An Introduction to the taxize package

The taxize authors

October 14, 2012

Contents

TODO: Introduction to package, explanation of function names,...

1 Phylomatic

1.1 Fetch a phylogenetic tree from Phylomatic

```
require(taxize)

## Loading required package:  taxize
## Loading required package:  rentrez
## Loading required package:  RCurl
## Loading required package:  bitops
## Loading required package:  XML
## Loading required package:  ritis

mytree <- "asteraceae/bidens/bidens_cernua\nasteraceae/madia/madia_exigua\nasteraceae/helian
```

1.1.1 Return a tree in the phylo object format

```
phylomatic_tree(mytree, "TRUE", "GET", "xml", "TRUE")

##
## Phylogenetic tree with 3 tips and 2 internal nodes.
##
## Tip labels:
```

```
## [1] "bidens_cernua"     "madia_exigua"      "helianthus_annuus"
##  Node labels:
## [1] "euphyllophyte" ""
##
## Rooted; includes branch lengths.
```

1.1.2 Return a tree in the phylo object format, using POST method

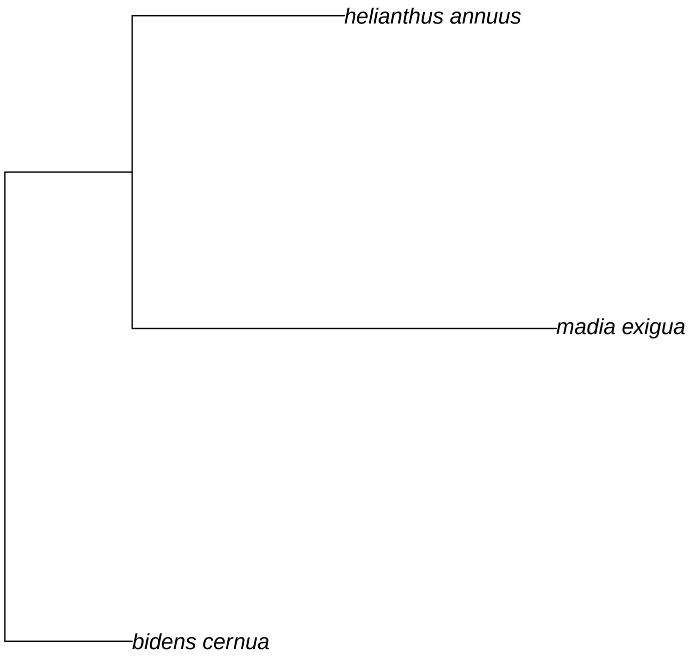
```
phylomatic_tree(mytree, "TRUE", "POST", "xml", "TRUE")

##
## Phylogenetic tree with 3 tips and 2 internal nodes.
##
## Tip labels:
## [1] "bidens_cernua"     "madia_exigua"      "helianthus_annuus"
##  Node labels:
## [1] "euphyllophyte" ""
##
## Rooted; includes branch lengths.
```

1.1.3 Return tree in newick format

```
phylomatic_tree(mytree, "TRUE", "GET", "new", "FALSE")

## [1] "(bidens_cernua:3.000000,(madia_exigua:10.000000,helianthus_annuus:5.000000):3.000000
```

1.2 Get a tree in newick format, read using ape::read.tree, and plot

```
tree <- phylomatic_tree(mytree, "TRUE", "POST", "new", "FALSE")
require(ape)

## Loading required package:  ape

treephylo <- read.tree(text = tree)
plot(treephylo)
```



2 Integrated Taxonomic Informaiton Service (ITIS)

2.1 Search by term and search type

Note: TSN stands for Taxonomic Serial Number

```
doc <- itis_get_xml("dolphin", "comname", "name")
doc <- itis_get_xml("grizzly bear", "comnameend", "name")
doc <- itis_get_xml("ursidae", "itistermssciname", "name")
```

2.2 Take advantage of default arguments, spaces accepted, and display data.frame output

```
doc <- itis_get_xml("Plethodon")
head(itis_parse(doc))
```

```
##                     sci    com    tsn
## 1 Monophyllus plethodon 180073 180073
## 2             Plethodon 173648 173648
## 3      Plethodon aeneus 208328 208328
## 4  Plethodon ainsworthi 668315 668315
## 5    Plethodon albagula 208278 208278
## 6      Plethodon amplus 668316 668316
```

2.3 Show higher taxonomy for a given TSN

```
classification(685566, ID = "tsn")
```

```
## [[1]]
##        rank            taxon    tsn
## 1   Kingdom         Animalia 685566
## 2    Phylum         Chordata 202423
## 3 Subphylum       Vertebrata 158852
## 4     Class         Amphibia 331030
## 5     Order          Caudata 173420
## 6    Family   Plethodontidae 173584
## 7 Subfamily   Plethodontinae 173631
## 8     Genus        Plethodon 550197
## 9   Species Plethodon asupak 173648
```

2.4 Get a TSN from a species name

```
tsn <- get_tsn(searchterm = "Quercus_douglasii", searchtype = "sciname")
```

```
##
## Retrieving data for species ' Quercus_douglasii '
```

```
# use classification with get_tsn together:
classification(tsn)
```

```
## $`1`
##             rank             taxon    tsn
## 1        Kingdom           Plantae  19322
## 2     Subkingdom    Viridaeplantae 202422
## 3   Infrakingdom      Streptophyta 846492
## 4       Division      Tracheophyta 846494
## 5    Subdivision   Spermatophytina 846496
## 6  Infradivision      Angiospermae 846504
## 7          Class     Magnoliopsida 846505
## 8     Superorder           Rosanae  18063
## 9          Order           Fagales 846548
## 10        Family          Fagaceae  19273
## 11         Genus           Quercus  19275
## 12       Species Quercus douglasii  19276
```

3 NCBI Taxonomy Browser

3.1 Get Unique Identifier

```
uids <- get_uid(c("Chironomus riparius", "Chaetopteryx"))
```

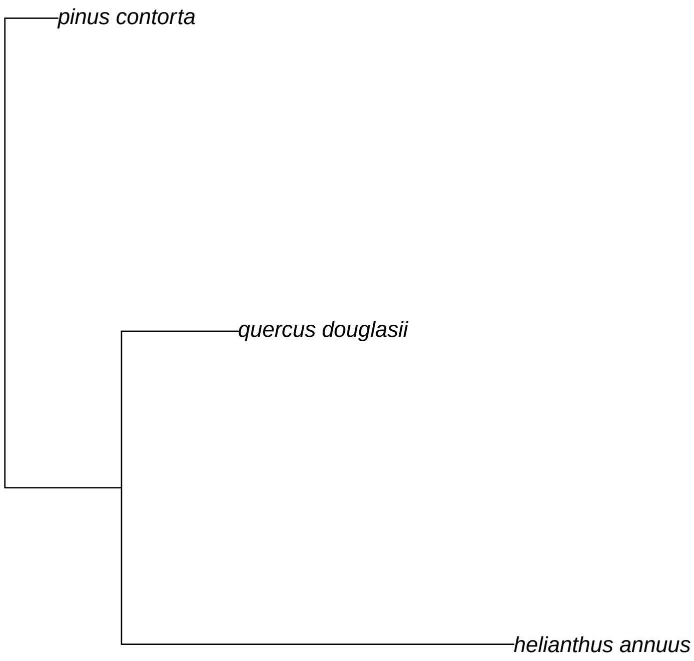
3.2 Retrieve classification for give UID

```
cl <- classification(uids)
cl

## [[1]]
##               ScientificName         Rank     UID
## 1         cellular organisms      no rank  131567
## 2                  Eukaryota superkingdom    2759
## 3               Opisthokonta      no rank   33154
## 4                    Metazoa      kingdom   33208
## 5                  Eumetazoa      no rank    6072
## 6                  Bilateria      no rank   33213
## 7                  Coelomata      no rank   33316
## 8                Protostomia      no rank   33317
## 9                  Ecdysozoa      no rank 1206794
## 10             Panarthropoda      no rank   88770
## 11                Arthropoda       phylum    6656
## 12               Mandibulata      no rank  197563
## 13              Pancrustacea      no rank  197562
## 14                  Hexapoda   superclass    6960
## 15                   Insecta        class   50557
## 16                Dicondylia      no rank   85512
## 17                 Pterygota      no rank    7496
## 18                  Neoptera     subclass   33340
## 19             Endopterygota   infraclass   33392
## 20                   Diptera        order    7147
## 21                Nematocera     suborder    7148
## 22              Culicimorpha   infraorder   43786
## 23             Chironomoidea  superfamily   41828
## 24              Chironomidae       family    7149
## 25              Chironominae    subfamily   54970
## 26               Chironomini        tribe   72530
## 27                Chironomus        genus    7150
## 28 Chironomus incertae sedis      no rank 1165752
##
## [[2]]
##        ScientificName         Rank     UID
## 1  cellular organisms      no rank  131567
## 2           Eukaryota superkingdom    2759
## 3        Opisthokonta      no rank   33154
## 4             Metazoa      kingdom   33208
## 5           Eumetazoa      no rank    6072
## 6           Bilateria      no rank   33213
## 7           Coelomata      no rank   33316
## 8         Protostomia      no rank   33317
## 9           Ecdysozoa      no rank 1206794
## 10      Panarthropoda      no rank   88770
## 11         Arthropoda       phylum    6656
## 12        Mandibulata      no rank  197563
## 13       Pancrustacea      no rank  197562
## 14           Hexapoda   superclass    6960
## 15            Insecta        class   50557
## 16         Dicondylia      no rank   85512
## 17          Pterygota      no rank    7496
## 18           Neoptera     subclass   33340
## 19      Endopterygota   infraclass   33392
## 20   Amphiesmenoptera   superorder   85604
## 21        Trichoptera        order   30263
## 22      Integripalpia     suborder   93875
## 23     Limnephiloidea  superfamily   41033
## 24      Limnephilidae       family   50645
## 25      Limnephilinae    subfamily  177669
## 26    Chaetopterygini        tribe  492564
```

4 Using ITIS and Phylomatic together

4.1 Use ITIS and itis_get_phymat_format to format a string for each species to be submitted to Phylomatic

```
require(plyr)

## Loading required package:  plyr

dat <- laply(list("36616", "19322", "183327"), itis_get_phymat_format, format = "rsubmit")
dat

## [1] "asteraceae%2Fhelianthus%2Fhelianthus_annuus"
## [2] "fagaceae%2Fquercus%2Fquercus_douglasii"
## [3] "pinaceae%2Fpinus%2Fpinus_contorta"

dat_mine <- paste(dat, collapse = "%0D%0A")  # collapse and replace \n's
dat_mine

## [1] "asteraceae%2Fhelianthus%2Fhelianthus_annuus%0D%0Afagaceae%2Fquercus%2Fquercus_dougla

tree <- phylomatic_tree(dat_mine, "FALSE", "GET", "new", "TRUE")
tree

##
## Phylogenetic tree with 3 tips and 2 internal nodes.
##
## Tip labels:
## [1] "helianthus_annuus" "quercus_douglasii" "pinus_contorta"
##  Node labels:
## [1] "euphyllophyte" ""
##
## Rooted; includes branch lengths.

plot(tree)
```



4.2 A somewhat realistic workflow

4.2.1 User's species list

```
splist <- c("annona cherimola", "annona muricata", "shorea robusta", "pandanus patina",
    "durio zibethinus")
```

4.2.2 Get TSN code for each species

```
tsns <- get_tsn(splist, "sciname")

##
## Retrieving data for species ' annona cherimola '
## Retrieving data for species ' annona muricata '
## Retrieving data for species ' shorea robusta '
## Retrieving data for species ' pandanus patina '
## Retrieving data for species ' durio zibethinus '

tsns

## [1] "506198" "18098"  "506787" "507376" "506099"
## attr(,"class")
## [1] "tsn"
```

4.2.3 Get Phylomatic formatted strings for each species

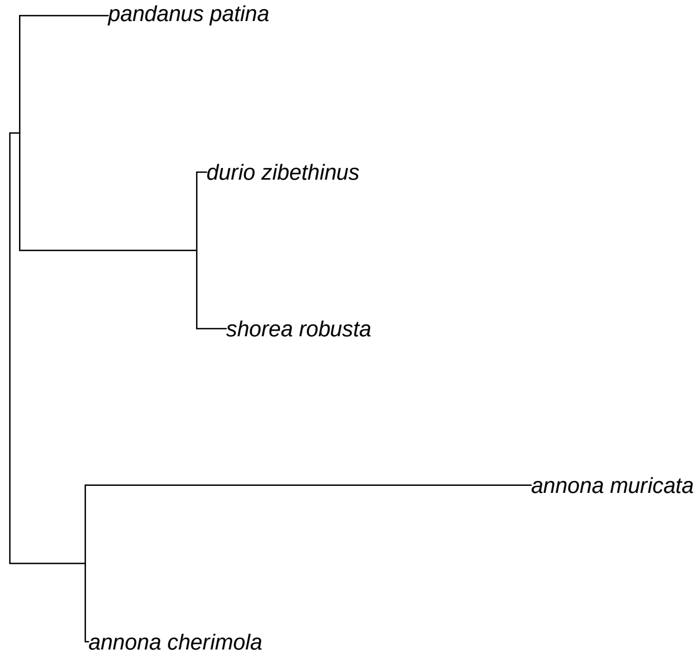
```
dat <- laply(tsns, itis_get_phymat_format, format = "rsubmit")
dat

## [1] "annonaceae%2Fannona%2Fannona_cherimola"
## [2] "annonaceae%2Fannona%2Fannona_muricata"
## [3] "dipterocarpaceae%2Fshorea%2Fshorea_robusta"
## [4] "pandanaceae%2Fpandanus%2Fpandanus_patina"
## [5] "malvaceae%2Fdurio%2Fdurio_zibethinus"
```

4.2.4 Send strings to Phylomatic and return newick tree

```
tree <- phylomatic_tree(dat, "TRUE", "GET", "new", "TRUE")
tree

##
## Phylogenetic tree with 5 tips and 4 internal nodes.
##
## Tip labels:
## [1] "annona_cherimola" "annona_muricata"  "shorea_robusta"
## [4] "durio_zibethinus" "pandanus_patina"
##  Node labels:
## [1] "euphyllophyte"       "annona"              "poales_to_asterales"
## [4] ""
##
## Rooted; includes branch lengths.
```

4.2.5 Plot tree, etc.

```
plot(tree)
```



5 Taxonomic Name Resolution Service

5.1 Name matching using the TNRS

```
iplant_tnrsmatch("best", taxnames = c("helianthus annuus", "acacia", "saltea"),
    output = "names")  # just best names

## Hitting the TNRS API and matching names...

##        AcceptedName   MatchFam MatchGenus MatchScore    Accept?
## 1 Helianthus annuus Asteraceae Helianthus          1 No opinion
## 2            Acacia   Fabaceae     Acacia          1 No opinion
## 3                                 Saltera       0.93 No opinion
##      SubmittedNames
## 1 helianthus annuus
## 2            acacia
## 3            saltea
```

6 Global Names Resolver (GNR)

6.1 Get just id's and names of sources in a data.frame

```
tail(gnr_datasources(todf = TRUE))

##     id                                title
## 82 164                            BioLib.cz
## 83 165 Tropicos - Missouri Botanical Garden
## 84 166                                nlbif
## 85 167  The International Plant Names Index
## 86 168              Index to Organism Names
## 87 169                        uBio NameBank
```

6.2 Give me the id for EOL (Encyclopedia of Life)

```
out <- gnr_datasources(todf = TRUE)
out[out$title == "EOL", "id"]

## [1] 12
```

6.3 Fuzzy search for sources with the word "zoo"

```
out <- gnr_datasources(todf = TRUE)
outdf <- out[agrep("zoo", out$title, ignore.case = TRUE), ]
outdf[1:2, ]

##     id             title
## 20 100 Mushroom Observer
## 25 105           ZooKeys
```

6.4 Resolve some names

6.4.1 Search for *Helianthus annuus* and *Homo sapiens*, return a data.frame

```
gnr_resolve(names = c("Helianthus annuus", "Homo sapiens"), returndf = TRUE)[1:2,
    ]

##   data_source_id    submitted_name       name_string score    title
## 1              4 Helianthus annuus Helianthus annuus 0.988     NCBI
## 3             10 Helianthus annuus Helianthus annuus 0.988 Freebase
```

6.4.2 Search for the same species, with only using data source 12 (i.e., EOL)

```
gnr_resolve(names = c("Helianthus annuus", "Homo sapiens"), data_source_ids = "12",
    returndf = TRUE)

##   data_source_id    submitted_name       name_string score title
## 1             12 Helianthus annuus Helianthus annuus 0.988   EOL
## 2             12      Homo sapiens      Homo sapiens 0.988   EOL
```

7 Tropicos

```
head(tp_getacceptednames(id = 25503923))

## http://services.tropicos.org/Name/25503923/AcceptedNames?apikey=e63f602b-f212-4c53-964c-

##                    variable             value category
## 1                    NameId          25503923  Synonym
## 2            ScientificName       Aira pumila  Synonym
## 3 ScientificNameWithAuthors Aira pumila Pursh  Synonym
## 4                    Family           Poaceae  Synonym
## 5                    NameId          25509881 Accepted
## 6            ScientificName         Poa annua Accepted

head(tp_getsynonyms(id = 25509881))

## http://services.tropicos.org/Name/25509881/Synonyms?apikey=e63f602b-f212-4c53-964c-94e70

##                    variable             value category
## 1                    NameId          25503923  Synonym
## 2            ScientificName       Aira pumila  Synonym
## 3 ScientificNameWithAuthors Aira pumila Pursh  Synonym
## 4                    Family          Poaceae  Synonym
## 5                    NameId         25509881 Accepted
## 6            ScientificName        Poa annua Accepted
```

8 uBio

8.1 Search uBio using the namebank_search API method

```
ubio_namebank(searchName = "elephant", sci = 1, vern = 0)

## Error:  object of type 'externalptr' is not subsettable

head(ubio_namebank(searchName = "Helianthus annuus", sci = 1, vern = 0)[, -c(2,
    3)])

## Error:  object of type 'externalptr' is not subsettable

out <- lapply(list("Helianthus debilis", "Astragalus aduncus"), function(x) ubio_namebank(se
    sci = 1, vern = 0))

## Error:  object of type 'externalptr' is not subsettable

head(out[[2]][, -c(2, 3)])  # just Astragalus aduncus output

## Error:  incorrect number of dimensions
```

9 IUCN Red List

9.1 Query the Red List API

```
ia <- iucn_summary(c("Panthera uncia", "Lynx lynx"))
```

9.2 Extract status

```
laply(ia, function(x) x$status)

## [1] "EN" "LC"
```

9.3 Extract Distribution

```
ia[["Panthera uncia"]]$distr

##  [1] "Afghanistan"        "Bhutan"             "China"
##  [4] "India"              "Kazakhstan"         "Kyrgyzstan"
##  [7] "Mongolia"           "Nepal"              "Pakistan"
## [10] "Russian Federation" "Tajikistan"         "Uzbekistan"
```