



## Case Report

# An online version and some updates of R package U.Taxonstand for standardizing scientific names in plant and animal species

Jian Zhang <sup>a,\*</sup>, Hong Qian <sup>b</sup>, Xinyang Wang <sup>c</sup>

<sup>a</sup> School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China

<sup>b</sup> Research and Collections Center, Illinois State Museum, 1011 East Ash Street, Springfield, IL 62703, USA

<sup>c</sup> School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China

## ARTICLE INFO

## Article history:

Received 25 July 2024

Received in revised form

8 September 2024

Accepted 9 September 2024

Available online 12 September 2024

## Keywords:

Biodiversity informatics

Species name matching

Taxonomic harmonization

Taxonomic tool

U.Taxonstand

We are now in the era of big data for biodiversity science. More and larger datasets on species geographic distributions, community composition, and functional traits are now becoming more increasingly than ever before. Correctly applying taxon names is a prerequisite for robust biodiversity studies of all taxonomic groups. It is very common that the same species may be documented using different scientific names in different data sources or have several synonyms in the same database. It is also common that a substantial proportion of taxon names in a species list cannot be directly matched with any names in a selected database due to spelling errors or variants. To standardize and harmonize scientific names of both plants and animals, we developed an R package 'U.Taxonstand' ('U' stands for 'universal'; 'Taxonstand' stands for 'taxonomic standardization') to resolve these issues at a high rate of matching success and a fast execution speed (Zhang and Qian, 2023). Since the R package was published in 2023, it has been cited over 60 times, and become a very useful tool for botanists, zoologists, ecologists and biogeographers.

Although the R package U.Taxonstand is easy to use for most researchers with some basic programming skills in R language, it may still need to take some time to install the package, format species lists,

or be familiar with several basic R functions (e.g., nameMatch) and their parameters, especially for the users who know little about R. Online tools for species name matching could provide such a solution for the users with the limited ability to access and standardize scientific names. However, only a few online tools are available now. For example, the Taxonomic Name Resolution Service (TNRS; <https://tnrs.biendata.org/>) can only match for vascular plants using two backbone databases (Boyle et al., 2013); Plantminer (<http://www.plantminer.com>; Carvalho et al., 2010) can only match with The Plant List, which hasn't been updated since 2013 and thus has become outdated. Therefore, it could be very handfull to provide an online version of the R package U.Taxonstand to meet the needs for name matching of different data sources and different taxonomic groups. Here we present 'U.Taxonstand Online' (<https://ecoinfor.shinyapps.io/UTaxonstandOnline>), a user-friendly web application developed utilizing R Shiny platform. In addition, we introduce some updates for the R package U.Taxonstand, including several new or revised R functions to clean and format the data inputs.

## 1. U.Taxonstand Online

The current version of 'U.Taxonstand Online' serves as a simple tool to standardize scientific names of seven taxonomic groups, including seed plants, bryophyte, amphibians, birds, mammals, reptiles, and fishes. For seed plants, five global databases (e.g., World Flora Online, and World Checklist of Vascular Plants) and one regional database (List of plant species in China - 2024 Edition) are provided for direct uses. For bryophyte, amphibians, mammals, and reptiles, the data from the Catalogue of Life Checklist are provided currently. The data sources of all these databases are listed in the webpage "Data Source & Citation" of the online application. To use this web application, you need to follow four steps (Fig. 1):

### 1) Choose one backbone database

Select the appropriate backbone database for your taxon name list. The dropdown list includes currently available databases to match with. This list might be expanded and updated when new datasets or new versions of these datasets are available. All the data

\* Corresponding author.

E-mail address: [zhangjian6@mail.sysu.edu.cn](mailto:zhangjian6@mail.sysu.edu.cn) (J. Zhang).

Peer review under the responsibility of Editorial Office of Plant Diversity.

**Step 1. Choose one backbone database to match with**  
Please make sure that you select the right database.  
More information in the page of the left sidebar 'Data Source and Citation'

Plants\_WFO (World Flora Online)

**Step 2a. Load your taxon name list (.csv or .xlsx)**  
The datafile should have three columns ('Name', 'Author', 'Rank'). The first column is required, and others are optional.

Browse... No file selected

Download CSV Download XLSX

**Step 2b. Paste your taxon names with or without authors**  
One name per line.

Hieracium aragonense Scheele  
Pilosella X subulatifolia (Zahn) Mateo  
Hieracium giganteum Sleumer  
Jurinea kapelkinki O.Fedtsch.  
Hieracium hypochaeroides Gibson  
Senecio marini Cabrera  
Hymenostephium tenue (A.Gray) E.E.Schill. & Panero  
Stevia pyrolifolia Schitdl.  
Vernonia morelana Gleason  
Mikania melissifolia (Lam.) Willd.

Download CSV Download XLSX

**Step 3. Processed taxa**  
Click the button 'RUN' and wait for the results showing here.

Show 10 entries

SORTER	Submitted_Name	Submitted_Author	Submitted_Genus	Submitted_Rank	Name_in_database	Author_in_database	Genus_in	
4	1	Hieracium aragonense	Scheele	Hieracium	2	Hieracium aragonense	Scheele	Hieracium
20	2	Pilosella X subulatifolia	(Zahn) Mateo	Pilosella	2	Pilosella subulatifolia	(Zahn) Mateo	Pilosella
6	3	Hieracium giganteum	Sleumer	Hieracium	2	Hieracium giganteum	Sleumer	Hieracium
27	4	Jurinea kapelkinki	O.Fedtsch.	Jurinea	2	Jurinea kapelkinki	O.Fedtsch.	Jurinea
23	5	Hieracium hypochaeroides	Gibson	Hieracium	2	Hieracium hypochaeroides	S.Gibson	Hieracium
32	6	Senecio marini	Cabrera	Senecio	2	Senecio marini	Cabrera	Senecio
8	7	Hymenostephium tenue	(A.Gray) E.E.Schill. & Panero	Hymenostephium	2	Hymenostephium tenue	(A.Gray) E.E.Schill. & Panero	Hymenos
13	8	Stevia pyrolifolia	Schitdl.	Stevia	2	Stevia pyrolifolia	Schitdl.	Stevia
15	9	Vernonia morelana	Gleason	Vernonia	2	Vernonia morelana	Gleason	Vernonia
10	10	Mikania melissifolia	(Lam.) Willd.	Mikania	2	Mikania melissifolia	(Lam.) Willd.	Mikania

Showing 1 to 10 of 30 entries

Previous 1 2 3 Next

**Step 4. Download the data**  
Click the button 'Download CSV' to save the results as CSV file.  
Click the button 'Download XLSX' to save the results as XLSX file.

Fig. 1. The screenshot of the web application 'U.Taxonstand Online'.

sources and related citation information are provided in the section "Data Source & Citation" of the web application.

## 2) Upload or paste taxon names

You have two options to upload the taxon names. One is to upload a data file containing a list of names in CSV or XLSX format. The data format is the same with the requirement of R package U.Taxonstand (Zhang and Qian, 2023). Another option is to directly paste taxon names into the text area (Fig. 1). The names could include authors or not. If the authors are included, the tool will format it using the R function "nameSplit" of U.Taxonstand automatically.

## 3) Run the matching process and check the results

Once you have uploaded or pasted the names, click the "Run" button to initiate the matching process. The results will be displayed in a table format, showing the standardized names and relevant details (Fig. 1).

## 4) Download the results in desired format

After reviewing the results, you can download them in your preferred file format. The available options are CSV and XLSX.

## 2. Some updates for U.Taxonstand

Since the early version of the R package U.Taxonstand was released, we have received a lot of feedbacks from users with different background. Based on these feedbacks and some bugs we detected, we make some improvements for U.Taxonstand.

- (1) The function 'nameMatch': This is the main function of our package. We added two arguments 'matchFirst' and 'Append'. By default, the 'matchFirst' only keeps the first 'BEST' matching result for each input taxon. If 'Append' is true, the matching results will add other columns (e.g., geographic distribution and common names in different language) in the backbone database.
- (2) The functions 'nameClean' and 'nameSplit': These two functions can use as both a subfunction of the main function

'nameMatch' and a separate function for data formatting (e.g., extract the genus information from species names, and separate the authority from binominal names).

- (3) The function 'familyMatch': This new function is used to add family information based on genus list if missing in the input data file.
- (4) The functions 'nameMatch\_WCVP' and 'nameMatch\_LPSC': For plants, the team of the World Checklist of Vascular Plants (WCVP) has been released R packages "rWCVP" and "rWCVPdata" (Brown et al., 2023), providing an easier way to use the newest version of the WCVP. The function 'nameMatch\_WCVP' serves to link these packages with U.Taxonstand. Similarly, the function 'nameMatch\_LPSC' serves to make the connection with R package 'LPSC' (Zhang, 2023) and the Checklist of Plant Species in China (2023 Edition).
- (5) The function 'synSearch' lists all the synonyms and accepted names for the input names. Note that it only works for the matched taxon names.

## 3. Citation

Because U.Taxonstand Online uses the functions reported in the R package U.Taxonstand (Zhang and Qian, 2023), when a user uses U.Taxonstand Online to standardize taxon names, both the current article and the article of Zhang and Qian (2023) should be cited.

## 4. Summary

By offering a web-based platform that simplifies the standardization and harmonization of scientific names across diverse plant and animal taxa, 'U.Taxonstand Online' serves as a valuable and dynamic resource for biodiversity studies. As a byproduct of R package U.Taxonstand (Zhang and Qian, 2023), the online tool reduces the efforts of downloading R packages and backbone databases, and learning some basic programming skills, although running U.Taxonstand Online may take a bit more computer time than running U.Taxonstand. Furthermore, it is worth noting that all tools for name matching (Grenié et al., 2023) should be used semi-automatically rather than automatically. All the users must check results and apply their judgement instead of blindly accepting matched results, particularly for those names that are matched by

fuzzy matching function (i.e. those with “TRUE” in the column “Fuzzy” in the file resulting from U.Taxonstand).

### CRediT authorship contribution statement

**Jian Zhang:** Writing – review & editing, Writing – original draft, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Hong Qian:** Writing – review & editing, Supervision, Conceptualization. **Xinyang Wang:** Methodology, Investigation, Formal analysis.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work was supported by the Innovation Program of Shanghai Municipal Education Commission (No. 2023ZKZD36).

### References

- Boyle, B., Hopkins, N., Lu, Z., et al., 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* 14, 16.
- Brown, M.J., Walker, B.E., Black, N., et al., 2023. rWCVP: a companion R package for the World checklist of vascular plants. *New Phytol.* 240, 1355–1365.
- Carvalho, G.H., Cianciaruso, M.V., Batalha, M.A., 2010. Plantminer: a web tool for checking and gathering plant species taxonomic information. *Environ. Model. Softw.* 25, 815–816.
- Checklist of Plant Species in China, 2023 Edition, 2023. Plant Data Center of Chinese Academy of Sciences. <https://doi.org/10.12282/plantdata.1390>. CSTR: 34735.11.PLANTDATA.1390.
- Grenié, M., Berti, E., Carvajal-Quintero, J., et al., 2023. Harmonizing taxon names in biodiversity data: a review of tools, databases and best practices. *Methods Ecol. Evol.* 14, 12–25.
- List of plant species in China, 2024. Plant Data Center of Chinese Academy of Sciences (CSTR), 2024 Edition. 34735.11.PLANTDATA.1476.
- Zhang, J., Qian, H., 2023. U.Taxonstand: an R package for standardizing scientific names of plants and animals. *Plant Divers.* 45, 1–5.
- Zhang, J., 2023. LPSC: tools for searching List of plant species in China. R package version 0.8.1. <https://github.com/helixcn/LPSC>.