

datelife class April 21, 2022

Resources:

- Collaborative document (the etherpad): pad.carpentries.org/datelife.
- DateLife documentation: phylotastic.org/datelife/index.html.

Install datelife

```
devtools::install_github("phylotastic/datelife")
```

```
## Skipping install of 'datelife' from a github remote, the SHA1 (0c39cee3) has not changed since last ...  
## Use `force = TRUE` to force installation
```

Get a list of taxonomic names that you are interested on

Note: it can't be common names (yet).

- We can create a list of species or more inclusive taxonomic group names.
- We can get tip labels from a tree.

Case when you are interested in one (or more) inclusive taxonomic groups, such as genus, family, order, etc.

DateLife will be able to get all species within the taxonomic group(s) you define here:

```
my_taxon <- "mantodea"
```

Case when we want taxon names from a tree

```
# read in your tree  
# if it is a nexus tree, use function ape::read.nexus()  
my_tree <- ape::read.tree(file = "data/krenek_etal_2013.tre")  
my_tree
```

```
##  
## Phylogenetic tree with 97 tips and 96 internal nodes.  
##  
## Tip labels:  
## 'Natronomonas pharaonis DnaK', 'Haloferax volcanii DnaK', 'Methanosarcina barkeri DnaK', 'Methanosarcina  
##  
## Rooted; includes branch lengths.
```

```
str(my_tree)
```

```
## List of 4  
## $ edge      : int [1:192, 1:2] 98 99 100 100 99 101 101 102 102 103 ...  
## $ edge.length: num [1:192] 0.0697 0.4477 0.1343 0.1035 0.0676 ...
```

```
## $ Nnode      : int 96
## $ tip.label  : chr [1:97] "'Natronomonas pharaonis DnaK'" "'Haloferax volcanii DnaK'" "'Methanosarcina barkeri DnaK'"
## - attr(*, "class")= chr "phylo"
## - attr(*, "order")= chr "cladewise"
```

```
# your taxon names will be in the tip.label part of the tree.
# you can access them as follows:
```

```
my_tree$tip.label
```

```
## [1] "'Natronomonas pharaonis DnaK'"
## [2] "'Haloferax volcanii DnaK'"
## [3] "'Methanosarcina barkeri DnaK'"
## [4] "'Methanosaeta thermophila DnaK'"
## [5] "'Nitrosopumilus maritimus DnaK'"
## [6] "'Aciduliprofundum boonei DnaK'"
## [7] "'Thermoplasma volcanium DnaK'"
## [8] "'Caenorhabditis elegans BiP2'"
## [9] "'Caenorhabditis elegans BiP1'"
## [10] "'Daphnia pulex BiP'"
## [11] "'Drosophila melanogaster BiP'"
## [12] "'Danio rerio HSPA5'"
## [13] "'Homo sapiens HSPA5'"
## [14] "'Mus musculus BiP'"
## [15] "'Gallus gallus BiP'"
## [16] "'Oxytricha trifallax BiP'"
## [17] "'Ichthyophthirius multifiliis BiP'"
## [18] "'Tetrahymena borealis BiP'"
## [19] "'Tetrahymena malaccensis BiP'"
## [20] "'Tetrahymena thermophila BiP'"
## [21] "'Paramecium caudatum PcHsp70ER1a/b'"
## [22] "'Paramecium tetraurelia Hsp70Pt07'"
## [23] "'Paramecium tetraurelia Hsp70Pt08'"
## [24] "'Paramecium tetraurelia Hsp70Pt09'"
## [25] "'Paramecium caudatum PcHsp70ER2a/b/c'"
## [26] "'Caenorhabditis elegans Hsp70-7'"
## [27] "'Caenorhabditis elegans Hsp70-9'"
## [28] "'Drosophila melanogaster Hsc70-2'"
## [29] "'Drosophila melanogaster Hsp70Ba'"
## [30] "'Drosophila melanogaster Hsp70Aa'"
## [31] "'Caenorhabditis elegans Hsp70-1'"
## [32] "'Drosophila melanogaster Hsc70-1'"
## [33] "'Drosophila melanogaster Hsc70-4'"
## [34] "'Danio rerio HSPA8'"
## [35] "'Homo sapiens HSPA8'"
## [36] "'Mus musculus HSPA8'"
## [37] "'Gallus gallus HSPA8Aa'"
## [38] "'Homo sapiens HSPA6'"
## [39] "'Mus musculus HSPA1A'"
## [40] "'Homo sapiens HSPA1A'"
## [41] "'Mus musculus HSPA1L'"
## [42] "'Homo sapiens HSPA1L'"
## [43] "'Danio rerio HSPA1B'"
## [44] "'Danio rerio HSPA1A'"
## [45] "'Homo sapiens HSPA2'"
```

```

## [46] "'Mus musculus HSPA2'"
## [47] "'Gallus gallus HSPA2'"
## [48] "'Daphnia pulex Hsp70-3'"
## [49] "'Daphnia pulex Hsp70-2'"
## [50] "'Daphnia pulex Hsp70-1'"
## [51] "'Oxytricha trifallax hsp70'"
## [52] "'Paramecium caudatum PcHsp70CY2b'"
## [53] "'Paramecium caudatum PcHsp70CY2a'"
## [54] "'Paramecium tetraurelia Hsp70Pt04'"
## [55] "'Paramecium tetraurelia Hsp70Pt03'"
## [56] "'Paramecium tetraurelia Hsp70Pt05'"
## [57] "'Paramecium caudatum PcHsp70CY1a/b'"
## [58] "'Paramecium caudatum PcHsp70CY1c'"
## [59] "'Paramecium tetraurelia Hsp70Pt01'"
## [60] "'Paramecium tetraurelia Hsp70Pt02'"
## [61] "'Ichthyophthirius multifiliis hsp70'"
## [62] "'Tetrahymena borealis hsp70-4'"
## [63] "'Tetrahymena thermophila hsp70-4'"
## [64] "'Tetrahymena malaccensis hsp70-4'"
## [65] "'Tetrahymena thermophila hsp70-1'"
## [66] "'Tetrahymena malaccensis hsp70-1'"
## [67] "'Tetrahymena borealis hsp70-1'"
## [68] "'Tetrahymena thermophila hsp70-3'"
## [69] "'Tetrahymena malaccensis hsp70-3'"
## [70] "'Tetrahymena borealis hsp70-3'"
## [71] "'Tetrahymena borealis hsp70-5'"
## [72] "'Tetrahymena thermophila hsp70-5'"
## [73] "'Tetrahymena malaccensis hsp70-5'"
## [74] "'Tetrahymena malaccensis hsp70-2'"
## [75] "'Tetrahymena thermophila hsp70-2'"
## [76] "'Oxytricha trifallax mtHsp70'"
## [77] "'Paramecium caudatum PcHsp70MT1a'"
## [78] "'Paramecium tetraurelia mtHsp70-2'"
## [79] "'Paramecium tetraurelia mtHsp70-1'"
## [80] "'Ichthyophthirius multifiliis mtHsp70-1'"
## [81] "'Tetrahymena thermophila mtHsp70'"
## [82] "'Tetrahymena malaccensis mtHsp70'"
## [83] "'Tetrahymena borealis mtHsp70'"
## [84] "'Ichthyophthirius multifiliis mtHsp70-2'"
## [85] "'Caenorhabditis elegans mtHsp70'"
## [86] "'Daphnia pulex mt Hsp70'"
## [87] "'Drosophila melanogaster mtHsp70'"
## [88] "'Danio rerio HSPA9B2'"
## [89] "'Danio rerio HSPA9B'"
## [90] "'Gallus gallus HSPA9A'"
## [91] "'Homo sapiens HSPA9B'"
## [92] "'Mus musculus HSPA9A'"
## [93] "'Wolbach endosymbiont of Drosophila melanogaster'"
## [94] "'Rhizobium etli DnaK'"
## [95] "'Agrobacterium tumefaciens DnaK'"
## [96] "'Rickettsia prowazekii DnaK'"
## [97] "'Methylobacterium nodulans DnaK'"

```

Our tip labels have some tags at the end that do not belong to the taxon names. We have to remove those,

so we only keep the genus and species epithet:

```
# extract your tip labels into an object of its own:
my_group <- my_tree$tip.label

# remove the tilde, you will have to run this twice:
my_group <- gsub("'", "", my_group)
my_group <- gsub("~", "", my_group)

# split your names by the blank space (or the underscore, if that is your case)
my_group_list <- strsplit(my_group, split = " ")

# Save the first two elements of the split, and paste them with an underscore:
elements_to_keep <- 1:2
paste_with <- "_"

my_group <- sapply(my_group_list, function(x) paste0(x[elements_to_keep],
                                                    collapse = paste_with))

# your names are now "clean" and ready to be processed by DateLife
my_group
```

```
## [1] "Natronomonas_pharaonis"      "Haloferax_volcanii"
## [3] "Methanosarcina_barkeri"      "Methanosaeta_thermophila"
## [5] "Nitrosopumilus_maritimus"    "Aciduliprofundum_boonei"
## [7] "Thermoplasma_volcanium"      "Caenorhabditis_elegans"
## [9] "Caenorhabditis_elegans"      "Daphnia_pulex"
## [11] "Drosophila_melanogaster"      "Danio_rerio"
## [13] "Homo_sapiens"                "Mus_musculus"
## [15] "Gallus_gallus"               "Oxytricha_trifallax"
## [17] "Ichthyophthirius_multifiliis" "Tetrahymena_borealis"
## [19] "Tetrahymena_malaccensis"      "Tetrahymena_thermophila"
## [21] "Paramecium_caudatum"           "Paramecium_tetraurelia"
## [23] "Paramecium_tetraurelia"        "Paramecium_tetraurelia"
## [25] "Paramecium_caudatum"           "Caenorhabditis_elegans"
## [27] "Caenorhabditis_elegans"        "Drosophila_melanogaster"
## [29] "Drosophila_melanogaster"       "Drosophila_melanogaster"
## [31] "Caenorhabditis_elegans"        "Drosophila_melanogaster"
## [33] "Drosophila_melanogaster"       "Danio_rerio"
## [35] "Homo_sapiens"                 "Mus_musculus"
## [37] "Gallus_gallus"                "Homo_sapiens"
## [39] "Mus_musculus"                 "Homo_sapiens"
## [41] "Mus_musculus"                 "Homo_sapiens"
## [43] "Danio_rerio"                  "Danio_rerio"
## [45] "Homo_sapiens"                 "Mus_musculus"
## [47] "Gallus_gallus"                "Daphnia_pulex"
## [49] "Daphnia_pulex"                "Daphnia_pulex"
## [51] "Oxytricha_trifallax"          "Paramecium_caudatum"
## [53] "Paramecium_caudatum"           "Paramecium_tetraurelia"
## [55] "Paramecium_tetraurelia"        "Paramecium_tetraurelia"
## [57] "Paramecium_caudatum"           "Paramecium_caudatum"
## [59] "Paramecium_tetraurelia"        "Paramecium_tetraurelia"
## [61] "Ichthyophthirius_multifiliis" "Tetrahymena_borealis"
## [63] "Tetrahymena_thermophila"       "Tetrahymena_malaccensis"
## [65] "Tetrahymena_thermophila"       "Tetrahymena_malaccensis"
```

```
## [67] "Tetrahymena_borealis"      "Tetrahymena_thermophila"
## [69] "Tetrahymena_malaccensis"   "Tetrahymena_borealis"
## [71] "Tetrahymena_borealis"      "Tetrahymena_thermophila"
## [73] "Tetrahymena_malaccensis"   "Tetrahymena_malaccensis"
## [75] "Tetrahymena_thermophila"   "Oxytricha_trifallax"
## [77] "Paramecium_caudatum"        "Paramecium_tetraurelia"
## [79] "Paramecium_tetraurelia"     "Ichthyophthirius_multifiliis"
## [81] "Tetrahymena_thermophila"   "Tetrahymena_malaccensis"
## [83] "Tetrahymena_borealis"      "Ichthyophthirius_multifiliis"
## [85] "Caenorhabditis_elegans"    "Daphnia_pulex"
## [87] "Drosophila_melanogaster"    "Danio_rerio"
## [89] "Danio_rerio"               "Gallus_gallus"
## [91] "Homo_sapiens"              "Mus_musculus"
## [93] "Wolbach_endosymbiont"      "Rhizobium_etli"
## [95] "Agrobacterium_tumefaciens" "Rickettsia_prowazekii"
## [97] "Methylobacterium_nodulans"
```

Process your taxon names with datelife

To perform the chronogram search in the DateLife database, we have to standardize our group of interest to the taxonomy.

For this, we will use the `make_datelife_query` function:

```
# when you have a list of species names, you can use all default arguments:
my_query_group <- datelife::make_datelife_query(input = my_group)
```

```
## ... Making a DateLife query.
```

```
## ... Phylo-processing 'input'.
```

```
## 'input' is not a phylogeny.
```

```
## |
```

```
## Warning: Wolbach_endosymbiont are not matched
```

```
## |
```

```
## Working with the following taxa:
```

```
## Natronomonas pharaonis | Haloferax volcanii | Methanosarcina barkeri | Methanotherix thermoacetophil
```

```
## DateLife query done!
```

```
# when you have an inclusive taxon name, you have to set the argument "get_spp_from_taxon" to TRUE:
```

```
my_query <- datelife::make_datelife_query(my_taxon, get_spp_from_taxon = TRUE)
```

```
## ... Making a DateLife query.
```

```
## ... Phylo-processing 'input'.
```

```
## 'input' is not a phylogeny.
```

```
## |
```

```
## |
```

```
## |
```

```
## |
```

```

## Working with the following taxa:
## Stenophylla gallardi | Stenophylla lobivertex | Mantoida schraderi | Metalliticus violacea | Metalliticus
## DateLife query done!
# look at the data you obtained by running datelife::make_datelife_query:
str(my_query)

## List of 3
## $ cleaned_names: Named chr [1:362] "Stenophylla gallardi" "Stenophylla lobivertex" "Mantoida schraderi" "Mantoida schraderi" ...
## ..- attr(*, "names")= chr [1:362] "Mantodea1" "Mantodea2" "Mantodea3" "Mantodea4" ...
## $ ott_ids       : Named int [1:362] 4176725 711174 1046541 235229 892960 711168 3484353 1012466 42381
## ..- attr(*, "names")= chr [1:362] "Stenophylla gallardi" "Stenophylla lobivertex" "Mantoida schraderi" "Mantoida schraderi" ...
## $ phy           : logi NA
## - attr(*, "class")= chr "datelifeQuery"
# check how many taxon names you got:
length(my_query$cleaned_names)

## [1] 362

```