Updating and checking the taxonomy

6 Oct 2015

The following R code illustrates how to update the taxonomy used in COMADRE using the R package **taxize** (Chamberlain & Szöcs 2013) to query the Catalogue of Life (http://www.catalogueoflife.org) with the names used by the source publication authors in COMADRE.

The function `classification` in `taxize` allows the programmatic querying of databases including the Catalogue of Life. For example:

```
library(taxize)
```

```
classification("Daphnia magna", db = 'col')
```

```
>
> Retrieving data for taxon 'Daphnia magna'

> $`Daphnia magna`
>               name     rank                               id
> 1       Animalia  Kingdom 5ede24b0534ebd5e1f552d5b9f874a6a
> 2     Arthropoda   Phylum 89ac18bfcf1654a9662a600ba06bb494
> 3  Branchiopoda    Class 1793fd966478a2d1c959221e2a711569
> 4   Diplostraca    Order 04df00f2834eac8dc22966b299226a13
> 5  Not assigned   Family                         24298066
> 6        Daphnia    Genus 1acee7e4189fdad4c8abf00900254b85
> 7  Ctenodaphnia Subgenus                         24532001
> 8 Daphnia magna  Species e44e0a35c2cdb085859074a9fe73abdf
>
> attr(,"class")
> [1] "classification"
> attr(,"db")
> [1] "col"
```

One can capitalise on this for COMADRE.

Then, the required package, `taxize`.

```
library(taxize)
```

Firstly, make a new vector called `SpeciesBinomial` by concatenate the accepted genus (`GenusAccepted`) and accepted species epithet (`SpeciesEpithetAccepted`) together. This is necessary, rather than simply using `SpeciesAccepted`, because `SpeciesAccepted` retains the infra-specific information which is not used by the following code:

```
comadre$metadata$SpeciesBinomial <-
paste(comadre$metadata$GenusAccepted, comadre$metadata$SpeciesEpithetAccepted)
```

Some species do not have an epithet (e.g. *Tribolium* sp.), for these the epithet is listed as `NA`. Therefore, to search the Catalogue of Life effectively, the `NA` needs to be removed using `gsub`:

```
comadre$metadata$SpeciesBinomial <- gsub("NA", "",
                                    comadre$metadata$SpeciesBinomial)
```

Because species appear in the database numerous times it is advisable for efficiency reasons to make a unique subset of the data:

```
temp <- unique(comadre$metadata[, c("SpeciesBinomial",
"GenusAccepted", "Family", "Order", "Class", "Phylum", "Kingdom")])
```

This dataset is still quite large (333 rows), so here I will just obtain the information for the first 5 rows.

```
temp <- temp[1:5,]
```

This is accomplished using the `classification` function which repeatedly queries the Catalogue of Life for each entry. Note that whenever there is an uncertainty as to which species is intended, the software prompts the user to select a species from a list.

```
x <- classification(temp$SpeciesBinomial,db='col')
```

```
>
> Retrieving data for taxon 'Acipenser fulvescens'
>
>
> Retrieving data for taxon 'Acipenser transmontanus'
>
>
> Retrieving data for taxon 'Ambloplites rupestris'
>
>
> Retrieving data for taxon 'Ammocrypta pellucida'
>
>
> Retrieving data for taxon 'Amphiprion percula'
```

The object produced by this code is a `list` with elements corresponding to each species. Thus `x[[1]]` will return the first species etc. One can also address the list by element name, which is defined by the query that was submitted e.g. `x[["Acipenser fulvescens"]]`

```
x[[1]]
```

```
>                    name     rank                               id
> 1              Animalia  Kingdom 5ede24b0534ebd5e1f552d5b9f874a6a
> 2              Chordata   Phylum 4313bc7637e1fc1feb316a4dea2b668b
> 3         Actinopterygii    Class 655300e5b96a3dc5e76e02175c3aa191
> 4      Acipenseriformes    Order 1b6074738c3c562418eee137a0369bdb
> 5         Acipenseridae   Family 5a7e5d2984e865f1aff640138c760d04
> 6             Acipenser    Genus b506984ec1a4c2f6191b842dcbce8fa7
> 7 Acipenser fulvescens  Species ebc713afa86464a0cfb313955a1db149
```

Thus, to acquire the taxonomic Order of the species of interest one would use:

```
x[[1]]$name[x[[1]]$rank == "Order"]
```

```
> [1] "Acipenseriformes"
```

One can run through this list object in a loop to compare the information held in COMADRE, with that produced by the query to Catalogue of Life. For example, to check the Orders of the 5 species held in `temp`

```
for(i in 1:nrow(temp)){
n <- temp$SpeciesBinomial[i]
dat <- x[[n]]
CoL_Order <- dat$name[dat$rank == "Order"]
if(temp$Order[i]==CoL_Order){temp$OrderOK[i] <- "OK"} else {temp$OrderOK[i] <- "Differs!"}
}
```

which will add an extra column called `OrderOK` to the `temp` data frame:

```
temp
```

```
>             SpeciesBinomial GenusAccepted        Family          Order
> 1       Acipenser fulvescens       Acipenser Acipenseridae Acipenseriformes
> 14 Acipenser transmontanus       Acipenser Acipenseridae Acipenseriformes
> 15    Ambloplites rupestris    Ambloplites Centrarchidae     Perciformes
> 18     Ammocrypta pellucida     Ammocrypta       Percidae     Perciformes
> 20       Amphiprion percula     Amphiprion Pomacentridae     Perciformes
>             Class    Phylum   Kingdom OrderOK
> 1  Actinopterygii Chordata Animalia      OK
> 14 Actinopterygii Chordata Animalia      OK
> 15 Actinopterygii Chordata Animalia      OK
> 18 Actinopterygii Chordata Animalia      OK
> 20 Actinopterygii Chordata Animalia      OK
```

This approach can naturally be repeated this for each part of the taxonomy.