

# taxize: taxonomic search and retrieval in R

Eduard Szöcs<sup>1</sup> and Scott A. Chamberlain<sup>2</sup>  
<sup>1</sup>University of Koblenz-Landau, <sup>2</sup>Simon Fraser University



## Summary

**Taxize** is a R package that provides an interface to various taxonomic data sources around the web. The functionality of taxize facilitates the data cleaning step before a statistical analysis and eases handling of taxonomic data in R.

## Data Sources

Taxize currently provides simple and programmatic access to taxonomic data from 14 data sources around the web.



## Features (cont.)

### Retrieve children taxa

Another common task is to retrieve the complete taxonomic hierarchy for a taxon:

```
downstream('Baetis', db = 'col', downto = 'Species')
##      childtaxa_name childtaxa_rank
## 1      Baetis acceptus           Species
## 2      Baetis aculeatus           Species
## 3      Baetis acuminatus           Species
## 4      Baetis adonis             Species
## 5      Baetis aeneus             Species
```

### Aggregate data to a specific taxonomic rank

Using the taxonomic information taxa can be easily aggregated to different levels, e.g. to study effects on different taxonomic levels. This is provided via the `tax_agg()` function:

```
tax_agg(dune, rank = 'family', db = 'ncbi')
```

### Match tables with different taxonomic resolution

Using the taxonomic information taxa can be easily aggregated to different levels, e.g. to study effects on different taxonomic levels. This is provided via the `tax_agg()` function:

```
tax_agg(dune, rank = 'family', db = 'ncbi')
```

## Features

### Resolve taxonomic names

We often have a list of species names and we want to know  
a) if we have the most up-to-date names,  
b) if our names are spelled correctly,  
c) and the scientific name for a common name.  
Taxize provides an interface to the EOL Global Names Resolver and Taxonomic Name Resolution Service, e.g.

```
gnr_resolve('Baetis roodani')
##      submitted_name      matched_name
## 1 Baetis roodani Baetis rhodani
```

### Retrieve higher taxonomic names

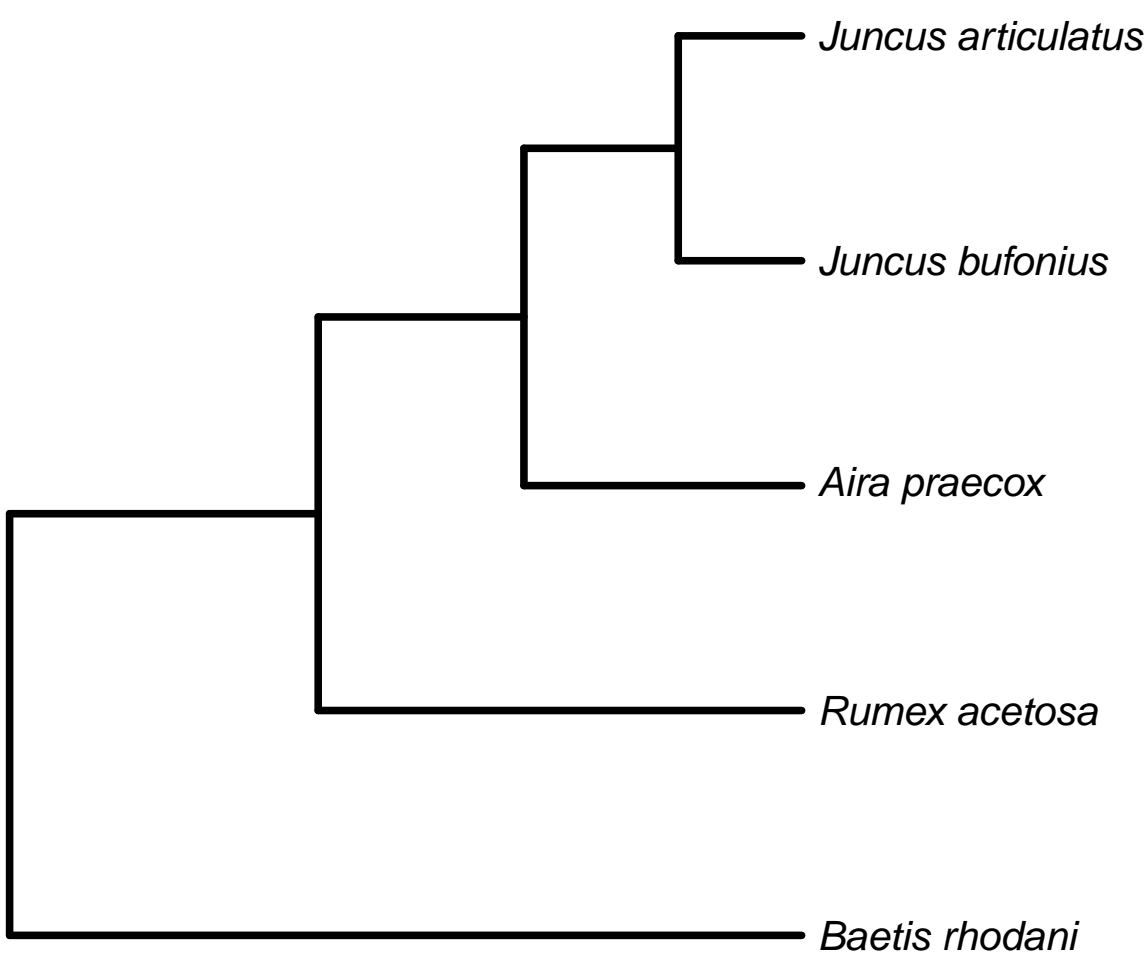
One can also search in the opposite direction, i.e. search species within a genus:

```
classification('Baetis rhodani', db = 'col')
##      name      rank
## 1  Animalia  Kingdom
## 2  Arthropoda  Phylum
## 3   Insecta   Class
## 4 Ephemeroptera Order
## 5   Baetoidea Superfamily
## 6   Baetidae  Family
## 7    Baetis   Genus
## 8 Baetis rhodani Species
```

### Building taxonomic trees

Using this taxonomic information we can build taxonomic trees. The can be used a surrogates when phylogenetic data is scarce.

```
species <- c("Juncus bufonius", "Juncus articulatus",
            "Aira praecox", "Rumex acetosa", "Baetis rhodani")
hier <- classification(species, db = 'ncbi')
plot(class2tree(hier))
```



## Under the hood

## Get involved!

taxize is currently developed collaboratively in github. Feature requests, bug reports and contributions are strongly encouraged!



<https://github.com/ropensci/taxize>

