# AvianLexiconAtlas: A database of descriptive categories of English-language bird names around the world

Erin S. Morrison[1¶*], Guinevere P. Pandolfi[1], Stepfanie M. Aguillon[2], Jarome R. Ali[3], Olivia Archard[1], Daniel T. Baldassarre[4], Illeana Baquero[1], Kevin F.P. Bennett[5], Kevin M. Bonney[1], Riley Bryant[1], Rosanne M. Catanach[6], Therese A. Catanach[7], Ida Chavoshan[1], Sarah N. Davis[8], Brooke D. Goodman[4], Eric R. Gulson-Castillo[9], Matthew Hack[9], Jocelyn Hudon[10], Gavin M. Leighton[11], Kira M. Long[12], Ziqi Ma[1], Dakota E. McCoy[13], J.F. McLaughlin[14], Gaia Rueda Moreno[1], Talia M. Mota[1], Lara Noguchi[1], Ugo Nwigwe[1], Teresa Pegan[9], Kaiya L. Provost[15], Shauna A. Rasband[5], Jessie Frances Salter[16], Lauren C. Silvernail[17], Jared A. Simard[1], Heather R. Skeen[18], Juliana Soto-Patiño[19], Young Ha Suh[20], Qingyue Wang[1], Matthew E. Warshauer[21], Sissy Yan[1], Betsy Zalinski[1], Ziqi Zhao[1], and Allison J. Shultz[20¶]

[1]Liberal Studies, New York University, New York, NY, USA

[2]University of California, Los Angeles, Los Angeles, CA, USA

[3]Princeton University, Princeton, NJ, USA

[4]SUNY Oswego, Oswego, NY, USA

[5]University of Maryland, College Park, College Park, MD, USA

[6]Trinity, AL, USA

[7]Academy of Natural Sciences of Drexel University, Philadelphia, PA, USA

[8]Carnegie Museum of Natural History, Pittsburgh, PA, USA

[9]University of Michigan, Ann Arbor, MI, USA

[10]Royal Alberta Museum, Edmonton, Alberta, Canada

[11]SUNY Buffalo State University, Buffalo, NY, USA

[12]University of Idaho, Moscow, ID, USA

[13]University of Chicago, Chicago, IL, USA

[14]University of Alaska Anchorage, Anchorage, AK, USA

[15]Adelphi University, Garden City, NY, USA

[16]Santa Monica College, Santa Monica, CA, USA

[17]Oswego, NY, USA

[18]Negaunee Integrative Research Center, Field Museum of Natural History, Chicago, IL, USA

[19]University of Illinois at Urbana-Champaign, Urbana, IL, USA

[20]Ornithology Department, Natural History Museum of Los Angeles County, Los Angeles, CA, USA

[21]Brooklyn, New York, USA

*Corresponding author

Email: erin.morrison@nyu.edu (ESM)

¶These authors contributed equally to this work.

# Abstract

Common names of species are important for communicating with the general public. In principle, these names should provide an accessible way to engage with and identify species. The common names of species have historically been labile without standard guidelines, even within a language. Currently, there is no systematic assessment of how often common names communicate identifiable and biologically relevant characteristics about species. This is a particular issue in ornithology, where common names are used more often than scientific names for species of birds in written and spoken English, even by professional researchers. To gain a better understanding of the types of terminology used in the common names of bird species and their frequency of use, we used a crowdsourcing approach and recruited 85 professional ornithologists and non-professional participants to classify unique descriptors in the English-language common names of all recognized bird species from around the world. Each species' common name was assigned to one of ten categories associated with aspects of avian biology, ecology, or human culture. Across 10,906 species of birds, 89% have names describing the biology of the species, while the remaining 11% of species have names derived from human cultural references or local non-English languages. Species with common names based on features of avian biology are more likely to be related to each other or be from the same geographic region. The crowdsourced data collection also revealed that many common names contained specialized or historic terminology unknown to many of the data collectors, and this was documented in a glossary and gazetteer alongside the dataset. As the first comprehensive assessment of the state of terminology in English-language common names of birds, the AvianLexiconAtlas database sheds light on historical approaches to nomenclature and provides insight into how the general public currently engages with species through their names.

# Introduction

Humans have been observing and classifying living organisms for thousands of years across cultures and languages [1]. The names of species can have consequences for how they are perceived by people, and this can, in turn, affect how people engage with species as part of education, research, and conservation efforts [2–6]. Constructing classification systems that delineate and name organisms can be subjective [5,7,8]. To standardize scientific animal names, the International Commission on Zoological Nomenclature (ICZN) has overseen the assignment of a unique universal scientific name for each recognized animal species since 1895. This scientific name is written as a Latin binomial and includes the genus and species names in the Linnaean taxonomy system [9]. The vernacular names of species, however, are not standardized across languages and have been labile across history, even within a single language or region [10–12]. Known as 'common' or, as argued by [13], 'standard' names, the use of these vernacular names can communicate about species in a way that is more accessible to a wider audience outside of the scientific community [3,14,15]. However, even these common names may be unique to a single region, species that span cultures may take on multiple common names in the same language, or the same common name may be unofficially used for multiple species.

In ornithology, the common names of birds tend to be used more frequently in spoken and written English than their scientific binomial names, even among professionals [16]. Despite this, the English-language common names of birds are only standardized regionally by professional ornithology societies [17], and there is not a universal set of rules for English-language common names [3,18–21]. Within the ornithological community, there is much reflection about what the terminology in standardized English-language common names should communicate about species [5,16,22–24]. Currently, some names directly describe a species

(e.g., Yellow-rumped Warbler, *Setophaga coronata*), while other names are ambiguous (e.g., Barnacle Goose, *Branta leucopsis*) or unrelated to the species' biology (e.g., Wilson's Warbler, *Cardellina pusilla*). As part of these deliberations, it is crucial to examine and analyze what the terminology used in English-language common names currently communicates about species.

While the history of English-language bird names has been extensively documented [25–28], there is currently no systematic analysis of the types of terminology used in the English-language common names of birds. We therefore endeavored to inventory and classify the terms currently used as the specific descriptor in English-language common names of all avian species. We assigned each descriptor into distinct categories associated with avian physical traits, avian natural history, or human culture. Using this dataset, we investigated frequency differences among different categories across species, and also examined trends in the terminology associated with species' close taxonomic relationships or with shared biogeography. We predicted that similarities among species due to their biology, life history, or the history of taxonomy in different parts of the world might be reflected in different naming trends across taxonomic groups or geographic regions (for example, species in the gulls family of birds or species from Africa would share common types of terminologies). The results of this work provide insight into the extent to which English-language common names of birds are associated with aspects of avian biology and human culture. This dataset could impact future discussions around changing common bird names, or be a resource for both professional ornithologists, linguists, or the general public to learn more about how humans identify species.

Additionally, since one goal of common names is to be useful for the general public, data collection for this project presented a unique opportunity to involve participants from both within and beyond the academic ornithology community. The inclusion of undergraduate students and

amateurs as data collectors was a chance to assess familiarity with the terminology used in English-language bird names. During data collection, participants were also asked to contribute to a glossary and a gazetteer when they encountered unfamiliar terminology. We make these resources and the dataset produced from their work available for future analyses on the state of avian nomenclature as the AvianLexiconAtlas.

# Materials and methods

## Selection of the eBird/Clements Checklist for species names

The English-language common names used for the study are based on the taxonomy of 10,906 species in the 2022 eBird/Clements Checklist [29]. At the time of the data collection, there were two additional annually curated global checklists with similar, but not identical, taxonomy to the eBird/Clements Checklist. The International Ornithological Congress (IOC) Checklist 12.2 contained 11,140 species, and 86.9% of these species had the same scientific and common names in the eBird/Clements 2022 Checklist [20]. The Handbook of Birds of the World (HBW) and BirdLife International Checklist 7.0 contained 11,170 species, and 79.7% of these species had the same scientific and common names in the eBird/Clements 2022 Checklist [30]. We made the decision to use the 2022 eBird/Clements checklist based on its alignment with the comprehensive *Birds of the World* online database [31], the main reference source for the data collection. The eBird/Clements taxonomy used in *Birds of the World* incorporates the HBW and BirdLife International taxonomy but makes more conservative distinctions between species and subspecies (refer to [32] and [33] for more details).

## Establishment of categories for species' English-language common names

A majority of bird species' English-language common names contain a unique descriptor followed by a group name it shares with closely related species. For example, the unique descriptors in Common Ostrich (*Struthio camelus*) and Somali Ostrich (*Struthio molybdophanes*) are 'Common' and 'Somali', respectively, while the shared group name is 'Ostrich'. This study is only focused on the unique descriptor in a species' English-language common name, not the shared group name.

Prior to the start of data collection, we ran a series of trial categorizations with a random sample of 30 species drawn from the eBird/Clements 2021 Checklist [34]. We started with a list of potential categories and three people independently attempted to categorize each of the species names without communicating with each other. Importantly, this group included two professional ornithologists and one undergraduate student without any formal training in ornithology. As a result of this trial, we established via consensus 11 categories for the unique descriptor to encompass aspects of avian physical traits (physical trait in both sexes, male-only physical trait, female-only physical trait, size), avian natural history (geographic location, life history, behavior), and human-constructed terminology independent of the biology of the species in the English language (local language, eponyms (named after a particular person), general descriptor, random) (see S1 Table for detailed descriptions of the categories). After completing the initial assignments for all 10,906 species into these 11 categories, eight professional ornithologists met and refined the names and descriptions of the categories into a final dataset consisting of 10 categories that align with independent categorizations for both common and scientific bird names [26,27] (Table 1, S1 Table).

**Table 1. Categories used to group unique descriptors in English common names of bird species.**

| Category | Description |
|---|---|
| Avian physical traits | |
| Both sexes physical trait | Physical characteristics that are observable in both males and females of a species |
| Male physical trait | Physical characteristics that are only observable and clearly identifiable in males of a species |
| Female physical trait | Physical characteristics that are only observable and clearly identifiable in females of a species |
| Size | Relative physical size of the species |
| Avian natural history | |
| Behavior | Specific behavior associated with the species |
| Geographic location | Specific geographic location |
| Natural history | Relates to habitat, nest type, or diet of the species |
| Human-constructed terminology | |
| Local language | Phonetic English pronunciation of species name in a non-English language |
| Miscellaneous | Descriptor that does not clearly fit into any of the other categories |
| Person | Name of a person |

See S1 Table for more detailed descriptions of each category.

# Categorization of species common names

## Participants

A total of 85 people participated in the data collection. Of these participants, 56 were pursuing undergraduate degrees (students), 7 had at least an undergraduate degree but no formal training in ornithology (amateurs), and 22 were pursuing graduate degrees or worked in

post-graduate careers related to ornithology (professionals). Of the 56 undergraduate students, 50 were enrolled in an introductory biology seminar course for non-science majors at New York University during the Fall 2022 semester. All participants who categorized at least 100 species as part of the data collection were eligible for authorship credit provided that they agreed to review the manuscript prior to submission for publication. Students in the course were given in-class participation credit for engaging in the data collection during two 75-minute class sessions. This participation grade counted towards 1.6% of their final grade. There was not enough time in class to complete all 100 species for authorship eligibility. Students pursuing authorship eligibility continued to engage in the data collection outside of class and did not receive any additional course credit for this extra work.

## Data collection

Participants were split into two groups (A and B; S1 Fig) and each group evaluated identical lists of 10,906 species. Each species name was therefore independently categorized by two different people. Each participant only worked on one of the two datasets and was not given access to the other dataset. Categorizing the names of species in duplicate provided an opportunity to identify descriptors that were difficult to classify and to identify potential mistakes by participants. Professionals, amateurs, and the six students not enrolled in the introductory biology course were randomly assigned to Dataset A or B based on the order in which they signed up for the project. The introductory biology course consisted of two sections of 25 students each, and each of these sections was assigned to a different dataset. A total of 39 participants worked on Dataset A (28 students, 2 amateurs, and 9 professionals) and a total of 47 participants worked on Dataset B (29 students, 5 amateurs, and 13 professionals). Participants

signed up for 10 species at a time, partitioned by the order they were listed in the 2022 eBird/Clements Checklist. Groups of 10 species had to be categorized in sequential order and participants were not allowed to selectively choose groups to categorize.

Categorization of the unique descriptor in a species' name first started by finding its account in the *Birds of the World* database and then the category assignment associated with the unique descriptor was determined using the provided text, images, and media in the species account. If the origin of the unique descriptor could not be determined from *Birds of the World*, participants would then conduct a broader internet search to find other potential sources. Any sources used outside of *Birds of the World* were recorded in the complete dataset. If the origin of the unique descriptor could not be determined based on information in *Birds of the World* or a subsequent internet search within 10 minutes, participants were asked to note that they could not identify the name within the given time limit and move on to the next species in their list.

To facilitate categorizations for the participants who were not familiar with common terms in ornithology, a glossary was developed for the project that included diagrams of common field markers and definitions of common ornithology descriptors of traits and colors. Participants were granted editing access to the glossary (S1 Appendix), and were asked to add additional terms and definitions they had to look up during the data collection period. A gazetteer (S1 Appendix) that included descriptions of all geographic locations encountered in species' names was also added to the glossary to help participants, particularly with outdated or obscure terminology. In addition to these resources, participants were provided with two lists of eponyms documented in species' English-language names compiled by Beolens and Watkins [25] and Bird Names For Birds [24].

## Dataset reconciliation

When the two duplicate datasets were compared, 906 out of the 10,906 species (8.4%) were assigned to different categories in the two datasets (mismatches). There were also 9 additional species that were designated as time limited by data collectors in both datasets. These 915 species were subsequently re-categorized independently in three identical datasets (C, D, E) by 10 professional ornithologists that participated in the first round, each of whom was assigned to one of the three datasets. When the 915 species' re-categorizations were compared across the three datasets, categories for 554 species (60.5%) matched across all 3 datasets and were retained for the final dataset. Categories for 324 (35.4%) of the re-categorized species matched in 2 out of 3 datasets, and the majority category was retained for these species in the final dataset. There was no category agreement across the three datasets for the remaining 37 (4.0%) re-categorized species. The final category for each of these species was jointly assigned by the lead authors of the study (ESM and AJS), together, based on the references provided by all those who had categorized the species in the first and second round.

# Phylogenetic and geographic trends

We predicted that similarities in categorization trends exist among species across taxonomic groups or geographic regions. To investigate this prediction, we mapped the general categories assigned to each species' name (i.e., avian physical traits, avian natural history traits, or human-constructed terminology; Table 1) to a time-calibrated maximum clade credibility global avian phylogeny of 10,824 species [35]. The taxonomy in this phylogeny is based on the eBird/Clements 2021 checklist. The package *clootl* 0.0.0.900 [36] in R 4.3 [37] was used to extract the phylogeny, and the final tree was pruned to only include the 10,775 species with

common names that occur in both the 2021 and 2022 checklists (S4 Fig.). We measured phylogenetic signals for each of these categories using Fritz & Pervis' *D* for binary traits (presence/absence) [38], using the package *caper* 1.0.3 [39] in R. *D* calculates the number of sister-clade differences in a binary trait for a given phylogeny [38]. An estimated *D* close to 1 represents a random distribution of a binary trait among related species on the phylogeny, while an estimated *D* close to 0 represents a clumped distribution of a binary trait among related species that would be expected under the Brownian motion model of evolution [38]. To test for significance, *D* was estimated for two different trait distributions for each trait that were simulated on the tips of the same phylogeny based on (1) randomly reshuffling the trait values and (2) trait evolution under Brownian motion. Each simulation was repeated 1,000 times.

To identify any geographic trends, we mapped all species names to the IOC World Bird List 13.2 [40]. We extracted the "general region" for the breeding range of each species, which is generally, but not always, at the continent level. Species with a breeding range in more than one general region were assigned to each region. We acknowledge the importance of the non-breeding range, but the lack of standardization of non-breeding range information across species precluded our ability to assign species to a general region for non-breeding range. We grouped all oceans together into a single category to increase interpretability, dropped species with a Worldwide distribution (n=10), an Antarctic distribution (n=12; sample size too small for analyses), and 110 species that could not be mapped due to taxonomic differences, for a total of 12,058 species-general region assignments in this dataset. To test whether general regions varied in the proportions of general descriptor categories (avian physical traits, avian natural history traits, or human-constructed terminology), we conducted a Chi-square test using the chisq.test function in R.

# Results

## Participation in data collection

In Datasets A and B, the 56 students categorized a total of 3,099 species, the 7 amateurs categorized a total of 1,444 species, and the 22 professional ornithologists categorized 17,269 species (S1 Fig). Students, amateurs, and professionals all contributed to the 915 species that had to be categorized a second time due to mismatched categorizations between Dataset A and Dataset B or time limits during the initial categorizations. We calculated the cumulative rate of mismatches and time limited species within each participant group in each dataset. This was measured by the proportion of species with a mismatched category between the two datasets or time limited, out of the total number of species categorized by members of the participant group in each dataset. The mismatch rate was lowest for professionals (Dataset A: 0.076; Dataset B: 0.073) compared to students (Dataset A: 0.112; Dataset B: 0.131) and amateurs (Dataset A: 0.139; Dataset B: 0.103). At the end of the first round of data collection, 15 of the 56 students, all 7 amateurs, and all 22 professionals categorized at least 100 species across the two datasets and were eligible for authorship credit.

## Dataset comparisons

The majority of the 915 species that were assigned to different categories in Datasets A and B were labeled as a physical trait appearing in both sexes in one dataset and as a male-only physical trait in the other dataset (S2 Fig). This was followed by mismatches in category assignments between geographic location and life history, general descriptor and a physical trait

in both sexes, and life history and behavior. As described in S1 Table, prior to the categorizations of these mismatched species in the second round of data collection, 'life history' was renamed 'natural history' to distinguish it more clearly from the 'geographic location' and 'behavior' categories. The 'General Descriptor' and 'Random' categories were also combined into a 'Miscellaneous' category to clarify the difference between a term that described a specific physical feature observed in a species and one associated with human culture that was unrelated to a species' biology.

## Dataset description

Of the 10,906 species categorized, 57% are named based on avian physical traits, 32% are named based on avian natural history, and 11% are named based on human-constructed terminology unrelated to the biology of the species (Fig 1; Table 2; S1 Appendix). Within avian physical traits, most species are named after traits present in both sexes, although nearly 1,000 species are named after traits only found in males, whereas only 20 species are named after traits only found in females (Fig 1; Table 2). Within avian natural history, most species are named after their geographic location compared to other natural history attributes or behavior (Fig 1; Table 2), and in human-constructed terminology, most species are named after people (eponyms), compared to local language or miscellaneous categorizations (Fig 1; Table 2). There is variation across categories for the number of times a term was repeated across species, with local language having the greatest proportion of distinct terms, and size having the smallest proportion of distinct terms (Table 2).

**Fig 1. Counts of descriptor categories of bird common names in English-language per category in the final dataset of 10,906 bird species.** Categories are colored by avian physical traits (blue), avian natural history (orange), and human-constructed terminology unrelated to the biology of a species (green).

**Table 2. Numbers and proportions of unique descriptors as common names in different categories.**

| Category | General category | N | N unique descriptors | Proportion unique descriptors | Mean times repeated if not unique | Median times repeated if not unique |
|---|---|---|---|---|---|---|
| **Both sexes physical trait** | **Avian physical trait** | 4926 | 1530 | 0.31 | 6.30 | 3 |
| **Male physical trait** | **Avian physical trait** | 933 | 517 | 0.55 | 3.62 | 3 |
| **Female physical trait** | **Avian physical trait** | 20 | 19 | 0.95 | 2.00 | 2 |
| **Size** | **Avian physical trait** | 355 | 28 | 0.08 | 22.80 | 13 |
| **Geographic location** | **Avian natural history** | 2852 | 739 | 0.26 | 7.02 | 4 |
| **Natural history** | **Avian natural history** | 455 | 196 | 0.43 | 4.32 | 2 |
| **Behavior** | **Avian natural history** | 197 | 145 | 0.74 | 3.00 | 2.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Person** | **Human-constructed terminology** | 810 | 505 | 0.62 | 3.19 | 2 |
| **Local language** | **Human-constructed terminology** | 102 | 101 | 0.99 | 2.00 | 2 |
| **Miscellaneous** | **Human-constructed terminology** | 256 | 112 | 0.44 | 7.26 | 4 |
| **Total dataset** | | 10906 | 3499 | 0.32 | 5.89 | 3 |

We next investigated the frequencies of the name descriptors (for this purpose, we considered the first word in the name as the descriptor) across the dataset and across categories. The five most common descriptors in the dataset are “African” (69 times), “Common” (68 times), “Lesser” (65 times), “Great” (65 times), and “Black” (64 times) (Fig 2). We note that after the size-related terms, many of the commonly used physical trait descriptors are related to plumage color (e.g., black, white-browed, spotted) (S3 Fig). The natural history-related common terms are geography related, especially regarding cardinal directions (e.g., northern, southern) (S3 Fig). The human-constructed terminology has very few commonly used terms, except for the term “Common” (S3 Fig).

**Fig 2. Word cloud of most common descriptors across the dataset.** Word sizes are scaled to the number of times they are repeated, and descriptors are colored according to their general category: avian physical trait (blue), avian natural history trait (orange), and human-constructed terminology (green).

## Phylogenetic and geographic trends

All three general categories of species names exhibit weak phylogenetic signal across 10,775 species (Fig 3, S4 Fig, S2 Table). Species names associated with avian physical traits ($D$ = 0.729) and avian natural history traits ($D$ = 0.745) have stronger phylogenetic signal than names associated with human-constructed terminology ($D$ = 0.875). Based on simulation tests of the trait distributions, the phylogenetic signal $D$ of each category differs significantly ($P < 0.001$) from what would be expected under Brownian motion ($D$ = 0), but also differs significantly ($P < 0.001$) from what would be expected from a random trait distribution on the phylogeny ($D$ = 1) for each category.

**Fig 3. Cladogram of 249 families of birds.** The location of each family branch was determined by a representative species in a time calibrated phylogeny of 10,775 species of birds [35]. At the tips, the shaded lines represent the proportion of species in each family with English common names associated with the general categories avian physical traits (blue), avian natural history (orange), and human-constructed terminology unrelated to the biology of the species (green) (S1 Table). See S4 Fig for species-specific name categories mapped to phylogenetic relationships.

The proportion of general descriptor categories is significantly different across geographic regions (Fig 4; $\chi^2$ = 940.43, df = 14, $P < 0.0001$). South America and Middle America show the highest proportions of species named after avian physical traits (0.74 and 0.72, respectively), and Oceans show the lowest proportion (0.26). Oceans show the highest proportion of species named after avian natural history (0.58) and Middle America and South America show the lowest (0.20 and 0.19, respectively). Africa and Oceans show the highest

proportions of species named after human-constructed terminology (both 0.17) and South America show the lowest (0.06).

**Fig 4. The frequency of general descriptor categories in species' breeding ranges.** General categories are avian physical traits (blue), avian natural history (orange), and human-constructed terminology unrelated to the biology of a species (green).

# Discussion

The AvianLexiconAtlas contains assignments for the descriptive English-language common names of 10,906 bird species into 10 categories. We find that 89% of species are named after some aspect of their biology, whether it be their appearance, ecology, behavior, or geographic location. We establish that the most common way to name species is after a physical trait, usually a physical trait that is found in both sexes. We also find a tendency for species that are related to each other or from similar geographic areas to share similar types of descriptors. This effort, which was a successful collaboration among professional ornithologists, amateurs, and students, has led to insights about the nature of term recognition, and the difficulty in some cases of understanding the meaning behind descriptors, including some that many professional ornithologists would have considered to be general knowledge. In addition to the category assignments themselves, the AvianLexiconAtlas contains all references for the assignments outside of the Birds of the World database, a glossary of common terminology and a gazetteer of geographic locations used in English-language common names.

One potential explanation for why physical traits are the most frequently applied category in common names is due to centuries of specimen-based taxonomic description. Many species

were originally described primarily by scientists working with specimens in collections, whether in museums or private collections. Specimens themselves typically only have information about physical traits and size, though well-curated metadata should also reveal geographic location, natural history, and, on occasion, behaviors. If the original historical basis of species taxonomy is specimen-based, then it is sensible that this inertia should impact both modern scientific nomenclature and English-language nomenclature.

The frequency with which types of physical trait descriptors are used in common names aligns with documented patterns of avian trait diversity. Among species that are named after their physical traits, the majority have common names that describe a trait that occurs in both sexes (Fig 1). This outcome aligns with documented evidence that most bird species do not display any sex differences in plumage brightness and/or pattern [41]. Most common names that do describe sex-specific traits of species, however, refer to traits that only occur in the males, and rarely identify traits unique to the females (Fig 1). One potential reason for this discrepancy is the tendency for humans to focus on more elaborate features that are more commonly found in males, which has been proposed as one explanation for the biases towards male specimens in museum collections [42].

Another explanation for the bias towards naming species after traits found only in males is that, due to intersexual selection, many males have more elaborate and faster evolving physical traits than females in dichromatic species [41,43–46]. Males of species in lineages characterized by a large degree of sexual dichromatism may have more unique and prominent traits that can be used as descriptive identifiers compared to females of the same lineage. Female plumage color does vary across species within lineages, but it tends to converge on cryptic colors and patterns that are difficult to distinguish and thus these traits may not be as easy to use for identification

markers between species [43,47–49]. Future work could compare naming patterns to existing data on the degree of sexual dimorphism in species [41,43,50]. It is likely that species that are named after single-sex traits will be more dimorphic than species that are named after both-sex traits, but it remains to be seen whether this pattern would hold for species named after behaviors or other potentially less sex-specific aspects of biology.

We find that avian-centered name categories have stronger phylogenetic signals than human-constructed name categories. This result aligns with previous work establishing that the types of traits used to name species often have phylogenetic signals, including color [51–54], song [55–57], and ecology [45,58,59]. Anecdotally, participants noticed that closely-related species have similar naming schemes, particularly in speciose groups found in the Neotropics. For example, in the 45 species of *Grallaria* antpittas, 22 and 18 species are named after physical traits or geographic locations, respectively. This is also demonstrated by geographic trends, which show that the large majority of species in Middle and South American geographic regions are named after physical traits. Conversely, oceanic species are most commonly named after their natural history (primarily geographic location), likely reflecting that many ocean-dwelling species are difficult to tell apart by plumage or other morphological traits, but live in different geographic areas. We anticipate that further work using the database investigating common group names might find very different patterns, as there are different goals for the components of a name (i.e., to tell species apart or to identify groups of species). Similarly, future work using the database could more specifically analyze different geographic trends in specific common names and the influence of human history and values on these patterns.

Including non-professionals in the data collection was a way to provide both students and amateurs the chance to learn more about ornithology and, for the students in particular, to gain

experience working on a research project alongside professional ornithologists. There is a wide gap between the general public's knowledge of species identification in birds and that of professionals, even for local species [60], and there is also evidence that this knowledge has declined over time in school children [61]. When people have the chance to learn about species, however, their interest in and connection to nature and biodiversity increases [60,62]. Working on this project exposed each of the students and amateurs to a diverse range of bird species that they did not get to select themselves and required them, in most cases, to learn about the biology of a species to make a category assignment. Furthermore, participating in community science projects, such as this one, are important opportunities for non-professionals to gain scientific literacy and can increase participants' interest and engagement in science [63–65].

Although professional ornithologists categorized a majority of the species in both datasets, the non-professionals involved in the data collection made meaningful contributions to the project (S1 Fig). Almost all of the students and amateurs who were eligible for authorship stopped participating in the data collection once they met the minimum threshold of 100 species for authorship credit, whereas most professionals continued to work on the project until all species were categorized. The number of species required for authorship credit was purposefully set low to encourage the participation of non-professionals, who may not have had the same interest or time availability as professional ornithologists. In the two class sessions provided for the data collection, for example, there was not enough time for students to categorize 100 species in the allotted time. Students motivated to pursue authorship credit thus had to spend time outside of class working on the dataset, which did not contribute in any way to their grade in the class. Despite these limitations, it was encouraging to see that all seven amateurs and 15 of the

56 students who participated in the data collection were engaged enough to categorize the necessary number of species to become eligible for authorship.

The non-professionals involved in the data collection also provided an important perspective into the accessibility of common names. While mismatched categorizations did occur more frequently among both amateurs and students compared to the professionals, expertise in ornithology did not exempt the professionals from contributing to mismatched category assignments for species (S1 Fig). Most of these mismatches arose because, in some cases, it is difficult to distinguish whether a physical trait occurs in both sexes of a species or only in males (S2 Fig). In the Maroon Pigeon (*Columba thomensis*), for example, parts of the male's plumage are described in *Birds of the World* as "rich maroon", while female coloration is described as "duller… with only hints of maroon", and this distinction is difficult to make out in the provided images [66]. For issues such as this, we had to decide the extent that a trait would have to occur in both sexes for it to be assigned to that category, and much of this was up to personal interpretation. The resolution to independently categorize the mismatched species again in triplicate was an opportunity for consensus agreement in most species, but disagreements across all three datasets still remained for some species' names. These issues highlight the lack of specificity that is inherent in many of the trait descriptors used in common names.

Another issue that likely contributed to category mismatches was the use of historical and specialized English-language terminology in species names. Diagrams of the names and locations of parts of the bird were provided to all data collectors at the start of the project, but it was the terminology used to describe colors and patterns that proved to be one of the biggest issues (for example, "painted", "festive", or "glowing"). These types of descriptors comprise a majority of the words in the glossary that all participants were asked to contribute to when they

encountered an unknown term. Similarly, the gazetteer was started during the first round of data collection when both professional and non-professional participants struggled to find information on outdated or rarely used names for geographic locations. This brings up a question of how useful specialized terms for color, pattern, and geographic location are in English-language common names, even if they are very specific descriptors, if many people do not know what they mean without further research.

The scope of this project represents a comprehensive starting point for a quantitative examination of the types of terminology used in the common names of birds. We acknowledge that the methodology we developed still limits the conclusions we can make from these data. First, the terminology categories (Table 1; S1 Table) are inherently biased by the goals for this study. The categories were intentionally designed to broadly capture distinct themes in the terminology used for English-language common names, with a particular focus on comparing how often terminology is associated with different properties of avian biology or human culture. There are other ways to investigate the meaning of common names using different types of categories or the inclusion of more specific subcategories, and this presents opportunities for future research projects. For example, Burrowing Owls (*Athene cunicularia*) and Barking Owls (*Ninox connivens*) are both named after behaviors, but the former describes nesting habits and the latter describes vocalizations. Likewise, Golden-crowned Sparrows (*Zonotrichia atricapilla*) and White-throated Sparrows (*Zonotrichia albicollis*) are both named after physical traits (both sexes), specifically color, but the colors themselves and body parts are different. Striped Sparrows (*Oriturus superciliosus*), on the other hand, are named after a pattern that does not appear to be specific to any given body part. The question thus remains – what types of natural history, behaviors, or traits are the birds named after?

The results are also constrained by the decision to use the eBird/Clements Checklist [29]. At the time of the data collection, there were two additional digitally accessible global checklists, one curated by the IOC [20] and the other by Handbook of Birds of the World and BirdLife International [30]. A consequence of this decision is the exclusion of alternative English-language common names associated with the same species, either through local names or alternative names in different checklists. For example, the species officially known as the Jamaican Spindalis (*Spindalis nigricephala*) in the eBird/Clements 2022 Checklist has several local Jamaican names, including Mark Head, Cashew Bird, Silver Head, Spanish Quail, and Champa Beeza [67]. Similarly, the English-language common name of the species *Anisognathus melanogenys* in the eBird/Clements 2022 Checklist is the Black-cheeked Mountain Tanager and describes a physical trait in both sexes. Its common name in the IOC 13.1 and HBW and Birdlife International 7.0 checklists, however, is the Santa Marta Mountain Tanager, which refers to a geographic location. Future work could compare the categorizations of species' officially recognized English-language names to local, or alternative, English-language names on a regional scale.

Lastly, the database is currently limited to the categorizations of common names of birds in American (U.S.) English, which is the standard language of the eBird/Clements checklist [33]. Historically, European and, later, American naturalists played outsized roles in the development of modern taxonomy that occurred alongside the expansion of Western imperialism around the world [1,16]. The results presented here therefore mostly capture the philosophy of Western science in species nomenclature. While it is beyond the scope of this project, a possible next step to expand the database would be to repeat the same methodology for common names in different languages or amongst different taxonomic groups. Comparing the frequency of different types of

descriptors used for the same species names across multiple languages would provide valuable insight into how approaches to avian nomenclature and the philosophy surrounding it vary across cultures and regions. For example, some species might have many names in other languages in culturally diverse regions [68].

The AvianLexiconAtlas demonstrates that while common names of birds serve many purposes in English, there is a strong emphasis on descriptive characteristics associated with avian biology. We also highlight gaps in naming conventions, particularly for descriptors that are not associated with all individuals in a population or include specialized English-language terminology. The English language is constantly evolving at regional and global scales [69,70], which affects the utility of terms in common names. Even the way colors are categorized and described, for example, can vary among people [71]. This database will enable future work that can thus address the extent to which common names represent how people interact with birds today compared with historical interactions and consider differences among cultures.

# Acknowledgements

We thank *Birds of the World* for providing complimentary 1-week subscriptions for the student and amateur participants. We also thank Eden Bayou, Jacob Bijou, Jakiya J. Campbell, Emily Chen, Cindy Cheng, Reema Demopoulos, Savannah Garza, Tamar Hadad, Yujing He, Mia Hejlsberg, Jikke Inia, Yoonseo Jang, Zilai Jin, Risa Kanai, Nick Kruczynski, Langrun Li, Yujia Liu, Queenie Liu, Ziqi Ma, Nyjur Majok, Cecilia Méndez, Kimi Modiri, Yazmin Munoz, Iyioluwa Okediji, Sophia Ordonez, Mohammed Osmanu, Yian Pan, Wend-manegde Pitroipa, Stephanie Salas, Rylie Shaeffer, Andrew Shafer, Paul Shen, Jesse Sivan, Ciaran Timlin, Bryant To, Alexander Valenzuela-Jones, Frances Vandervoort, Nicholas Walsh, Qingyue Wang, Ruoxi

Ye, Yawei Zhang, Guicheng Zhang, Yuchong Zhu, Brian Zou for their contributions to the data collection. Thanks to Emily Webb, Kayce Bell, Jann Vendetti, and the NHMLAC Urban Nature Research Center members for providing comments on a draft of the manuscript.

# References

1. Raven PH, Berlin B, Breedlove DE. The Origins of Taxonomy. Science. 1971;174(4015):1210–1213.
2. Karaffa PT, Draheim MM, Parsons ECM. What's in a Name? Do Species' Names Impact Student Support for Conservation? Hum Dimens Wildl. 2012;17(4):308–310.
3. Ehmke G, Fitzsimons JA, Garnett ST. Standardising English names for Australian bird subspecies as a conservation tool. Bird Conserv Int. 2018;28(1):73–85.
4. Gregg EA, Bekessy SA, Martin JK, Garrard GE. Many IUCN red list species have names that evoke negative emotions. Hum Dimens Wildl. 2020;25(5):468–477.
5. Heard SB. The Name of the Rose (and Everything Else): How Codes and Practices in Naming Biological Species Reflect Cultural Identities. In: Nick IM, editor. Names Naming Law. Routledge; 2023.
6. Heard SB, Mlynarek JJ. Naming the menagerie: creativity, culture and consequences in the formation of scientific names. Proc R Soc B Biol Sci. 2023;290(2010):20231970.
7. Bailenson JN, Shum MS, Atran S, Medin DL, Coley JD. A bird's eye view: biological categorization and reasoning within and across cultures. Cognition. 2002;84(1):1–53.
8. Thiele KR, Conix S, Pyle RL, Barik SK, Christidis L, Costello MJ, et al. Towards a global list of accepted species I. Why taxonomists sometimes disagree, and why this matters. Org Divers Evol. 2021;21(4):615–622.

9. Ride WDL, Cogger HG, Dupuis C, Kraus O, Minelli A, Thompson FC, et al. Int. Code Zool. Nomencl [Internet]. 2012 [cited 2024 Nov 25]. Available from: https://www.iczn.org/the-code/the-international-code-of-zoological-nomenclature/.
10. McAtee WL. Longevity of Bird Names. Names. 1953;1(2):85–102.
11. Kitson PR. Old English bird-names. Engl Stud. 1997;78(6):481–505.
12. Caramaschi U, De Sá RO, Heyer WR. Common names for the frog genus Leptodactylus (Amphibia, Anura, Leptodactylidae). Herpetol Rev. 2005;36(2):119–120.
13. Crother BI. Standard language (insert language of choice) names versus common names. Herpetol Rev. 2007;38(2):143.
14. Link-Pérez MA, Dollo VH, Weber KM, Schussler EE. What's in a Name: Differential labelling of plant and animal photographs in two nationally syndicated elementary science textbook series. Int J Sci Educ. 2010;32(9):1227–1242.
15. García-de-Lomas J, Clavero M, García CM, Alba D, Torres JM. From Linderiella baetica to gambilusa: Involving children in conservation by giving a new species a common name. Aquat Conserv Mar Freshw Ecosyst. 2021;31(6):1543–1547.
16. Driver RJ, Bond AL. Towards redressing inaccurate, offensive and inappropriate common bird names. Ibis. 2021;163(4):1492–1499.
17. Schulenberg TS, Iliff MJ. Methods: Updating the eBird/Clements Checklist 6th Edition [Internet]. Clements Checklist. 2014 [cited 2024 Nov 25]. Available from: https://www.birds.cornell.edu/clementschecklist/about/methods/.
18. Schodde R, Glover B, Kinsky FC, Marchant S, McGill AR, Parher SA. Recommended English Names for Australian Birds. Emu - Austral Ornithol. 1978;77(sup1):245–307.
19. Chesser RT, Billerman SM, Burns KJ, Cicero C, Dunn JL, Hernández-Baños BE, et al.

Check-list of North and Middle American Birds [Internet]. Am. Ornithol. Soc. 2024 [cited 2024 Nov 25]. Available from: https://checklist.americanornithology.org/taxa/.

20. Gill F, Donsker D, Rasmussen P. English Names: Principles [Internet] IOC World Bird List v14.2. 2024 Aug 17 [cited 2024 Nov 25]. Available from: https://www.worldbirdnames.org/new/english-names/principles/.

21. Remsen JV Jr, Areta JI, Bonaccorso E, Claramunt S, Del-Rio G, Jaramillo A., et al. A classification of the bird species of South America [Internet]. Mus. Nat. Sci. La. State Univ. 2024 Nov 18 [cited 2024 Nov 25]. Available from: https://www.museum.lsu.edu/~Remsen/SACCBaseline.htm.

22. Guedes P, Alves-Martins F, Arribas JM, Chatterjee S, Santos AMC, Lewin A, et al. Eponyms have no place in 21st-century biological nomenclature. Nat Ecol Evol. 2023;7(8):1157–1160.

23. Liu IA, Gulson-Castillo ER, Wu JX, Demery A-JC, Cortes-Rodriguez N, Covino KM, et al. Building bridges in the conversation on eponymous common names of North American birds. Ibis. 2024;166(3):1092–1102.

24. Bird Names For Birds [Internet]. 2020 [cited 2024 Nov 25]. Available from: https://birdnamesforbirds.wordpress.com/.

25. Beolens B, Watkins M. Whose Bird?: Common Bird Names and the People They Commemorate. Yale University Press; 2004.

26. Jobling JA. The Helm Dictionary of Scientific Bird Names: From Aalge to Zusii. London: Christopher Helm; 2010.

27. Meiter GH. Bird is the Word: An Historical Perspective on the Names of North American Birds. Newark, Ohio: The McDonald & Woodward Publishing Company;

2020.

28. Myers S. The Bird Name Book: A History of English Bird Names. Princeton, NJ: Princeton University Press; 2022.

29. Clements JF, Schulenberg TS, Iliff MJ, Federicks TA, Lepage D, Billerman SM, et al. The eBird/Clements checklist of Birds of the World: v2022 [Internet]. 2022 Oct [cited 2024 Nov 25]. Available from: https://www.birds.cornell.edu/clementschecklist/updateindex/october-2022/.

30. HBW and BirdLife International. Handbook of the Birds of the World and BirdLife International Digital Checklist of the Birds of the World [Internet]. 2022 [cited 2024 Nov 25]. Available from: https://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v7_Dec22.zip.

31. Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS., editors. Birds of the World [Internet]. Ithaca, NY: Cornell Laboratory of Ornithology; 2024 [cited 2024 Nov 25]. Available from: https://birdsoftheworld.org/bow/home.

32. Birds of the World. Taxonomy [Internet]. Ithaca, NY: Cornell Laboratory of Ornithology. 2024 [cited 2024 Nov 25]. Available from: https://birdsoftheworld.org/bow/content/taxonomy.

33. eBird. Bird Names in eBird [Internet]. Ithaca, NY: Cornell Laboratory of Ornithology. 2024 Sep 6 [cited 2024 Nov 25]. Available from: https://support.ebird.org/en/support/solutions/articles/48000804865-bird-names-in-ebird.

34. Clements JF, Schulenberg TS, Iliff MJ, Federicks TA, Gerbracht JA, Lepage D, et al.

The eBird/Clements checklist of Birds of the World: v2021 [Internet]. EBirdClements Checklist. Birds World V2021. 2021 [cited 2024 Nov 25]. Available from: https://www.birds.cornell.edu/clementschecklist/updateindex/august-2021/.

35. McTavish EJ, Gerbracht JA, Holder MT, Iliff MJ, Lepage D, Rasmussen P, et al. A complete and dynamic tree of birds. BioRxiv [Preprint]. 2024 bioRxiv 05.20.595017 [posted 2024 May 22; cited 2024 Nov 25]. Available from: https://www.biorxiv.org/content/10.1101/2024.05.20.595017v1.

36. Miller E, McTavish EJ. clootl [Internet]. 2024 [cited 2024 Nov 25]. Available from: https://github.com/eliotmiller/clootl/tree/master.

37. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2023 [cited 2024 Nov 25]. Available from: https://www.r-project.org/.

38. Fritz SA, Purvis A. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. Conserv Biol. 2010;24(4):1042–1051.

39. Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, et al. caper: Comparative analyses of phylogenetics and evolution in R [Internet]. 2023 [cited 2024 Nov 25]. Available from: https://cran.r-project.org/web/packages/caper/index.html.

40. Gill F, Donsker D, Rasmussen P, editors. IOC World Bird List (v13.2) [Internet]. 2023 [cited 2024 Nov 25]. Available from: https://doi.org/10.14344/IOC.ML.13.0.

41. Gonzalez-Voyer A, Thomas GH, Liker A, Krüger O, Komdeur J, Székely T. Sex roles in birds: Phylogenetic analyses of the influence of climate, life histories and social environment. Ecol Lett. 2022;25(3):647–660.

42. Cooper N, Bond AL, Davis JL, Portela Miguez R, Tomsett L, Helgen KM. Sex biases in bird and mammal natural history collections. Proc R Soc B Biol Sci. 2019;286(1913):20192025.

43. Dale J, Dey CJ, Delhey K, Kempenaers B, Valcu M. The effects of life history and sexual selection on male and female plumage colouration. Nature. 2015;527(7578):367–370.

44. Delhey K. The colour of an avifauna: A quantitative analysis of the colour of Australian birds. Sci Rep. 2015;5(1):18514.

45. Cooney CR, Varley ZK, Nouri LO, Moody CJA, Jardine MD, Thomas GH Sexual selection predicts the rate and direction of colour divergence in a large avian radiation. Nat Commun. 2019;10:1773.

46. Nolazco S, Delhey K, Nakagawa S, Peters A. Ornaments are equally informative in male and female birds. Nat Commun. 2022;13(1):5917.

47. Price JJ, Eaton MD. Reconstructing the evolution of sexual dichromatism: Current color diversity does not reflect past rates of male and female change. Evolution. 2014;68(7):2026–2037.

48. Shultz AJ, Burns KJ. The role of sexual and natural selection in shaping patterns of sexual dichromatism in the largest family of songbirds (Aves: Thraupidae). Evol Int J Org Evol. 2017;71(4):1061–1074.

49. Delhey K, Valcu M, Muck C, Dale J, Kempenaers B. Evolutionary predictors of the specific colors of birds. Proc Natl Acad Sci. 2023;120(34):e2217692120.

50. Owens IPF, Hartley IR. Sexual dimorphism in birds: why are there so many different forms of dimorphism? Proc R Soc Lond B Biol Sci. 1998;265(1394):397–407.

671 51. Stoddard MC, Prum RO. Evolution of Avian Plumage Color in a Tetrahedral Color
672 Space: A Phylogenetic Analysis of New World Buntings. Am Nat.
673 2008;171(6):755–776.
674 52. Shultz AJ, Burns KJ. Plumage evolution in relation to light environment in a novel
675 clade of Neotropical tanagers. Mol Phylogenet Evol. 2013;66(1):112–125.
676 53. Marcondes RS, Brumfield RT. Fifty shades of brown: Macroevolution of plumage
677 brightness in the Furnariida, a large clade of drab Neotropical passerines. Evolution.
678 2019;73(4):704–719.
679 54. Merwin JT, Seeholzer GF, Smith BT. Macroevolutionary bursts and constraints
680 generate a rainbow in a clade of tropical birds. BMC Evol Biol. 2020;20(1):32.
681 55. Freeman BG, Montgomery GA, Schluter D. Evolution and plasticity: Divergence of
682 song discrimination is faster in birds with innate song than in song learners in
683 Neotropical passerine birds. Evolution. 2017;71(9):2230–2242.
684 56. Mason NA, Burns KJ, Tobias JA, Claramunt S, Seddon N, Derryberry EP. Song
685 evolution, speciation, and vocal learning in passerine birds. Evolution.
686 2017;71(3):786–796.
687 57. Demery A-JC, Burns KJ, Mason NA. Bill size, bill shape, and body size constrain bird
688 song evolution on a macroevolutionary scale. Ornithology. 2021;138(2):ukab011.
689 58. Phillimore AB, Freckleton RP, Orme CDL, Owens IPF. Ecology Predicts Large-Scale
690 Patterns of Phylogenetic Diversification in Birds. Am Nat. 2006;168(2):220–229.
691 59. Hawkins BA, Diniz-Filho JAF, Jaramillo CA, Soeller SA. Climate, Niche
692 Conservatism, and the Global Bird Diversity Gradient. Am Nat.
693 2007;170(S2):S16–S27.

60. Hooykaas MJD, Schilthuizen M, Aten C, Hemelaar EM, Albers CJ, Smeets I. Identification skills in biodiversity professionals and laypeople: A gap in species literacy. Biol Conserv. 2019;238:108202.

61. Gerl T, Randler C, Jana Neuhaus B. Vertebrate species knowledge: an important skill is threatened by extinction. Int J Sci Educ. 2021;43(6):928–948.

62. Cox DTC, Gaston KJ. Likeability of Garden Birds: Importance of Species Knowledge & Richness in Connecting People to Nature. PLOS ONE. 2015;10(11):e0141505.

63. Oberhauser K, LeBuhn G. Insects and plants: engaging undergraduates in authentic research through citizen science. Front Ecol Environ. 2012;10(6):318–320.

64. Bonney R, Phillips TB, Ballard HL, Enck JW. Can citizen science enhance public understanding of science? Public Underst Sci. 2016;25(1):2–16.

65. Ruiz-Mallén I, Riboli-Sasco L, Ribrault C, Heras M, Laguna D, Perié L. Citizen Science: Toward Transformative Learning. Sci Commun. 2016;38(4):523–534.

66. Baptista LF, Trail PW, Horblit HM, Boesman PFD, Garcia EFJ. Maroon Pigeon (Columba thomensis), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E, editors. Birds of the World [Internet]. Ithaca, NY: Cornell Lab of Ornithology; 2020 [cited 2024 Nov 25]. Available from :https://doi.org/10.2173/bow.marpig1.01.

67. Raffaele H, Wiley J, Garrido O, Keith A, Raffaele J. A Guide to the Birds of the West Indies. Princeton, NJ: Princeton University Press; 1998.

68. Sicard-Ayala AM, Jaramillo-Mejía L, Ayerbe-Quiñones F. Un ave, muchos nombres: un pluriverso. Ornitol Colomb. 2019;17(17):eC01.

69. Galloway N, Rose H. The “New” Englishes. In: Introducing Global Englishes. Oxford, UK: Taylor & Francis Group; 2015. p. 95–123.

70. McWhorter JH. Words on the move: Why English won't- and can't- sit still (like, literally). New York: Henry Holt and Co.; 2016.

71. Lindsey DT, Brown AM. The color lexicon of American English. J Vis. 2014;14(2):17.

# Supporting information

**S1 Fig. Summary of participation in crowdsourced data collection based on the data collectors' expertise in ornithology.**

**S2 Fig. Summary of mismatched category assignments in the 915 species common names assigned to different categories in Dataset A and Dataset B.**

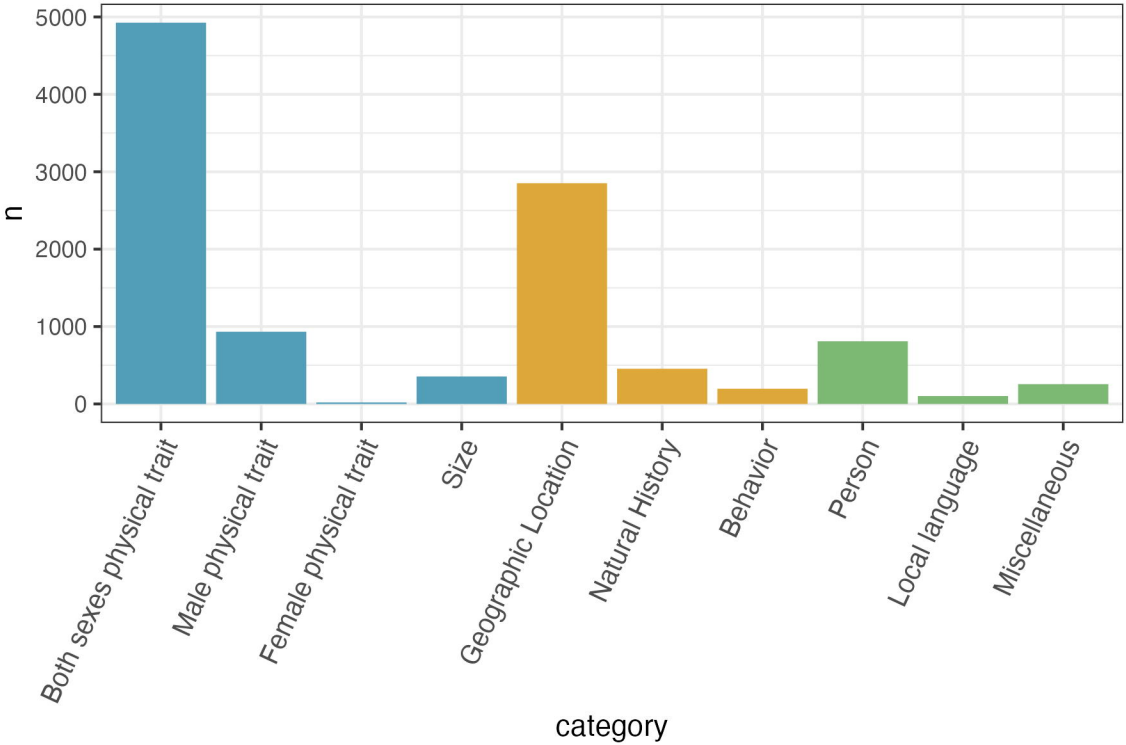
**S3 Fig. Word cloud frequencies of terminology in common names.**

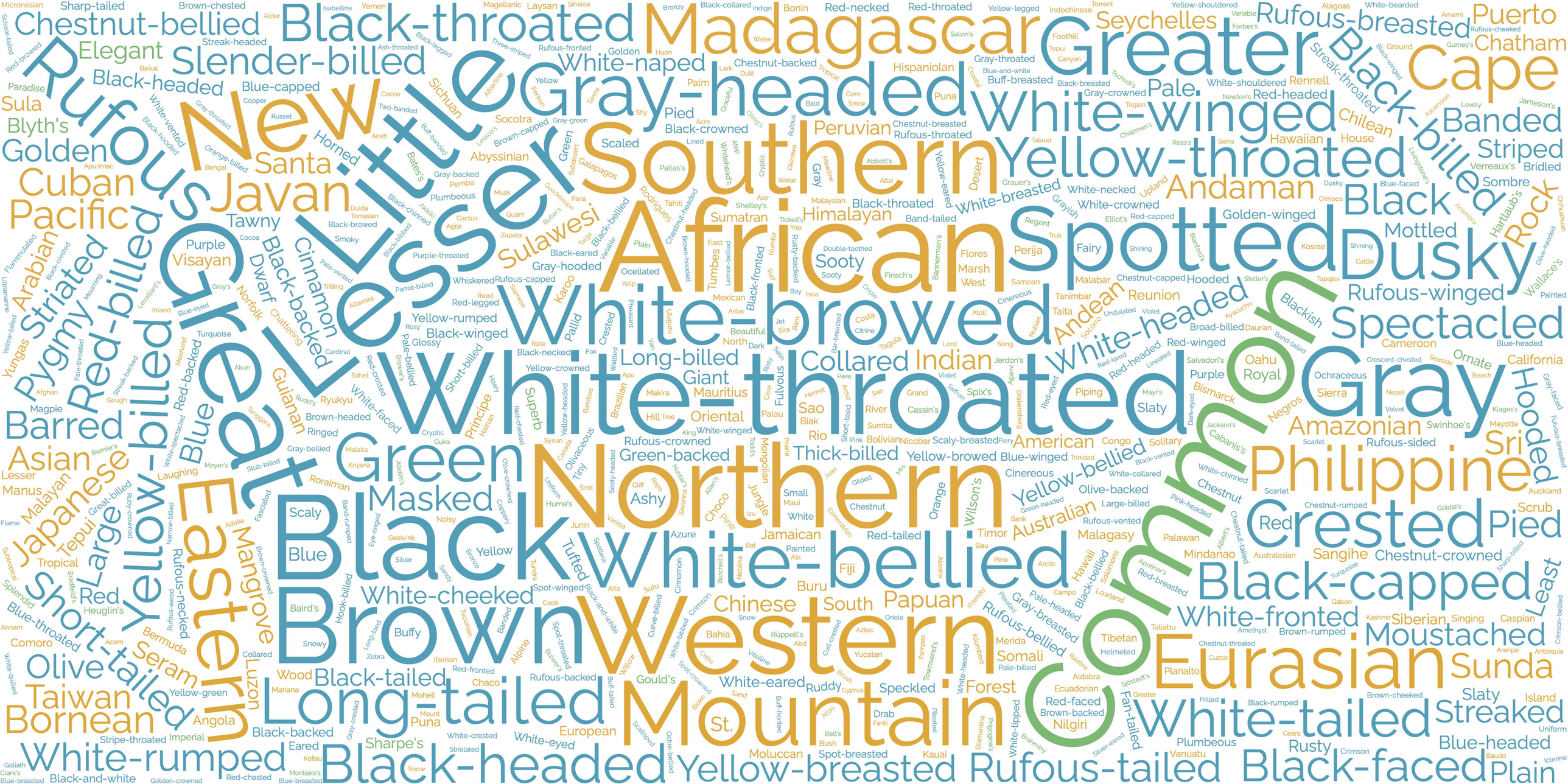
**S4 Fig. Cladogram of the categories assigned to the English common names of 10,775 species of birds.**

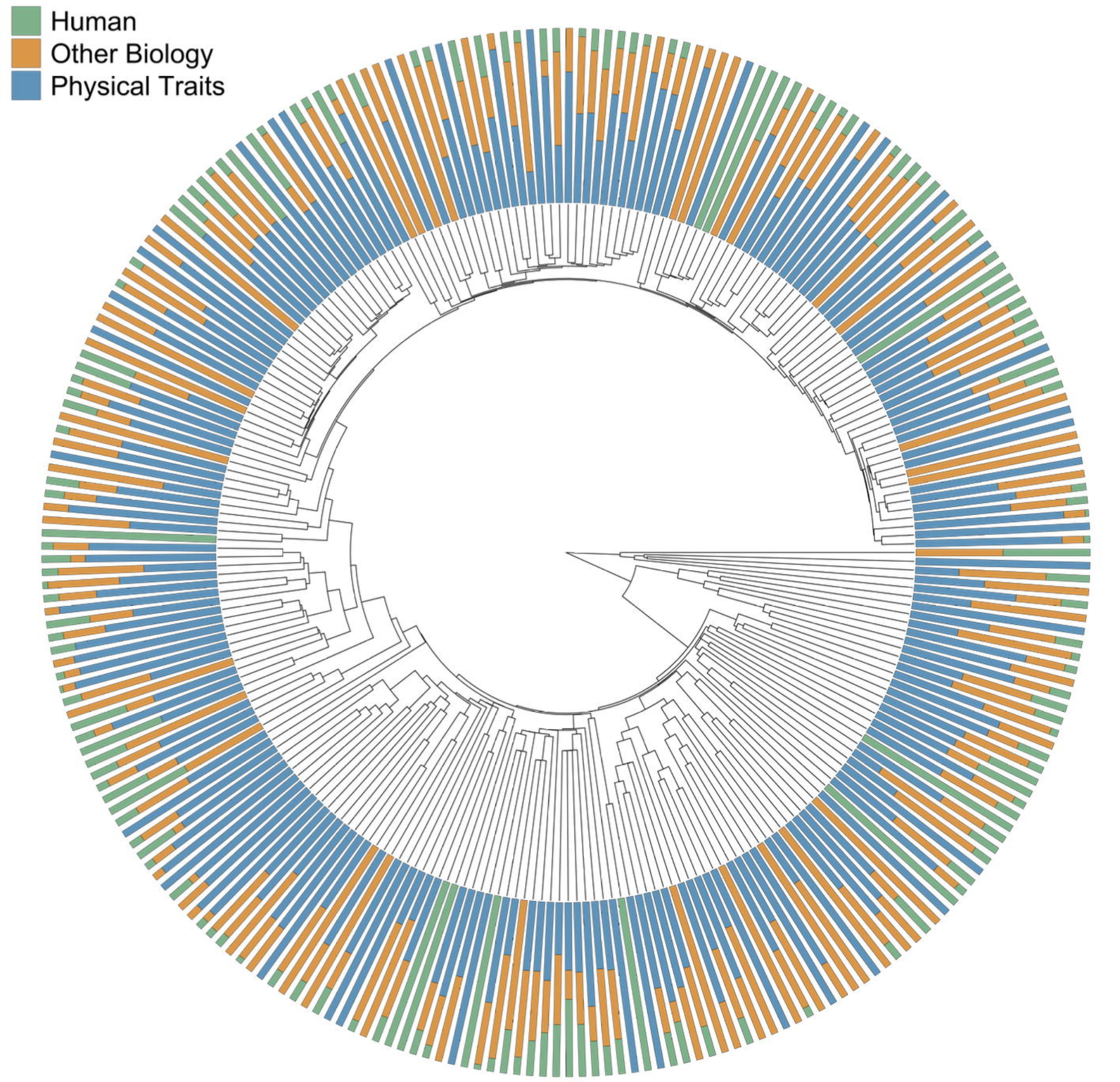
**S1 Table. Categories of unique descriptors in English common names of birds.**

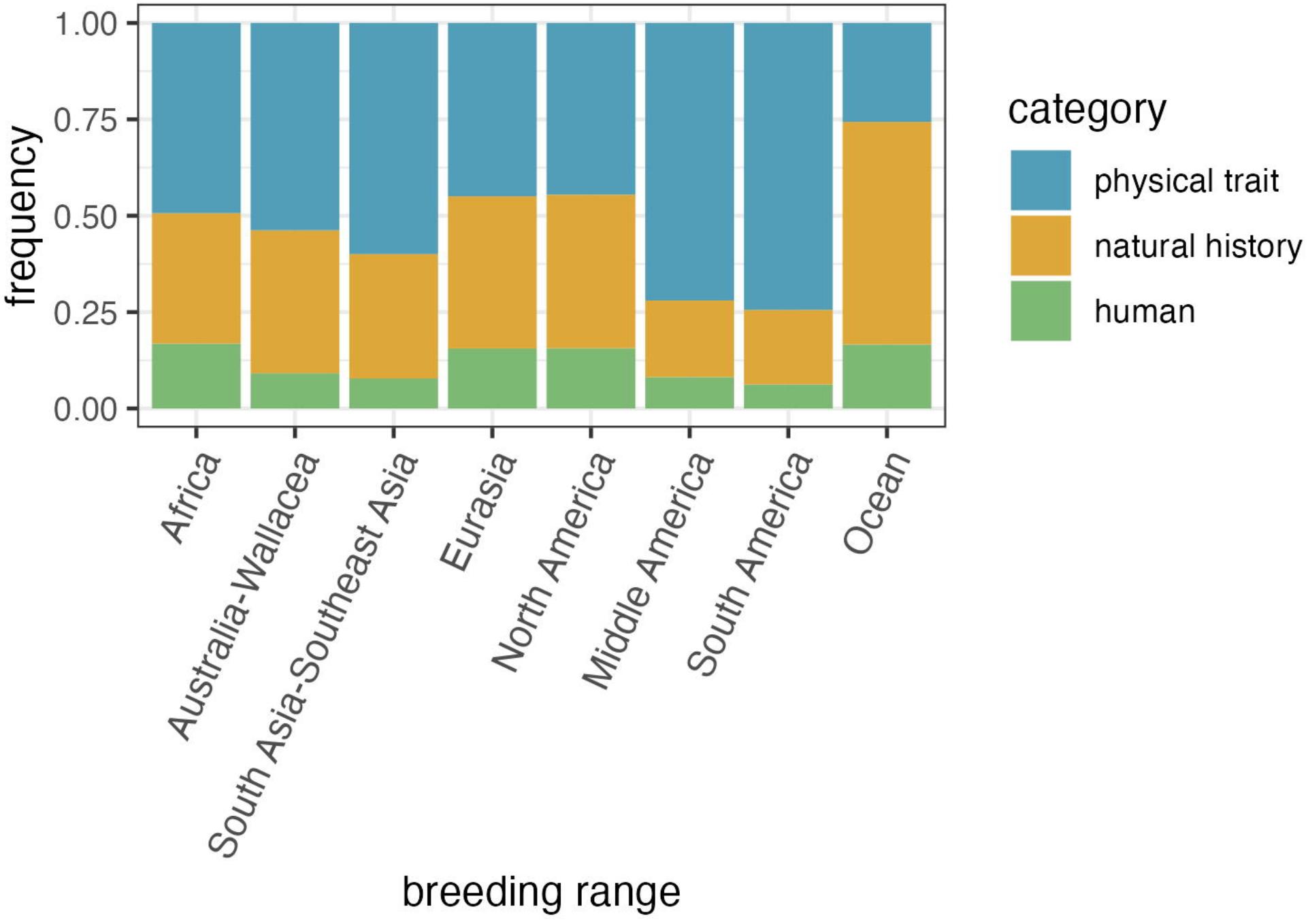
**S2 Table. Calculation of D statistic for the phylogenetic structure of categories.**

**S1 Appendix. AvianLexiconAtlas Database Files.** The data, glossary, and gazetteer reported in this article can be accessed at https://github.com/ajshultz/AvianLexiconAtlas.


5000
4000
3000
2000
1000
0
n
Both sexes physical trait
Male physical trait
Female physical trait
Size
Geographic Location
Natural History
Behavior
Person
Local language
Miscellaneous
category


Madagascar
Greater
Black-throated
Chestnut-bellied
Slender-billed
Elegant
Rufous-breasted
Seychelles
Puerto
Chatham
Cape
Black-billed
Streak-throated
Rufous
New
Little
Gray-headed
White-winged
Southern
Yellow-throated
Banded
Striped
Golden
Black-headed
Blue-capped
Sula
Blyth's
Santa
Horned
Sichuan
Socotra
Abyssinian
Peruvian
Pied
Pale
Andaman
Black
Rock
Cuban
Javan
Pacific
Tawny
Sulawesi
Sumatran
Himalayan
Lesser
Great
African
Spotted
Dusky
Mottled
Purple
Visayan
Cinnamon
Black-backed
Dwarf
Arabian
Striated
Pygmy
Red-billed
Yellow-billed
White-browed
Rufous-winged
Spectacled
Andean
White-headed
Blackish
Reunion
Long-billed
Collared
Indian
White-throated
Gray
Common
Oahu
Royal
Ornate
California
Hooded
Sri
Philippine
Amazonian
Barred
Blue
Guianan
Superb
Principe
Giant
Mauritius
Oriental
Green
Green-backed
Thick-billed
American
Yellow-bellied
Asian
Japanese
Malayan
Tepui
Large
Masked
Northern
Black
Eastern
Short-tailed
Olive
Taiwan
Bornean
Seram
Manglove
Red
Brown
Western
White-bellied
Black-capped
Crested
Pied
Least
Australian
Malagasy
Jamaican
Chinese
South
Papuan
White-cheeked
White-fronted
Moustached
Eurasian
Sunda
Long-tailed
Black-tailed
Mountain
White-eared
Ruddy
Speckled
Forest
Gould's
Tibetan
Helmeted
Red-faced
Brown-backed
Nilgiri
Fan-tailed
White-tailed
Streaked
Slaty
Blue-headed
Rusty
Plumbeous
Black-faced
Plain
Rufous-tailed
Yellow-breasted
Black-headed
White-rumped
Sharpe's
White-eyed
European
Moluccan
Spot-breasted
Kauai
Vanuatu
Sangihe
Chestnut-crowned
Mindanao
Palawan
Siberian
Singing
Caspian
Rufous-bellied
Gray-breasted
Pale-headed
Somali
Merida
Yucatan
Aldabra
Ecuadorian
Red-throated
Rufous-throated
Chestnut-breasted
Black-throated
Band-tailed
Sooty
Flores
Marsh
West
Perija
Fairy
Malabar
Chestnut-capped
Hooded
Golden-winged
White-crowned
White-necked
Upland
Grayish
Regent
Truk
Elliot's
Red-capped
Blue-faced
Sombre
Bridled
Verreaux's
Hawaiian
House
Chilean
Ross's
Serra
Talaud
Chapman's
Yellow-eared
Desert
Grauer's
White-breasted
Gray
Meadow
Okinawa
Rufous
Whitehead's
Cryptic
Abbott's
Altai
Alor
Shelley's
Malaysian
Black-throated
Band-tailed
Tickell's
Yap
Numfor
Rusty-backed
Bannerman's
Oriole
Bay
Inca
Double-toothed
Sooty
Finsch's
Black-fronted
Lemon-bellied
Brown-hooded
Tumbes
East
Chestnut-headed
Rodrigues
Tahiti
Galapagos
Green
Subdesert
Scaled
Black-crowned
Lined
Acre
Palm
Tanna
Two-banded
Cocos
Aceh
Buff-banded
Lesson's
Brown-capped
Gray-green
Bates's
Black-bellied
Black-chinned
Guadeloupe
Paria
Duida
Torresian
Black-browed
Smoky
Cocoa
Zapata
Taiga
Black-ribbed
Variable
Black-eared
Purple-throated
Gray-hooded
Rufous-capped
Whiskered
Parrot-billed
Altamira
Karoo
Pallid
Crested
Pleasant
Uluguru
Mexican
Arfak
Beautiful
North
Dark
Black-necked
Yellow-crowned
Watlled
Apo
Brewer's
Short-billed
Heavy
Rudd's
Ryukyu
Brown-headed
Ringed
Gray-bellied
Malaita
Knysna
Canada
Olivaceous
Tiny
Sooty-headed
Hunter's
Savanna
Allen's
Todd's
Prairie
Mongolian
Jungle
Choco
Pine
Izu
Cave
Fulvous
Rio
Bolivian
Nicobar
Scaly-breasted
Fiery
Yellow-browed
Blue-winged
Trinidad
Cinereous
Olive-backed
Large-billed
Rufous-vented
Malagasy
Palawan
Mindanao
Australasian
Sangihe
Chestnut-crowned


Human
Other Biology
Physical Traits


1.00
0.75
0.50
0.25
0.00
frequency
Africa
Australia-Wallacea
South Asia-Southeast Asia
Eurasia
North America
Middle America
South America
Ocean
breeding range
category
physical trait
natural history
human