

Retrieving Accepted Names from Kew's Plants of the World and Assigning an IPNI Identification Number

Kelsey Brock

2/18/2020

Part 1: Getting POW Accepted Names and IPNI IDS for each Taxon

Step 1: Set Up R

```
require(knitr)
```

```
## Loading required package: knitr
```

```
knitr::opts_knit$set(root.dir = "C:/Users/Kelsey/Documents/rwork/hapi") #set the working directory
```

```
#get the needed packages  
if(!require("pacman")){  
  install.packages("pacman")  
  library(pacman)}
```

```
## Loading required package: pacman
```

```
p_load("taxize", "dplyr", "reshape2", "data.table",  
       "profvis", "magrittr", "knitr", "kableExtra")
```

Step 2: Read in a List of Species Names

```
### Read in a csv file with your species on it.  
spplist <- read.csv("sample_spp_list.csv", header=T, sep=',', stringsAsFactors=F)  
head(spplist, 25)
```

```
##                               Taxa
## 1                             Bignonia tweediana
## 2                             Dolichandra unguiscati
## 3                             Acacia farnesiana
## 4                             Acacia koa
## 5                             Aristolochia philippinensis
## 6                             Breynia vitis-idaea
## 7                             Citrus "Shasta Gold"
## 8                             Citrus unknownium
## 9                             Colysis ampla
## 10                            Chamaecrista nictitans subsp. patellaria
## 11 Chamaecrista nictitans subsp. patellaria var. glabrata
## 12                            Crinum asiaticum
## 13                            asystasia gangetica subsp micrantha
## 14                            Elaeocarpus angustifolius
## 15                            Fouquieria columnaris (Kellogg) Kellogg ex Curran
## 16                            Heliconia hirsuta
## 17                            Hibiscus columnaris
## 18                            Jasminum rex
## 19                            Melia azedarach
## 20                            Odontonema cuspidatum
## 21                            Phragmanthera capitata
## 22                            Pithecellobium dulce
## 23                            Stylurus banksii
## 24                            Tarenna cumingiana
## 25                            Vigna sandwicensiss
```

- In this sample data set of 25 taxa, I have included misspellings (e.g. *Vigna sandwicensiss*), subspecies (*Chamaecrista nictitans* subsp. *patellaria*), varieties (*Chamaecrista nictitans* subsp. *patellaria* var. *glabrata*), cultivars (*Citrus* 'Shasta Gold'), and things that likely won't get a database hit (*Citrus unknownium*). There's also an example of bad capitalization and punctuation (*asystasia gangetica* subsp *micrantha*) and an instance where the authority is included (*Fouquieria columnaris* (Kellogg) Kellogg ex Curran).

Step 3: Make sure each name is a valid synonym (even if it's not the “accepted” name that you ultimately want)

- This step will also take care of misspellings and capitalizations of genus versus specific epithets

Checking for taxonomic reference databases; we're looking for The International Plant Names Index

```
gnr_datasources()
```

```
## # A tibble: 100 x 12
##   created_at data_hash data_url description    id logo_url
##   <chr>      <lgl>      <lgl>      <chr>      <int> <lgl>
## 1 2012-07-0~ NA        NA        "[\"This r~    1 NA
## 2 2012-07-0~ NA        NA        The free s~    2 NA
## 3 2012-02-0~ NA        NA        The White ~    3 NA
## 4 2012-02-0~ NA        NA        The Nation~    4 NA
## 5 2012-02-0~ NA        NA        Classifica~    5 NA
## 6 2012-02-0~ NA        NA        GRIN taxon~    6 NA
## 7 2012-02-0~ NA        NA        This class~    7 NA
## 8 2012-02-0~ NA        NA        The Interi~    8 NA
## 9 2012-02-0~ NA        NA        An authori~    9 NA
## 10 2012-02-0~ NA        NA        An entity ~   10 NA
## # ... with 90 more rows, and 6 more variables: name_strings_count <int>,
## #   refresh_period_days <int>, title <chr>, unique_names_count <int>,
## #   updated_at <chr>, web_site_url <chr>
```

```
# IPNI is # 167
```

Now we run the Global Names Resolver, asking it only return the nearest matching IPNI name

```
gnr.long <- spplist$Taxa %>% gnr_resolve(data_source_ids = 167, best_match_only=T, with_canonical_ranks=T)
```

```
head(gnr.long)
```

```
## # A tibble: 6 x 5
##   user_supplied_na~ submitted_name   data_source_title score matched_name2
##   <chr>             <chr>             <chr>             <dbl> <chr>
## 1 Bignonia tweedia~ Bignonia tweedi~ The Internationa~ 0.75  Bignonia twee~
## 2 Dolichandra ungu~ Dolichandra ung~ The Internationa~ 0.75  Dolichandra u~
## 3 Acacia farnesiana Acacia farnesia~ The Internationa~ 0.988 Acacia farnes~
## 4 Acacia koa        Acacia koa        The Internationa~ 0.988 Acacia koa
## 5 Aristolochia phi~ Aristolochia ph~ The Internationa~ 0.988 Aristolochia ~
## 6 Breynia vitis-id~ Breynia vitis-i~ The Internationa~ 0.988 Breynia vitis~
```

- You can see that it returns a column called “matched_name2” with the viable synonym.
- Because the Global Names Resolver can be time consuming for long species lists, it's probably wise to save the output as a csv.

```
#writing to csv file
write.csv(gnr.long,
          "samplegnr.csv",
          row.names = F, quote = F)
```

```
#reading it in again, in case you had to restart R for some reason
gnr.long <- read.csv("samplegnr.csv",
                    header=T, stringsAsFactors = F)
```

Step 4: Return an ID Number that Matches the Accepted Name (According to Plants of the World)

- this step will prompt you if the database is unsure about which name to return. You have to specify the row number of taxa you want to select as it runs through them.

```
#fetching IPNI ids from POW
pow.output <- get_pow(gnr.long$matched_name2, db = "pow", accepted = TRUE, ask = TRUE)
```

```
head(pow.output)
```

```
## [1] "urn:lsid:ipni.org:names:77100466-1"
## [2] "urn:lsid:ipni.org:names:77100466-1"
## [3] NA
## [4] "urn:lsid:ipni.org:names:470659-1"
## [5] "urn:lsid:ipni.org:names:93225-1"
## [6] "urn:lsid:ipni.org:names:340112-1"
```

-You can see that and IPNI id has been returned in the “ids” column. -Because the get_pow() function can be time consuming for long species lists, it's probably wise to save the output as a csv.

```
#writing to csv file
write.csv(as.data.frame(pow.output),
          "sample_pow_output.csv",
          row.names = F, quote = F)
```

```
#reading it in again, in case you had to restart R for some reason
pow.output <- read.csv("sample_pow_output.csv",
                    header=T, stringsAsFactors = F)
```

-Now to combine the IPNI IDs back to the original names

```
inputname_ID <- as.data.frame(cbind(gnr.long$user_supplied_name, gnr.long$matched_name2, pow.output$ids))
```

```
colnames(inputname_ID) = c("user_supplied_name", "matched_name2", "IPNI_ID")
head(inputname_ID)
```

```
##          user_supplied_name          matched_name2
## 1      Bignonia tweediana      Bignonia tweediana
## 2      Dolichandra unguiscati      Dolichandra unguis-cati
## 3      Acacia farnesiana      Acacia farnesiana
## 4      Acacia koa      Acacia koa
## 5      Aristolochia philippinensis      Aristolochia philippinensis
## 6      Breynia vitis-idaea      Breynia vitis-idaea
##          IPNI_ID
## 1 urn:lsid:ipni.org:names:77100466-1
## 2 urn:lsid:ipni.org:names:77100466-1
## 3      <NA>
## 4 urn:lsid:ipni.org:names:470659-1
## 5 urn:lsid:ipni.org:names:93225-1
## 6 urn:lsid:ipni.org:names:340112-1
```

Step 5: Separate the troublemakers

```
#these names have returned an IPNI ID
inputname_ID.NAsdrop <- subset(inputname_ID, (!is.na(inputname_ID$IPNI_ID)))
nrow(inputname_ID.NAsdrop)
```

```
## [1] 18
```

```
#these names have NOT returned an IPNI ID and need special human attention
inputname_ID.NAonly <- subset(inputname_ID, (is.na(inputname_ID$IPNI_ID)))
nrow(inputname_ID.NAonly)
```

```
## [1] 7
```

```
#writing the trouble makers to csv file (deal with them later)
write.csv(inputname_ID.NAonly,
          "NA_NeedsHumanAttention.csv",
          row.names = F, quote = F)
```

Step 6: Use the IPNI ID to return data (name, rank, family, genus, native range, etc)

First, we'll get taxon name

```
#writing a function that will return the taxon name associated with the IPNI ID according to POW
Get_taxonname <- function(spnames){
  temp <- c()
  temp <- pow_lookup(spnames)
  ifelse(is.null(temp$meta$name) == "TRUE", "ERROR", temp$meta$name)
}
```

```
# applying the function to your list
POW.taxonname <- sapply(inputname_ID.NAsdrop$IPNI_ID[], Get_taxonname)
```

Second, we'll get family

```
#writing a function that will return the family name associated with the IPNI ID according to POW
Get_family <- function(spnames){
  temp <- c()
  temp <- pow_lookup(spnames)
  ifelse(is.null(temp$meta$family) == "TRUE", "ERROR", temp$meta$family)
}
```

```
# applying the function to your list
POW.family <- sapply(inputname_ID.NAsdrop$IPNI_ID[], Get_family)
```

Third, we'll get genus

```
#writing a function that will return the genus name associated with the IPNI ID according to POW
Get_genus <- function(spnames){
  temp <- c()
  temp <- pow_lookup(spnames)
  ifelse(is.null(temp$meta$genus) == "TRUE", "ERROR", temp$meta$genus)
}
```

```
# applying the function to your list
POW.genus <- sapply(inputname_ID.NAsdrop$IPNI_ID[], Get_genus)
```

Fourth, we'll get rank

```
#writing a function that will return the taxon rank associated with the IPNI ID according to POW
Get_taxonrank <- function(spnames){
  temp <- c()
  temp <- pow_lookup(spnames)
  ifelse(is.null(temp$meta$rank) == "TRUE", "ERROR", temp$meta$rank)
}
```

```
# applying the function to your list
# this will return a bunch of red text warnings, but disregard
POW.taxonrank <- sapply(inputname_ID.NAsdrop$IPNI_ID[], Get_taxonrank)
```

Fifth, we'll get author

```
#writing a function that will return the taxon rank associated with the IPNI ID according to POW
Get_author <- function(spnames){
  temp <- c()
  temp <- pow_lookup(spnames)
  ifelse(is.null(temp$meta$author) == "TRUE", "ERROR", temp$meta$author)
}
```

```
# applying the function to your list
# this will return a bunch of red text warnings, but disregard
POW.author <- sapply(inputname_ID.NAsdrop$IPNI_ID[], Get_author)
```

Step 7: combining the original names, IPNI IDs and other data into a single spreadsheet

```
#getting the data attached to the IPNI ID in one place
POWinput <- cbind(POW.family, POW.genus, POW.taxonname, POW.author, POW.taxonrank)
POWinput <- as.data.frame(POWinput)
head(POWinput)
```

```
##          POW.family    POW.genus          POW.taxonname
## 1  Bignoniaceae Dolichandra Dolichandra unguis-cati
## 2  Bignoniaceae Dolichandra Dolichandra unguis-cati
## 3    Fabaceae      Acacia      Acacia koa
## 4 Aristolochiaceae Aristolochia Aristolochia philippinensis
## 5  Phyllanthaceae      Breynia      Breynia vitis-idaea
## 6  Polypodiaceae Leptochilus      Leptochilus amplus
##          POW.author POW.taxonrank
## 1          (L.) L.G.Lohmann      SPECIES
## 2          (L.) L.G.Lohmann      SPECIES
## 3              A.Gray      SPECIES
## 4              Warb.      SPECIES
## 5 (Burm.f.) C.E.C.Fisch.      SPECIES
## 6      (F.Muell.) Noot.      SPECIES
```

```
POWAcc_IPNIid <- cbind(inputname_ID.NAsdrop, POWinput)
head(POWAcc_IPNIid)
```

```
##          user_supplied_name          matched_name2
## 1          Bignonia tweediana          Bignonia tweedieana
## 2          Dolichandra unguiscati          Dolichandra unguis-cati
## 4          Acacia koa          Acacia koa
## 5 Aristolochia philippinensis Aristolochia philippinensis
## 6          Breynia vitis-idaea          Breynia vitis-idaea
## 9          Colysis ampla          Colysis ampla
##          IPNI_ID          POW.family          POW.genus
## 1 urn:lsid:ipni.org:names:77100466-1          Bignoniaceae          Dolichandra
## 2 urn:lsid:ipni.org:names:77100466-1          Bignoniaceae          Dolichandra
## 4   urn:lsid:ipni.org:names:470659-1          Fabaceae          Acacia
## 5   urn:lsid:ipni.org:names:93225-1 Aristolochiaceae Aristolochia
## 6   urn:lsid:ipni.org:names:340112-1          Phyllanthaceae          Breynia
## 9 urn:lsid:ipni.org:names:17568150-1          Polypodiaceae          Leptochilus
##          POW.taxonname          POW.author POW.taxonrank
## 1          Dolichandra unguis-cati          (L.) L.G.Lohmann          SPECIES
## 2          Dolichandra unguis-cati          (L.) L.G.Lohmann          SPECIES
## 4          Acacia koa          A.Gray          SPECIES
## 5 Aristolochia philippinensis          Warb.          SPECIES
## 6          Breynia vitis-idaea (Burm.f.) C.E.C.Fisch.          SPECIES
## 9          Leptochilus amplus          (F.Muell.) Noot.          SPECIES
```

Step 8: labelling the names that returned a duplicated IPNI ID and POW taxonname

- this is especially useful if you have many synonymous names in your user supplied names list

```
POWAcc_IPNIid <- data.table( POWAcc_IPNIid )
POWAcc_IPNIid<- POWAcc_IPNIid[ , duplicateCount := 1:.N , by = "IPNI_ID" ]
head(POWAcc_IPNIid)
```

```
##          user_supplied_name          matched_name2
## 1:          Bignonia tweediana          Bignonia tweedieana
## 2:          Dolichandra unguiscati          Dolichandra unguis-cati
## 3:          Acacia koa          Acacia koa
## 4: Aristolochia philippinensis Aristolochia philippinensis
## 5:          Breynia vitis-idaea          Breynia vitis-idaea
## 6:          Colysis ampla          Colysis ampla
##          IPNI_ID          POW.family          POW.genus
## 1: urn:lsid:ipni.org:names:77100466-1          Bignoniaceae          Dolichandra
## 2: urn:lsid:ipni.org:names:77100466-1          Bignoniaceae          Dolichandra
## 3:   urn:lsid:ipni.org:names:470659-1          Fabaceae          Acacia
## 4:   urn:lsid:ipni.org:names:93225-1 Aristolochiaceae Aristolochia
## 5:   urn:lsid:ipni.org:names:340112-1          Phyllanthaceae          Breynia
## 6: urn:lsid:ipni.org:names:17568150-1          Polypodiaceae          Leptochilus
##          POW.taxonname          POW.author POW.taxonrank
## 1:          Dolichandra unguis-cati          (L.) L.G.Lohmann          SPECIES
## 2:          Dolichandra unguis-cati          (L.) L.G.Lohmann          SPECIES
## 3:          Acacia koa          A.Gray          SPECIES
## 4: Aristolochia philippinensis          Warb.          SPECIES
## 5:          Breynia vitis-idaea (Burm.f.) C.E.C.Fisch.          SPECIES
## 6:          Leptochilus amplus          (F.Muell.) Noot.          SPECIES
##          duplicateCount
## 1:          1
## 2:          2
## 3:          1
## 4:          1
## 5:          1
## 6:          1
```

Done! ... wrting to csv file

```
#writing to csv file
write.csv(POWAcc_IPNIid,
          "POWAcc_IPNIid.csv",
          row.names = F)
```

Part 2: Getting IPNI IDs for Families and Genera in your species list

Step 1: Compile a list of both families and genera using your previous output

```
families <- as.character(POWAcc_IPNIid$POW.family)
genera <- as.character(POWAcc_IPNIid$POW.genus)
head(families,44)
```

```
## [1] "Bignoniaceae"      "Bignoniaceae"      "Fabaceae"
## [4] "Aristolochiaceae"  "Phyllanthaceae"    "Polypodiaceae"
## [7] "Fabaceae"          "Amaryllidaceae"    "Acanthaceae"
## [10] "Elaeocarpaceae"   "Fouquieriaceae"    "Malvaceae"
## [13] "Oleaceae"          "Acanthaceae"       "Loranthaceae"
## [16] "Proteaceae"       "Rubiaceae"         "Fabaceae"
```

```
#remove duplicates
fam.dupsremoved <- unique( families[] )
gen.dupsremoved <- unique( genera[] )
fam.gen <- c(fam.dupsremoved,gen.dupsremoved)
```

```
length(fam.dupsremoved)
```

```
## [1] 14
```

```
length(gen.dupsremoved)
```

```
## [1] 17
```

Step 2: Fetch IPNI IDs and associated data for families

```
fam.IDs <- get_pow(fam.dupsremoved, db = "pow", rank_filter = "family", accepted = TRUE, ask = TRUE)
```

Step 3: Fetch IPNI IDs and associated data for genera

```
gen.IDs <- get_pow(gen.dupsremoved, db = "pow", rank_filter = "genus", accepted = TRUE, ask = TRUE)
```

```
fam.IDs.df <- as.data.frame(fam.IDs)
fam.IDs.df <- cbind(fam.dupsremoved, fam.IDs.df$ids)
colnames(fam.IDs.df) <- c("names", "IPNI_ID")
gen.IDs.df <- as.data.frame(gen.IDs)
gen.IDs.df <- cbind(gen.dupsremoved, gen.IDs.df$ids)
colnames(gen.IDs.df) <- c("names", "IPNI_ID")
head(gen.IDs.df)
```

```
##      names      IPNI_ID
## [1,] "Dolichandra" "urn:lsid:ipni.org:names:4063-1"
## [2,] "Acacia"      NA
## [3,] "Aristolochia" "urn:lsid:ipni.org:names:330834-2"
## [4,] "Breyenia"    "urn:lsid:ipni.org:names:327612-2"
## [5,] "Leptochilus" "urn:lsid:ipni.org:names:17132410-1"
## [6,] "Chamaecrista" "urn:lsid:ipni.org:names:329389-2"
```

```
#combine genera and families into a single dataframe
fam.gen.IDs.df <- rbind(fam.IDs.df, gen.IDs.df)
fam.gen.IDs.df <- as.data.frame(fam.gen.IDs.df)
nrow(fam.gen.IDs.df)
```

```
## [1] 31
```

Step 4: Separate the troublemakers

```
#these names have returned an IPNI ID
fam.gen.ID.NAsdrop <- subset(fam.gen.IDs.df, (!is.na(fam.gen.IDs.df$IPNI_ID)))
nrow(fam.gen.ID.NAsdrop)
```

```
## [1] 20
```

```
#these names have NOT returned an IPNI ID and need special human attention
fam.gen.IDs.NAonly <- subset(fam.gen.IDs.df, (is.na(fam.gen.IDs.df$IPNI_ID)))
nrow(fam.gen.IDs.NAonly)
```

```
## [1] 11
```

```
#writing the trouble makers to csv file (deal with them later)
write.csv(fam.gen.IDs.NAonly,
          "NA_Fam_Genera_NeedsHumanAttention.csv",
          row.names = F, quote = F)
```

Step 5: Fetch data from POW database using IPNI ID

```
POW.fam.gen.name <- sapply(fam.gen.ID.NAsdrop$IPNI_ID[], Get_taxonname)
POW.fam.gen.author <- sapply(fam.gen.ID.NAsdrop$IPNI_ID[], Get_author)
POW.fam.gen.rank <- sapply(fam.gen.ID.NAsdrop$IPNI_ID[], Get_taxonrank)
POW.fam.gen.family <- sapply(fam.gen.ID.NAsdrop$IPNI_ID[], Get_family)
```

```
#getting the data attached in one place
POWinput.fam.gen <- cbind(POW.fam.gen.family, POW.fam.gen.name, POW.fam.gen.author, POW.fam.gen.rank)
POWinput.fam.gen <- as.data.frame(POWinput.fam.gen)
head(POWinput.fam.gen)
```

```
## POW.fam.gen.family POW.fam.gen.name POW.fam.gen.author POW.fam.gen.rank
## 1      Bignoniaceae      Bignoniaceae      Juss.      FAMILY
## 2  Aristolochiaceae  Aristolochiaceae      Juss.      FAMILY
## 3    Polypodiaceae    Polypodiaceae J.Presl & C.Presl      FAMILY
## 4    Amaryllidaceae    Amaryllidaceae      J.St.-Hil.      FAMILY
## 5      Acanthaceae      Acanthaceae      Juss.      FAMILY
## 6      Oleaceae       Oleaceae Hoffmanns. & Link      FAMILY
```

```
POWAcc_IPNIid.fam.gen <- cbind(fam.gen.ID.NAsdrop, POWinput.fam.gen)
head(POWAcc_IPNIid.fam.gen)
```



```
##          names          IPNI_ID POW.fam.gen.family
## 1      Bignoniaceae urn:lsid:ipni.org:names:30000204-2      Bignoniaceae
## 3  Aristolochiaceae urn:lsid:ipni.org:names:30000909-2  Aristolochiaceae
## 5      Polypodiaceae urn:lsid:ipni.org:names:30000471-2      Polypodiaceae
## 6      Amaryllidaceae urn:lsid:ipni.org:names:30000959-2  Amaryllidaceae
## 7      Acanthaceae urn:lsid:ipni.org:names:30000618-2      Acanthaceae
## 11      Oleaceae urn:lsid:ipni.org:names:30001336-2      Oleaceae
## POW.fam.gen.name POW.fam.gen.author POW.fam.gen.rank
## 1      Bignoniaceae          Juss.          FAMILY
## 3  Aristolochiaceae          Juss.          FAMILY
## 5      Polypodiaceae  J.Presl & C.Presl      FAMILY
## 6      Amaryllidaceae          J.St.-Hil.      FAMILY
## 7      Acanthaceae          Juss.          FAMILY
## 11      Oleaceae  Hoffmanns. & Link      FAMILY
```

Done! ... wrting to csv file

```
#writing to csv file
write.csv(POWAcc_IPNIid.fam.gen,
          "POWAcc_IPNIid.fam.gen.csv",
          row.names = F)
```