## SUPPORTING INFORMATION

## Is there really such a thing as *Tropical* Biology?

Emilio M. Bruna 1,2

1 Department of Wildlife Ecology and Conservation, University of Florida, PO Box 110430, Gainesville, FL 32611-0430, USA

2 Center for Latin American Studies, University of Florida, PO Box 115530, Gainesville, FL 32611-5530, USA

Corresponding author; email: [embruna@ufl.edu](mailto:embruna@ufl.edu).

## 1. Collection, processing, and visualization of bibliometric data

To identify the conceptual domains studied by researchers working in ‘Tropical’ and “non-Tropical’ locations, I used information extracted from the bibliographic records of articles published from 1990-2022 in N = 10 journals (*Journal of Evolutionary Biology, Ecology, Journal of Applied Ecology, Evolution, Biotropica, Journal of Ecology, Tropical Conservation Science, American Naturalist, Tropical Ecology, Journal of Tropical Ecology, Revista de Biologia Tropical*). Specifically, I extracted and summarized the information from two structural components used by authors to describe the subject of their articles: the title and keywords. These provide distinct but complementary information, and so they are often analyzed both independently and in unison. Below I describe how the article records were identified, downloaded, processed, and assigned to the ‘Tropical’ and”non-Tropical’ categories using code written in the R programming language (R Core Team 2023).

On 8 February 2023, I downloaded all bibliographic data available in SCOPUS and the Web of Science ‘Core Collection’ for all articles published in the focal journals; both SCOPUS and the Web of Science were queried because they differ in the years indexed for each journal. I then used the refsplitr package (Fournier *et al.* 2020) to process the records and remove any duplicates. After removing all stopwords (Benoit *et al.* 2021) from article titles and keywords, I spell-checked, stemmed, and lemmatized all of the keywords and title words and extracted the bigrams (i.e., pairs of sequential words, e.g., *seed predation*, *species diversity*) from titles with the tidytext library (Silge & Robinson 2016). Finally, I identified each article as either ‘Tropical’ or ‘non-Tropical’; all articles published in (*Journal of Evolutionary Biology, Ecology, Journal of Applied Ecology, Evolution, Biotropica, Journal of Ecology, Tropical Conservation Science, American Naturalist, Tropical Ecology, Journal of Tropical Ecology, Revista de Biologia Tropical*) were assigned to the ‘Tropical’ category, while articles published in the other journals were assigned to one of these categories based on a search of the titles, keywords, or abstracts for a list of domain-specific terms (e.g., tropical: *amazon*, *andes*, *congo*, *bci*, *chamela*; non-tropical: *finland*, *boreal*, *eastern decid*, *arctic*, *polar*). These procedures resulted in N = 37,807 total articles published, of which N = 11,210 reported research conducted in the tropics and N = 26,597 were based on work conducted in other locations. Collectively, the N = 41,219 contained a total of N = 126,796 bigrams. Not all of the articles included keywords, however; from the the N = 37,807 that did I was able to extract a total of N = 62,883. There were N = 41,142 articles from which I was able to extract both title bigrams and keywords. I used these sets of articles to conduct three geographic comparisons: (1) title bigrams, (2) keywords, and (3) title bigrams + keywords (hereafter, ‘terms’).

The number of articles varies widely between journals, as does the number of keywords per article or title length. Comparing counts of keyword, bigram, or term frequency in tropical and non-tropical articles could therefore bias results towards the content published a journals allowing more keywords or journals publishing more articles. To correct for this, I calculated the percentage of articles in each geographic category that used each keyword, title bigram, or term. I then selected the N = 50 most frequently used in each geographic category, and identified (a) any keywords, bigrams, or terms that ‘tropical’ and ‘non-tropical’ articles had in common, and (b) any keywords, bigrams, or terms that were unique to each article category.

## 2. Data and Code

The data used in this publication are available at Dryad <*DOI added upon acceptance*>. The code used to import, organize, and analyze these data, along with the code for preparing the figures, tables, and manuscript, are available at Zenodo <*DOI added upon acceptance*>.  
The data used in this paper are part of a larger data set collected for a longitudinal study of research in the tropics. That data set and the code used to harvest, clean, and organize it are available at Github <https://github.com/BrunaLab/tropical_bibliometrics>. Questions regarding the data or code, or suggestions for improvement should be posted as Issues on that repository or referred to E. M. Bruna. That repository also includes a NEWS.md file summarizing any post-publication updates.

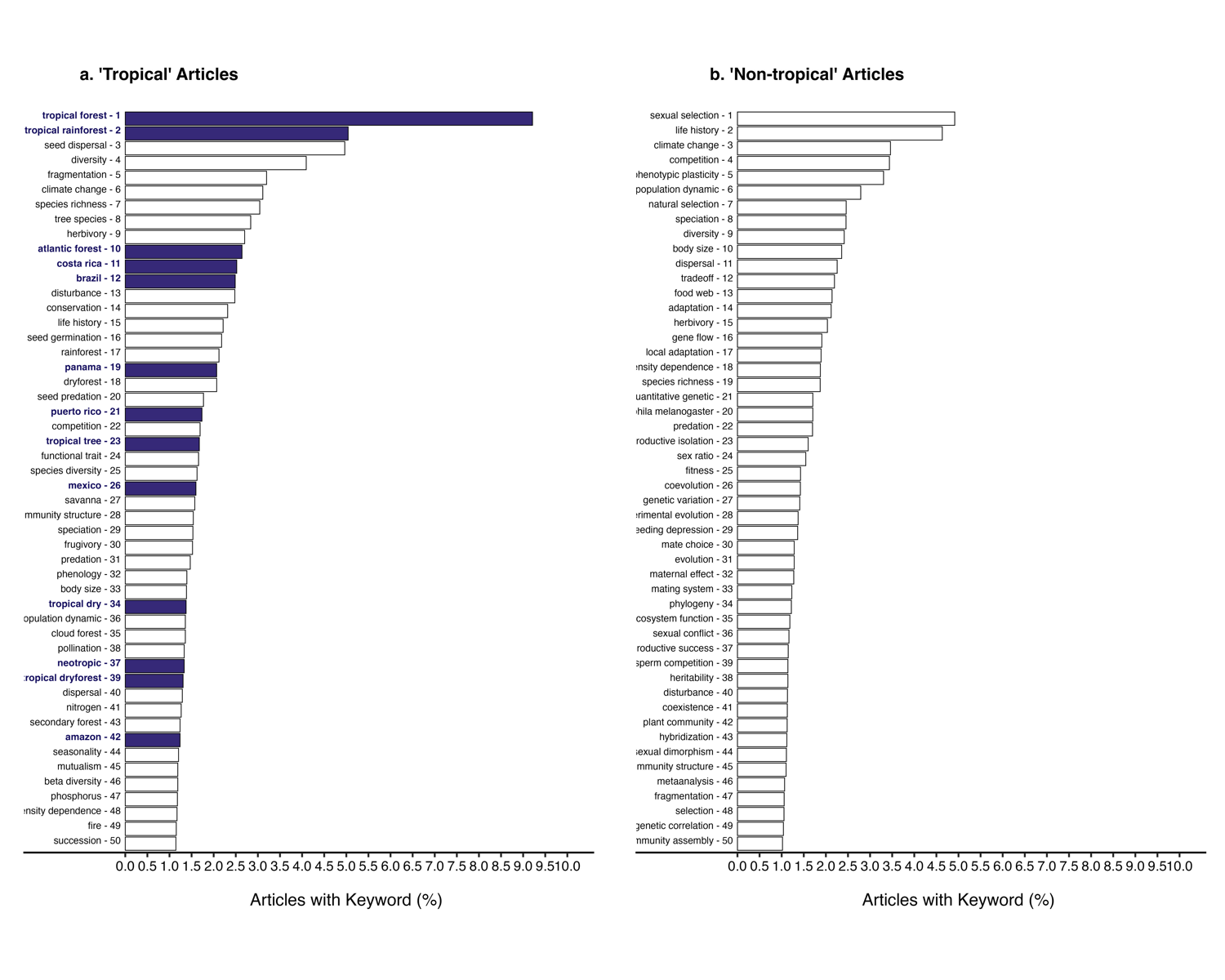
## REFERENCES

Benoit, K., D. Muhr, and K. Watanabe. 2021. Stopwords: Multilingual stopword lists. <https://CRAN.R-project.org/package=stopwords>

Fournier, A. M. V., M. E. Boone, F. R. Stevens, and E. M. Bruna. 2020. Journal of Open Source Software 5: 2028.

R Core Team. 2023. R: {A} language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Silge, J., and D. Robinson. 2016. . Journal of Open Source Software 1(3).



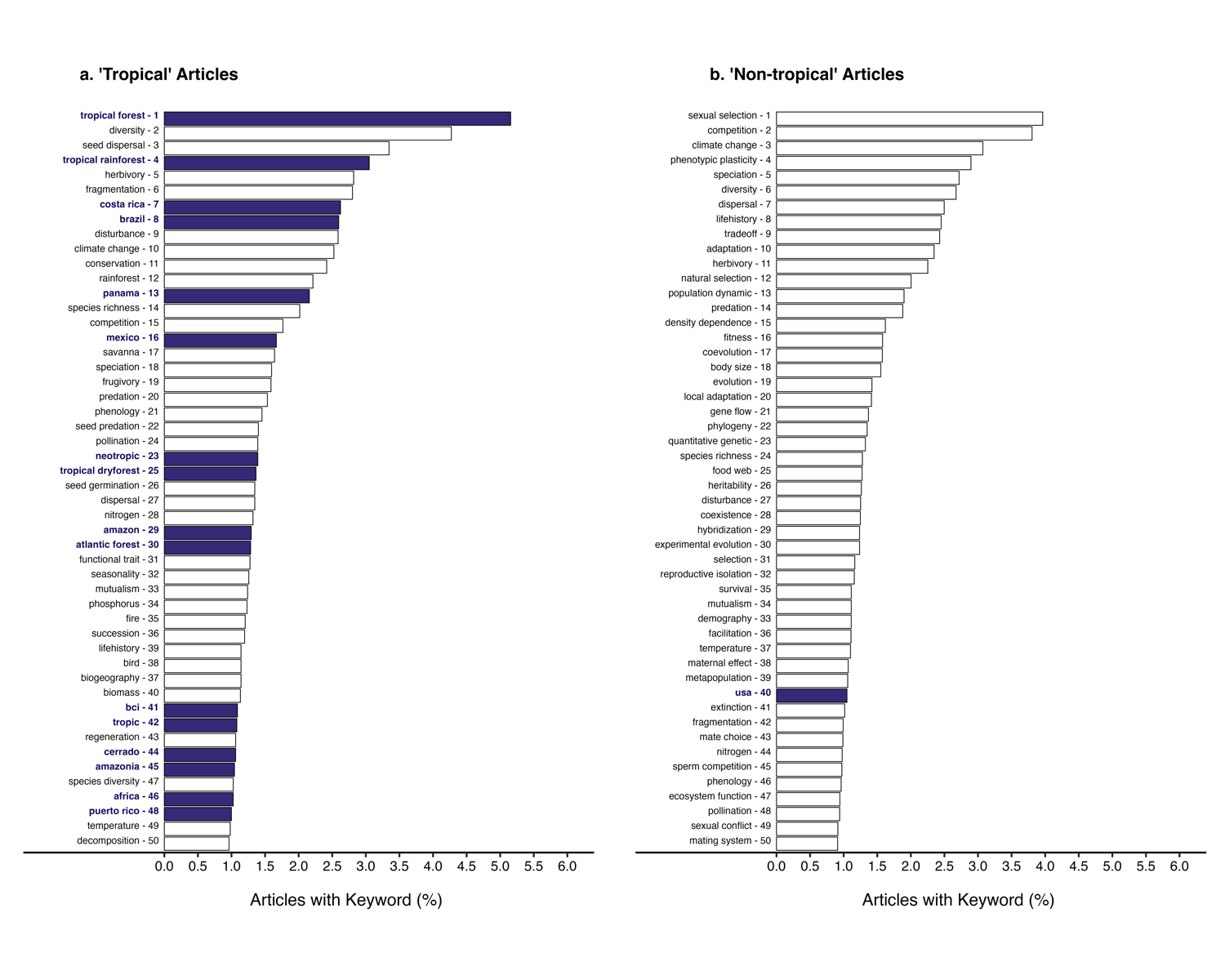


Figure 5: The N = 50 most common keywords from articles based on research conducted in (a) the tropics and (b) non-tropical regions. The rank of these words is based on the percentage of articles in each category that included them. Terms reflecting geography (e.g., ) are indicated in bold and with filled bars.

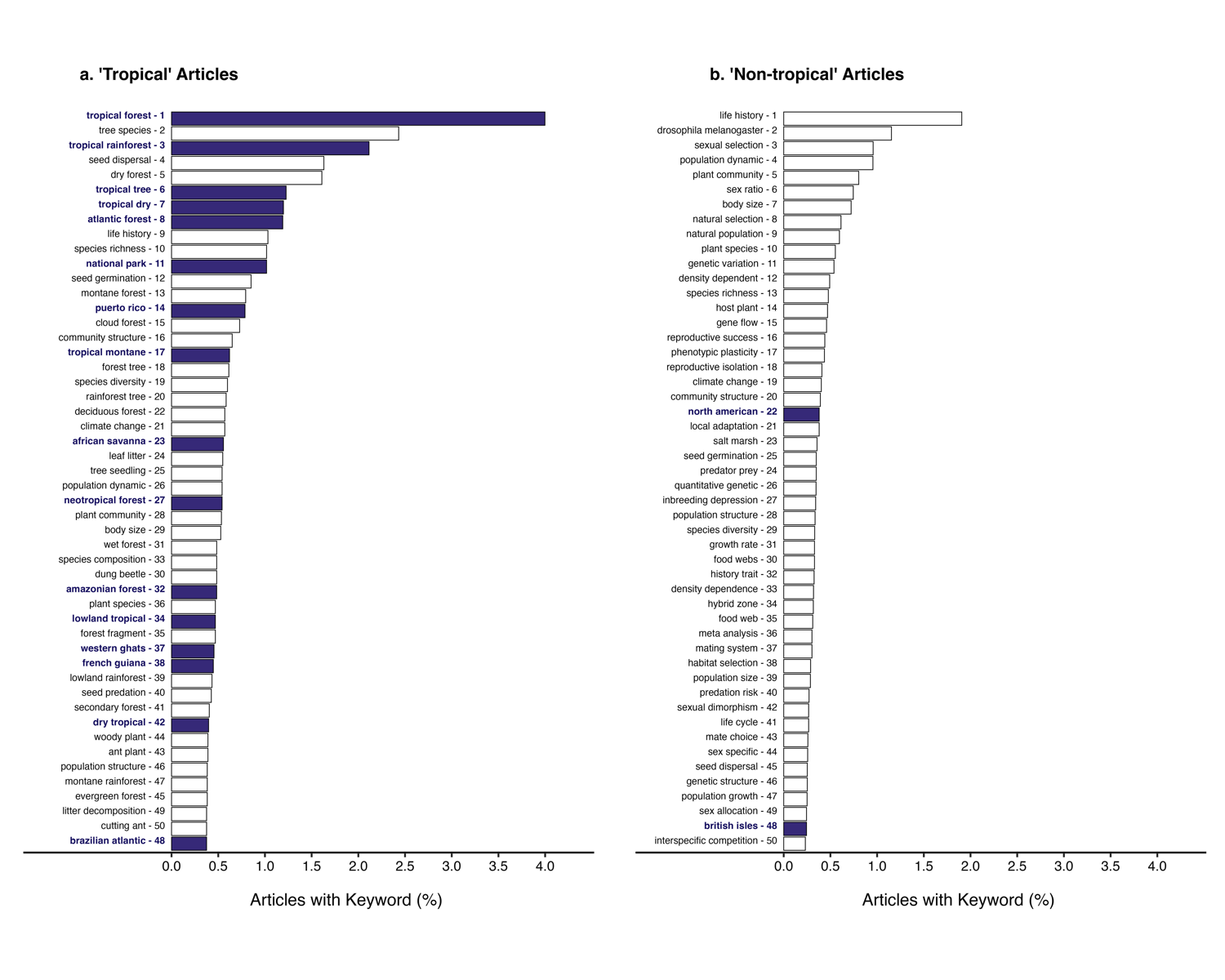
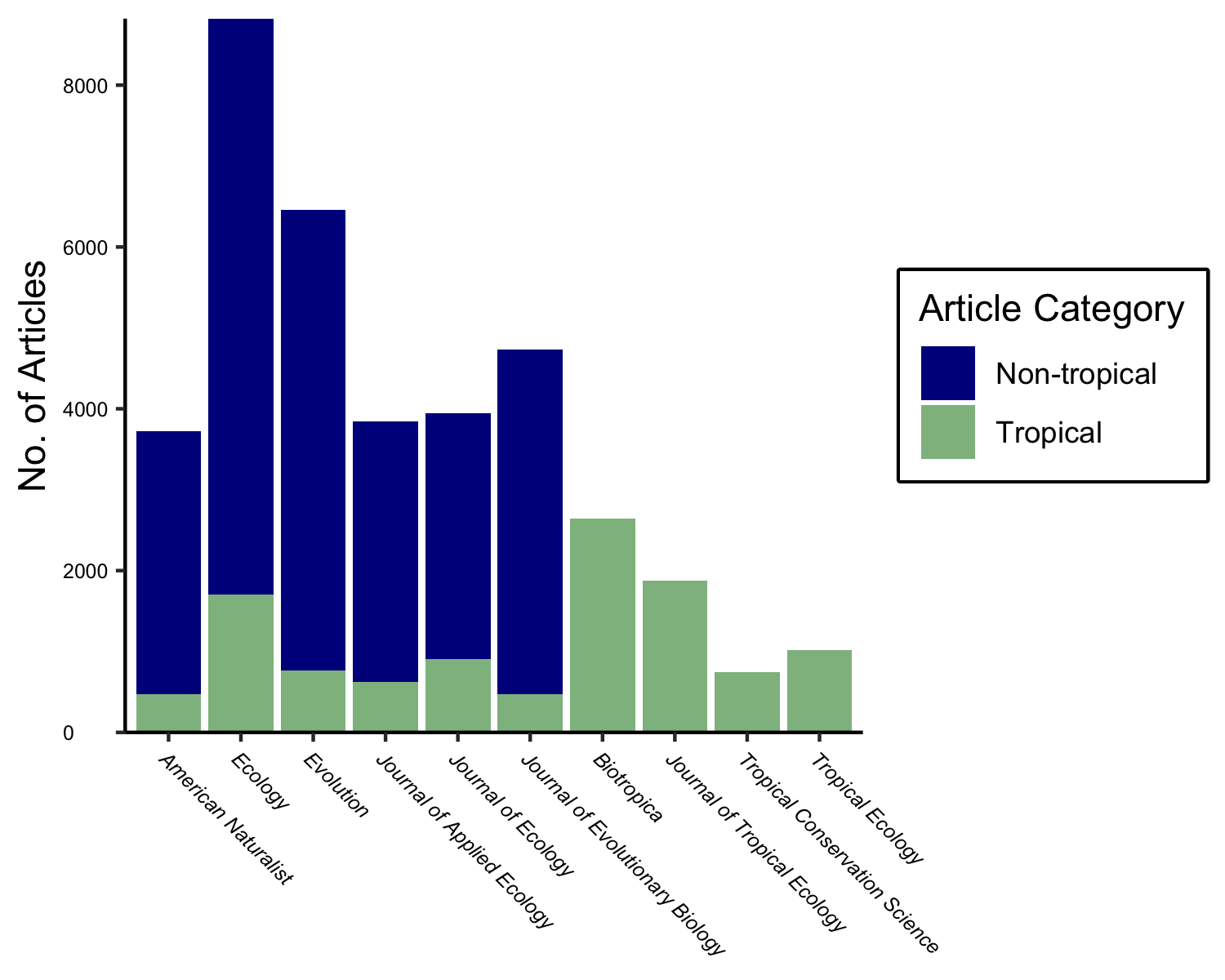


Figure 6: The N = 50 most common bigrams in titles of articles based on research conducted in (a) the tropics and (b) non-tropical regions. The rank of these words is based on the percentage of article titles in each category that included those words. Bigrams reflecting geography (e.g., ) are indicated in bold and with filled bars.



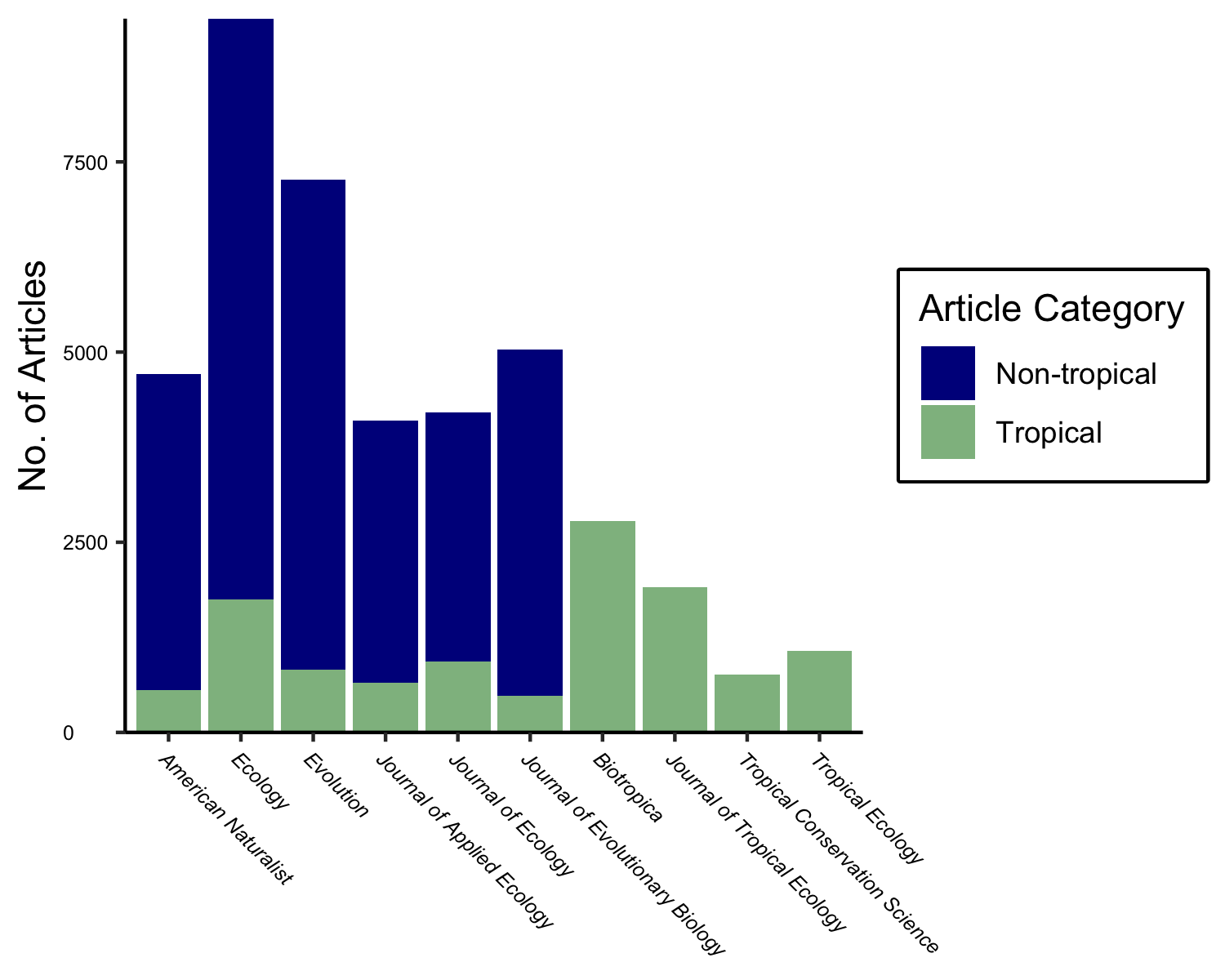


Figure 8: The number of articles from each journal and geographic category that were used in the analysis of title words and title bigrams.