Taxonomic Name Resolution Service

Barb Banbury, University of Tennessee, darwinthesun@gmail.com

June 19, 2014

1 Introduction

The Taxonomic Name Resolution Service (TNRS) is a service hosted by iPlant to name match plant taxa (`http://tnrs.iplantcollaborative.org/`). This service will check taxonomic names against a database of published plant names, so that users can ensure that they have the valid status and spelling. Similarly, Phylotastic has a name resolution service, Taxosaurus - The Taxonomic Thesaurus, that works on both plant and animal taxa (`http://api.phylotastic.org/tnrs`). Both sites act to standardize scientific nomenclature.

rPlant has several functions that serve to interact with both the iPlant TNRS and the Phylotastic TNRS.

2 Name Checking

2.1 iPlant TNRS

The `ResolveNames` function will take a list of taxonomic names and pass them to the iPlant TNRS. This only works for plant taxa. The TNRS does not currently require authorization to use. As an example, we can use the DNA.fasta data that comes with *rPlant*.

```
> library(rPlant)
> data("DNA.fasta")
> TNRSnames <- ResolveNames(names(DNA.fasta))
> TNRSnames

 [1] "Asplenium"
 [2] "Corsinia_coriandrina"
 [3] "Colysis_longipes"
 [4] "Greenwayodendron_suaveolens"
```

```
 [5] "Sloanea_latifolia"
 [6] "Uvaria_anonoides"
 [7] "Leiocolea_heterocolpos"
 [8] "Heliconia_irrasa"
 [9] "Aristolochia_maxima"
[10] "Gollania_splendens"
```

2.2   Phylotastic Taxosaurus

Taxosaurus also accepts and returns a list of names. However, the service works a little different as it is all done by http submission. For this service, a vector of names can be passed to the server via the function `GetPhylotasticToken`, and then a http token is returned with the matching names. The service returns a website with a JSON script that includes all the taxonomic names matching several sources (iPlant and NCBI).

Once the matching is complete, these names can be retrieved using the function `RetrieveTNRSNames`. Because the server returns several sources, you will have to define which one to use.

```
> myToken <- GetPhylotasticToken(names(DNA.fasta))

Your request is being processed. You can retrieve the results at http://taxosaurus.org/

OR use the rPlant function RetrieveTNRSNames to pull directly into R

> NCBINames <- RetrieveTNRSNames(names(DNA.fasta),
+             myToken, source="NCBI")
> NCBINames

 [1] "Asplenium_shuttleworthianum"
 [2] "Corsinia_coriandra"
 [3] "Colysis_longipes"
 [4] "Greenwayodendron_suaveolens"
 [5] "Sloanea_latifolia"
 [6] "Uvaria_anonoides"
 [7] "Leiocolea_heterocolpos"
 [8] "Heliconia_irrasa"
 [9] "Aristolochia_maxima"
[10] "Gollania_splenden"

> iPlantNames <- RetrieveTNRSNames(names(DNA.fasta),
+             myToken, source="iPlant_TNRS")
> iPlantNames
```

```
 [1] "Asplenium_shuttleworthianum"
 [2] "Corsinia_coriandra"
 [3] "Colysis_longipes"
 [4] "Greenwayodendron_suaveolens"
 [5] "Sloanea_latifolia"
 [6] "Uvaria_anonoides"
 [7] "Leiocolea_heterocolpos"
 [8] "Heliconia_irrasa"
 [9] "Aristolochia_maxima"
[10] "Gollania_splenden"
```

3 Compare Names

We also have a function for comparing the list of two names, CompareNames. This function will return a comment about how many taxon names have changed between the two lists.

```
> CompareNames(names(DNA.fasta), TNRSnames)

[1] "3 taxa changed names according to TNRS"
```

4 sessionInfo

```
> sessionInfo()

R version 2.15.2 (2012-10-26)
Platform: x86_64-pc-linux-gnu (64-bit)

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=C                 LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] tools     stats     graphics  grDevices utils
[6] datasets  methods   base

other attached packages:
[1] rPlant_2.10.4  seqinr_3.0-7   RCurl_1.95-4.1
[4] bitops_1.0-6   rjson_0.2.13
```