www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

The Avian Diet Database as a source of quantitative information on bird diets

Allen H. Hurlbert[1,2] ✉, **Aaron M. Olsen**[3], **Melissa M. Sawyer**[1] & **Patrick M. Winner**[1,4]

This data paper describes a compilation of 73,075 quantitative diet data records for 759 primarily North American bird species, providing standardized information not just on the diet itself, but on the context for that diet information including the year, season, location, and habitat type of each study. The methods used for collecting and cleaning these data are described, and we present tools for summarizing and visualizing diet information by bird species or prey.

Background & Summary

Diet is one of the most fundamental aspects of an organism's ecology, and a knowledge of diet is required to better understand connections within food webs, potential determinants of population size and distribution, and potential conservation threats. For birds, existing compilations exist that classify diet broadly into categories such as insectivore, omnivore, and frugivore[1,2], but finer resolution information about the specific prey items used and their relative importance are typically scattered throughout the literature in studies focused on only one or a few species at a time. One exception is the effort by the United States Department of Agriculture in the early twentieth century to quantify the stomach contents of over 100 North American bird species, identifying prey items to species where possible[3–10]. Nevertheless, these data are often reported in verbal summaries, and the taxonomy of both birds and prey items is in many cases outdated.

This data paper describes a compilation of 73,075 quantitative diet data records for 759 bird species, providing standardized information not just on the diet itself, but on the context for that diet information including the year, season, location, and habitat type of each study. While the species in this compilation are primarily from North America, this dataset will continue to grow and may expand in geographic coverage in the future. Below we describe our procedures for identifying datasets, standardizing those data, and conducting quality control and quality assurance checks.

Methods

Data discovery. We searched the literature for studies with quantitative data on diet. For North American bird species, we examined every citation we could access from the "Diet and Foraging" section of each relevant Birds of North America[11] (now Birds of the World, birdsoftheworld.org) species account. In addition, we performed literature searches on Google Scholar and Web of Science using the keywords "[common name] AND (diet OR foraging OR food)", replacing [common name] with the name of the bird species under investigation. For Anseriformes, studies were located using Google Scholar searches and Birds of North America species accounts, as described previously, in addition to ornithology handbooks[12–14].

Data entry. For each species for which quantitative diet data were found, we recorded details surrounding the study as described in Online-only Table 1. Researchers have historically used three different methods for quantifying the prevalence of a prey type. The fraction of the diet by number of items, the fraction of the diet by weight or volume, and the percent occurrence (i.e., the fraction of birds examined that consumed at least one member of a given a prey type). While the first two measures sum to 100% across all diet items in a study, the last one does not. In just a handful of studies, the authors quantified diet in terms of percent total energy content, or percent surface area of stomach contents examined on a miscroscope stage. For our standardization purposes, these were

[1]Department of Biology, University of North Carolina, Chapel Hill, NC, 27599, USA. [2]Energy, Ecology and Environment Program, University of North Carolina, Chapel Hill, NC, 27599, USA. [3]Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, 02912, USA. [4]Present address: NOVA Engineering and Environmental, 2201 Brentwood Rd Suite 105, Raleigh, NC, 27604, USA. ✉e-mail: hurlbert@bio.unc.edu

both classified as fraction of diet by weight or volume, but the exact method was recorded in the Notes field. In some cases, while quantitative information on diet was provided, the authors were unclear on which measure of prevalence was used, and we have recorded the Diet Type as "Unspecified".

Some studies reported a non-zero, or "trace", amount of a particular prey type in the diet, or reported an upper bound (e.g., "< 0.01"). In these cases, we entered a value of half of the lowest reported numeric diet fraction, or half of the upper bound. For example, if a study calculated that the prey item with the lowest representation in the diet that was still represented numerically was 0.014, and another prey item was reported as "trace", that trace item would be entered as having a diet fraction of 0.007.

In some cases, researchers reported diet fractions for plant and animal portions of the diet separately (e.g., "beetles make up 30% of all animal food, grass seeds make up 10% of all plant food"). When possible, we used reported sample sizes of each diet portion to recalculate values that reflected the prevalence of those items in the diet as a whole.

Bird taxonomy. Bird taxonomy in the database follows the 2019 Clements/eBird Checklist[15]. Species names as used in older published studies that did not match any names in the most recent Clements/eBird checklist at the time of data entry were searched in the online database AviBase (https://avibase.bsc-eoc.org/) which has features that allow one to map old names onto current taxonomic concepts. While simple name changes (e.g. genus *Dendroica* changing to *Setophaga*) were straightforward to incorporate, taxonomic splits were more complicated. For example, "Western Flycatcher" *Empidonax difficilis* was split into two species, Pacific-slope Flycatcher which retained the original scientific name, and Cordilleran Flycatcher, *Empidonax occidentalis*. Where possible, we used the geographic location of the study to determine which of the two split species was likely being referred to by the original study. When this was not possible due to range overlap of the two split species, the name would be reported as, for example, Pacific-slope/Cordilleran Flycatcher and *Empidonax difficilis/occidentalis*. In the case of taxonomic lumps, the relevant subspecies was listed for formerly distinct species when possible.

Data Records

The latest official release of the database can be found in the file AvianDietDatabase.txt archived on Zenodo[16], while the latest unofficial version to which new studies are continually added can be found in the AvianDietDatabase.txt file on Github (https://github.com/hurlbertlab/dietdatabase). Each record corresponds to a trophic link between one bird species and one prey taxon, along with the metadata detailing the location, date, habitat, and strength of this observation as characterized by the fields described in Online-only Table 1.

Technical Validation

Data quality checking. After initial data entry for a study, a cleaning script was run on the data to perform the following checks. 1) Bird common names, scientific names, and family names were checked against the 2019 Clements/eBird checklist[15]. 2) Data fields for which numeric values were expected were checked to ensure that values were indeed numeric, and potential outlier values were flagged based on accepted ranges of values for each field (Online-only Table 1). 3) Any values for Location_Region within North America that were not the names of U.S. or Mexican states or Canadian provinces were flagged, while outside of North America any values that were not country names were flagged. 4) Values for Habitat_type, Prey_Stage, Prey_Part, Diet_Type, and Study_Type were checked for allowable terminology (Online-only Table 1). 5) Finally, for each analysis (often a study, but occasionally a single study reported several diet analyses by season or habitat) broken down by number of items, weight or volume, or unspecified (but not by fraction occurrence), we checked that the Fraction_Diet values summed to 1. Analyses with sums that deviated by more than 0.03 (some small deviation is expected due to the summing of rounding errors) were compared to the original study for typos, and any records for which the analysis sum deviated from 1 by more than 0.03 have the phrase "values as reported do not sum to 100%" in the Notes field.

Prey taxonomy and name cleaning. We recorded the names of prey items with as much taxonomic resolution as the original study allowed, following the Integrated Taxonomic Information System (ITIS; https://www.itis.gov) taxonomy where possible. If the reported prey name did not match any ITIS taxonomic concept, then it was entered into the Global Names Resolver (https://resolver.globalnames.org) to see whether it matched an outdated name that could be mapped to a current taxonomic concept in ITIS. We list the ITIS taxonomic serial number in the Prey_Name_ITIS_ID field and report the Prey_Name_Status as "verified" when a prey name was successfully matched to a concept in ITIS. Some names had no match in ITIS, but were regarded as valid taxonomic concepts by other authorities (e.g. Encyclopedia of Life, National Center for Biotechnology Information, World Register of Marine Species, etc.). In such cases, the Prey_Name_ITIS_ID was assigned 'NA' and we reported the Prey_Name_Status as "accepted".

In some cases, researchers utilized an "other" or "unidentified" taxonomic category which required a particular form of encoding. For example, a study might report the diet fraction for "Ants", "Vespid wasps", and "Other wasps" or "Unidentified wasps". These would be reported as "Hymenoptera: Formicidae", "Hymenoptera: Vespidae", and "Hymenoptera", respectively. However, in this example, the last entry of "Hymenoptera" really refers to "all Hymenoptera not in Formicidae or Vespidae", and is not comparable with the "Hymenoptera" of another study which was taxonomically inclusive. As such, we created a separate field in the database called "Prey_Taxon_Inclusive". In this example, "yes" would be entered for both ants and vespid wasps because the diet fractions presented refer to the inclusive set of prey delimited by that prey name reported, but for the Hymenoptera record corresponding to "Unidentified wasps" the "Prey_Taxon_Inclusive" field would be "no".

Taxonomic changes within the common insect order Hemiptera merit specific attention. Up until the late twentieth century, "Hemiptera" was used to refer to the True Bugs, which are now defined as the suborder

Heteroptera within order Hemiptera. "Homoptera" had been used to refer to aphids, psyllids, leafhoppers, treehoppers and cicadas, but is now no longer considered a valid name. Instead, aphids, psyllids and scale insects make up the suborder Sternorrhyncha within Hemiptera, while leafhoppers, treehoppers, and cicadas make up the suborder Auchenorrhyncha within Hemiptera. Thus, an older study that reported prey as "Hemiptera" and "Homoptera" were entered as "Hemiptera: Heteroptera" and "Hemiptera", with the latter listed as not inclusive. In some cases, studies provided additional information (e.g., "Homoptera (principally leafhoppers)") allowing us to specify an inclusive suborder designation.

Usage Notes

In addition to the raw data, we provide two means of exploring the Avian Diet Database and extracting species- or prey-specific summaries. The first is through the website https://aviandiet.unc.edu where users can enter a bird species name to explore a summary of diet information known for that species, or a prey name to explore which bird species are known to eat that prey taxon. We also provide an R package ('aviandietdb') for exploring the database, which should be loaded in R by typing:

```
install.packages("devtools")
library(devtools)
devtools::install_github("ahhurlbert/aviandietdb")
library(aviandietdb)
```

Three useful R functions for summarizing records in the database are detailed below.

dbSummary().

Example usage:

```
dbSummary()
```

This function returns the total number of database records, the unique number of bird species, and the unique number of publications summarized in the Diet Database. In addition, it provides a tally of the number of records by bird species listed in alphabetical order, as well as a summary for each bird family in the American Birding Association (ABA) Checklist (version 8.0.6a) of 1) the number of species in the family in the database, 2) the total number of species in the family based on the ABA checklist, and 3) the percent of the family represented based on the species expected in North America. This information on taxonomic coverage is also provided in Online-only Table 2.

speciesSummary().

Example usage:

```
speciesSummary("Bald Eagle", by="Order")
```

This function provides a summary of the total number of records and total number of studies available in the database for this species, along with a summary of how those records are distributed across seasons, years, and geographic regions. The number of records are also summarized by taxonomic level to which prey were identified and by analysis type (by number of items, weight or volume, occurrence, or unspecified). Finally, for each analysis type, the mean fraction of diet is given for each prey category at the hierarchical taxonomic level specified with the "by" argument. This is an overall mean, averaged across year, region, and season. If the original data source indicated that specific parts of the prey taxon were consumed (e.g. fruit, seed, vegetation, etc.) then they are listed in the Prey_Part field.

dietSummary().

Example usage:

```
dietSummary("Bald Eagle", season="summer", region="California", year-
Range=c(1940, 1970), by="Order", dietType="Items")
```

This function allows one to specify season, region, a year range, analysis type, and taxonomic level for prey summarization, and then provides the mean fraction of diet information based on all studies meeting the stated criteria.

dietSummaryByPrey().

Example usage:

```
dietSummaryByPrey("Lepidoptera", preyLevel="Order", dietType="Items",
yearRange=c(1985, 2000), season="summer", preyStage="larva",
speciesMean=TRUE)
```

This function provides a list of all bird species that consume a particular prey taxon in decreasing order of importance. In addition to providing the prey taxon name, you must also specify the taxonomic level (*preyLevel*) of that name. Like *dietSummary()*, this function allows one to specify season, region, a year range, and analysis type. There are two additional arguments not present in dietSummary(). One is *preyStage*, which specifies the life stage of the prey item (if applicable) for which a summary should be conducted. By default ('any'), diet records will be included regardless of prey stage. Alternatively, one can specify that the summary should only be conducted for records including the terms 'larva', 'adult', or 'pupa' in the Diet Database's 'Prey_Stage' field. This is most relevant for Lepidoptera and a few other insect groups, where one might want to single out the importance of caterpillars or other larvae, for example.

By specifying *speciesMean* = TRUE, only a single value is returned for each bird species that is known to consume a specified prey taxon which represents the average across all analyses meeting the season, region, and year criteria. If *speciesMean* is FALSE, then each analysis of a bird species which meets the specified criteria will be listed separately.

Example code and output is available in the Github README.md document (https://github.com/ahhurlbert/aviandietdb).

Code availability

R code for quality assurance/quality control and prey taxonomic name cleaning can be found in the "database_error_checking.R" script in the database building project's Github repository (https://github.com/hurlbertlab/dietdatabase). An R package with functions for loading and querying the database is called 'aviandietdb' and is available at https://github.com/ahhurlbert/aviandietdb.

Received: 5 March 2021; Accepted: 30 August 2021;

References

1. Sekercioglu, C. H., Daily, G. C. & Ehrlich, P. R. Ecosystem consequences of bird declines. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 18042–18047 (2004).
2. Wilman, H. *et al*. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027–2027 (2014).
3. Beal, F. E. L. *Food habits of the swallows: A family of valuable native birds*. (US Department of Agriculture, 1918).
4. Beal, F. E. L. *Birds of California in relation to the fruit industry*. (US Department of Agriculture, Biological Survey, 1907).
5. Beal, F. E. L. *Food of the woodpeckers of the United States*. (US Department of Agriculture, Biological Survey, 1911).
6. Beal, F. E. L. *Food of the robins and bluebirds of the United States*. (US Department of Agriculture, 1915).
7. Beal, F. E. L. *Cuckoos and Shrikes in Relation to Agriculture*. (US Department of Agriculture, Division of Biological Survey, 1898).
8. Beal, F. E. L. *Food of the bobolink, blackbirds, and grackles*. (US Department of Agriculture, Division of Biological Survey, 1900).
9. Beal, F. E. L. *Food of our more important flycatchers*. (US Government Printing Office, 1912).
10. Beal, F. E. L., McAtee, W. L. & Kalmbach, E. R. *Common birds of southeastern United States in relation to agriculture*. (US Government Printing Office, 1941).
11. Rodewald, P. *The Birds of North America*. Cornell Laboratory of Ornithology, Ithaca, NY. https://birdsna.org (2019).
12. Johnsgard, P. A. *Ducks, geese, and swans of the world*. (University of Nebraska Press, 1978).
13. Merchant, S. & Higgins, P. J. *Handbook of Australian, New Zealand and Antarctic Birds*. (Oxford University Press: Melbourne, 1990).
14. Kear, J. & Hulme, M. *Ducks, geese and swans: Species accounts (Cairina to Mergus)*. vol. 2 (Oxford University Press, 2005).
15. Clements, J. F. *et al*. The eBird/Clements Checklist of Birds of the World: v2019. https://www.birds.cornell.edu/clementschecklist/download/. (2019).
16. Hurlbert, AH., Olsen, AM., Sawyer, MM. & Winner, PM. The Avian Diet Database as a source of quantitative information on bird diets, *Zenodo*, https://doi.org/10.5281/zenodo.5523102 (2021).

Acknowledgements

Many individuals contributed to this compilation of diet records. Specifically we would like to thank M. Auriemma, J. Barnes, M. Blawas, C. Bridges, J. Cook, R. Cooper, A. Duprey, I. Edwards, J. Fox, G. Glass, T. Hennessey, J. Knowlton, E. Kremer, W. Larsen, E. McGeady, T. McKinnon, G. Ni, L. Peeler, R. Quindlen, A. Richards, D. Ryu, M. Sandoval, C. Seyfried, S. Swinea, C. Whitener, J. Willard, and A. Zachman. M. Pan, T. Randby, and T. Le developed the web interface for exploring the database. AHH also acknowledges support from NSF DEB-1354563 and NSF EF-1702708.

Author contributions

A.H.H. designed the database, coordinated data entry, entered data, cleaned data, wrote code for data summary and analysis, and wrote the manuscript. A.M.O. entered data, cleaned data, and wrote the manuscript. P.W. and M.M.S. entered data and cleaned data, and provided comments on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.H.H.