

# Obtaining host-parasite occurrence data from Genbank

Andrew Park

```
library(genbankr)
library(rentrez)
library(stringr)
library(tibble)
library(dplyr)
library(magrittr)

findHostsbySeq <- function(x) {
  # FN: finds hosts of parasite based on genbank sequences
  # ARGS: x=character parasite latin binom RTN:
  # df=tibble(para.name,host.name,uid,acc.num)
  df <- tibble(para.name = character(), host.name = character(),
               uid = character(), acc.num = character())
  para.uid <- entrez_search(db = "nucleotide", term = x)
  for (i in para.uid$ids) {
    para.acc <- entrez_fetch(db = "sequences", id = i, rettype = "acc")
    para.acc <- str_replace_all(para.acc, "\n", "")
    host <- ifelse(para.acc != "", readGenBank(GBAccession(para.acc),
                                                         ret.seq = F, partial = T, verbose = F)@sources@elementMetadata@listData$host,
                  "")
    tmp <- tibble(para.name = x, host.name = host, uid = i,
                  acc.num = para.acc)
    df <- bind_rows(df, tmp)
  }
  df %<>% filter(host.name != "" & host.name != x)
  return(df)
}
```

```
q<-findHostsbySeq("Amblyomma ovale") %>% print
```

```
## # A tibble: 7 x 4
##   para.name      host.name      uid      acc.num
##   <chr>         <chr>         <chr>    <chr>
## 1 Amblyomma ovale Akodon montensis 1147163806 KX137898.1
## 2 Amblyomma ovale Akodon montensis 1021221716 KU894385.1
## 3 Amblyomma ovale Euryoryzomys russatus 1021221715 KU894384.1
## 4 Amblyomma ovale Euryoryzomys russatus 1021221714 KU894383.1
## 5 Amblyomma ovale Euryoryzomys russatus 1021221713 KU894382.1
## 6 Amblyomma ovale Hylaeamys megacephalus 1021221712 KU894381.1
## 7 Amblyomma ovale Hylaeamys megacephalus 1021221711 KU894380.1
```