LET THE RECORDS SHOW: ATTRIBUTION OF SCIENTIFIC CREDIT IN NATURAL HISTORY COLLECTIONS

Rebecca B. Dikow,[1,*] Jenna T. B. Ekwealor,* William J. B. Mattingly,* Michael G. Trizna,* Elizabeth Harmon,† Torsten Dikow,‡ Carlos F. Arias,*,§ Richard G. J. Hodel,*,‖ Jennifer Spillane,* Mirian T. N. Tsuchiya,* Luis Villanueva,# Alexander E. White,* Madeline G. Bursell,*,** Tiana Curry,*,†† Christelle Inema,*,‡‡ and Kayla Geronimo-Anctil*,§§

*Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, USA; †Smithsonian Libraries and Archives, Smithsonian Institution, Washington, DC, USA; ‡Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA; §Smithsonian Tropical Research Institute, Smithsonian Institution, Panama City, Panama; ‖Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA; #Digitization Program Office, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, USA; **North Carolina State University, Raleigh, North Carolina, USA; ††Viterbi School of Engineering, University of Southern California, Los Angeles, California, USA; ‡‡Harris School of Public Policy, University of Chicago, Chicago, Illinois, USA; and §§California State University, Channel Islands, Camarillo, California, USA

Guest Editor: Mauricio Bonifacino

Premise of research. Natural history collections are essential resources for taxonomy, systematics, and ecological and climate change research. Mass digitization of these collections provides the opportunity to study broad biological patterns among specimens and their associated metadata at a scale that was previously impossible. The specimen metadata can also be used to study the contributions of the people that collected and identified these specimens. A proper accounting of these contributions impacts our understanding of the history of these collections and who played a role in their growth.

Methodology. Here, we provide an assessment of the scientific contributions of past women in science at the Smithsonian Institution, focusing on their specimen collections and identifications. We evaluate natural history specimen collections records available from the Global Biodiversity Information Facility and Smithsonian annual reports, volumes dating to the founding of the Smithsonian in 1846.

Pivotal results. We identify 40 women with specimen collections or identifications, with a total of more than 120,000 total specimens attributed to them. In cases where specimens are not yet digitized, we are able to learn more about the women's contributions using annual reports, which provide a richer picture of their work at the Smithsonian. This work relies on collaboration as well as deep institutional knowledge. We also release a semantic search application, which allows users to search the Smithsonian annual reports.

Conclusions. Collections records are a rich resource, but there are significant barriers to accurate specimen attribution, which disproportionately affect women collectors and determiners. We propose ways that we might document these problems at scale and remedy cases of misattribution in digital repositories of record.

Keywords: Smithsonian Institution, collections records, specimens, women in science, natural history museums.

Introduction

Natural history collections (NHCs) have been amassed and studied for centuries and are essential resources for biodiversity and global change research (Meineke et al. 2019). Indeed, the list of 100 uses of a herbarium (well, at least 72) described by Vicki Funk (2004) continues to grow; one can now add numerous recent advances in molecular and computational methods. Digitization and advanced computation, combined with the improved use of DNA and even RNA from museum specimens, allow us to compare and summarize vast numbers of specimens and their biological traits of interest (e.g., evolutionary relationships, ecological similarities, phenological histories; Short et al. 2018; Soltis et al. 2020; Folk et al. 2021; Speer et al. 2022). However, because these collections do not exist independently of the political and social systems that enabled their collection and production (Wintle 2016; Daru et al. 2018; Vogel 2019; Hughes et al. 2021; Park et al. 2021), we can also leverage NHCs databases to assess the people and communities that contribute to the history of science and evaluate the social and political dynamics that influence their participation.

Even basic summary statistics of these collections can reveal significant historical forces. For example, geographic disparities

[1] Author for correspondence; email: dikowr@si.edu.

Manuscript received September 2022; revised manuscript received January 2023; electronically published May 12, 2023.

  https://doi.org/10.1086/724949

in where specimens are held (overwhelmingly in institutions in the Global North) versus where biodiversity occurs are stark reminders of the continued impact of colonialism (Park et al. 2021; Raja et al. 2021). These disparities perpetuate the systemic barriers to participation in science for people of color. While any plan to counter the negative impact of such disparities is surely complex, we contend that any remedies must begin with a basic proper accounting of the ways that NHCs reveal a history of marginalization for historically excluded groups. Here, we examine how women are represented in the historical record of NHCs and how historical institutional and societal forces disguise their full contribution to the Smithsonian Institution and their scientific impact as members of the Smithsonian community.

Scientific productivity is most often measured in terms of grant funding awarded and articles published, but these metrics often conceal the support of many other people, including technicians, local field guides, students, and, in particular for NHCs, specimen preparators, data and collections managers, and collections staff. It is easy to lose the names and stories of those working in support roles over time because they are not as prominent and because they may not be present in the publication record. For historical women, it can be even harder for their names and stories to be preserved because of the social and political climates of their time. For example, many of the women working in science who were affiliated with the Smithsonian in its first ~100 years were there as volunteers or working alongside their spouses and were not officially employed. This is not only true at the Smithsonian. In one notable case at the University of Massachusetts, Amherst, prolific entomologist C. P. Alexander, who described more than 11,000 species of crane flies, was reliant on the work of his wife Mabel Alexander. Mabel was involved in every part of the process of collecting and processing specimens and describing species (Heard 2020).

Section 213 of the 1932 Economy Act required that government workers whose spouses also held federal jobs should be let go first when staff reductions were necessary. This law disproportionately affected women, whose salaries were usually less than their husbands' salaries. This happened to Doris Blake, an entomologist with the USDA, whose husband was a botanist there. She was let go and wrote a letter to the secretary of agriculture to protest (Harmon 2021). Blake subsequently worked at the National Museum of Natural History (NMNH) for the rest of her career, mostly unpaid (Froeschner et al. 1981). If women were not officially employed while they were affiliated with the Smithsonian, we cannot use employment records to learn more about their tenures and work and must rely on information from archives and the institutional knowledge passed down within science units and departments.

The Funk List

The challenges related to documenting the work done by women over the Smithsonian's history directly led to the creation of a "living" list called the "Funk List," which we use as a starting point for the analyses presented here. The Funk List is a list of past and present women in science affiliated with the Smithsonian (AWHI 2019). It began as one of the early activities of the Smithsonian American Women's History Initiative (AWHI), the precursor to the Smithsonian American Women's History Museum. The AWHI curatorial committee, including NMNH Botanist and Senior Scientist Vicki Funk, led a collaborative effort to gather data from all Smithsonian science units, including the NMNH, Smithsonian Tropical Research Institute, Smithsonian Astrophysical Observatory, National Zoological Park and Conservation Biology Institute, Museum Conservation Institute, National Air and Space Museum, and other units where people may perform or study science, including the National Museum of American History and the National Asian Art Museum. Funk was instrumental in going door-to-door to ask current staff about the women who historically worked or contributed to science in their departments, whether as staff scientists, curators, assistants, technicians, aides, volunteers, research associates, or collaborators. When Funk passed away in October 2019, the list was named the "Funk List" in her honor. Curation continues to refine this list—both to add names to the list and to add additional data about the women already on it.

Because many of the women on the Funk List were engaged in collections-based research, we can leverage the work being done to improve specimen attributions and the linking of biodiversity data in order to learn more about their scientific impact at the Smithsonian (e.g., Page 2019; Shorthouse and Page 2019; Groom et al. 2020). Bionomia (https://bionomia.net/), a website for attributing specimens to collectors and determiners, is particularly useful because it links specimen records from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/), using Wikidata and ORCID identifiers. For women on the Funk List who do not appear on specimen labels, we can use other documents, including the Smithsonian annual reports, to learn more about their work. The annual reports contain staff lists, accounts of expeditions and collecting trips, and descriptions of activities across the institution. Advances in natural language processing allow us to build semantic searching tools. Instead of being limited by exact name searches, we can use machine learning to build tools to find nuanced topics in the annual reports and other documents. For some women on the Funk List, the annual reports are one of the only places where we find their names in the documents that are digitized to date.

Collections Digitization and Specimen Attribution

Collections digitization has accelerated because of improvements in technology and also investments, including the Integrated Digitized Biocollections, the National Resource for Advancing Digitization of Biodiversity Collections funded by the National Science Foundation (https://www.idigbio.org/). At the Smithsonian, the digitization program office, part of the office of the chief information officer, undertakes collaborative mass digitization projects across Smithsonian museums. One of these projects has been the digitization of the herbarium at NMNH, which is now fully digitized (both images of sheets and label transcriptions for more than 4.5 million specimens). Botanical specimens lend themselves to mass digitization projects because they are mostly flat and the sheets on which they are mounted are consistent in size. A custom conveyor-belt system was operated by collections staff to photograph specimens. Volunteers (known at the Smithsonian as "volunpeers") transcribed many of the specimen labels through the Smithsonian transcription center (https://transcription.si.edu/) and a contracted transcription service completed the rest. Transcriptions and images were then imported into the NMNH collections information system by

NMNH data managers and collections staff. NMNH contributes their specimen data (from botany, entomology, invertebrate zoology, and vertebrate zoology departments) to the GBIF, which aggregates specimen data from NHCs across the world. We leverage these data to learn more about the collections work of the women on the Funk List.

While we are fortunate to have access to these resources, digitization of other NMNH collections is proceeding at a slower pace because of challenges in the digitization process. Insects on pins, for example, many with labels obscured until unpinning, require significantly more person-time to unpin and place to photograph. Multiple views for each specimen may also be necessary to capture traits of interest. The NMNH entomology collection houses approximately 35 million specimens, of which 19 million are on pins, and its digitization will be a larger and more complex undertaking. When specimens are not digitized, however, the primary biodiversity data are only accessible to those able to visit the collection or request loans of specimens. We will not have the full picture of who built the collections without these data.

There also continue to be significant challenges in accurately attributing specimens to people, even for fully digitized collections, and these challenges are amplified for women. In 2019 we found specimens in the NMNH collections database that were collected by Mary Vaux Walcott, a naturalist and botanical illustrator, but they were attributed to her husband, Charles Doolittle Walcott, a paleontologist and former Smithsonian secretary (Dikow and Glenn 2020). We noticed these specimens because they were collected after Charles's death, and we were able to confirm that they were indeed collected by Mary by visually inspecting the digital image of the specimens. In this case, these specimens were labeled as either "Mrs. C. D. Walcott" or "Mrs. Walcott," and "Mrs." was not included in the digital record. This underscores the importance of the specimen image, as without it, we might have assumed that the date was wrong. This also led us to think about how we could document the frequency of this type of misattribution, how many records were affected, and what else we could learn about challenges with NHCs data in our aim to better document women in science at the Smithsonian.

Humans in the Loop

While data science tools aid us in the effort to find and document the work of women in Smithsonian collections, this work must be human centered. We aim to avoid misgendering people, and we do not use software that attempts to assign gender based on first names. We instead use first-person accounts in which people self-identify, and if that is not possible, we use documents written about the person and use the pronouns present in these documents. This approach is not perfect, as social and political climates may have prevented people from using the pronouns or presenting as the gender for which they identified. Nonetheless, we have made the decision to take these accounts at face value, and for this study, we considered all individuals on the Funk List to be women. We are very fortunate to have access to the personal papers of many Smithsonian-affiliated scientists in our Smithsonian Libraries and Archives (SLA), which are used by SLA archivists and historians to bring the stories of many of these women to Smithsonian audiences through their blog series "Wonderful Women Wednesday." Wikipedia pages have been created for many of the women on the Funk List, and we know a lot about many of their lives and careers. What is missing, however, is a deep dive into their contributions to NHCs. It is truly inspiring to read about the lives of these women. Learning about them has led us to want to learn more about their networks and the other women with whom they worked. We include compelling biographical details of a subset of these women in table 1.

Material and Methods

We used two data sets of NMNH collections data from GBIF, the NMNH extant specimens data set and the NMNH paleobiology specimens data set (GBIF 2022*a*, 2022*b*; Orrell 2022*a*, 2022*b*). These data sets conform to the Darwin Core standard (https://github.com/tdwg/dwc; Wieczorek et al. 2012). We considered the following Darwin Core fields: recordedBy (the name(s) of the specimen collectors), identifiedBy (the determiner(s) of the specimen collectors), institutionCode (the institution where the specimen resides), family (the taxonomic family of the specimen), collectionCode (the department or collection where the specimen resides), year (the year in which the specimen was collected), and countryCode (the country where the specimen was collected). It is important to consider how these data make their way to GBIF. The data in the NMNH collections information system (an instance of EMu; https://emu.axiell.com) are pushed to GBIF in Darwin Core Archive format via the Integrated Publishing Toolkit (IPT; https://www.gbif.org/ipt). Not all fields from EMu map cleanly to Darwin Core fields, so the Darwin Core Archive is not a full representation of the EMu database. After GBIF receives the Darwin Core Archive export from the NMNH IPT, GBIF runs the data set through an extensive quality control pipeline before the records are incorporated into the GBIF data. We downloaded the data from GBIF to leverage these quality control steps.

We used the Funk List as a starting point to search for women in NMNH collections data. The Funk List includes names, birth and death dates when known, educational history, spouse's name, and other information added as SLA staff members continue their research. For most of the women on the Funk List, SLA staff have created Wikidata entries. We used these Wikidata ID numbers to add members of the Funk List that had a connection to natural history specimens to Bionomia if they were not already there. For this project, we only considered individuals from the Funk List that are deceased in order to capture their complete careers. In Bionomia, we attributed specimens (both to collectors and determiners) to all Funk List individuals for which we could find specimens, as well as for other Smithsonian collectors and determiners, focusing particularly on those other individuals who were connected either personally or professionally to the Funk List individuals. For some NMNH departments, in particular entomology, for which few specimens are digitized in relation to the whole collection, we used a literature search to look for mentions of specimens collected or identified (or species described) by Funk List individuals. We then located some of these specimens found in the literature in the collection and visually inspected the labels to confirm attribution.

Following attribution, "specimen" CSV files were downloaded from Bionomia for all Funk List individuals. We used R (ver. 4.2.0; R Core Team 2021) to plot the collected and determined specimens for all Funk List individuals. For a subset

Table 1

Biographical Highlights from Selected Women on the Funk List

Name	Years of life	Biographical highlights
Mary Jane Rathbun	1860–1943	Rathbun gave up her position (second assistant curator in the division of marine invertebrates) in invertebrate zoology at the National Museum of Natural History (NMNH) so that Waldo Schmitt could be hired (McCain 1943); more than half of her publications (84 of 166) were published after she "retired" to offer her position to Schmitt (Henson 2014; Farabaugh 2015)
Mary Agnes Chase	1869–1963	Chase was a botanist who collected and identified thousands of specimens; she wanted to travel to Panama to collect plants on Barro Colorado Island (BCI) but was not allowed to because only men were allowed on BCI (Henson 2003); decades later, Blake and Cochran were able to travel to BCI
Doris Holmes Blake	1892–1978	Blake was an entomologist who for most of her career was a volunteer and honorary research associate without an office space; she described 818 species of beetles and published 100 papers (Froeschner et al. 1981); she also traveled with Cochran
Doris Mable Cochran	1898–1968	Cochran was an NMNH herpetologist who rose to the rank of curator near the end of her career and traveled with Blake to Central and South America to collect specimens and visit museum collections, funded by the National Science Foundation; Cochran's travelog of their trip is available online and was recently transcribed by volunteers (https://transcription.si.edu/project/6618)
Regina Olson Hughes	1895–1993	Hughes worked at the USDA and as a translator at the State Department and then volunteered at NMNH as a botanical illustrator after "retiring"; she was also deaf and received an honorary doctorate from Gallaudet University in Washington, DC (https://siarchives.si.edu/blog/regina-hughes)
Sophie Lutterlough	1910–2009	Lutterlough was the first woman elevator operator at NMNH; she was Black and began working at NMNH at a time when it was legal to discriminate on the basis of race and Black people were prevented from working in curatorial roles (Sayah 2016); she eventually became a scientific assistant in entomology
Margaret S. Collins	1922–1996	Collins was a research associate at NMNH and one of if not the first Black woman entomologist; she traveled and taught extensively and was an active civil rights activist (Lewis 2016)

of Funk List individuals (botany: Calderón, Chase, Funk, Rudd; invertebrate zoology: McLaughlin, Pettibone, Rathbun, Rice) we plotted their collections and identifications grouped by world region. We aggregated countries into world regions by first translating country names to English with the R package googleLanguageR (Edmondson 2020) and the Google Translate API () and then classifying countries into regions with the countrycode package (Arel-Bundock et al. 2018), using the seven world regions defined by the World Bank development indicators. We added Antarctica as an eighth world region and manually added any countries that were not automatically placed into a region. We used the Bionomia CSV specimen files to compile a list of all taxonomic families collected by Funk List individuals and highlighted these branches on a family-level tree of life built using the Open Tree of Life (OpenTreeOfLife et al. 2019) R package, rotl (Michonneau et al. 2016). Additional R packages for data analysis and visualization include gdata (Warnes et al. 2022), dplyr (Wickham et al. 2022), maps (Becker et al. 2021), and ggplot2 (Wickham 2016).

To query the NMNH Darwin Core Archives for Funk List individuals, we generated regular expressions formulas for each individual's name that accounted for different combinations of name arrangements (initials, first/middle/last, last name first, etc.). We wrote Python scripts to query the Darwin Core Archives using these formulas, both for Funk List individuals and their spouses and separately for recordedBy and identifiedBy fields, and captured the results in CSV files.

In addition to the Darwin Core Archives, we looked for mentions of Funk List individuals in historical Smithsonian annual reports. These publications are called either *Smithsonian Year* or *Annual Report of the Board of Regents of the Smithsonian Institution*. We also used United States National Museum annual report documents as well as *The National Museum of Natural History: 75 Years in the Natural History* by Ellis L. Yochelson (published in 2000) and volumes of *Explorations and Field-Work of the Smithsonian Institution* (years 1921–1940). We downloaded the JPG versions of these documents from https://library.si.edu/digital-library/collection/smithsonian-legacy-publications except for a few volumes of the Smithsonian annual reports that were unavailable there and that we downloaded instead from the Biodiversity Heritage Library (https://biodiversitylibrary.org). We ran optical character recognition (OCR) across all documents using Tesseract (ver. 5.2.0; https://github.com/tesseract-ocr/tesseract) using PDF and TXT flags to produce new PDF and TXT files. OCR results were cleaned using Python scripts. We developed a custom spaCy pipeline in Python for querying the annual reports and *Explorations* text OCR output files for each Funk List individual as well as spouses when known, as above. The output from the pipeline identifies and extracts all variant names for Funk List individuals and their spouses. The data are stored in the spaCy Doc container object as custom attributes. We captured results in JSON files indicating the file name, name string found, sentence identified with the name mentioned, and the Funk List individual's name. We filtered out mentions that were from staff lists without other context. We also used txtai (https://github.com/neuml/txtai) to build a search application implemented in Streamlit, with links to the document PDFs.

Results

Specimen Attributions

As of the time of download, the NMNH extant specimens data set contained 9,074,413 occurrence records, and the NMNH paleobiology data set contained 718,471 occurrence

Table 2

Women on the Funk List with Specimens Attributed in Bionomia

Name	Bionomia profile	No. specimen attributions
Aime M. Awl	https://bionomia.net/Q109588174	2
Jessie G. Beach	https://bionomia.net/Q108536179	2
Jean Milton Berdan	https://bionomia.net/Q6171199	741
Tillie Berger	https://bionomia.net/Q109588177	8
Doris Holmes Blake	https://bionomia.net/Q5297944	466
Pearl Lee Boone	https://bionomia.net/Q56289965	339
Cleofé Calderón	https://bionomia.net/Q8347261	8552
Mary Agnes Chase	https://bionomia.net/Q3822242	27,735
May Belle Hutson Chitwood	https://bionomia.net/Q110242995	2035
Doris Mable Cochran	https://bionomia.net/Q521351	853
Margaret James Collins	https://bionomia.net/Q19662936	7
Serena Dandridge	https://bionomia.net/Q56041985	230
Frances Densmore	https://bionomia.net/Q469150	12
Marie Poland Fish	https://bionomia.net/Q41486582	1
Vicki Ann Funk	https://bionomia.net/Q19060876	16,466
Louisa Bernie Gallaher	https://bionomia.net/Q107589105	4
Julia Anna Gardner	https://bionomia.net/Q15999449	115
Grace E. Glance	https://bionomia.net/Q69787738	2
Roxie Collie Layborune	https://bionomia.net/Q15920982	293
Eula Davis McEwan	https://bionomia.net/Q108535415	57
Patsy Ann McLaughlin	https://bionomia.net/Q21340495	5825
Mary Miller	https://bionomia.net/Q67186411	1757
Sophy Ivanova Parfin	https://bionomia.net/Q69787671	123
Kittie Fenley Parker	https://bionomia.net/Q21522642	2281
Isabel C. Perez Farfante	https://bionomia.net/Q6077744	542
Marian H. Pettibone	https://bionomia.net/Q55264645	10,246
Mary Jane Rathbun	https://bionomia.net/Q2679156	22,999
Mary E. Rice	https://bionomia.net/Q34953876	1476
Suzanne Ripley	https://bionomia.net/Q110445706	479
Velva Elaine Rudd	https://bionomia.net/Q6160123	12,473
Marie-Hélène Sachet	https://bionomia.net/Q5998141	4128
Viola Schantz	https://bionomia.net/Q7933039	13
Harriet Richardson Searle	https://bionomia.net/Q16750599	2874
Lucile E. St. Hoyme	https://bionomia.net/Q70139671	14
Matilda Coxe Stevenson	https://bionomia.net/Q6787501	112
Ellen Powell Thompson	https://bionomia.net/Q62570763	93
Mary Vaux Walcott	https://bionomia.net/Q6780882	466
Rose Ella Warner Spilman	https://bionomia.net/Q110315492	235
Mildred Stratoon Wilson	https://bionomia.net/Q27829065	601

Note. This table includes names, links to Bionomia profiles, and number of specimens attributed and identified (as found to date). "Q" numbers in the Bionomia links are the Wikidata identifiers. Specimens that were both collected and identified appear twice in the final counts.

records (GBIF 2022*a*, 2022*b*). For 40 individuals on the Funk List, we were able to attribute specimens in Bionomia, either as collections or identifications. These are shown in table 2. Following specimen attribution, CSV files downloaded from Bionomia were edited to split the "recorded,identified" action into two, so that these specimens are recorded in both the recorded and the identified categories. These edited CSVs, as well as code to generate the figures, are available on GitHub (https://github.com/sidatasciencelab/Funk_List_Collections_Records). It is important to note that specimen attribution can be challenging because of missing data—for example, if no label image is present or if metadata fields, in particular "year," are missing. We attributed specimens to the best of our abilities and using all the data we had, including dates of affiliation with the Smithsonian, birth and death dates, and spouse's name. These attributions are undoubtedly incomplete for most individuals and will change with increased digitization of the NMNH collections.

Figure 1*A* shows the number of specimen attributions for Funk List individuals with the breakdown by Smithsonian units, and figure 1*B* shows the breakdown by department for NMNH-affiliated people. These include both specimen collections and determinations. It is not surprising that the bulk of attributions occur for NMNH scientists, as the data set is contributed to GBIF from NMNH and NMNH is the largest in terms of staff and the oldest science unit at the Smithsonian. It is also not surprising that botany and invertebrate zoology are the best-represented departments, as these departments have the most specimens digitized and historically also had significantly more women working in them than the other departments did. It is important to note that while anthropology is plotted in figure 1*B*,

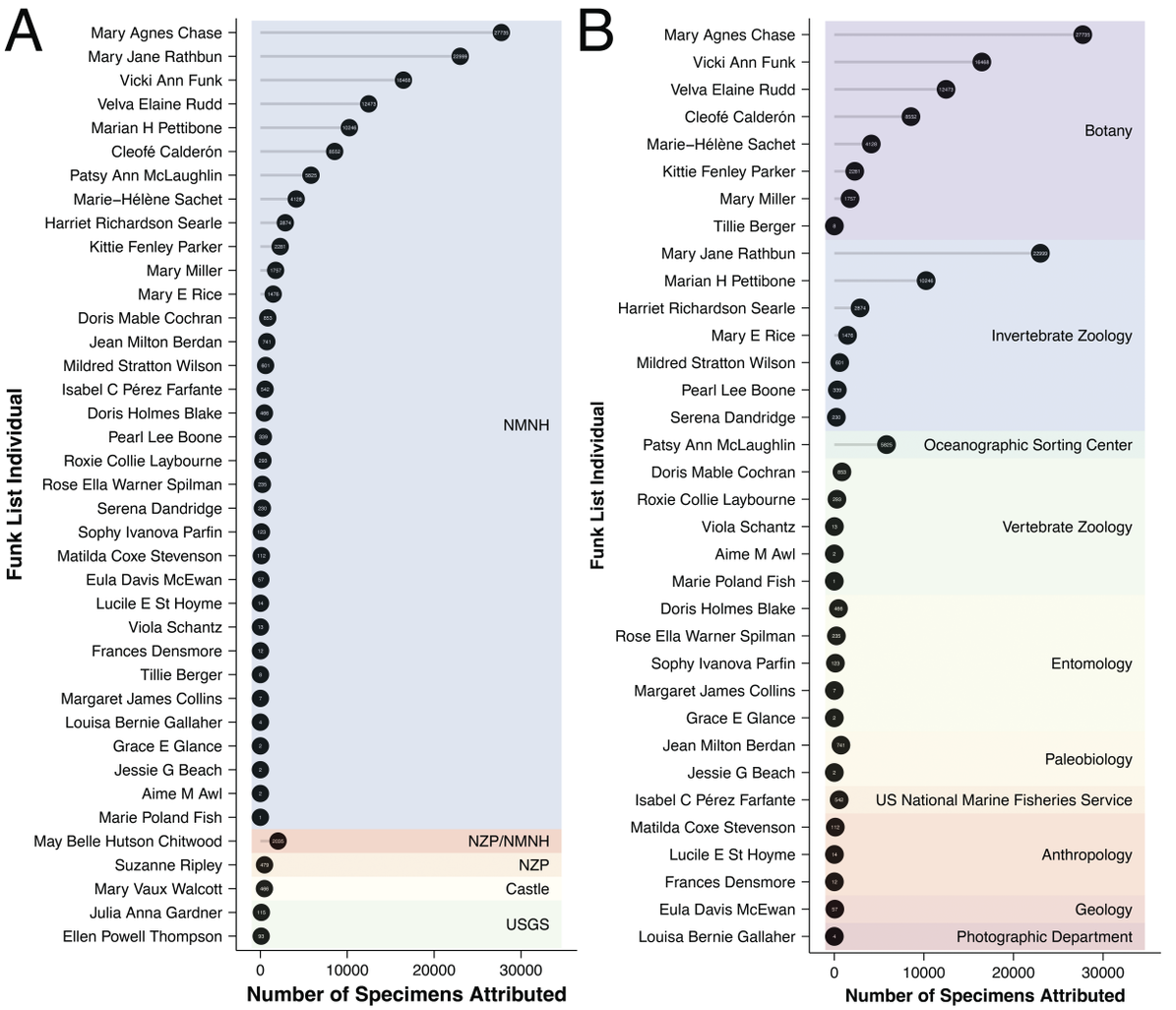

Fig. 1 Counts of Funk List specimen attributions, both collections and identifications. *A*, Members of the Funk List separated by Smithsonian unit. *B*, Those associated with National Museum of Natural History (NMNH) are separated by department. NZP = National Zoological Park; USGS = US Geological Survey.

NMNH anthropological collections data are not in GBIF. The specimens plotted here represent specimens collected by individuals affiliated with NMNH anthropology for other NMNH departments. Considerations of the anthropological collections are outside the scope of this article. The specimen attribution counts shown in figure 1 include specimens housed at NMNH and other institutions. In figure 2, we break this down by NMNH collection code and "other," which refers to other institutions. We felt it was important to not only consider the work the women on the Funk List did for NMNH collections but also consider their impact on global collections.

We were also interested in understanding the taxonomic scope of the specimens collected or identified by women on the Funk List. This is plotted on a tree of life at the family level in figure 3. Vicki Funk herself collected specimens from at least 267 families over the course of her career. Because there were three families of plants that were extreme outliers (more than 10,000 specimens each for Poaceae, Asteraceae, and Fabaceae), we used the $\log_{10}$ number of specimens per family collected or identified by deceased members of the Funk List. Taxonomic coverage by the 40 women on the Funk List to whom we were able to attribute specimens is remarkably broad, particularly when one takes into account that specimen digitization is not nearly complete. There were 1482 families present in the 40 CSV files from Bionomia, and 1132 of these are plotted on the tree. There were 350 that did not make it into the tree because of taxonomy issues (e.g., nonmonophyly or misspelled or otherwise inaccurate name) or because there was not a tree with that tip found in the Open Tree of Life database, including three Protozoa collections.

Figures 4 and 5 highlight particular women who had dramatic impacts on NMNH collections in terms of the numbers of specimens collected and identified. Figure 4 highlights four botanists: Mary Agnes Chase, Velva Rudd, Cleofé Calderón, and Vicki

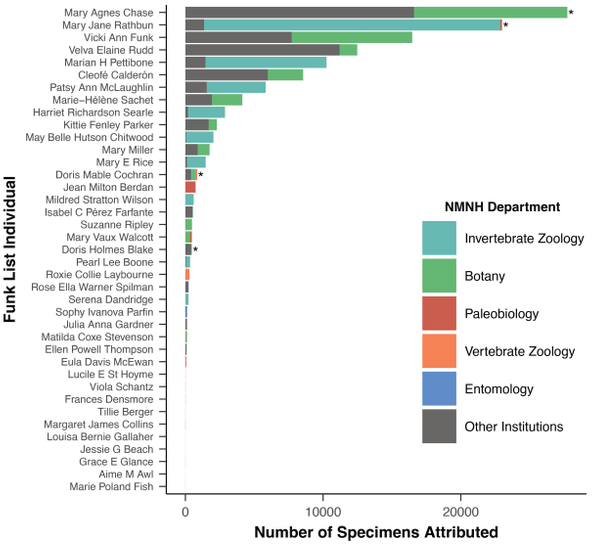

Fig. 2 Plot of specimen attributions per person, broken out by institution and National Museum of Natural History (NMNH) collection code. Asterisks indicate broad collectors that collected for at least four different NMNH departments. Those without visible bars have fewer than 15 specimens attributed to them.

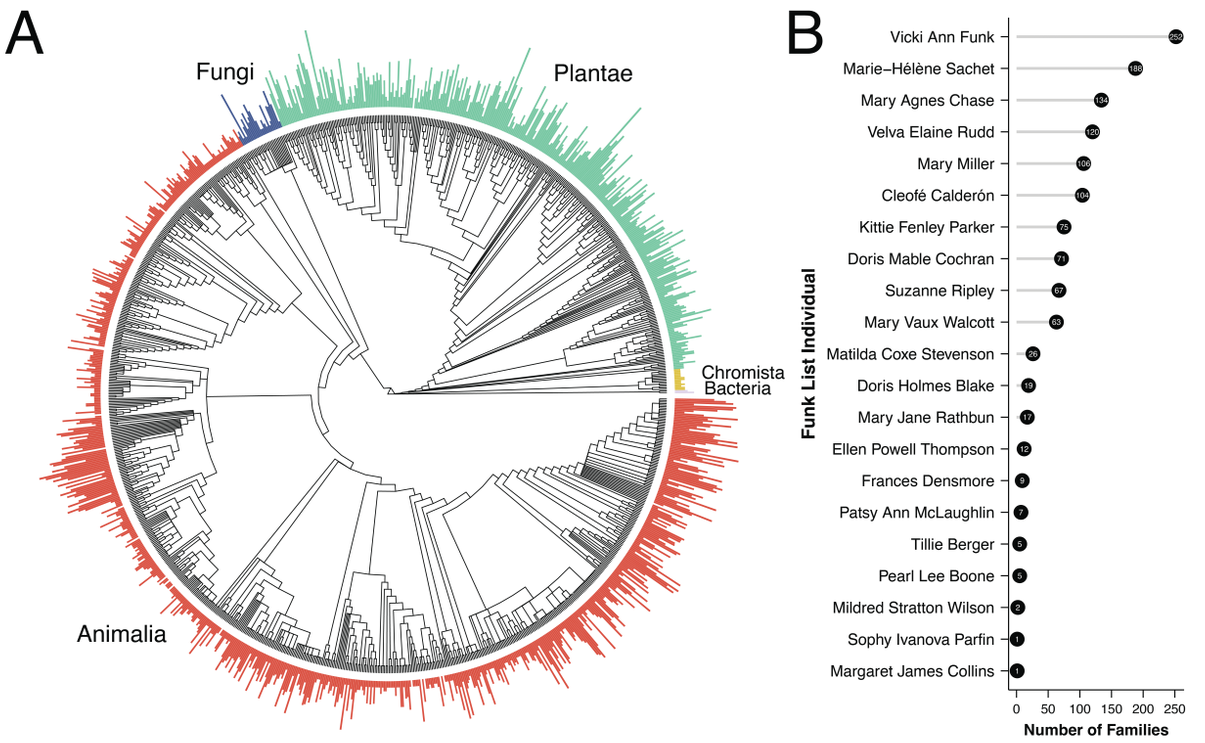

Fig. 3 *A*, Tree of life with $\log_{10}$ number of specimens per family attributed to Funk List members. *B*, Count of families with specimens either collected or identified by a member of the Funk List.

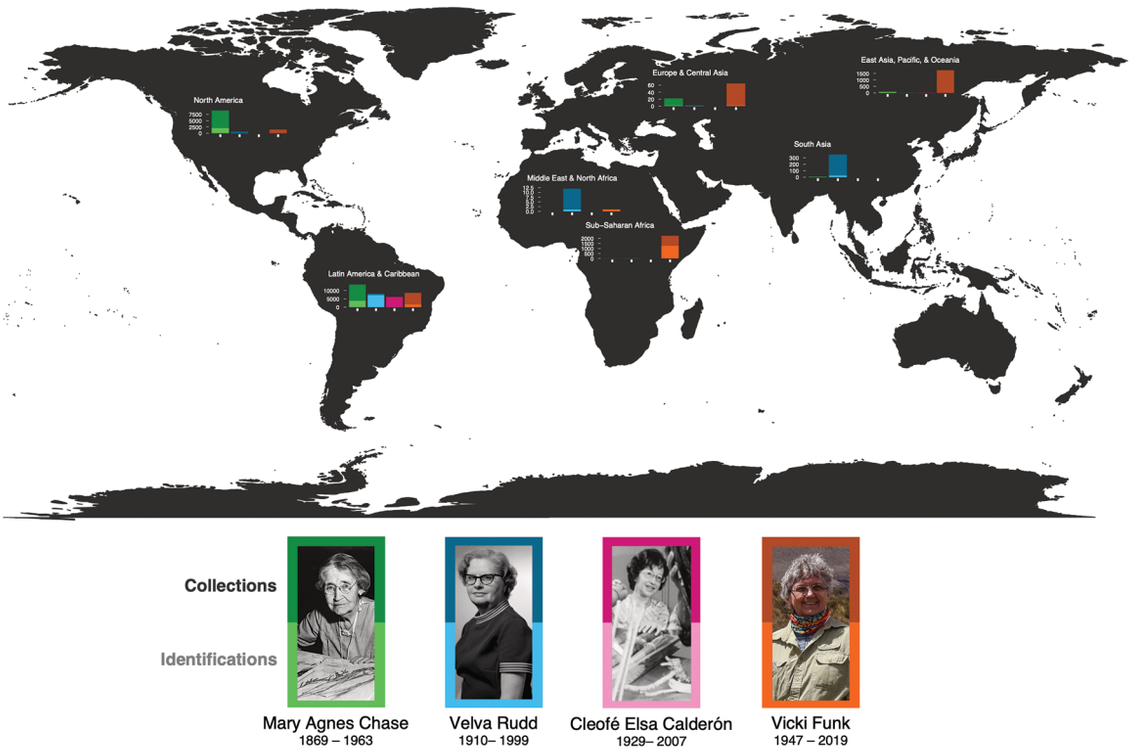

Fig. 4 World map showing geographic distributions of specimen attributions binned into world regions by botanists Mary Agnes Chase, Velva Rudd, Cleofé Calderón, and Vicki Funk. For each researcher, the darker shade indicates collections, and the lighter shade indicates identifications. Photo of Funk taken by M. Bonifacino and published in Susanna et al. (2020). Other photos are from the Smithsonian Institution.

Funk. Rudd and Funk were staff curators of botany at NMNH. Chase was a USDA employee and considered the custodian of the grasses collection, even after her retirement (Henson 2003; Smith 2018). Calderón was funded by the Smithsonian as a collaborator of Thomas Soderstrom, curator of grasses at NMNH (Clark et al. 2008). Funk, a senior scientist, was a valued mentor to many young researchers, instrumental in the development of modern phylogenetic methods, and a proponent of digitization at NMNH (Susanna et al. 2020).

Figure 5 highlights four invertebrate zoologists: Mary Jane Rathbun, Marian Pettibone, Mary Rice, and Patsy Ann McLaughlin. Rathbun had a long career at NMNH, beginning as a copyist, and was eventually promoted to assistant curator before resigning to release her funding to allow the department to hire an assistant (Schmitt 1973). Pettibone, a polychaete expert, was an NMNH curator. Rice was an NMNH research zoologist and curator and later founder and director of the Smithsonian Marine Station. McLaughlin was supervisor for invertebrates at the Smithsonian Oceanographic Sorting Center from 1965 to 1968 (Lemaitre 2012). Even though her official tenure with the Smithsonian was short, she had an outsize impact on its collections. McLaughlin was hired by Waldo Schmitt, who was Rathbun's mentee years earlier. We have also found a barnacle specimen originally identified by Rathbun, whose identification was changed by McLaughlin later (https://www.gbif.org/occurrence/1317447220). There are many similar instances, for example, a crab collected by Pearl Boone in 1914, identified by Rathbun (e.g., https://www.gbif.org/occurrence/2571473095, https://www.gbif.org/occurrence/2571448009). Additional examples include a specimen collected by Calderón and identified by Rudd (https://www.gbif.org/occurrence/2997354038), a specimen collected by Boone and identified by Pettibone (https://www.gbif.org/occurrence/1318256847), and a specimen collected by Boone and later painted by Mary Vaux Walcott (https://www.gbif.org/occurrence/1322922031).

In the NMNH entomology collection, it is common to find specimens in collections without an identification label, especially when they are part of a series of specimens determined by the same person. Only the first specimen receives an identification label, and all specimens in the same row or unit tray can be assumed to have been determined by the same person at the same time. This is shown in figure 6. In addition to the specimen attributions in Bionomia, we searched for all of the women in the Funk List in the NMNH GBIF data sets, and the code used for these searches is on GitHub. Because these searches account for multiple name permutations and spousal names, there are false positives in the results, in particular for those women with common last names (e.g., Smith). While we focus on the Bionomia results here, because we have filtered out any false positives and can easily share these with the larger community, we will use the results of searching the entire NMNH GBIF data sets to report suggested changes to NMNH data managers.

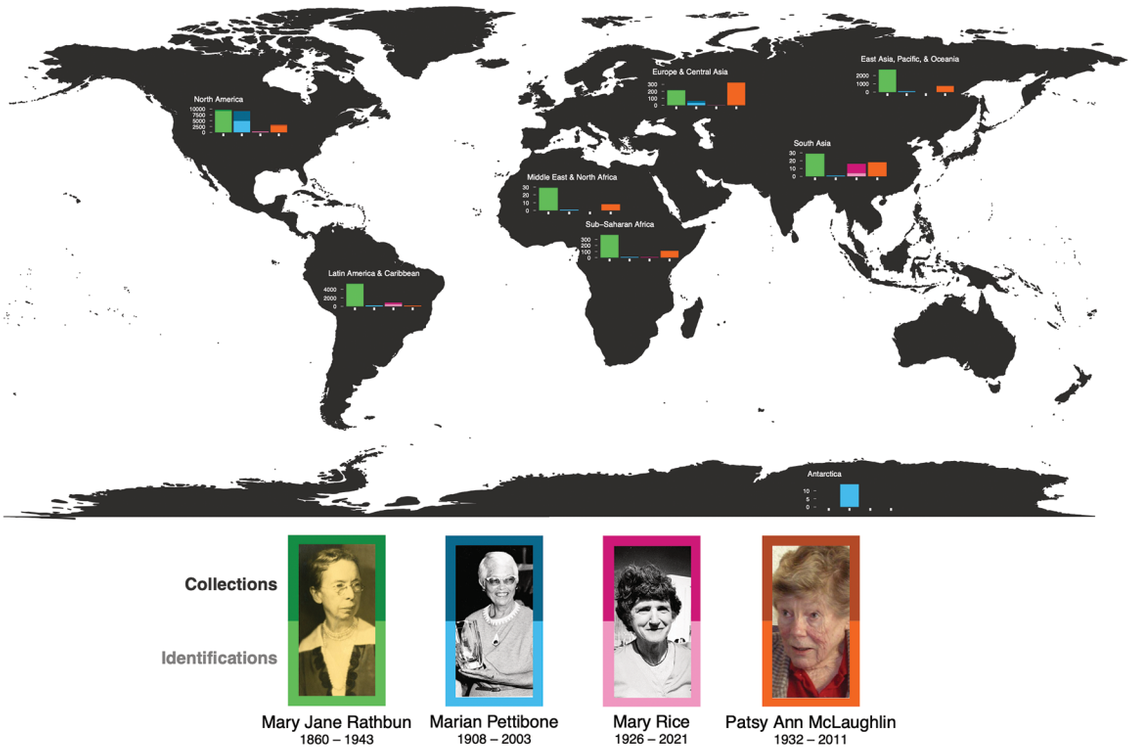

Fig. 5 World map showing geographic distributions of specimen attributions binned into world regions by invertebrate zoologists Mary Jane Rathbun, Marian Pettibone, Mary Rice, and Patsy Ann McLaughlin. For each researcher, the darker shade indicates collections, and the lighter shade indicates identifications. Photo of McLaughlin taken by Lemaitre (2012). Other photos are from the Smithsonian Institution.

Semantic Search of the Smithsonian Annual Reports

We leveraged machine learning language models, specifically transformers (which can parse noisy documents and understand the larger semantic meanings of texts), to vectorize all sentences with more than seven words in the annual reports. Using the Python library txtai, we created a searchable index of nearly 550,000 sentences from the 199 annual reports and *Explorations* documents. Through this approach, users can query the indexed sentences semantically. In other words, they can go beyond simple keyword searching and instead search for abstract ideas or concepts whose keywords may be too numerous or unknown to account for fully in a traditional keyword search. We made this index accessible for all via a beta Streamlit application available at https://sidatasciencelab.github.io/Funk_List_Collections_Records/. The application allows a user to specify a specific query, an example of a search might be "women on a research trip," which is vectorized by txtai. The user can also specify the number of results they wish to see populated. The larger the number, the longer the results will take to populate. The results are ranked based on semantic similarity to the initial query. Additionally, the results are color coordinated so that the user can see which words led to the results.

Discussion

While working with the NMNH Darwin Core data sets and performing specimen attributions in Bionomia, we came across challenges with the data. We detail these in table 3 and discuss them below.

Titles and Inconsistent Names

We were able to find 7264 records in the NMNH extant specimen data set and the paleobiology data set that contained Miss, Mrs., or Ms. in the recordedBy or identifiedBy fields. This set of records is on GitHub. Of these entries, there are a total of 140 unique women, most of whom are not on the Funk List and were potentially never associated directly with the Smithsonian. Many may be difficult to identify because they are without first names or initials. This is particularly true for women identified as "Miss," whose first names or initials are rarely included. Titles are most often trimmed out of the transcribed data. We found multiple examples of a woman's specimens being attributed to her spouse, but unfortunately, this pattern is not consistent enough to propose a single solution. In cases where specimen or label images are present, we can use these to look for titles and reattribute the specimen to the correct person, but this process is largely manual. Using regular expressions formulas to account for name permutations is helpful in catching more of these issues but likewise cannot solve every problem.

We expect names to be written inconsistently on labels, given that specimens have been collected over the course of more than a century and processed by thousands of different people. In addition to differences in the names written on labels, data management processes can cause us to miss (or almost miss) specimens

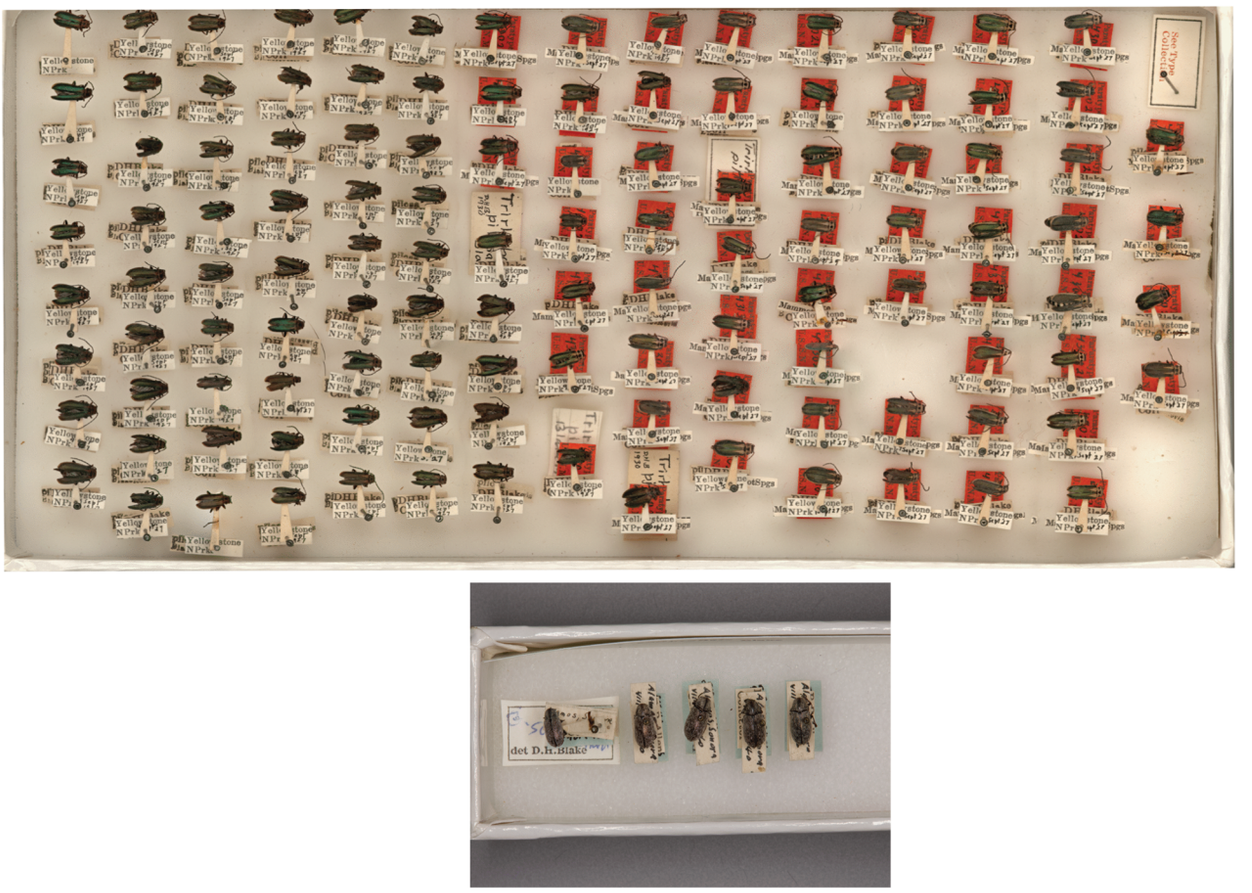

Fig. 6 Leaf beetles collected (*top*) and identified (*top* and *bottom*) by Doris Holmes Blake. Note that the set of five specimens on the bottom includes only a single identification label with Holmes's name on the first specimen (larger white label), while all five specimens are believed to have been studied when so arranged—a practice common in entomological collections.

belonging to Funk List individuals. One instance is a specimen collected by Suzanne Ripley. Suzanne was a primatologist affiliated with the National Zoo, and she was also married to F. R. Fosberg, a botanist at the US Geological Survey and NMNH (Primate Conservation 2008). Suzanne went on an expedition to Sri Lanka in 1968–1970 and while there collected more than 400 legume specimens. In the digitized records, her specimens are mostly listed as "S. Ripley." Her name is often typewritten (as "Suzanne Ripley") on these specimen labels however, which makes it easy, though time consuming, to visually confirm her attributions. During this same time period, the secretary of the Smithsonian was Sidney Dillon Ripley, an ornithologist, who remained an active collector during his tenure as secretary. Most of his specimens are in the database as "S. D. Ripley." We found a specimen attributed to S. D. Ripley, but since it was collected in Sri Lanka during the time period of the expedition, we checked this specimen visually, and indeed, the label had "Suzanne Ripley" printed on it. While in this example we are describing just one specimen, this problem is widespread. Because S. D. Ripley is a very well-known name at the Smithsonian (it is even the name of a building), the attribution of an S. Ripley collection from the late 1960s to Secretary Ripley likely seemed to make sense to the person databasing the transcriptions.

From these examples, it is clear that data are being lost by going from the paper label to the digitized occurrence. This could happen at the time of transcription, when transcriptions are entered into the collections database, or during a name-merging process. The Suzanne Ripley example helped us to reflect back to when we first noticed the Mrs. <spouse's name> problem with Mary Vaux Walcott's specimens. Whether the cause is a systematic process, where women are referred to as "Mrs. <spouse's name>" and then "Mrs." is removed, or an issue of switching S. Ripley for S. D. Ripley, the effect is the same—women's contributions to collections data are erased. The "raw" data we send to a global aggregator like GBIF should be as much an exact reflection of what is written on the paper labels as possible so that these kinds of mix-ups can be identified programmatically. In addition, as suggested by others, global unique identifiers such as Wikidata IDs or ORCIDs should be listed to allow true aggregation of specimens across departments and institutions (Shorthouse and Page 2019).

Rejected Dates

Accurate collection dates are essential for many applications of collections data. In the case of our work for this project (and notably in the Mary Vaux Walcott example), we used

Table 3

Summary of Collections Data Challenges and Proposed Solutions

Challenge	Recommendation(s)
Titles (e.g., Mrs., Miss) present on paper labels are often not included in transcribed data; Mrs. <spouse's name> becomes just <spouse's name>	Look at images of paper labels when available Use regular expressions to search for multiple name permutations
Collector and identifier names are written inconsistently (e.g., Walcott, C. Walcott, C. D. Walcott, Charles Walcott, Charles Doolittle Walcott)	Encourage use of ORCID or Wikidata numbers in recordedByID and identifiedByID fields Use regular expressions to search for multiple name permutations
When dates sent to Global Biodiversity Information Facility (GBIF) do not have a month and year specified or have 0 for either month or day, they can be rejected	Look at images of paper labels when available Look at institution CIS data directly Institutions should conduct quality checks of their data in GBIF 0 values should not be used for month or day
Specimens part of a series (particularly for entomology specimens) may not all have labels, so determiner may not be captured when digitizing	When digitizing, pay attention to order of specimens, which can provide evidence of who identified them
Type specimens do not consistently include species author under identifiedBy	Check other fields for whether specimens are holotypes and add describer to identifiedBy field
Collections not digitized	Use other documents and expedition and travel logs to gather information about a collector or researcher's work, use these data to prioritize digitization efforts

collection dates to help identify specimens that may have been collected by women on the Funk List but attributed to their spouses or others. When dates are not present, this work is much more difficult. For some specimens, we may know only the year of collection and not the month or day. In these cases, it is important to decide on a convention that is accepted by GBIF so that at least the year information is present in the record. For some NMNH specimens, 0 has been entered for month and day (indicating the absence of this information in the label), and this entire date is then rejected by GBIF. For these specimens, we can examine the collecting year/date either on the NMNH collections portal or in some instances by visually inspecting the specimen image, but this is not efficient.

Type Specimens and Specimen Series in Entomology

Primary type specimens (holotypes) may not consistently include the species author under identifiedBy in databases even though these specimens have been studied and determined by the species author (fig. 6). The absence of a determination label on a particular specimen poses a substantial challenge to attributing identifications to persons in Bionomia or GBIF in general. During mass digitization workflows, the specimen series might be broken up so that the reference to the "1st specimen in a row or unit tray" might be lost.

Doris Holmes Blake was one of the most productive beetle taxonomists at the NMNH. She described 25 genera and 818 species of beetles in her career for which 441 holotypes are deposited at the NMNH alone. She studied leaf beetle species (Chrysomelidae) in numerous NHCs, and some 538 species that she had described are housed in the NMNH collection today. While she worked for the USDA based at the NMNH in the 1920s and was temporarily employed by the Smithsonian for a beetle curation project in the 1950s (Froeschner et al. 1981), Doris Holmes Blake was never a Smithsonian staff member. However, in terms of taxonomic output of species described and revisions published, she outshines most if not all of the other staff scientists in entomology at that time. Early revisions were published with the institutional affiliation of assistant entomologist (USDA; e.g., Blake 1931), while she used honorary research associate (Smithsonian NMNH) later (e.g., Blake 1967, 1974). The holotypes of species at the NMNH that Doris Holmes Blake described have been digitized, and the data and images are accessible at GBIF. However, the many specimens studied and identified as well as specimens she collected in the general Chrysomelidae collection have not been digitized. Blake's broader impact of shaping the collection through her work (identifiedBy and recordedBy) cannot be fully appreciated at this time, but we have some more information about the number of specimens she collected from the annual reports. The 1963 report states that "1,454 miscellaneous insects from South America [were] collected by Mrs. Doris M. Blake, 1962–63." In 1964, "as a result of field work conducted by members of the Smithsonian staff the following were acquired: 1,100 miscellaneous South American beetles from Mrs. Doris H. Blake and Dr. Doris M. Cochran." The 1981 report states, "Doris H. Blake (Deceased): 12 stuffed and mounted turtles."

Importance of Digitization

We are missing information about many women because collections are not digitized. Margaret S. Collins is one notable case. She studied termites and described a single species from Florida (Nickle and Collins 1989). The NMNH termite collection has not been digitized, so the impact of her sorting and identification work as well as collecting efforts cannot be documented at this time. Our hope is that digitization can be prioritized for specimens collected and identified by women on the Funk List as well as other collectors and identifiers from marginalized identities. This work is not straightforward; it requires studying a variety of documents, including publications, Smithsonian newsletters, field books, and personal papers, and talking to collections staff to find the specimens of interest. This is one reason why this work focuses mostly on Smithsonian-affiliated researchers. We were able to build on the work being done as part of the AWHI. This foundational work to document the

women working and volunteering at our institutions, including their backgrounds, affiliations, spouses, and networks, will also need to be done at other institutions before the biodiversity community as a whole can begin taking a comprehensive look at how women have shaped collections at institutions across the world.

Of course, Smithsonian researchers engage in many different kinds of research, only some of which is specimen based. Here, we focus on the collections angle, but we have tremendous appreciation for the other kinds of research in which the women on the Funk List were involved. Continued focus on digitization and transcription of archival material will help identify both tangible and intangible contributions they made throughout their careers. We encourage anyone with an interest in this topic to dive into Bionomia to attribute specimens (for an excellent guide to getting started, see Leachman 2020) and to work with the Smithsonian Transcription Center to transcribe historical documents—it is a rewarding activity that helps us better document the history of the Smithsonian and other natural history and cultural heritage institutions. We hope that the annual reports search application provides a glimpse into what might be possible as historical documents continue to be digitized. We look forward to seeing what we can learn about the networks of Funk List women and how we can grow the list by looking at as many resources as are available. This application is not limited to natural history endeavors, and we encourage users to contact us if their queries lead to new potential research areas.

Acknowledgments

We thank Jessica Bird, Dave Furth, Erin Kolski, Sasha Konstantinov, Sue Lutz, Eric Schuettpelz, Sylvia Orli, Tammy Peters, Mariah Wahl, and Kelly Doyle at SI and everyone who has created Wikidata and Wikipedia content and Bionomia attributions for women on the Funk List. We also thank two anonymous reviewers for their constructive comments. This work was funded in part by a Smithsonian Women's Committee Grant to R. B. Dikow and an American Women's History Initiative Grant to R. B. Dikow and M. G. Trizna. Portions of the computations performed for this study were conducted on the Smithsonian Institution High-Performance Cluster (https://doi.org/10.25572/SIHPC). While this work was in review, we lost Effie Kapsalis, champion of open-access data at the Smithsonian. This work is in many ways inspired by Effie's commitment to improving gender representation in Smithsonian data. This article is dedicated to Vicki Funk for her commitment to open science, mentorship, and inclusive collaborations. R. B. Dikow asks, "What would Vicki do?" at least weekly.

Literature Cited

Arel-Bundock V, N Enevoldsen, CJ Yetman 2018 countrycode: an R package to convert country names and country codes. J Open Source Softw 3:848. https://doi.org/10.21105/joss.00848.

AWHI (American Women's History Initiative) 2019 Because of her story: the Funk List. Smithsonian American Women's History Museum. https://womenshistory.si.edu/stories/2019/11/because-her-story-funk-list

Becker AR, AR Wilks, R Brownrigg, TP Minka, A Deckmyn 2021 maps: draw geographical maps. R package version 3.4.0. https://CRAN.R-project.org/package=maps.

Blake DH 1931 Revision of the species of beetles of the genus *Trirhabda* north of Mexico. Proc US Natl Mus 79:1–36. https://doi.org/10.5479/si.00963801.79-2868.1.

——— 1967 Revision of the beetles of genus *Glyptoscelis* (Coleoptera: Chrysomelidae). Proc US Natl Mus 123:1–53. https://doi.org/10.5479/si.00963801.123-3604.1.

——— 1974 The costate species of *Colaspis* in the United States (Coleoptera: Chrysomelidae). Smithson Contrib Zool 181:1–24. https://doi.org/10.5479/si.00810282.181.

Clark LG, EJ Judziewicz, X Londono, TS Filgueiras 2008 Cleofé E. Calderón (1929–2007). Bamboo Sci Cult 21:1–8.

Daru BH, DS Park, RB Primack, CG Willis, DS Barrington, TJS Whitfield, TG Seidler, et al 2018 Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytol 217:939–955.

Dikow RB, M Glenn 2020 What's in a name? OCIO Data Science Lab. https://datascience.si.edu/news/whatsinaname.

Edmondson M 2020 googleLanguageR: call Google's "Natural Language" API, "Cloud Translation" API, "Cloud Speech" API and "Cloud Text-to-Speech" API. R package version 0.3.0. https://CRAN.R-project.org/package=googleLanguageR.

Farabaugh F 2015 The inspiring Mary Jane Rathbun—Women's History Month highlight. Department of Invertebrate Zoology News: No Bones. https://nmnh.typepad.com/no_bones/2015/03/womens-history-month-highlight-mary-jane-rathbun.html.

Folk RA, HR Kates, R LaFrance, DE Soltis, PS Soltis, RP Guralnick 2021 High-throughput methods for efficiently building massive phylogenies from natural history collections. Appl Plant Sci 9:e11410

Froeschner RC, EML Froeschner, OL Cartwright 1981 Doris Holmes Blake. Proc Entomol Soc Wash 83:544–564. https://www.biodiversitylibrary.org/page/16364889.

Funk VA 2004 100 uses for an herbarium: well at least 72. Am Soc Plant Taxon Newsl 17:17–19. https://repository.si.edu/handle/10088/11385.

GBIF (Global Biodiversity Information Facility) 2022*a* NMNH extant specimen records (USNM, US). GBIF occurrence download (23 August 2022). https://doi.org/10.15468/dl.92brp2.

——— 2022*b* NMNH paleobiology specimen records (USNM). GBIF occurrence download (23 August 2022). https://doi.org/10.15468/dl.a48h5p.

Groom Q, A Güntsch, P Huybrechts, N Kearney, S Leachman, N Nicolson, RDM Page, DP Shorthouse, AE Thessen, E Haston 2020 People are essential to linking biodiversity data. Database 2020: baaa072. https://doi.org/10.1093/database/baaa072.

Harmon E 2021 Doris Holmes Blake and the fight for women's right to paid employment. Smithsonian Institution Archives. https://siarchives.si.edu/blog/doris-holmes-blake-and-fight-women's-right-paid-employment.

Heard SG 2020 Charles Darwin's barnacle and David Bowie's spider. Yale University Press, New Haven, CT.

Henson PM 2003 "What holds the Earth together": Agnes Chase and American agrostology. J Hist Biol 36:437–460. https://doi.org/10.1023/B:HIST.0000004568.11609.2d.

——— 2014 Diminutive but determined: Mary Jane Rathbun. Smithsonian Institution Archives. https://siarchives.si.edu/blog/diminutive-determined-mary-jane-rathbun

Hughes AC, MC Orr, K Ma, MJ Costello, J Waller, P Provoost, Q Yang, C Zhu, H Qiao 2021 Sampling biases shape our view of the natural world. Ecography 44:1259–1269.

Leachman S 2020 Auckland Museum volunteer instructions for Bionomia. Version 1. Zenodo, https://doi.org/10.5281/zenodo.3908727.

Lemaitre R 2012 Patsy Ann McLaughlin, May 27, 1932–April 4, 2011. J Crustac Biol 32:991–1002. https://doi.org/10.1163/1937240X-00002095.

Lewis VR 2016 Child prodigy, pioneer scientist, and women and civil rights advocate: Dr. Margaret James Strickland Collins (1922–1996). Fla Entomol 99:334–336. https://doi.org/10.1653/024.099.0235.

McCain L 1943. Mary Jane Rathbun. Science 97:2524. https://doi.org/10.1126/science.97.2524.435.

Meineke EK, TJ Davies, BH Daru, CC Davis 2019 Biological collections for understanding biodiversity in the Anthropocene. Philos Trans R Soc B 374:20170386.

Michonneau F, JW Brown, DJ Winter 2016. rotl: an R package to interact with the Open Tree of Life data. Methods Ecol Evol 7:1476–1481. https://doi.org/10.1111/2041-210X.12593.

Nickle DA, MC Collins 1989 Key to the Kalotermitidae of eastern United States with a new Neotermes from Florida (Isoptera). Proc Entomol Soc Wash 91:269–285. https://www.biodiversitylibrary.org/page/16135011.

OpenTreeOfLife, B Redelings, LL Sanchez Reyes, KA Cranston, J Allman, MT Holder, EJ McTavish 2019 Open Tree of Life Synthetic Tree. Version 12.3. Zenodo, https://doi.org/10.5281/zenodo.3937742.

Orrell T, Informatics Office 2022*a* NMNH extant specimen records (USNM, US). Version 1.58. National Museum of Natural History, Smithsonian Institution. Occurrence data set, https://doi.org/10.15468/hnhrg3.

——— 2022*b* NMNH paleobiology specimen records (USNM). Version 1.59. National Museum of Natural History, Smithsonian Institution. Occurrence data set, https://doi.org/10.15468/7m0fvd.

Page RDM 2019 Wikidata and the biodiversity knowledge graph. Biodivers Inf Sci Stand 3:e34742. https://doi.org/10.3897/biss.3.34742.

Park DS, X Feng, S Akiyama, M Ardiyani, N Avendaño, Z Barina, B Bärtschi, et al 2021 The colonial legacy of herbaria. bioRxiv, https://doi.org/10.1101/2021.10.27.466174.

Primate Conservation 2008 Suzanne Ripley. Primate Conserv 23:146–147.

Raja NB, EM Dunne, A Matiwane, TM Khan, PS Nätscher, AM Ghilardi, D Chattopadhyay 2021 Colonial history and global economics distort our understanding of deep-time biodiversity. Nat Ecol Evol 6:145–154. https://doi.org/10.1038/s41559-021-01608-8.

R Core Team 2021 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Sayah O 2016 "It won't be easy to leave after 40 years": Sophie Lutterlough's career at the National Museum of Natural History. Smithsonian Issues Archives. https://siarchives.si.edu/blog/sophie-lutterlough.

Schmitt WL 1973 Mary J. Rathbun 1860–1943. Crustaceana 24:283–297.

Short AEZ, T Dikow, CS Moreau 2018 Entomological collections in the age of big data. Annu Rev Entomol 65:513–530. https://doi.org/10.1146/annurev-ento-031616-035536.

Shorthouse DP, RDM Page 2019 Quantifying institutional reach through the human network in natural history collections. Biodivers Inf Sci Stand 3:e35243. https://doi.org/10.3897/biss.3.35243.

Smith JP Jr 2018 Mary Agnes Chase. Bot Stud 81:1–4. https://digitalcommons.humboldt.edu/botany_jps/81.

Soltis PS, G Nelson, A Zare, EK Meineke 2020 Plants meet machines: prospects in machine learning for plant biology. Appl Plant Sci 8: e11371. https://doi.org/10.1002/aps3.11371.

Speer KA, MTR Hawkins, MFC Flores, MR McGowen, RC Fleischer 2022 A comparative study of RNA yields from museum specimens, including an optimized protocol for extracting RNA from formalin-fixed specimens. Front Ecol Evol 10:953131. https://doi.org/10.3389/fevo.2022.953131.

Susanna A, BG Baldwin, RJ Bayer, JM Bonifacino, N Garcia-Jacas, SC Keeley, JR Mandel, S Ortiz, H Robinson, TF Stuessy 2020 The classification of the Compositae: a tribute to Vicki Ann Funk (1947–2019). Taxon 69:807–814. https://doi.org/10.1002/tax.12235.

Vogel G 2019 Natural history museums face their own past. Science 363:1371–1372.

Warnes GR, B Bolker, G Gorjanc, G Grothendieck, A Korosec, T Lumley, D MacQueen, et al 2022 gdata: various R programming tools for data manipulation. R package version 2.18.0.1. https://CRAN.R-project.org/package=gdata.

Wickham H 2016 ggplot2: elegant graphics for data analysis. Springer, New York.

Wickham H, R François, L Henry, K Müller 2022 dplyr: a grammar of data manipulation. R package version 1.0.10. https://CRAN.R-project.org/package=dplyr.

Wieczorek J, D Bloom, R Guralnick, S Blum, M Döring, R Giovanni, T Robertson, D Vieglais 2012 Darwin Core: an evolving community-developed biodiversity data standard. PLoS ONE 7:e29715. https://doi.org/10.1371/journal.pone.0029715.

Wintle C 2016 Decolonizing the Smithsonian: museums as microcosms of political encounter. Am Hist Rev 121:1492–1520.