specieshindex: How scientifically popular is a species?

Jessica Tam[1], **Malgorzata Lagisz**[1], **Shinichi Nakagawa**[1], **and Will Cornwell**[1]

1 Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia

DOI: 10.21105/joss.03776

Software

- Review
- Repository
- Archive

Editor: Hugo Gruson

Reviewers:

- @kamapu
- @tom-gu
- @DrMattG

Submitted: 21 September 2021
Published: 10 November 2021

Summary

Conservation efforts rely heavily on the body of scientific research on the relevant species. Reintroduction projects, pest eradication, breeding programs and habitat restoration, all require sound scientific knowledge for their design and implementations. Unfortunately, there is an uneven spread of research among species, taxonomic bias (Bonnet et al., 2002), such that certain species receive more research interest than the others (Donaldson et al., 2017; Santos et al., 2020; Seddon et al., 2005; Titley et al., 2017; Troudet et al., 2017). Over time, taxonomic bias in research has widened the gap between well- and poorly-studied species. Quantifying taxonomic bias is the first step towards allowing the scientific community to fill the gaps and to direct scientific efforts towards where they are most needed for conservation.

Over the last decade, the quantification of research biases have gained popularity, as shown by the growing number of studies on taxonomic bias among species of plants and animals (Adamo et al., 2021; Donaldson et al., 2017; Ducatez, 2019; Fleming & Bateman, 2016; McKenzie & Robertson, 2015; Robertson & Mckenzie, 2015; Santos et al., 2020; Schiesari et al., 2007; Tensen, 2018; Trimble & van Aarde, 2010). Most of these studies evaluated small groups of species, no more than a few hundred. Additionally, the majority of the studies used publication or citation counts to identify the bias. However, publication count can only capture the total research productivity. While citation count can capture research interest, it can be easily inflated by highly cited papers (Hirsch, 2005), without equally considering each publication. Here, we propose and implement the use of Hirsch's *h*-index (Hirsch, 2005) as a better alternative in quantifying research interest for taxonomic groups of species including genus and species.

Statement of need

`specieshindex` aims to make streamline the calculation of *h*-index in the context of measuring popularity of species in the scientific literature. The *h*-index was introduced by Hirsch (Hirsch, 2005) to compare the influence of academics with a single number (Hirsch & Buela-Casal, 2014). The *h*-index is calculated as the total number of publications (n) that have at least been cited n times, after ranking the publications in a descending order by their number of citations. The *h*-index is now also being used to measure the research interest and influence of the publications of different academic disciplines (Banks, 2006; Harzing & Alakangas, 2016), journals (Braun et al., 2006), countries (Csajbók et al., 2007), species of animals (Fleming & Bateman, 2016; McKenzie & Robertson, 2015; Robertson & Mckenzie, 2015), plants (Adamo et al., 2021), and pathogens (Cox et al., 2016). `specieshindex` calculates the *h*-index of different species and genera.

The package `specieshindex` connects to the Scopus and Web of Science databases (interdisciplinary broad-range databases of academic literature) and extracts citation records of relevant publications, as identified via a search query. It returns blibliometric information including the publication title, number of citations, publication type, etc. The binomial or genus names of the species should be used in the search query, instead of their common names, since the latter are less specific and may refer to more than 1, and even unrelated species. An example of such is 'pig,' which can refer to the domestic pig (*Sus scrofa* - an ungulate), or the guinea pig (*Cavia porcellus* - a rodent). Although this package also connects to Bielefeld Academic Search Engine (BASE), article extraction is not available, hence *h*-index calculation is unavailable with BASE.

There are two types of functions that connect to the literature databases. The count functions, e.g. `CountSpT()` and `CountSpTAK()` for species, and `CountGenusT()` and `CountGenusTAK()` for genus, return the total publication count without extracting any citation records. Whereas the fetch functions, e.g. `FetchSpT()` and `FetchSpTAK()` for species, and `FetchSpT()` and `FetchSpTAK()` for genus, extract citation records for index calculations. These functions are further distinguished by their suffixes "T" and "TAK." "T" functions only find publications with the species' name in the publication title, whereas "TAK" functions query articles' title, abstract and keywords. Apart from the *h*-index, `specieshindex` can also compute other established influence indices, including the *h5* index, the *m*-index, and the *i10* index. The *h5* index is the *h*-index of articles published in the most recent 5 years (Crotty, 2017). The *m*-index is the *h*-index divided by the number of years since the first publication (Hirsch, 2005). The *i10* index is the total number of articles with 10 or more citations; it is currently used by Google Scholar (Jacso, 2012).

Implementation

The following packages are required for `specieshindex` to work.

```
# Installation from GitHub
install.packages("rscopus")
install.packages("wosr")
install.packages("rbace")
install.packages("taxize")
install.packages("XML")
install.packages("jsonlite")
install.packages("httr")
install.packages("dplyr")
install.packages("data.table")
install.packages("tidyr")
devtools::install_github("jessicatytam/specieshindex", force = TRUE, build_vignett

# Load the library
library(specieshindex)
```

Connecting to Scopus

An API key from Scopus is required to extract citation records from their database legally. Here are the steps to obtain the key.

1. Go to https://dev.elsevier.com/ and click on the button `I want an API key`.
2. Create an account and log in.
3. Go to the `My API Key` tab on top of the page and click `Create API Key`.
4. Read the legal documents and check the boxes.

Afterwards, run the following line of code to gain access to the Scopus database:

```
# Setup API key
apikey <- "your_api_key_from_scopus"
```

Connecting to Web of Science

You do not need a key to extract data from Web of Science if you are using this package from your institution's location. To gain access to the Web of Science database, run the following line of code:

```
# Setup session ID
sid <- auth(username = NULL, password = NULL)
```

Connecting to BASE

Having a whitelisted IP address is essential when gaining access to the BASE database, which you can get on https://www.base-search.net/about/en/contact.php. A token or API key, however, is not required. Only count functions, e.g. `CountSpT()` and `CountSpTAK()`, are available as BASE does not return citation counts. Hence, index calculations will also be unavailable using this database.

Example

The following example will demonstrate the calculation of various indices with citation records from Scopus. The species' binomial name (e.g. *Sarcophilus harrisii*) or genus name (e.g. *Sarcophilus*) is required for the extraction:

```
# Extract citation data
Woylie <- FetchSpTAK(db = "scopus', genus = "Bettongia", species = "penicillata")
Quokka <- FetchSpTAK(db = "scopus', genus = "Setonix", species = "brachyurus")
Platypus <- FetchSpTAK(db = "scopus', genus = "Ornithorhynchus", species = "anatin
Koala <- FetchSpTAK(db = "scopus', genus = "Phascolarctos", species = "cinereus")
```

These four datasets are readily available within the package. An efficient way to calculate all of the indices at once is to use the function `Allindices()`.

```
# Calculate indices
W <- Allindices(data = Woylie, genus = "Bettongia", species = "penicillata")
Q <- Allindices(data = Quokka, genus = "Setonix", species = "brachyurus")
P <- Allindices(data = Platypus, genus = "Ornithorhynchus", species = "anatinus")
K <- Allindices(data = Koala, genus = "Phascolarctos", species = "cinereus")

# Combine citation records into a single dataframe
CombineSp <- rbind(W, Q, P, K)
CombineSp
```

```
##                genus_species     species           genus publications citations
## 1     Bettongia_penicillata penicillata       Bettongia          113      1903
## 2        Setonix_brachyurus  brachyurus         Setonix          242      3427
## 3 Ornithorhynchus_anatinus    anatinus Ornithorhynchus          321      6365
```

The Journal of Open Source Software

```
## 4   Phascolarctos_cinereus    cinereus   Phascolarctos            773     14291
##    journals articles reviews years_publishing  h      m i10 h5
## 1        55      110       3               43 26 0.605  54  7
## 2       107      237       5               66 29 0.439 121  4
## 3       153      308      13               67 41 0.612 177  7
## 4       227      744      29              139 53 0.381 427 14
```

Plots

This package has 2 built-in plotting functions. They are `plotAllindices()` and `plotPub()`. `plotAllindices()` plots the *h*-index, *m*-index, *i10* index, and *h5* index in the same plot.

```
# Plot the indices
plotAllindices(CombineSp)
```

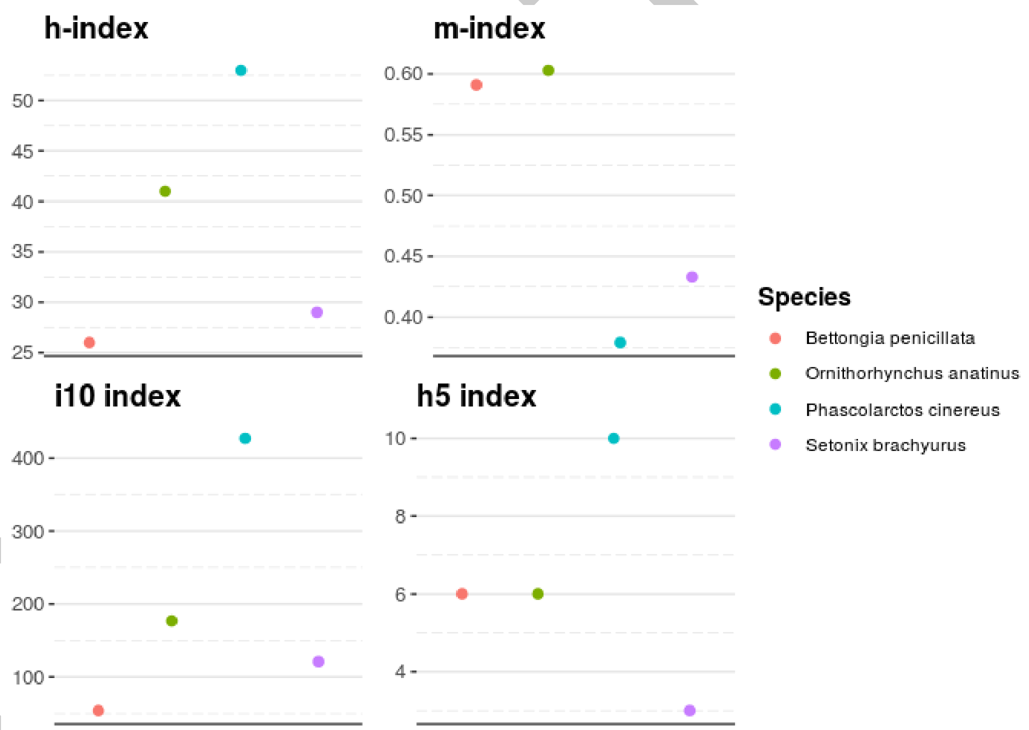

Figure 1. The *h*-index, *m*-index, *i10* index, and *h5* index of the Woylie (*Bettongia penicillata*), Platypus (*Ornithorhynchus anatinus*), Koala (*Phascolarctos cinereus*), and Quokka (*Setonix brachyurus*).

`plotPub()` plots the total publications per year for species, after extracting data on years and number of publications per year with the `getYear()` function.

```
# Extract year and frequency
extract_year_W <- getYear(data = Woylie, genus = "Bettongia", species = "penicilla
extract_year_Q <- getYear(data = Quokka, genus = "Setonix", species = "brachyurus"
extract_year_P <- getYear(data = Platypus, genus = "Ornithorhynchus", species = "a
extract_year_K <- getYear(data = Koala, genus = "Phascolarctos", species = "cinere

# Combine year and frequency into a single dataframe
Combine_pub <- rbind(extract_year_W, extract_year_Q, extract_year_P, extract_year_
```

```
# Plot the number of publications by year
plotPub(Combine_pub)
```

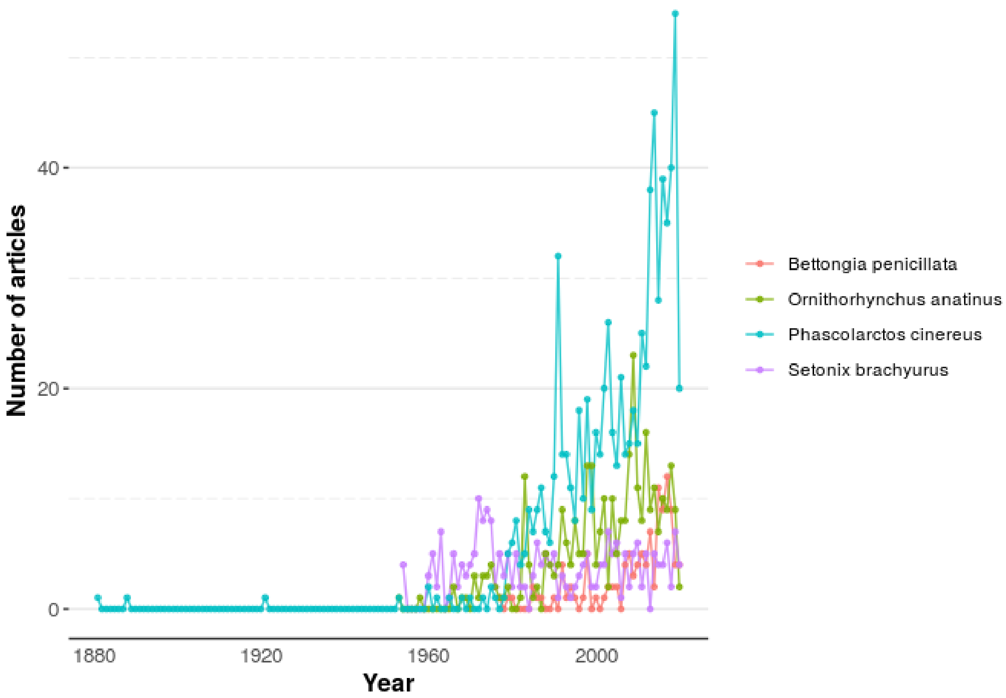

Figure 2. The total number of publications per year of the Woylie (*Bettongia penicillata*), Platypus (*Ornithorhynchus anatinus*), Koala (*Phascolarctos cinereus*), and Quokka (*Setonix brachyurus*).

Acknowledgements

I acknowledge the contributions of the authors of the dependence packages: `rscopus`, `wosr`, `rbace`, `taxize`, `XML`, `jsonlite`, `httr`, `dplyr`, `data.table`, `tidyr`. `specieshindex` is enabled by Scopus, Web of Science, and BASE (Pieper & Summann, 2006).

References

Adamo, M., Chialva, M., Calevo, J., Bertoni, F., Dixon, K., & Mammola, S. (2021). Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. *Nature Plants*, *7*. https://doi.org/10.1038/s41477-021-00912-2

Banks, M. G. (2006). An extension of the Hirsch index: Indexing scientific topics and compounds. *Scientometrics*, *69*(1), 161–168. https://doi.org/10.1007/s11192-006-0146-5

Bonnet, X., Shine, R., & Lourdais, O. (2002). Taxonomic chauvinism. *Trends in Ecology & Evolution*, *17*(1), 1–3. https://doi.org/10.1016/S0169-5347(01)02381-3

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69*(1), 169–173. https://doi.org/10.1007/s11192-006-0147-4

Cox, R., Mcintyre, K. M., Sanchez, J., Setzkorn, C., Baylis, M., & Revie, C. W. (2016). Comparison of the h-Index Scores Among Pathogens Identified as Emerging Hazards in North America. *Transboundary and Emerging Diseases*, *63*(1), 79–91. https://doi.org/10.1111/tbed.12221

Crotty, D. (2017). Other metrics: Beyond the impact factor. *European Heart Journal*, *38*(35), 2646–2647. https://doi.org/10.1093/eurheartj/ehx446

Csajbók, E., Berhidi, A., Vasas, L., & Schubert, A. (2007). Hirsch-index for countries based on Essential Science Indicators data. *Scientometrics*, *73*(1), 91–117. https://doi.org/10.1007/s11192-007-1859-9

Donaldson, M. R., Burnett, N. J., Braun, D. C., Suski, C. D., Hinch, S. G., Cooke, S. J., & Kerr, J. T. (2017). Taxonomic bias and international biodiversity conservation research. *FACETS*, *1*(1), 105–113. https://doi.org/10.1139/facets-2016-0011

Ducatez, S. (2019). Which sharks attract research? Analyses of the distribution of research effort in sharks reveal significant non-random knowledge biases. *Reviews in Fish Biology and Fisheries*, *29*. https://doi.org/10.1007/s11160-019-09556-0

Fleming, P. A., & Bateman, P. W. (2016). The good, the bad, and the ugly: Which Australian terrestrial mammal species attract most research? *Mammal Review*, *46*(4), 241–254. https://doi.org/10.1111/mam.12066

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804. https://doi.org/10.1007/s11192-015-1798-9

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Hirsch, J. E., & Buela-Casal, G. (2014). The meaning of the h-index. *International Journal of Clinical and Health Psychology*, *14*(2), 161–164. https://doi.org/10.1016/S1697-2600(14)70050-X

Jacso, P. (2012). Google scholar author citation tracker: Is it too little, too late? *Online Information Review*, *36*, 126–141. https://doi.org/10.1108/14684521211209581

McKenzie, A. J., & Robertson, P. A. (2015). Which Species Are We Researching and Why? A Case Study of the Ecology of British Breeding Birds. *PLOS ONE*, *10*(7), e0131004. https://doi.org/10.1371/journal.pone.0131004

Pieper, D., & Summann, F. (2006). Bielefeld academic search engine (BASE) an end-user oriented institutional repository search service. *Library Hi Tech*, *24*. https://doi.org/10.1108/07378830610715473

Robertson, P. A., & Mckenzie, A. J. (2015). The scientific profiles of terrestrial mammals in Great Britain as measured by publication metrics. *Mammal Review*, *45*(2), 128–132. https://doi.org/10.1111/mam.12038

Santos, J. W. dos, Correia, R. A., Malhado, A. C., Campos-Silva, J., Teles, D., Jepson, P., & Ladle, R. (2020). Drivers of taxonomic bias in conservation research: A global analysis of terrestrial mammals. *Animal Conservation*. https://doi.org/10.1111/acv.12586

Schiesari, L., Britta, G., & Grillitsch, H. (2007). Biogeographic biases in research and their consequences for linking amphibian declines to pollution. *Conservation Biology : The Journal of the Society for Conservation Biology*, *21*, 465–471. https://doi.org/10.1111/j.1523-1739.2006.00616.x

Seddon, P. J., Soorae, P. S., & Launay, F. (2005). Taxonomic bias in reintroduction projects. *Animal Conservation*, *8*(1), 51–58. https://doi.org/10.1017/S1367943004001799

Tensen, L. (2018). Biases in wildlife and conservation research, using felids and canids as a case study. *Global Ecology and Conservation*, *15*, e00423. https://doi.org/10.1016/j.gecco.2018.e00423

Titley, M. A., Snaddon, J. L., & Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLOS ONE*, *12*(12), e0189577. https://doi.org/10.1371/journal.pone.0189577

Trimble, M., & van Aarde, R. (2010). Species inequality in scientific study. *Conservation Biology : The Journal of the Society for Conservation Biology*, *24*, 886–890. https://doi.org/10.1111/j.1523-1739.2010.01453.x

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-09084-6