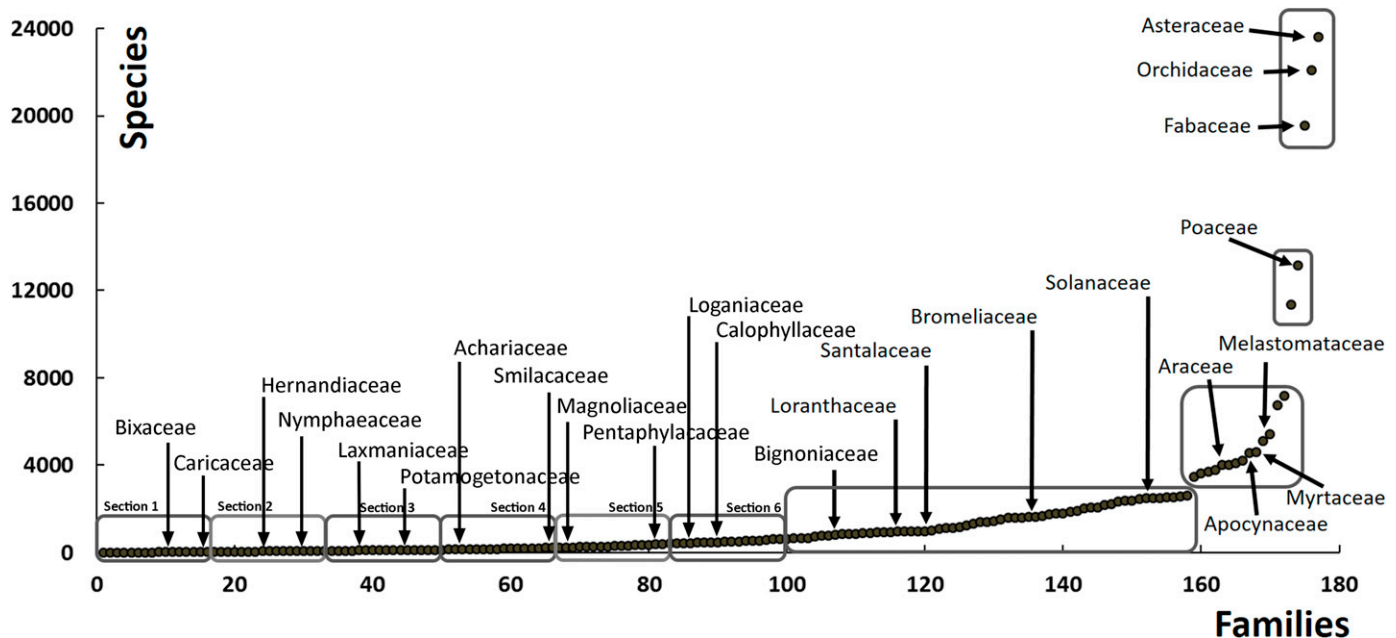


FIGURE 1 Distribution of global numbers of species in families occurring in the Cerrado. Groupings and families sampled are indicated by arrows.

The distribution of taxa in this new sample of 246 species was tested using a χ^2 test, and 23 of the 25 families did not differ significantly in size from a random sample of Cerrado species. Araceae and Bromeliaceae were significantly overrepresented; therefore, numbers of species of Araceae and Bromeliaceae were reduced at random until they equaled expected χ^2 values. The final sample was 214 species. This sample represents 14% of the 178 angiosperm families that occur in the Cerrado and ca. 2% of the estimated 12,000 species (Mendonça et al., 2008).

Species information—Species data (Appendix S1, see Supplemental Data with the online version of this article) were gathered from The International Plant Names Index (<http://www.ipni.org>), TROPICOS (2015), The Plant List (2013) version 1.1, original publications in which names were established, JSTOR Global Plants (2012–2015), CRIA (2015), FloResCer (Universidade de Brasília, 2014), and Herbário Virtual Reflora (Jardim Botânico do Rio de Janeiro, 2015).

Collector and author information—Collector and author data (Appendix S2, see online Supplemental Data) were gathered from the above references, and the Flora Brasiliensis List of Collectors (von Martius et al., 1904), Taxonomic Literature II (IAPT, 2012–2015), Harvard University Index of Botanical Databases (Harvard University Herbaria, 2013), and specific literature (Gardner, 1846; Freyreiss, 1907; Saint Hilaire, 1937; Archer, 1962; Pohl, 1976; Kury, 2001; Moreira and Quinteros, 2008; Heizer, 2010).

Variables—Twenty-two biological, ecological, social, and historical variables were used as described below and following the recommendations of Cronk (1989).

Biological and ecological variables—(1) Family richness—Family is usually the first level of identification that botanists aim for when confronted with an unidentified flowering plant (Raven et al., 1971).

Generally, the larger the family the more difficult it is to identify material at the generic and specific level (Walters, 1961) and to recognize a specimen as belonging to an undescribed species, perhaps delaying the second phase, identification as an undescribed species. Families were classified into one of four size classes: 1 (small) = up to 999 spp.; 2 (medium) = 1000–2499 spp.; 3 (large) = 2500–4999 spp.; and 4 (very large) = more than 5000 spp.

(2) Habit-imposed difficulties—Habit summarizes the amount and spatial distribution of a plant's biomass (Proença et al., 2010). Very large trees may be difficult to collect (Archer, 1962), while very small plants may be overlooked, causing a delay in phase one (collection). For the final analysis, habit-imposed difficulty was defined as the distance (in metres) from the tip of an average-sized specimen to human breast height (1.3 m). The average size of each species was calculated by compiling specialist-determined collections in online herbarium databases and calculating the average; 15–20 specimens was the target, but in some cases only a few specimens had this information. For hemiparasites, these data were so scant that we consulted an expert (C. S. Caires, Universidade Estadual do Sudoeste da Bahia, personal communication) who suggested group 3 as a conservative estimate. Groups were 1, easily seen and collected (0.65 m to 1.95 m); 2, some difficulty in seeing and collecting (ground or water level up to 0.65 m, or between 1.96 m and 2.6 m); 3, moderately difficult to see and collect (2.6 to 5.2 m); and 4, very difficult to see and collect (>5.3 m).

(3) Habitat—Habitat is the vegetation type or types in which the species occurs. According to Ter Steege et al. (2011), botanists tend to collect each species only once per field excursion and to collect in as many different types of habitats as possible. Some habitats present more difficulties than others, but ranking habitat-imposed difficulties proved challenging. Six habitat groups (as in Proença et al., 2010) were adopted: 1, anthropic/ruderal; 2, aquatic; 3, swamp; 4, grassland; 5, forest; and 6, savannah.

(4) Global distribution—For a species' global geographic range (Forzza et al., 2012), a colored star system was adopted (Hawthorne and Abu-Juam, 1995), ordered from narrow to wide ranging: 1, black star "Narrow-endemic" species, restricted to one Brazilian state that intercepts the Cerrado contour or in islands of discontinuous cerrado, in fewer than three noncontiguous municipalities, or three or fewer contiguous municipalities; 2, gold star "Local", found in two or more states that intercept the cerrado contour or in islands of discontinuous cerrado and also in two states with cerrado; or three or more noncontiguous municipalities or more than three contiguous municipalities, in combinations of the state of Minas Gerais and Maranhão, or Piauí and Mato Grosso do Sul, or Piauí and the Distrito Federal, or Maranhão and Mato Grosso do Sul, or Maranhão and the Distrito Federal; 3, blue star "Regional" species, found with discontinuous New World distribution, that is, Brazil and Mesoamerica, or Brazil and Mexico, or Brazil and the Caribbean or Brazil and North America; found in two contiguous or one Brazilian state but with a wide distribution across that state or found in at least four Brazilian states with Cerrado as the dominant vegetation; 4, green star "Widespread" species, introduced species, cultivated or naturalized in Brazil, with at least part of their native distribution out of the New World, or with a wide distribution in South or Central America, the Caribbean, or North America; or widely distributed in two or more Brazilian States and not restricted to the Cerrado.

(5) Cerrado occupancy—The number of approximately equal-sized quadrats of 30' latitude \times 30' longitude (ca. 3000 km²) in which species have been collected at least once, considering the whole Cerrado area in Brazil and Bolivia (Proença et al., 2010; Villarroel, 2011), a total of 1012 quadrats. This estimate is of true occupancy (quadrats in which a species does occur). Estimated occupancy is affected by collection effort, i.e., a species will be more likely to be collected in a well-collected area than in a poorly collected area (Funk, 1997; Simon and Proença, 2000); also, the chance of a species being collected will increase as its local abundance increases. Uncommon species or those restricted to poorly collected areas will show lower than true occupancy values. This bias is more or less equally distributed among species, since localized poor collection effort affects all the species present in that area.

(6) Pollination syndrome—This variable is related to the probable pollinator and is an attempt to rank the appeal of a flowering specimen in the field (online Appendix S3). Large, colorful flowers are easier to see from a distance (Schippmann et al., 2006), enhancing their chance of being collected. Showy inflorescences or flowers are usually pollinated by animals. Species were classed following Cox and Grubb (1991), Proença et al. (2010), and Willmer (2011). Five categories were adopted: 1, wind-pollinated; 2, self-pollinated; 3, insect-pollinated; 4, mammal-pollinated; and 5, bird-pollinated.

(7) Dispersal syndrome—This variable is related to the probable disperser and is an attempt to rank the appeal of a fruiting specimen in the field (online Appendix S4). Specimens with fleshy fruits may be more attractive because they can be easier to see at the distance, are more colorful, or are edible and better known, or seen as potentially more useful; large, often colorful fruits are usually animal-dispersed. Conversely, large, fleshy fruits are more difficult to dry or preserve, time-consuming to prepare for the plant press, and cumbersome to transport. Species were classed according to that of

Howe and Smallwood (1982) and van der Pijl (1982) and when available as listed in Proença et al. (2010). Six categories were adopted: 1, anemochory; 2, autochory; 3, entomochory; 4, hydrochory; 5, mammalochory; and 6, ornithochory.

(8, 9) Conspicuousness (size-related)—Conspicuousness of flowers and fruits in the field are a function of flower/fruit size and the number of flowers in an inflorescence or fruits in an infructescence. Four size groups (≤ 7 mm diameter, 8–33 mm, 34–88 mm, > 88 mm) and four number groups (≤ 7 , 8–54, 55–144, > 144) were defined based on average reported values from floras, herbarium specimens, and descriptions. These data were used to define three conspicuousness categories: 1, inconspicuous (size ≤ 7 mm diameter and number ≤ 54 , or size 8–33 mm and number ≤ 7); 2, moderately conspicuously (size ≤ 7 mm diameter and number ≥ 55 , size 8–33 mm and number 8–144, or size 34–88 mm and number ≤ 54); 3, conspicuous (size 8–33 mm diameter and number > 144 , size 34–88 mm diameter and number ≤ 54 , or size > 88 mm diameter and number 1 to > 144).

(10, 11) Conspicuousness (color-related)—Colors present in flower or fruit display were ranked in five categories of conspicuousness: 1, green, yellow-green, beige, straw, light to dark brown; 2, white, cream, or pale to light hues of pink, blue, lilac, yellow, orange, or any of these combined with type 1 colors; 3, medium to deep hues of pink, yellow, magenta, blue, purple, reddish brown, wine-colored, black, or any of these combined with type 1 colors; 4, deep hues of yellow-orange, orange, or red, or two or more strongly contrasting colors in categories 2, 3, or 4.

Social-historical variables—(12) Age of binomial 2015—The number of years elapsed since description and 2015. Ranging from 0 for a species described in 2015 to a maximum of 262 for a species described in Linnaeus' (1753) *Species Plantarum*.

(13) Collector starting date—This variable quantifies the period, in 25-yr periods, when a collector's career started: period 1 = 1650–1674, period 2 = 1675–1699, ..., period 15 = 2000–2024. The average time between angiosperm collection and description has been estimated as 33 yr, and herbaria house many undescribed species (Bebber et al., 2010, 2012).

(14) Collector interests—The breadth of a collector's interests were classified as: 1, specialist botanists (seed plants only); 2, general botanists (plants in general); and 3, naturalists (plants and at least one of animals, fossils, or rocks).

(15) Collector range—The ranges of a collector's fieldwork and, indirectly, their experience. So-called "big hitting collectors", who make a disproportionately large contribution to the description of new species (Bebber et al., 2012), visit many countries (although specializing in one) and collect for many years. Four categories were adopted for this variable: 1, global, the collector botanized in at least two continents in addition to their continent of birth; 2, continental, the collector botanized in one other continent in addition to their continent of birth (collecting in three or more countries); 3, regional, the collector botanized widely in Brazil/Bolivia, or Brazil/Bolivia and up to two adjacent countries; and 4, restricted, the collector botanized in a single region of Brazil/Bolivia.

(16) Collector specialization—An indicator of whether plant collecting was the collector's main or professional occupation: 1, professional collector, the collector's main or frequent occupation; and 2, amateur collector, when collecting plants was a secondary occupation, hobby, or undertaken for a short period.

(17) Collector duration—This variable expresses how long the collector's career lasted in years from first to last known collection. A long collecting career should theoretically increase quantitative (more collections per time in the field) and qualitative (better preparation and field identification) efficiency, and foster postcollection contacts and hasten distribution and incorporation into herbaria. Long collecting careers foster species description, at least for so-called "big hitting" collectors (Bebber et al., 2012).

(18) Author starting date—The 25-yr period when an author started their career: period 1 = 1650–1674, period 2 = 1675–1699, ..., period 15 = 2000–2024.

(19) Author occupation—Expresses whether the author was: 1, specialized as a botanist, describing only flowering plants; or 2, general botanist.

(20) Author legacy—The number of species an author published during his or her career, assuming experience hastens phases 2 (identification as an undescribed species), 3 (description), and 4 (publication): 1, minor, up to 10 spp.; 2, small, 11–100 spp.; 3,

medium, 101–500 spp.; 4, large, 501–1000 spp.; 5, very large, 1001–5000 spp.; and 6, vast, >5000 spp.

(21) Author duration—The length of an author's career in years. Long careers allow accumulation of experience and time for more species to be described at an accelerated pace than short careers.

(22) Collector/author communication barrier—A measure of communication difficulties, involving author access, interest, and exchange of information, possibly affecting phase 2 (identification as an undescribed species) and phase 3 (description): 1, lowest level of communication barrier, when collector and author were one and the same; 2, collector and single author were contemporaries and from the same country; or when the collector was one of two or more authors; 3, collector and one or more authors either not contemporary or not from the same country, but not both; 4, collector and single author neither contemporary nor from the same country.

Analyses—Descriptive statistics and graphs were produced using Microsoft Excel v. 11 (Redmond, Washington, USA). Since the exact year of collection was often unknown, particularly for older collections, four sets of analyses were done: (1) only species with exact year of collection; (2) all species considering the earliest possible collection year; (3) all species considering the latest possible collection year; (4) all species considering the middle collection year of the interval from (2) and (3).

Principal component analyses (PCA) was done initially using 43 randomly selected species (chosen from the intermediate sample of 251 species) and 18 variables using the program PAST v. 2.7 (Hammer et al., 2001). The objective was to eliminate uninformative variables ($p \leq 0.4$). Following the pilot screen, the PCA was done again with 17 variables and a final sample of 214 species using PAST and PCORD version 6.0 (McCune and Mefford, 2011). One-way ANOVAs were done using PAST v. 2.7 to test whether there were differences in the age of binomial and first collection–description delay between groups of categorical variables that were found to be significantly correlated with the former in PCA. This test was essential for confidence in the results, since four of our variables were nonbinary and categorical, i.e., authors decided ranks subjectively. All other variables were either numerical, categorical and binary, or categorical and multistate but objectively ranked in a numerical order.

RESULTS

The average Cerrado hotspot species is a savanna-living, insect-pollinated, wind-dispersed, Green Star (wide-ranging) herb, with a Cerrado occupancy value of 11 quadrats, first collected in 1852 and described in 1881. Two thirds of the species were published in the 19th century (Figs. 2, 3; average age of binomial = 134 yr).

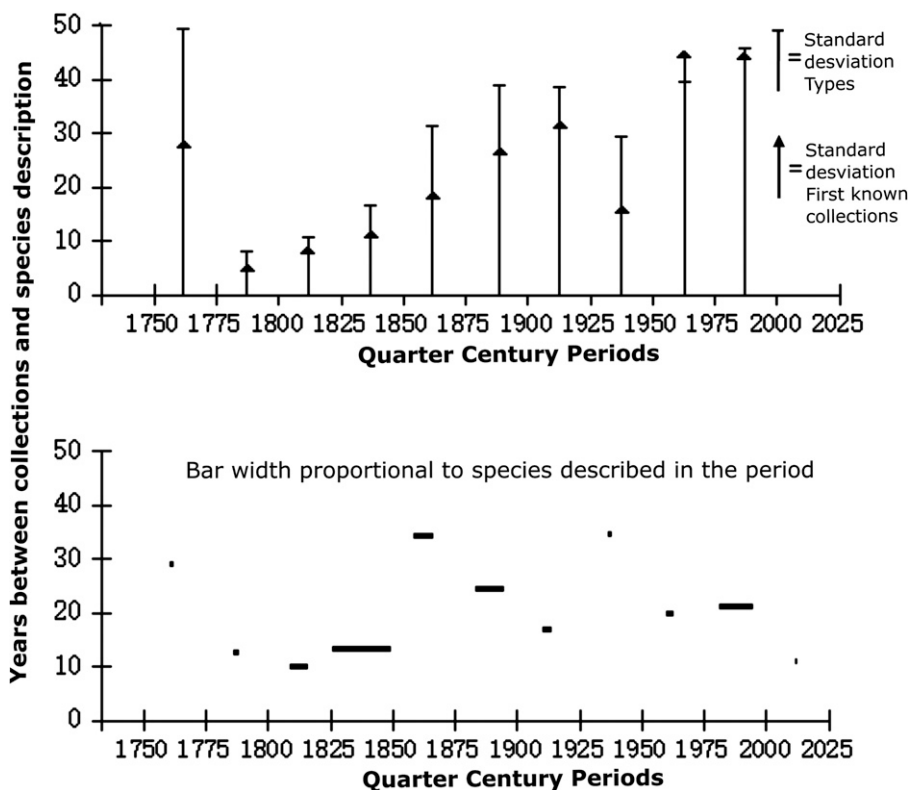


FIGURE 2 Average number of years between first collections and description for sampled taxa in 25-yr periods. (A) Standard deviation showing progressive increase over time. (B) Number of species described vs. average delay between first collection and description per period.

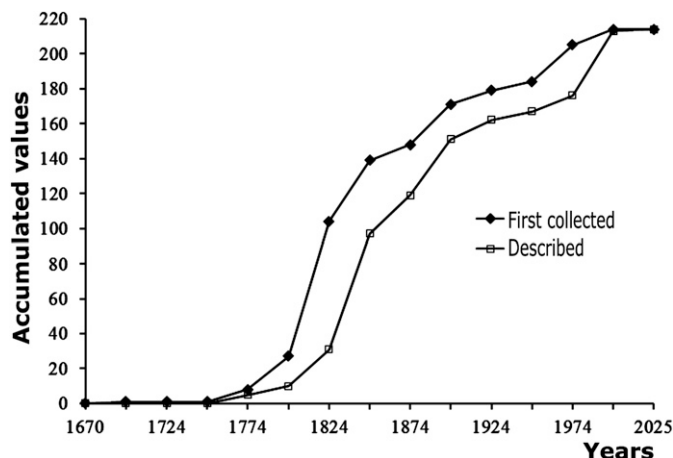


FIGURE 3 Species accumulation curves for first known collections and descriptions in 25-yr periods.

The average author was a 19th-century, professional German botanist, dedicated to the study of angiosperms, whose career lasted 41 yr (Appendix S1).

Across the four sets of analyses of first collections and type collections, using the oldest possible date or the most recent (when the exact date of a collection was unknown but there was a known interval) had small impact, of less than a year, so the middle year of the interval was adopted. The average delay between first collection and species description was 28.6 yr, with a standard deviation of 28.5 yr. The average delay between type collection and description was 19.05 yr, with a standard deviation of 19.8 yr. The shortest delays between type collection (virtually all also first collections) and description (ca. 10 yr) occurred between 1775 and 1824.

The most first collections (77 species, over a third) occurred in the 1800–1824 quarter century. The highest numbers of future types collected and of new species described (60 and 66, respectively) occurred between 1825 and 1849. For over half the species, the first known collection was also the type or cited in the prologue (63.1%), but for over a third (36.9%) it was not, and were possibly unknown to the author of the binomial.

Thirteen (15.5%) collectors were responsible for over half (53%) of the types or oldest syntypes, the cut-off being four or more type collections. In chronological order, with numbers of species types or oldest syntypes in parentheses, these were Alexander von Humboldt (9), Carl Friederich von Martius (15), Joseph Emmanuel Pohl (9), Friederich Sellow (16), Ludwig Riedel (8), William Burchell (7), Georg Blanchet (4), George Gardner (10), Richard Spruce (4), Ernst Heinrich Georg Ule (7), Antoine Marie François Glaziou (4), Gerdt Hatschbach (4), and Howard Irwin (16). Three collectors (Humboldt, Blanchet, and Spruce) never set foot within the Cerrado biodiversity hotspot (Kohlhepp, 2005; von Martius et al., 1904) but collected species that occur within its boundaries; only one collector (Hatschbach) was born in Brazil.

PCA analysis (pilot screen)—Two variables (habitat; collector range) were nonsignificant and uncorrelated with age of binomial and were excluded. Three variables (family richness, pollination syndrome, collector interests) were marginally below significance or nonsignificant but showed a sum of loadings comparable to those of significant variables. The variables family richness and

collector interests were maintained in the final analysis. The variables pollination syndrome and dispersal syndrome were substituted by groups of conspicuousness of the flower and fruit display, ranked by size vs. numbers and by colors. Habit was substituted by the variable habit-imposed difficulties. The variable author starting point was highly significant but autocorrelated with the variable age of binomial due to the limitations of the human life span and so was not included in the final PCA analysis.

PCA analysis (final)—The first three PCA axes explained 42.7% of the variance (Fig. 4). The first axis explained 19.9% of the variance, and the highest correlations (by order of value >0.15) were with variables age of binomial, collector starting date, author occupation, global distribution, cerrado occupancy, collector interests, author legacy, habit-imposed difficulties, and floral display (size). The second axis explained 10.4% of the variance, and the highest correlations (by order of value) were with fruit display (color), fruit display (size), flower display (size), family richness, author legacy, floral display (color), author duration, habit-imposed difficulties, and global distribution (>0.15). The third axis explained 11.1% of the variance, and the highest correlations (by order of value) were with author duration, author legacy, fruit display (color), collector duration, fruit display (size), and family richness (>0.15).

Axis 1 appears to have successfully summarized most of the factors that influence the rate of the descriptive process, as demonstrated by the significant association of all its main vectors with age of binomial. Axis 2 appears to have mainly summarized the taxonomic and ecological variability within our sample. Geographic distribution variables (global distribution and cerrado occupancy) and author legacy were present in both axis 1 and 2.

Age of binomial as response variable—PCA—Age of binomial showed the strongest correlation with axis 1 in the final PCA ($+0.50$). The seven next most highly correlated variables with axis 1 (listed above) were also significantly correlated with age of binomial (Table 1). Collector starting date showed the highest (-0.856) correlation with age of binomial. The second highest was author occupation ($+0.626$). The third was global distribution ($+0.599$) and the fourth, cerrado occupancy ($+0.372$); see Fig. 5 for adopted grid—both related to geographic distribution. The fifth was author legacy (0.359); the remaining significant variables had values below <0.3 and, by order of value, were collector interests and habit-imposed difficulties. Five biological variables and four social-historical variables had nonsignificant ($p > 0.05$) correlation with age of binomial. The nonsignificant biological variables were flower display (color), flower display (size), fruit display (color), fruit display (size), and family richness; floral display (size) was near significant ($p = 0.07$). The four nonsignificant social-historical variables were collector specialization, collector duration, author duration, and collector/author barriers.

One-way analysis of variance—The significance of the categorical variables used for the final PCA (Table 1) was tested per group by one-way analysis of variance (ANOVA). For age of binomial PCA and ANOVA were congruent, except for collector/author communication barriers (Fig. 6) that showed a significant difference between group pairs that was not found in PCA (Fig. 6). The significance of all categorical variables was also tested by analysis of variance for the variable first collection-description delay; the only significant differences between the means were found between classes of author occupation and global distribution.

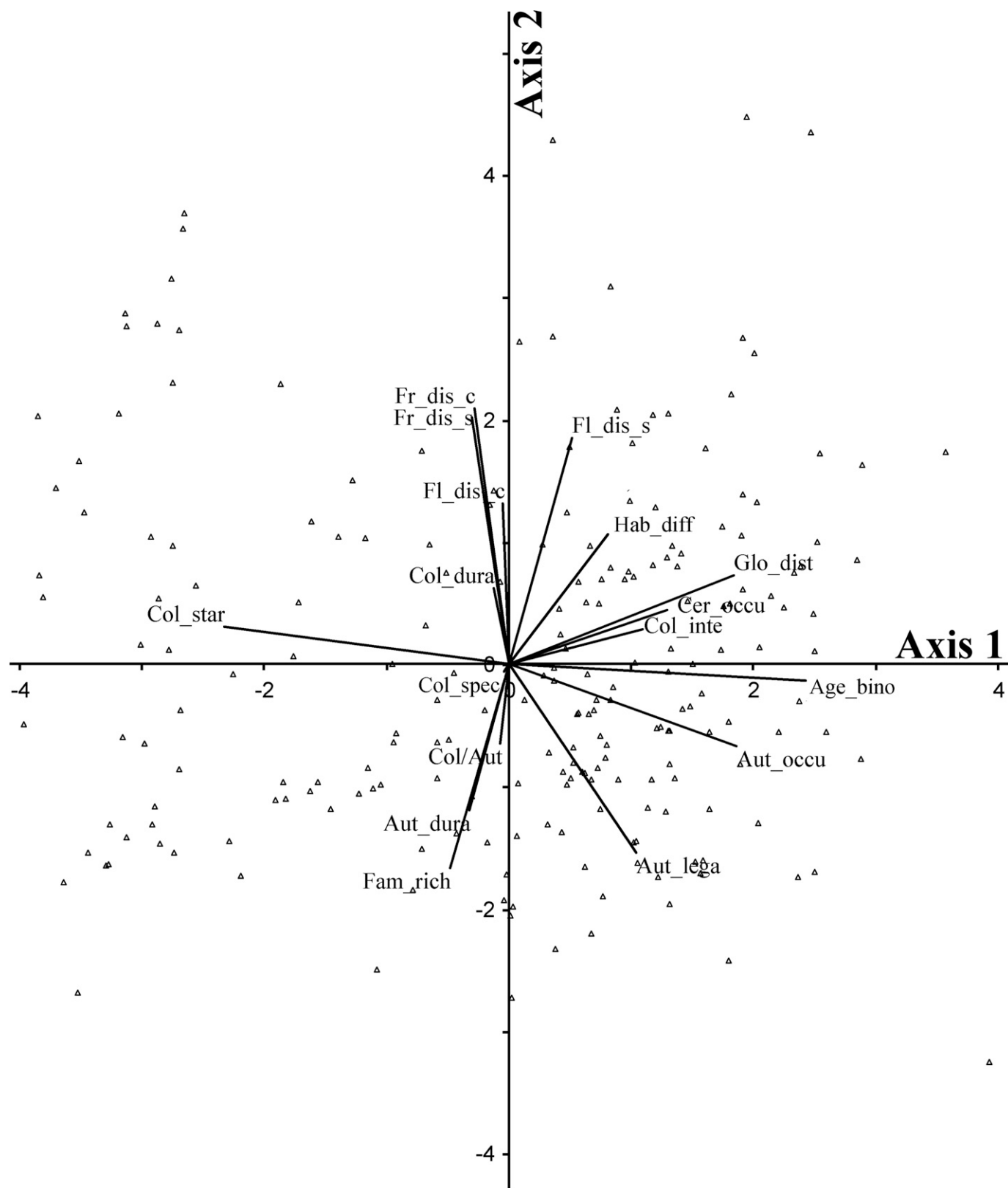


FIGURE 4 Principal component analysis of the 214 species and 17 final variables. Age_bino = age of binomial; Aut_dura = author duration; Aut_lega = author legacy; Aut_occu = Author occupation; Cer_occu = Cerrado occupancy; Col/Aut = collector/Author barriers; Col_spec = collector specialization; Col_dura = collector duration; Col_inte = collector interests; Col_star = collector starting date; Fam_rich = family richness; Fl_dis_c = floral display (color); Fl_dis_s = floral display (size); Fr_dis_c = fruit display (color); Fr_dis_s = fruit display (size); Glo_dist = global distribution; Hab_diff = habit-imposed difficulties.

TABLE 1. PCA correlation table. Correlation coefficient below diagonal, significance above diagonal (ns, nonsignificant; * $0.05 \geq p \geq 0.01$; ** $0.01 \geq p \geq 0.001$; *** $p \leq 0.001$). One-way analysis of variance, using age of binomial as the explanatory variable (bold), are identical except that collector/author barrier is significant ($0.05 \geq p \geq 0.01$). Variable codes are as in Fig. 4. Lines separate explanatory variable (1), biological variables (6), geographic variables (2) and social-historical variables (8). Significant correlations are in boldface.

Variable	Age_bino	Age_diff	Fl_dis_c	Fl_dis_s	Fr_dis_c	Fr_dis_s	Fam_rich	Glo_dist	Cer_occu	Col/Aut	Col_inte	Aut_occu	Aut_lega	Col_spec	Col_star	Col_dura	Aut_dura
Age_bino	1																
Hab_diff	0.205	1															
Fl_dip_c	0.007	-0.087	1														
Fl_dis_s	0.138	0.027	0.042	1													
Fr_dis_c	-0.068	0.035	0.222	0.288	1												
Fr_dis_s	-0.087	0.036	0.201	0.075	0.644	1											
Fam_rich	-0.075	-0.380	-0.140	-0.337	0.053	-0.063	1										
Glo_dist	0.599	0.255	-0.016	0.192	0.109	0.034	-0.126	1									
Cer_occu	0.372	0.037	0.027	0.025	-0.011	0.044	-0.167	0.298	1								
Col/Aut	-0.019	0.011	-0.006	0.073	-0.077	-0.106	-0.040	-0.065	-0.167	1							
Col_inte	0.279	-0.045	-0.105	0.123	-0.044	0.102	-0.019	0.202	0.151	-0.016	1						
Aut_occu	0.626	0.077	0.036	-0.035	-0.104	-0.075	0.015	0.317	0.292	-0.119	0.163	1					
Aut_lega	0.359	0.079	-0.141	-0.069	0.0001	-0.081	0.192	0.142	0.067	0.177	0.020	0.419	1				
Col_spec	0.072	0.048	-0.038	-0.044	0.008	-0.043	-0.063	0.004	0.100	-0.222	-0.577	0.043	-0.009	1			
Col_star	-0.856	-0.185	0.014	-0.108	0.158	0.137	0.092	-0.523	-0.315	-0.060	-0.402	-0.535	-0.034	-0.034	1		
Col_dura	-0.101	0.078	-0.081	0.191	0.171	0.032	-0.022	-0.022	0.033	-0.211	0.106	-0.019	0.056	-0.002	0.114	1	
Aut_dura	-0.145	-0.031	-0.182	-0.050	0.083	0.009	0.176	-0.133	-0.071	0.080	-0.146	-0.009	0.536	0.008	0.127	0.200	1

DISCUSSION

The scientific process and those who engage in it have changed dramatically since Linnean times (Ogilvie, 2006; Endersby, 2010). The significant correlations observed in the lower right half of Table 1 are related to social historical trends. The correlation between age of binomial and author occupation for example (recent binomials tend to be described by angiosperm specialists and old binomials by general botanists), and collector interests (recent binomials tend to be described by narrow-interest collectors and old species by wide-interest collectors) are not cause-and-effect relationships but rather reflect progressive specialization in science (Higham, 1980). Classical naturalists of the 18th century, interested in geology, zoology, botany, and sometimes anthropology, have been superseded by specialized botanists who study single families or even genera.

The significant positive correlation of author legacy with age of binomial showed that recent binomials tend to be described by authors whose legacy is or was minor or small and old binomials by authors with large or vast legacies. This result is probably a reflection of the dwindling number of undescribed species and the fact that collections of known species operate as noise in phase 2 of describing a new species, i.e., recognizing a collection as an undescribed species. The identification of a known species may be as difficult and time-consuming as confirming that a collection belongs to a new species, yet the effort results in a correct identification and not in a new description. In later periods, authors' career profiles changed to involve many other activities besides plant description. Of course, recent authors that are still alive may be productive for decades to come, i.e., their author legacy values may be underestimates of potential true values.

Two-thirds of the species were described in the 19th century, with a peak in 1825–1849 period (Fig. 2), a reflection of the “golden era” of taxonomy; most non-European species were still undescribed, and species description was seen as highly prestigious science (Shteir, 1989). Collection expeditions, fostered by the search for new drugs and useful plants (Miller and Reill, 2010; Brandão et al., 2011) or adventure and botanical novelties resulted in rapidly growing herbaria that fuelled descriptive activity (Endersby, 2010). At this time, collecting expeditions were filled with danger, disease, and misadventure, and often took several years (Kury, 2001). The collecting expeditions of today are (usually) safer and quicker, but as countries attempt to limit access to their natural resources, the bureaucracy associated with plant collecting has mushroomed and permits often take months to negotiate, particularly for non-nationals (see The Access and Benefit Sharing Clearing House, 2016).

Collector starting date showed the strongest correlation with age of binomial, suggesting that the availability of the collection is not only a *sine qua non* condition, but also a strong incentive to description. A plant cannot be described before it has been collected, but there can be a centuries-long delay between collection and description (Landrum and de Oliveira, 2010). The average lag between collection of the actual type and species description was ca. 19 yr, well below the global average of 33 yr found by Bebbier et al. (2010). The average delay between first known collection and description, however, was 28.6 yr, much closer to the value found by Bebbier et al. (2010). Furthermore, the standard deviation around the average delay between first known collection and description and also between type collection and description, rose steadily over time (Fig. 2), suggesting there is an increasing lag in the process. For the lag between first known collection and description, SD = 5

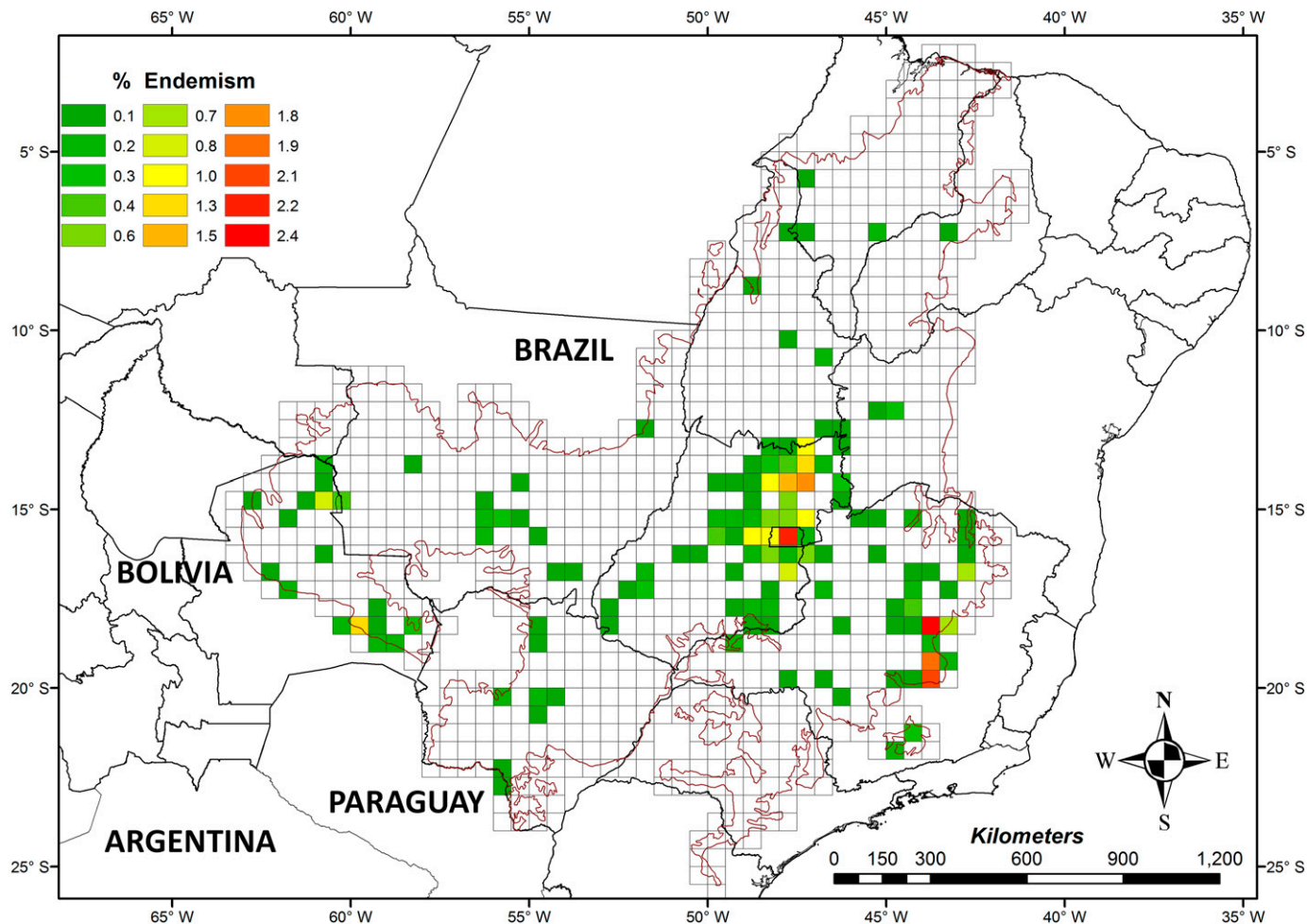


FIGURE 5 The Cerrado biome in Brazil and Bolivia with a 30' latitude \times 30' longitude grid superimposed; grid squares in which narrow-endemic species of the taxa sampled by Proença et al. (2010) and Villarroel (2011) occur are highlighted. Narrow-endemic taxa defined as restricted to an area smaller than 2° latitude \times 3° longitude (Simon and Proença, 2000).

yr in the 1775–1799 period, going up to SD = ca. 43 yr in 1950–1974 and 1975–1999. For the lag between type collection and description, SD = ca. 5 yr in 1775–1799 going up to ca. 27 yr in 1975–1999. Our study thus confirms that undescribed species accumulate in herbaria, some recognized as new and bearing manuscript names, but yet to be formally described and published (Bebber et al., 2010).

The 1925–1949 period is atypical. In absolute values, this period had the lowest number of types (4) collected and species (5) described in the data set. It was also the only period in which the standard deviation in delay between first known collection and description fell significantly. This low standard deviation could be an artifact of the low number of species, but the low number of new descriptions is real and might be explained by two factors. The first factor is the European effects of the post-World War I economic depression and World War II (only one species was both collected and described by a European botanist). The second factor is the rise of Brazilian botanists or botanists naturalized in Brazil who were concentrating their efforts in the Amazon and the Atlantic Forest (Heizer, 2010; CRIA, 2015). The species of this period were mostly either collected or described by one of the two most important Brazilian botanists of this period, Hoehne and Kuhlmann; if collected by neither, types were collected by botanists that were either born

or long established in Brazil. The type localities from this period were in the Amazon (Amapá), the Atlantic Forest (Rio de Janeiro) or the Atlantic Forest/Cerrado transitional zone (Itacolomi in Minas Gerais).

Age of binomial becomes significantly greater as an angiosperm species' range and occupancy increases. Green Star (Wide-ranging) species binomials are on average 167 yr old (average description during 1848), Blue Star (Regional) species binomials are on average 109 yr old (average description during 1906), Gold Star (Local) species binomials are on average 83.5 yr old (average description between 1931 to 1932) and Black Star (Narrow-endemic) species binomials are an average 59 yr old (average description during 1956). Green Star (Wide-ranging) species were described significantly earlier ($p < 0.001$) than all other categories. Blue Star (Regional) species were described significantly ($p < 0.001$) earlier than Black Star (Narrow-endemic) species, but there was no significant difference between time of description of Blue Star (Regional) and Gold Star (Local), or between Gold Star (Local) and Black Star (Narrow-endemic) species. This pattern has been demonstrated in British plants (Bebber et al., 2007a), and in invertebrates (Allsopp, 1997; Gaston et al., 1995) and vertebrates (Blackburn and Gaston, 1995; Collen et al., 2004; Diniz Filho et al., 2005). Global distribution

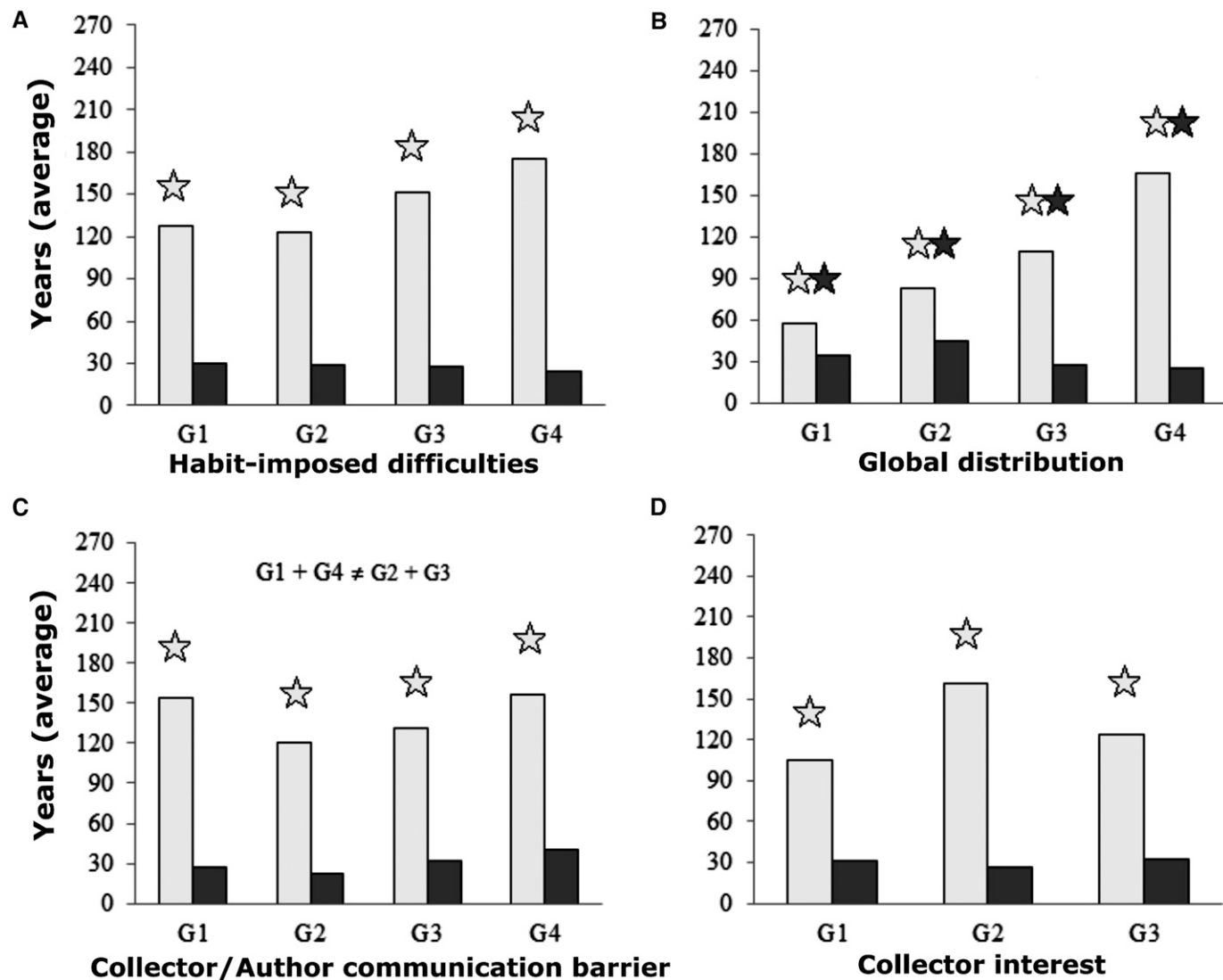


FIGURE 6 Group means and significance of multistate categorical variables in ANOVA. (A) Habit-imposed difficulties. G1 = 0.65 m to 1.95 m; G2 = ground or water level up to 0.65 m, or between 1.96 m and 2.6 m; G3 = 2.6 to 5.2 m; G4 > 5.3 m. (B) Global distribution: G1 = Green Star species (Wide-spread); G2 = Blue Star species (Regional); Gold Star species (Local); Black Star species (Narrow-endemic). (C) Collector/author communication barrier: G1 = collector and author are the same; G2 = collector and author contemporary and from the same country or collector is one of the authors; G3 = collector and author are either contemporary or from the same country but not both; G4 = collector and author are neither contemporary nor from the same country. (D) Collector interests: G1 = seed plants only; G2 = plants in general; G3 = plants and at least one of animals, fossils or minerals. Pale gray bars = age of binomial; Black bars = First collection-description delay. Pale gray stars = significant differences in age of binomial between groups. Black stars = significant differences in first collection-description delay between groups.

was, as might be expected, highly correlated with Cerrado occupancy, and these variables also showed the fourth and sixth strongest correlations with axis 1 of the PCA.

Visibility and attractiveness, as evaluated by size, number of units in the display, and colors of flowers and fruits apparently have no influence on the age of binomial at least in the Cerrado, although floral display (size) was very nearly significant. Habit-imposed difficulties, however, was weakly (but significantly) correlated with age of binomial, suggesting that plants at or around breast height (easy to see and collect) are actually described later than those that require stooping, digging, bending, pushing, pulling, or climbing. Habit-imposed difficulties, however, were also positively correlated with size of floral display and global distribution, so this might be

an indirect effect of the association between habit and global distribution (a strong driver of rapid description) or the fact that visibility of the display outweighs the difficulties of collection, not forgetting that collectors of the past often had large parties of servants. Collector duration and collector specialization showed no correlation with age of binomial.

Habitat, habit, and pollination and dispersal syndromes are complex concepts that we found difficult to rank meaningfully, a *sine qua non* condition for PCA. Habitat appeared as nonsignificant at the pilot stage and was abandoned in the final analysis. Although we attempted to logically rank collecting difficulties associated with habitat, it is possible that another ranking rationale would produce a different result. A more detailed study of habitat

difficulties would be desirable before affirming that it does not influence collection and description. Pollination syndrome (nonsignificant in the pilot screen) and dispersal syndrome (significant in the pilot screen) and habit (significant in the pilot screen) are also complex concepts and, for this reason, were deconstructed into groups of size and color of floral or fruit display for the final analysis. Although more objective and easier to rank, groups ranked by size and colors are probably an over-simplification of the visibility and attractiveness of flower and fruit displays to collectors; unexplored aspects of position, shape, or aroma may also play a part in the process. Results for categorical variables where ranking was subjective was considered tentative and were confirmed by ANOVA which does not require category ranking (see Table 1).

Collector/author barriers were nonsignificant in explaining age of binomial in the PCA, but ANOVA showed groups 1 and 4 together were significantly older than groups 2 and 3. Group 1 was of species collected and described by the same person and group 4 of species described by authors that were neither contemporary nor from the same country as the collectors, while groups 2 and 3 were species that showed various levels of collaboration (see Methods). Since both group 1 and 4 allow “solo” descriptions while group 2 and 3 tend to involve consultations and ethical decisions, we tentatively propose that these latter considerations slow down the descriptive process.

There was no significant difference in first collection-description delay between groups of categorical variables, except for two. Binomials described by nonspecialized botanists had much shorter delays (average ca. 25 yr) than those described by flowering plant specialists (ca. 35 yr), probably a historical trend associated to progressive specialization in science in parallel to dwindling collections of undescribed species. Binomials of widespread species had significantly shorter first collections-description delays (average ca. 57 yr) than those more narrowly distributed, going progressively up for regional species (average ca. 83 yr), local species (average ca. 110 yr) and narrow-endemic species (average 165 yr). This may be associated to the fact that species with small ranges usually have few available collections, hence diminishing the chances that a collection will be identified as an undescribed species.

What could hasten species description in the Cerrado biodiversity hotspot?—Two different strategies are likely to be effective: one field-based and the other herbarium-based. Access, by whatever means, to unidentified specimens at family or generic level in large herbaria that house collections from past generations of botanists by well-trained taxonomists should prove effective in identifying undescribed species. Taxonomic groups that are below 2.6 m should be preferentially targeted as they are more likely to be undescribed; another possible target group would be species with small- or medium-sized flowers in few-flowered inflorescences. Herbarium material has been collected and is available for immediate study and this may be the quickest, cheapest way to accelerate the description of new species (Bebber et al., 2010).

A second, latter (or simultaneous) strategy would be intensive collecting in poorly known areas of the cerrado biodiversity hotspot, known to have narrow-endemic species (Fig. 5). Although most collections are concentrated along major roads and near big cities (Tobler et al., 2007), there is also a trend for modern botanists to target known centers of endemism (Nelson et al., 1990; Schatz, 2002; van Gerner et al., 2005) that are likely to produce new species, even if these areas are in remote, inaccessible places (Stannard,

1995). As above, the focus should be on taxonomic groups with habits below 2.6 m with small flowers and few-flowered inflorescences, from known centers of endemism.

In short, for immediate returns in terms of new species descriptions, it is better to finance herbarium-based study, and field expeditions led by knowledgeable general botanists with significant collecting experience to under-collected putative centers of endemism, followed by time to focus on descriptive activity of the collections they have made, rather than separating field and herbarium study by different types of funding or job descriptions. Two highly productive careers followed this model: Carl Friederich von Martius in the nineteenth century and Howard Irwin in the 20th century. May's (2004) intuitive prediction in the title of his essay that “collecting new species in the field will remain the rate-limiting step” to species description will probably ultimately prove true.

ACKNOWLEDGEMENTS

The authors thank Project “Conservation of the Cerrados of Eastern Bolivia” (University of Oxford/Museo Noel Kempff Mercado, financed by The Darwin Initiative) that supported fieldwork in Bolivia (2007–2010). Project “Considerações de Biodiversidade na seleção de locais para cultivo de cana de açúcar para biocombustível” (UnB/University of Oxford, financed by BP) supported star ratings for the Cerrado flora. E.K.S.C. thanks Project “FloResCer: Flora Integrada da Região Centro-oeste, Cerrado e Pantanal” (UnB/Embrapa/UFG/UFMT, financed by Programa Reflora/CNPq) for a grant, D.V. thanks CAPES, Pós-Graduação em Botânica da Universidade de Brasília for a grant.

LITERATURE CITED

- Allsopp, P.G. 1997. Probability of describing an Australian scarab beetle: Influence of body size and distribution. *Journal of Biogeography* 24: 717–724.
- Archer, W. A. 1962. Adolpho Ducke, botanist of the Brazilian Amazon (1876–1959). *Taxon* 11: 233–242.
- Areia, M. L. R., M. A. Miranda, and T. Hartmann. 1991. Memória da Amazônia: Alexandre Rodrigues Ferreira e a Viagem Philosophica pelas capitanias do Grão-Pará, Rio Negro, Mato Grosso e Cuyabá. 1783–1792. Museu e Laboratório Antropológico da Universidade de Coimbra, Coimbra.
- Batalha, M. A. 2011. The Brazilian Cerrado is not a biome [O cerrado não é um bioma]. *Biota Neotropica* 11: 21–24. Available at <http://www.scopus.com/inward/record.url?eid=2-s2.0-84856294488&partnerID=40&md5=fa978fc07bc4df307b967c2ef97c86c>.
- Bebber, D. P., M. A. Carine, G. Davidse, D. J. Harris, E. M. Haston, M. G. Penn, S. Cafferty, et al. 2012. Big hitting collectors make massive and disproportionate contribution to the discovery of plant species. *Proceedings of the Royal Society, B, Biological Sciences* 279: 2269–2274.
- Bebber, D. P., M. A. Carine, J. R. I. Wood, A. H. Wortley, D. J. Harris, G. T. Prance, G. Davidse, et al. 2010. Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences* 107: 22169–22171.
- Bebber, D. P., S. A. Harris, K. J. Gaston, and R. W. Scotland. 2007a. Ethnobotany and the first printed records of British flowering plants. *Global Ecology and Biogeography* 16: 103–108.
- Bebber, D. P., F. H. C. Marriott, K. J. Gaston, S. A. Harris, and R. W. Scotland. 2007b. Predicting unknown species numbers using discovery curves. *Proceedings of the Royal Society, B, Biological Sciences* 274: 1651–1658.
- Blackburn, T. M., and K. J. Gaston. 1995. What determines the probability of discovering a species—A study of South-American oscine passerine birds. *Journal of Biogeography* 22: 7–14.
- Brandão, M. G. L., C. F. F. Graell and C. W. Fagg. 2011. European naturalists and medicinal plants of Brazil. In O. Grillo and G. Venora [eds.], *Biological diversity and sustainable resources use*, 101–120. Intech, Rijeka, Croatia.

- Collen, B., A. Purvis, and J. L. Gittleman. 2004. Biological correlates of description date in carnivores and primates. *Global Ecology and Biogeography* 13: 459–467.
- Cox, P. A., and P. J. Grubb. 1991. Abiotic pollination: An evolutionary escape for animal-pollinated angiosperms [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 333: 217–224.
- CRIA [Centro de Referência em Informação Ambiental]. 2015. SpeciesLink. CRIA, Campinas, Brazil. Available at <http://splink.cria.org.br/> [accessed July to November 2015].
- Cronk, Q. C. B. 1989. Measurement of biological and historical influences on plant classifications. *Taxon* 38: 357–370.
- Cruz, A. L. R. B. 2002. As viagens são os viajantes: Dimensões identitárias dos viajantes naturalistas brasileiros do século XVIII. *História. Questões & Debates: Revista da Associação Paranaense de História* 36: 61–98.
- de León, P. C. 1962. Historiadores primitivos de Indias II: La crónica del Peru. *Biblioteca de Autores Españoles* 26: 354–458. M. Rivadeneyra, Madrid, Spain.
- de Oliveira Filho, A. T., and J. A. Ratter. 2002. Vegetation physiognomies and woody flora of the cerrado biome. In P. S. Oliveira and J. R. Marquis [eds.], *The Cerrados of Brazil: Ecology and natural history of a neotropical savanna*, 91–120. Columbia University Press, New York, New York, USA.
- Dean, W. 1991. A botânica e a política imperial: A introdução e a domesticação de plantas no Brasil. *Estudos Historicos (Rio de Janeiro, Brazil)* 4: 216–228.
- Diniz-Filho, J. A. F., R. P. Bastos, T. F. L. V. B. Rangel, L. M. Bini, P. Carvalho, and R. J. Silva. 2005. Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Global Ecology and Biogeography* 14: 469–477.
- Endersby, J. 2010. Imperial nature. Joseph Hooker and the practices of Victorian science. University of Chicago Press, Chicago, Illinois, USA.
- Felippe, G., and L. P. Zaidan. 2008. Do Éden ao Éden: Jardins botânicos e a aventura das plantas. SENAC, São Paulo, Brazil.
- Forzza, R. C., P. M. Leitman, A. F. Costa, A. A. Carvalho Jr., A. L. Peixoto, B. M. T. Walter, C. Cicudo, et al. 2012. Lista de espécies da flora do Brasil. Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil.
- Franco, J. L. A., and J. A. Drummond. 2009. Proteção à natureza e identidade nacional no Brasil, anos 1920–1940. Fiocruz, Rio de Janeiro, Brazil.
- Freyreiss, G. W. 1907. Viagem ao interior do Brasil: 1814–1815. Typographia do Diário Oficial, São Paulo, Brazil.
- Funk, V. A. 1997. Using collections data and GIS to examine biodiversity information levels in Guyana. In K. E. Hoagland and A. Y. Vrijji [eds.], *Global genetic resources: Access, ownership, and intellectual property rights*, 117–128. U.S. Agricultural Research Service and Association of Systematics Collections, Washington, D.C., USA.
- Gardner, G. 1846. Travels in the interior of Brazil, principally through the Northern Provinces, and the gold and diamond districts, during the years 1836–1841. Reeves Brothers, London, UK.
- Gaston, K. J., T. M. Blackburn, and N. Loder. 1995. Which species are described first?: The case of North-American butterflies. *Biodiversity and Conservation* 4: 119–127.
- Giulietti, A. M., R. M. Harley, L. P. de Queiroz, M. das G. L. Wanderely, and C. van den Berg. 2005. Biodiversidade e conservação das plantas no Brasil. *Megadiversidade* 1: 52–61.
- Hammer, Ø., D. A. T. Harper, and P. D. Ryan. 2001. Past: Paleontological statistics software package for education and data analysis. Available at <http://folk.uio.no/ohammer/past/> [accessed November 2015].
- Harvard University Herbaria. 2013. Index of botanists. Harvard University Herbaria, Cambridge, Massachusetts, USA. Available at <http://kiki.huh.harvard.edu/databases/> [accessed 2013–2015].
- Hawthorne, W., and M. Abu-Juam. 1995. Forest protection in Ghana—with particular reference to vegetation and plant species. International Union for Conservation of Nature, Cambridge, UK.
- Heizer, A. 2010. João Geraldo Kuhlmann e a Comissão de Defesa da Borracha de 1912. In A. Heizer and A. P. Videira [eds.], *Ciência, civilização e república nos trópicos. Mauad/Faperj*, Rio de Janeiro, Brazil.
- Hemming, J. 2007. Ouro vermelho: A conquista dos índios brasileiros. Editora Universidade de São Paulo, São Paulo, Brazil.
- Higham, J. 1980. The matrix of specialization. *Bulletin of the American Academy of Arts and Sciences* 33: 9–29.
- Howe, H., and J. Smallwood. 1982. Ecology of seed dispersal. *Annual Review of Ecology and Systematics* 13: 201–228.
- Jardim Botânico do Rio de Janeiro. 2015. Herbário Virtual Re flora. Rio de Janeiro, Rio de Janeiro, Brazil. Available at <http://reflora.jbrj.gov.br/reflora/herbarioVirtual> [accessed December 2015].
- Jarvis, C. 2007. Order out of chaos: Linnaean plant names and their types. Linnean Society of London in association with the Natural History Museum, London, UK.
- Joppa, L. N., D. L. Roberts, N. Myers, and S. L. Pimm. 2011. Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences, USA* 108: 13171–13176.
- JSTOR Global Plants. 2012–2015. Ithaka, New York, New York, USA. Available at <http://plants.jstor.org/> [accessed August 2012 to December 2015].
- Klink, C. A., and R. B. Machado. 2005. Conservation of the Brazilian Cerrado. *Conservation Biology* 19: 707–713.
- Kohlhepp, G. 2005. Scientific findings of Alexander von Humboldt's expedition into the Spanish-American Tropics (1799–1804) from a geographical point of view. *Anais da Academia Brasileira de Ciências* 77: 325–342.
- Kury, L. 2001. Viajantes-naturalistas no Brasil oitocentista: Experiência, relato e imagem. *História, Ciências, Saúde-Manguinhos* 8: 863–880.
- Landrum, L. R., and M. I. U. de Oliveira. 2010. A new species of *Campomanesia* (Myrtaceae) from Bahia, Brazil, based on specimens collected by J.S. Blanchet over 150 years ago. *Journal of the Botanical Research Institute of Texas* 4: 603–607.
- Leite, M. L. M. 1995. Naturalistas viajantes. *História, Ciências, Saúde-Manguinhos* 1: 7–19.
- Linnaeus, C. 1753. Species plantarum: Exhibentes plantas rite cognitadas ad genera relatas cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus secundum systema sexuale digestas. Available at <http://books.google.com/books?hl=en&lr=&id=jQ0AAAAQAAJ&oi=fnd&pg=PA1&dq=Species+plantarum:+exhibentes+plantas+rite+cognitas,+ad+genera+relatas,+cum+differentiis+specificis,+nominibus+trivialibus,+synonymis+selectis,+locis+natalibus,+secundum+systema+sexu>.
- Marx, R. B. 2004. Arte & Paisagem. Studio Nobel, São Paulo, Brazil.
- May, R. M. 2004. Tomorrow's taxonomy: Collecting new species in the field will remain the rate-limiting step. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 359: 733–734.
- McCune, B., and M. J. Mefford. 2011. PC-ORD: Multivariate analysis of ecological data, version 6.0. MjM Software, Gleneden Beach, Oregon, USA.
- McNeill, J., F. R. Barrie, W. R. M. Buck, V. Demoulin, W. Greuter, D. L. Hawksworth, P. S. Herendeen, et al. 2012. International Code of Botanical Nomenclature (Melbourne Code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011. Available at <http://www.iapt-taxon.org/nomen/main.php> [accessed May 2014].
- Mendonça, R. C., J. M. Felfili, B. M. T. Walter, M. C. Silva Júnior, A. V. Rezende, T. S. Filgueiras, P. E. Nogueira, and C. W. Fagg. 2008. Flora vascular do bioma cerrado: Checklist com 12.356 espécies. In S. M. Sano, S. P. Almeida, and J. F. Ribeiro [eds.], *Cerrado: Ecologia e flora*, 421–442. EMBRAPA Informação Tecnológica, Brasília, Brazil.
- Miller, D. P., and P. H. Reill. 2010. Visions of empire: Voyages, botany, and representations of nature. Cambridge University Press, Cambridge, UK.
- Moreira, L. F. V., and M. C. Quinteros. 2008. Em busca de Aimé Bonpland. *Revista Fronteiras* 10: 221–236.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.
- Nelson, B. W., C. A. C. Ferreira, M. F. da Silva, and M. L. Kawasaki. 1990. Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature* 345: 714–716.
- Ogilvie, B. W. 2006. The science of describing: Natural history in Renaissance Europe. University of Chicago Press, Chicago, Illinois, USA.
- Parkinson, P. 1975. The International Code of Botanical Nomenclature: An historical review and bibliography. *Tane* 21: 153–173.

- Paton, A. J., N. Brummitt, R. Govaerts, K. Harman, S. Hinchcliffe, B. Allkin, and E. N. Lughadha. 2008. Towards Target 1 of the Global Strategy for Plant Conservation: A working list of all known plant species—Progress and prospects. *Taxon* 57: 602–611.
- Pohl, J. E. 1976. Viagem no interior do Brasil: 1817–1821. Editora Universidade de São Paulo, São Paulo, Brazil.
- Proença, C. E. B., L. H. Soares-Silva, V. L. Rivera, M. F. Simon, R. C. Oliveira, I. A. Santos, J. A. N. Batista, et al. 2010. Regionalização, centros de endemismos e conservação com base em espécies de Angiospermas indicadoras da biodiversidade do Cerrado brasileiro. In I. R. Diniz, J. Marinho Filho, R. B. Machado, and R. B. Cavalcanti [eds.], *Cerrado: Conhecimento científico quantitativo como subsídio para ações de conservação*, 91–148. Thesaurus, Brasília, Brazil.
- Raven, P. H., B. Berlin, and D. E. Breedlove. 1971. The origins of taxonomy. *Science* 174: 1210–1213.
- Reveal, J. L., and C. E. Jarvis. 2009. Typification of names of temperate North American plants proposed by Linnaeus. *Taxon* 58: 977–984.
- Saint-Hilaire, A. F. C. P. 1937. Viagem às nascentes do Rio São Francisco e pela província de Goyaz. Companhia Editora Nacional (Coleção Brasileira), Rio de Janeiro, Brasil.
- Schatz, G. E. 2002. Taxonomy and herbaria in service of plant conservation: Lessons from Madagascar's endemic families. *Annals of the Missouri Botanical Garden* 89: 145–152.
- Schippmann, U., D. Leaman, and A. B. Cunningham. 2006. A comparison of cultivation and wild collection of medicinal and aromatic plants under sustainability aspects. In R. J. Bogers, L. E. Craker, and D. Lange [eds.], *Medicinal and aromatic plants: Agricultural, commercial, ecological, legal, pharmacological and social aspects*, 75–95. Springer, Wageningen, Netherlands.
- Sequeira, M. M., A. Santos-Guerra, C. E. Jarvis, A. Oberli, M. A. Carine, M. Maunder, and J. Francisco-Ortega. 2010. The Madeiran plants collected by Sir Hans Sloane in 1687 and his descriptions. *Taxon* 59: 598–612.
- SIEG [Sistema Estadual de Geoinformação]. 2015. Instituto Mauro Borges de Estatísticas e Estudos Socioeconômicos, Goiânia, Brazil. Available at <http://www.sieg.go.gov.br/> [accessed July 2016].
- Shteir, A. B. 1989. Botany in the breakfast room: Women and early nineteenth-century British plant study. In P. G. Abir-Am and D. Outram [eds.], *Uneasy careers and intimate lives: Women in science, 1789–1979*, 31–44. Rutgers University Press, New Brunswick, New Jersey, USA.
- Simon, M. F., and C. E. B. Proença. 2000. Phytogeographic patterns of *Mimosa* (Mimosoideae, Leguminosae) in the Cerrado biome of Brazil: An indicator genus of high-altitude centers of endemism? *Biological Conservation* 96: 279–296.
- Spix, J. B., and C. F. P. von Martius. 1938. Através da Bahia, São Paulo, Rio de Janeiro, Recife e Porto Alegre. Companhia Editora Nacional, Rio de Janeiro, Brazil.
- Stannard, B. 1995. Introduction. In B. Stannard [ed.], *Flora of the Pico das Almas*, 43–79. Royal Botanical Gardens, Kew, UK.
- Ter Steege, H., P. P. Haripersaud, O. S. Bánki, and F. Schieving. 2011. A model of botanical collectors behavior in the field: Never the same species twice. *American Journal of Botany* 98: 31–37.
- The Access and Benefit-Sharing Clearing-House. 2016. Secretariat of the Convention on Biological Diversity, Montreal. Available at <https://absch.cbd.int/> [accessed July 2016].
- The Plant List version 1.1. 2013. Royal Botanic Gardens, Kew and Missouri Botanical Garden. Available at <http://www.theplantlist.org/> [accessed April 2016].
- Tobler, M., E. Honorio, J. Janovec, and C. Reynel. 2007. Implications of collection patterns of botanical specimens on their usefulness for conservation planning: An example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. *Biodiversity and Conservation* 16: 659–677.
- Tropicos. 2015. Missouri Botanical Garden, St. Louis, Missouri, USA. Available at <http://www.tropicos.org/> [accessed April 2015].
- Universidade de Brasília. 2014. FloResCer. Available at <http://www.florescer.unb.br/> [accessed July 2014].
- van der Pijl, L. 1982. Principles of dispersal in higher plants. Springer-Verlag, Berlin, Germany.
- van Gernerden, B. S., R. S. Etienne, H. Olff, P. W. F. M. Hommel, and F. Van Langevelde. 2005. Reconciling methodologically different biodiversity assessments. *Ecological Applications* 15: 1747–1760.
- Villarroel, D. 2011. Aportes ao conhecimento da flora e diversidade do bioma Cerrado do Brasil e da Bolívia. Universidade de Brasília, Brasília, Brazil.
- von Martius, C. F. P., A. W. Eichler, and I. Urban. 1904. Flora brasiliensis, vol. 1 [online version]. Frid. Fleischer In Comm., Leipzig, Germany. Available at www.florabrasiliensis.cria.org.br/ [accessed 2012 to 2015].
- Walters, S. M. 1961. The shaping of angiosperm taxonomy. *New Phytologist* 60: 74–84.
- Warming, E. 1908. Lagoa Santa: Contribuição para a geographia phytobiologica. Imprensa Official do Estado de Minas Geraes, Belo Horizonte, Brazil.
- Wied-Neuwied, M.A.P. 1942. Viagem ao Brasil. São Paulo, Rio de Janeiro, Recife e Porto Alegre. Companhia Editora Nacional (Coleção Brasileira), Rio de Janeiro, Brazil.
- Willmer, P. 2011. Pollination and floral ecology. Princeton University Press, Princeton, New Jersey, USA.