

SSB 2020 Workshop: Using the Open Tree of Life for your Research

Get a dated tree for your *taxa of interest* (using R) with `datelife` and the Open Tree of Life

Luna L. Sánchez-Reyes

2019-12-30

Requirements for this tutorial

From R, install the packages `rotl`, `ape`, `devtools` and `stringr` with the function `install.packages()`, and the package `datelife` with the function `install_github()`. Load them into your workspace with `library()` or `require()`. If you do not want to load the packages, you can call functions specifying their package using two colons and the syntax `package_name::function_name()`. This implies more typing, but gives more clarity to reproduce the workflow later. So we will use that syntax for this tutorial. An exception to this are functions from packages that are “preloaded” (such as `library()` from `base` or `install.packages()`, from `utils`) that can be simply called by their name.

```
install.packages(c("rotl", "ape", "devtools", "stringr"))
library(rotl)
library(devtools)
library(stringr)
devtools::install_github("phylotastic/datelife")
library(datelife)
```

Find your taxa in the Open Tree of Life Taxonomy

Get the *ott ids* for your taxa with `tnrs_match_names()`. This will generate a “match_names” object.

```
my_taxa <- c("amphibians", "canis", "felis", "delphinidae", "spheniscidae")
resolved_names <- rotl::tnrs_match_names(my_taxa)
class(resolved_names)
> [1] "match_names" "data.frame"
resolved_names
>   search_string unique_name approximate_match ott_id is_synonym flags
> 1 amphibians    Amphibia          TRUE 544595      FALSE
> 2 canis         Canis          FALSE 372706      FALSE
> 3 felis         Felis          FALSE 563165      FALSE
> 4 delphinidae   Delphinidae     FALSE 698406      FALSE
> 5 spheniscidae  Spheniscidae     FALSE 494367      FALSE
>   number_matches
> 1              6
> 2              2
> 3              1
> 4              1
> 5              1
```

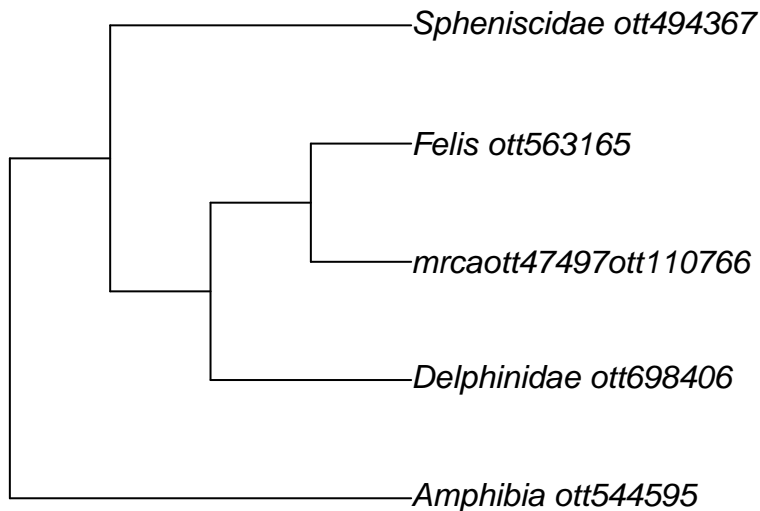
There are two ways to extract the *ott ids* from a “match_names” object:

```
rotl::ott_id(resolved_names) # as a list
resolved_names$ott_id # as a vector
```

Get an induced subtree for your set of taxa

```
mytree <- rotl::tol_induced_subtree(resolved_names$ott_id)
> Warning in collapse_singles(tr, show_progress): Dropping singleton nodes
> with labels: Mammalia ott244265, Theria (subclass in Deuterostomia)
> ott229558, Eutheria (in Deuterostomia) ott683263, Boreoeutheria ott5334778,
> Laurasiatheria ott392223, mrcaott1548ott6790, mrcaott1548ott3607484,
> mrcaott1548ott4942380, mrcaott1548ott4942547, mrcaott1548ott3021, Artiodactyla
> ott622916, mrcaott1548ott21987, mrcaott1548ott5256, mrcaott5256ott4944931,
> Whippomorpha ott7655791, Cetacea ott698424, mrcaott5256ott3615450,
> mrcaott5256ott44568, Odontoceti ott698417, mrcaott5256ott5269,
> mrcaott5269ott6470, mrcaott5269ott47843, mrcaott47843ott194312,
> mrcaott4697ott263949, Carnivora ott44565, Caniformia ott827263,
> Canidae ott770319, mrcaott47497ott3612617, mrcaott47497ott3612529,
> mrcaott47497ott3612596, mrcaott47497ott3612516, mrcaott47497ott3612589,
> mrcaott47497ott3612591, mrcaott47497ott3612592, mrcaott47497ott77889,
> Feliformia ott827259, mrcaott6940ott19397, mrcaott19397ott194349, Felidae
> ott563159, mrcaott54737ott660452, mrcaott54737ott86170, mrcaott54737ott86175,
> mrcaott54737ott442049, mrcaott54737ott86162, mrcaott54737ott86166, Sauropsida
> ott639642, Sauria ott329823, mrcaott246ott4128455, mrcaott246ott4127082,
> mrcaott246ott4129629, mrcaott246ott4142716, mrcaott246ott4126667,
> mrcaott246ott1662, mrcaott246ott2982, mrcaott246ott31216, mrcaott246ott4947920,
> mrcaott246ott4127428, mrcaott246ott4126230, mrcaott246ott4127421,
> mrcaott246ott664349, mrcaott246ott4126505, mrcaott246ott4127015,
> mrcaott246ott4129653, mrcaott246ott4127541, mrcaott246ott4946623,
> mrcaott246ott4126482, mrcaott246ott4128105, mrcaott246ott4127288,
> mrcaott246ott4132146, mrcaott246ott3602822, mrcaott246ott4143599,
> mrcaott246ott3600976, mrcaott246ott4132107, Aves ott81461, Neognathae
> ott241846, mrcaott246ott5481, mrcaott246ott5021, mrcaott246ott7145,
> mrcaott246ott5272, mrcaott5272ott9830, mrcaott9830ott86672, mrcaott9830ott90560,
> mrcaott9830ott18206, mrcaott18206ott60413, Sphenisciformes ott494366
```

```
ape::plot.phylo(mytree, cex = 1)
```



This is cool!

Now, what if you want a synthetic tree containing some or all descendants from your taxa of interest?

Get a subtree of one taxon

We can extract a subtree of all descendants of one taxon at a time using the function `tol_subtree()`.

Try to extract a subtree of all amphibians. Get the *ott id* first. It is already stored in the `resolved_names` object, but you can run the function `tnrs_match_names()` again if you want.

```
amphibia_ott_id <- resolved_names$ott_id[1] # extract the ott id from resolved_names
# amphibian_ott_id <- rotl::tnrs_match_names("amphibians")$ott_id # OR run tnrs again
```

Now, extract the subtree from the Open Tree synthetic tree using `tol_subtree()`.

```
amphibia_subtree <- rotl::tol_subtree(ott_id = amphibian_ott_id)
```

```
amphibia_subtree
>
> Phylogenetic tree with 10012 tips and 3100 internal nodes.
>
> Tip labels:
>   Odorrana_geminata_ott114, Odorrana_supranarina_ott14375, Odorrana_narina_ott14379, Odorrana_amamiensis_ott14380, ...
> Node labels:
>   Amphibia ott544595, Batrachia ott471197, Anura ott991547, , , , ...
>
> Unrooted; no branch lengths.
```

This is a large tree!

Try to extract a subtree for the genus *Canis*. It should be way smaller!

```
canis <- rotl::tnrs_match_names("canis") # get the ott id of the genus Canis (dogs)
canis
>   search_string unique_name approximate_match ott_id is_synonym flags
> 1         canis         Canis             FALSE 372706         FALSE
>   number_matches
> 1             2
```

```
subtree <- rotl::tol_subtree(canis$ott_id) # get the subtree
> Error: HTTP failure: 400
> list(contesting_trees = list(`ot_278@tree1` = list(attachment_points = list(list(children_from_taxon :
>   mrca = "mrcaott47497ott110766")[/v3/tree_of_life/subtree] Error: node_id was not found (broken ta
```

Why am I getting an error saying that the *node id* was not found and that my taxon is “broken”?? Generally, this happens when phylogenetic information does not match taxonomic information. For example, extinct lineages are sometimes phylogenetically included within the taxon but are taxonomically excluded, making the taxon appear as paraphyletic. On the browser, you can still get to the subtree. From R, you need to do something else first.

Get a subtree of a “broken” taxon

There is a way to find out that the group is “broken” before trying to get the subtree and getting an error.

```
rotl::is_in_tree(canis$ott_id)
```

We say that a taxon is not in the tree when its *ott id* is not assigned to a node in the synthetic tree. This is the reason why we get an error when we try to get a subtree using the *ott id*.

Then, how do I get a subtree of my “broken” taxon? There are some options.

a) Get a subtree using the *node id* instead of the *ott id*

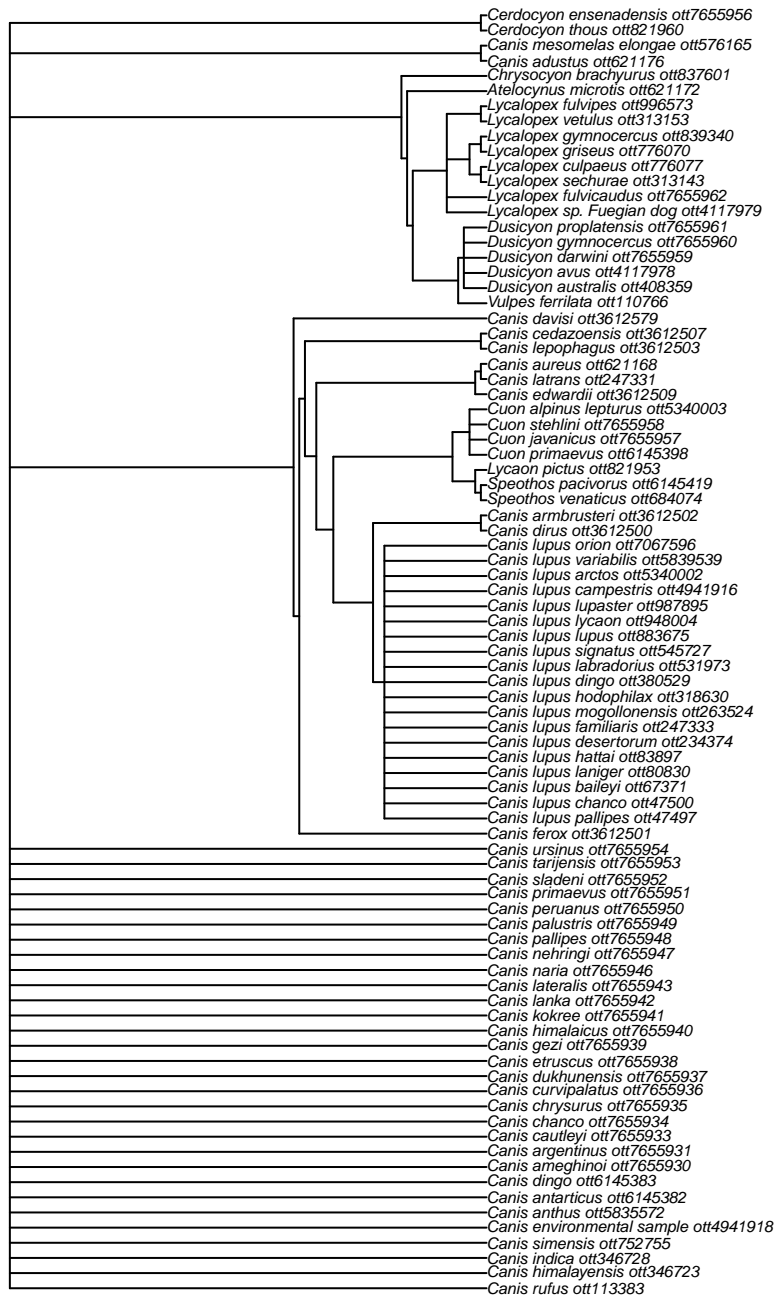
rotl has a function that gets for you all info from the node containing a taxon. That includes the *node id*.

```
canis_node_info <- rotl::tol_node_info(canis$ott_id)
canis_node_info
>
> OpenTree node.
>
> Node id: mrcaott47497ott110766
> Number of terminal descendants: 85
> Is taxon: FALSE
```

We can use that *node id* to get a subtree with `tol_subtree()`

```
canis_node_subtree <- rotl::tol_subtree(node_id = canis_node_info$node_id)
```

```
ape::plot.phylo(canis_node_subtree, cex = 0.5)
```



Nice! We got a subtree of 85 tips, containing all descendants from the node that also contains *Canis*.

b) Get an induced subtree of taxonomic children

We can get the *ott ids* of the taxonomic children of our taxon of interest and use the function `tol_induced_subtree()`.

First, get the taxonomic children.

```
canis_taxonomy <- rotl::taxonomy_subtree(canis$ott_id)
```

canis_taxonomy

\$tip_label [1] "Canis_dirus_ott3612500"
 [2] "Canis_anthus_ott5835572"
 [3] "Canis_rufus_ott113383"
 [4] "Canis_simensis_ott752755"
 [5] "Canis_aureus_ott621168"
 [6] "Canis_mesomelas_elongae_ott576165"
 [7] "Canis_adustus_ott621176"
 [8] "unclassified_Canis_ott7655955"
 [9] "Canis_latrans_ott247331"
 [10] "Canis_lupus_baileyi_ott67371"
 [11] "Canis_lupus_laniger_ott80830"
 [12] "Canis_lupus_orion_ott7067596"
 [13] "Canis_lupus_hodophilax_ott318630"
 [14] "Canis_lupus_signatus_ott545727"
 [15] "Canis_lupus_arctos_ott5340002"
 [16] "Canis_lupus_mogollonensis_ott263524"
 [17] "Canis_lupus_variabilis_ott5839539"
 [18] "Canis_lupus_lupus_ott883675"
 [19] "Canis_lupus_campestris_ott4941916"
 [20] "Canis_lupus_lycaon_ott948004"
 [21] "Canis_lupus_pallipes_ott47497"
 [22] "Canis_lupus_chancho_ott47500"
 [23] "Canis_lupus_x_Canis_lupus_familiaris_ott4941915" [24] "Canis_lupus_desertorum_ott234374"
 [25] "Canis_lupus_familiaris_ott247333"
 [26] "Canis_lupus_dingo_ott380529"
 [27] "Canis_lupus_labradorius_ott531973"
 [28] "Canis_lupus_hattai_ott83897"
 [29] "Canis_lupus_lupaster_ott987895"
 [30] "Canis_himalayensis_ott346723"
 [31] "Canis_indica_ott346728"
 [32] "Canis_environmental_samples_ott4941917"
 [33] "Canissp.KEB-2016ott5925604"
 [34] "Canis_sp._CANInt1_ott470950"
 [35] "'Canissp.Russia/33"
 [36] "500ott5338950"
 [37] "Canis_sp._ott247325"
 [38] "'Canissp.Belgium/36"
 [39] "000ott5338951"
 [40] "Canis_environmental_sample_ott4941918"
 [41] "Canis_morenis_ott6145387"
 [42] "Canis_niger_ott6145388"
 [43] "Canis_palaeoplatensis_ott6145390"
 [44] "Canis_osorum_ott6145389"
 [45] "Canis_thooides_ott6145392"
 [46] "Canis_antarcticus_ott6145381"
 [47] "Canis_proplatensis_ott6145391"
 [48] "Canis_feneus_ott6145384"
 [49] "Canis_geismarianus_ott6145385"
 [50] "Canis_ameghinoi_ott7655930"
 [51] "Canis_nehringi_ott7655947"
 [52] "Canis_palustris_ott7655949"
 [53] "Canis_lanka_ott7655942"
 [54] "Canis_pallipes_ott7655948"
 [55] "Canis_gezi_ott7655939"

```
[56] "Canis_montanus_ott7655945"
[57] "Canis_primaevus_ott7655951"
[58] "Canis_chrysurus_ott7655935"
[59] "Canis_dukhunensis_ott7655937"
[60] "Canis_kokree_ott7655941"
[61] "Canis_sladeni_ott7655952"
[62] "Canis_himalaicus_ott7655940"
[63] "Canis_chancho_ott7655934"
[64] "Canis_curvipalatus_ott7655936"
[65] "Canis_lateralis_ott7655943"
[66] "Canis_argentinus_ott7655931"
[67] "Canis_tarijensis_ott7655953"
[68] "Canis_naria_ott7655946"
[69] "Canis_peruanus_ott7655950"
[70] "Canis_cautleyi_ott7655933"
[71] "Canis_ursinus_ott7655954"
[72] "Canis_armbrusteri_ott3612502"
[73] "Canis_ferox_ott3612501"
[74] "Canis_lepophagus_ott3612503"
[75] "Canis_edwardii_ott3612509"
[76] "Canis_apolloniensis_ott3612508"
[77] "Canis_cedazoensis_ott3612507"
[78] "Canis_primigenius_ott3612506"
[79] "Canis_lydekkeri_ott7655944"
[80] "Canis_arnensis_ott7655932"
[81] "Canis_antarticus_ott6145382"
[82] "Canis_dingo_ott6145383"
[83] "Canis_etruscus_ott7655938"
[84] "Canis_spelaesus_ott3612504"
```

\$edge_label [1] "Canis_mesomelas_ott666235" "Canis_lupus_ott247341"

[3] "Canis_ott372706"

Now, extract the *ott* ids.

```
canis_taxonomy_ott_ids <- datelife::extract_ott_ids(x = canis_taxonomy$tip_label)
> After extracting ott ids, there are some non numeric elements:
> Canissp.KEB-2016ott5925604
> 'Canissp.Russia/33
> 500ott5338950'
> 'Canissp.Belgium/36
> 000ott5338951'
>
> NAs removed
```

Try to get an induced subtree of *Canis* taxonomic children.

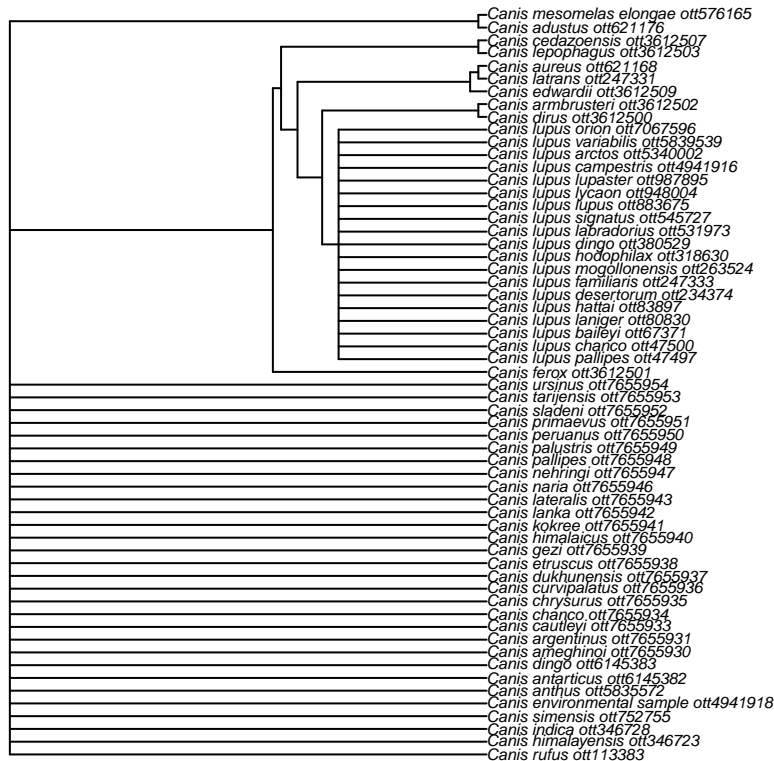
```
canis_taxonomy_subtree <- rot1::tol_induced_subtree(canis_taxonomy_ott_ids)
> Error: HTTP failure: 400
> [/v3/tree_of_life/induced_subtree] Error: node_id 'ott3612504' was not found!list(ott247325 = "pruned,
> ott7655944 = "pruned_ott_id", ott7655945 = "pruned_ott_id", ott7655955 = "pruned_ott_id")
```

It is often not possible to get an induced subtree of all taxonomic children from a taxon, because some of them are not in the tree.

To verify which ones are giving us trouble, we can use the function `is_in_tree()` again.

```
in_tree <- sapply(canis_taxonomy_ott_ids, rotl::is_in_tree)
canis_taxonomy_ott_ids_intree <- canis_taxonomy_ott_ids[in_tree]
canis_taxonomy_subtree <- rotl::tol_induced_subtree(canis_taxonomy_ott_ids_intree)

ape::plot.phylo(canis_taxonomy_subtree, cex = 0.5)
```



There! We have a synthetic subtree (derived from phylogenetic information) of the taxonomic children of *Canis* only.

It is sometimes thought that flags can help to detect taxa that are not in the tree, but they can be misleading. Get the flags of *Canis* taxonomic children.

```
canis_taxonomy_ott_ids_taxon_info <- rotl::taxonomy_taxon_info(canis_taxonomy_ott_ids)
canis_taxonomy_ott_ids_taxon_info_flags <- rotl::flags(canis_taxonomy_ott_ids_taxon_info)
in_tree <- sapply(canis_taxonomy_ott_ids, rotl::is_in_tree)
flags_not_in_tree <- canis_taxonomy_ott_ids_taxon_info_flags[!in_tree]
flags_not_in_tree <- unique(unlist(flags_not_in_tree))
flags_in_tree <- canis_taxonomy_ott_ids_taxon_info_flags[in_tree]
flags_in_tree <- unique(unlist(flags_in_tree))

flags_not_in_tree
> [1] "not_otu" "was_container" "hybrid" "infraspecific"
> [5] "incertae_sedis" "extinct" "hidden"
flags_in_tree
> [1] "infraspecific" "extinct"
```

Some flags such as “extinct” are used in taxa that are **not** in the tree and in taxa that **are** in the tree.

Get a subtree of a rank

In datelife, we have a function that gets all valid children from a rank. By default, we get all species.


```
canis_species <- datelife::get_ott_children(ott_ids = canis$ott_id)
```

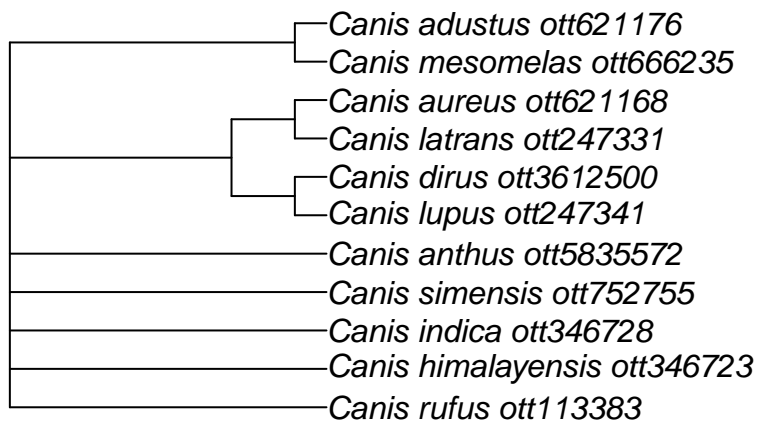
```
canis_species
> $Canis
>
>      ott_id  rank
> Canis dirus   3612500 species
> Canis anthus  5835572 species
> Canis rufus   113383 species
> Canis simensis 752755 species
> Canis aureus  621168 species
> Canis mesomelas 666235 species
> Canis adustus 621176 species
> Canis latrans 247331 species
> Canis lupus   247341 species
> Canis himalayensis 346723 species
> Canis indica  346728 species
```

Get an induced subtree of these *ott ids*.

```
canis_species_subtree <- rotl::tol_induced_subtree(canis_species$Canis$ott_id)
```

```
canis_species_subtree
>
> Phylogenetic tree with 11 tips and 5 internal nodes.
>
> Tip labels:
>   Canis_lupus_ott247341, Canis_dirus_ott3612500, Canis_latrans_ott247331, Canis_aureus_ott621168, Canis_mesomelas_ott666235, Canis_adustus_ott621176
> Node labels:
> [1] "mrcaott47497ott110766" "mrcaott47497ott247331" "mrcaott47497ott3612500"
> [4] "mrcaott247331ott621168" "mrcaott576165ott621176"
>
> Unrooted; no branch lengths.
```

```
ape::plot.phylo(canis_species_subtree, cex = 1)
```



Try to get all families of Amphibia.

```
amphibia_families <- datelife::get_ott_children(ott_ids = amphibia_ott_id, ott_rank = "family")
```

```
amphibia_families
> $Amphibia
>
>      ott_id  rank
```

> Caeciliidae	118029	family
> Typhlonectidae	639647	family
> Ichthyophiidae	639653	family
> Dermophiidae	654645	family
> Rhinatrematidae	128153	family
> Indotyphlidae	114139	family
> Siphonopidae	114359	family
> Scolecomorphidae	861429	family
> Herpelidae	379929	family
> Chikilidae	4948197	family
> Leiopelmatidae	611960	family
> Ascaphidae	1013114	family
> Bombinatoridae	1013112	family
> Alytidae	991545	family
> Plethodontidae	515329	family
> Proteidae	630300	family
> Dicamptodontidae	60819	family
> Rhyacotritonidae	459010	family
> Salamandridae	566011	family
> Amphiumidae	566022	family
> Ambystomatidae	984723	family
> Hynobiidae	336749	family
> Cryptobranchidae	336750	family
> Sirenidae	515352	family
> Pelodytidae	509554	family
> Megophryidae	828160	family
> Pelobatidae	485821	family
> Rhinophrynidae	459016	family
> Pipidae	465087	family
> Hemiphractidae	245330	family
> Leptodactylidae	414584	family
> Heleophrynidae	127159	family
> Alsodidae	533096	family
> Cycloramphidae	533097	family
> Hylidae	535782	family
> Strabomantidae	199520	family
> Brachycephalidae	719424	family
> Rhinodermatidae	4133476	family
> Rheobatrachidae	918183	family
> Hylodidae	1059295	family
> Eleutherodactylidae	63340	family
> Craugastoridae	63341	family
> Dendrobatidae	581837	family
> Bufonidae	889358	family
> Ceratophryidae	1008932	family
> Batrachylidae	1008933	family
> Petropedetidae	337155	family
> Ceratobatrachidae	1081207	family
> Micrixalidae	1081209	family
> Mantellidae	38969	family
> Nyctibatrachidae	1081210	family
> Ptychadenidae	1081211	family
> Ranidae	364560	family

```

> Ranixalidae          403946 family
> Rhacophoridae       432783 family
> Phrynobatrachidae   504589 family
> Pyxicephalidae      504591 family
> Odontobatrachidae   5536254 family
> Dicroglossidae      1081208 family
> Myobatrachidae      940181 family
> Calyptocephalellidae 346415 family
> Nasikabatrachidae   542087 family
> Sooglossidae        882346 family
> Hyperoliidae        535796 family
> Hemisotidae         165721 family
> Brevicipitidae      790993 family
> Microhylidae        1062202 family
> Arthroleptidae      968547 family
> Allophrynidae       57740 family
> Centrolenidae       467442 family

```

Get the subtree.

```

amphibia_families_subtree <- rotl::tol_induced_subtree(amphibia_families$Amphibia$ott_id)

ape::plot.phylo(amphibia_families_subtree, cex = 0.5)

```

