

# Phylo\_reutils

Anna\_Balan

2023-02-11

```
if (!("reutils" %in% installed.packages()))  
  install.packages("reutils")  
library(reutils)  
options(reutils.email = "your_email@gmail.com")
```

1. searches PubMed for articles of interest to abstracts articles in plain text format;

```
esearch(db = "pubmed", term = "crustacyanin")
```

```
## Object of class 'esearch'  
## List of UIDs from the 'pubmed' database.  
## [1] "35247793" "35010161" "34436301" "33919403" "33465290" "32851672"  
## [7] "32596057" "32236233" "31813041" "30860355" "29683674" "29178679"  
## [13] "28851818" "26220698" "25797168" "25605312" "24782450" "23570752"  
## [19] "23510436" "23441225" "22869108" "22428138" "22189778" "21391640"  
## [25] "21169698" "19579223" "19416706" "19414522" "19317475" "19299880"  
## [31] "19058530" "18667761" "17374944" "17188641" "17124125" "17124122"  
## [37] "17028694" "16833638" "16407115" "15686376" "15644340" "14993674"  
## [43] "14770227" "14646064" "12876374" "12832753" "12782314" "12777800"  
## [49] "12123366" "12119396" "11526314" "11526313" "11341939" "10944355"  
## [55] "10604288" "9761813" "9200677" "11540431" "8931133" "15299714"  
## [61] "7698348" "11542700" "1548709" "1935978" "2026162" "2001254"  
## [67] "2306227" "4033433" "6202261" "6627105" "7419516" "760804"  
## [73] "830471" "5644143" "6078541" "4959560" "5971798" "14234502"  
## [79] "18933429"
```

```
ms <- esearch(db = "pubmed", term = "crustacyanin")  
abstr <- efetch(ms, rettype = "abstract")  
abstr
```

```
## Object of class 'efetch'  
## 1. Comp Biochem Physiol Part D Genomics Proteomics. 2022 Jun;42:100977. doi:  
## 10.1016/j.cbd.2022.100977. Epub 2022 Feb 16.  
##  
## Searching and identifying pigmentation genes from Neocaridina denticulate  
## sinensis via comparison of transcriptome in different color strains.  
##  
## Lin S(1), Zhang L(2), Wang G(1), Huang S(1), Wang Y(1).  
##
```

```
## Author information:
## (1)Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of
## Agriculture, Fisheries College, Jimei University, Xiamen 361021, China.
## (2)Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of
## ...
## EFetch query using the 'pubmed' database.
## Query url: 'https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?efe...'
## Retrieval type: 'abstract', retrieval mode: 'text'

write(content(abstr), "abstracts.txt")
```

2. request to the nucleotide database sequences all in the footsteps by gene name for an organism by name view and returns a list of identities or information about their number;

```
esearch(db = "nucleotide", term = "crustacyanin")
```

```
## Object of class 'esearch'
## List of UIDs from the 'nucleotide' database.
## [1] "2312022563" "2312022559" "2312022555" "2311876279" "2311876276"
## [6] "2311876274" "2311876272" "2311863192" "2311863190" "2311863187"
## [11] "2311863185" "2311848461" "2307929728" "2307929724" "2307929643"
## [16] "2307928020" "2311226022" "2301310459" "2277724840" "2277724838"
## [21] "2231603375" "2231601330" "2231597514" "2231595668" "2231593297"
## [26] "2231593226" "2214798008" "2214798006" "2214797999" "2214797990"
## [31] "2214797986" "2214797970" "2175922880" "2175922879" "2171713983"
## [36] "2171713967" "2171713949" "2171713932" "2171713927" "2171713898"
## [41] "2154992186" "2154992038" "2128399489" "2096064420" "2077536330"
## [46] "2077535960" "2077535433" "2077534997" "2077534241" "2077530957"
## [51] "2077530956" "2077530797" "2077525862" "2077525742" "2077524696"
## [56] "2077524339" "2065206281" "2065193120" "2065190079" "2065188392"
## [61] "2065186734" "2065171964" "2065159855" "2065028193" "2056514813"
## [66] "2056498811" "2056493211" "2056492151" "2056488359" "2056483089"
## [71] "2056467578" "2056465480" "2056439345" "2037090538" "2032923631"
## [76] "1953685317" "1953635727" "1950720059" "1950720057" "1941200435"
## [81] "1935954895" "1935954881" "1933347811" "1933347799" "1933347780"
## [86] "1933347777" "1933347776" "1933347770" "1933329650" "1838624700"
## [91] "1511198136" "1511198134" "1511198132" "1721459068" "1721459065"
## [96] "1721459062" "1511198335" "1511198333" "1644883274" "1595304136"
```

```
esearch(db = "nucleotide", term = "crustacyanin AND human[orgn]") #human doesnt have this protein
```

```
## Object of class 'esearch'
## List of UIDs from the 'nucleotide' database.
## [1] "NA"
```

```
esearch(db = "nucleotide", term = "crustacyanin AND lobster[orgn]") #it doesnt know lobster
```

```
## Error(s):
## PhraseNotFound lobster[orgn]
```

```
## Warning(s):
##   OutputMessage   No items found.
```

```
## Object of class 'esearch'
##   PhraseNotFound
## "lobster[orgn]"
##   OutputMessage
## "No items found."
```

```
esearch(db = "nucleotide", term = "crustacyanin AND Homarus americanus[orgn]")
```

```
## Warning: HTTPS error: Status 429;
```

```
## Object of class 'esearch'
## [1] "HTTPS error: Status 429; "
```

```
crnc <- esearch(db = "nucleotide", term = "crustacyanin AND Homarus americanus[orgn]")
```

### 3. searches for an organism ID by name on the base;

```
esearch(db = "taxonomy", term = "Homarus americanus")
```

```
## Object of class 'esearch'
## List of UIDs from the 'taxonomy' database.
## [1] "6706"
```

```
esearch(db = "taxonomy", term = "Human")
```

```
## Object of class 'esearch'
## List of UIDs from the 'taxonomy' database.
## [1] "9606"
```

```
esearch(db = "taxonomy", term = "Homo sapiens")
```

```
## Warning: HTTPS error: Status 429;
```

```
## Object of class 'esearch'
## [1] "HTTPS error: Status 429; "
```

```
esearch(db = "taxonomy", term = "Mouse") #why two species??? we dont know
```

```
## Object of class 'esearch'
## List of UIDs from the 'taxonomy' database.
## [1] "10090" "10088"
```

```
esearch(db = "taxonomy", term = "Ape") #why two species??? we dont know
```

```
## Object of class 'esearch'
## List of UIDs from the 'taxonomy' database.
## [1] "314295" "4456"
```

```
#efetch(db = "taxonomy", uid = apes) #doesnt work for some reason
```

4. requests to protein databases or nucleotide sequences by name of the gene, after which it returns table with UID (in XML this field is called Id), inventory number (in XML this field is called Caption), long in direction (Slen);

```
crcnp <- esearch(db = "protein", term = "crustacyanin AND Homarus americanus[orgn]")
su <- esummary(crcnp)
```

```
## Warning: HTTPS error: Status 429;
```

```
cosu <- content(su, "parsed")
```

```
## Warning: Errors parsing DocumentSummary
```

```
as.data.frame(cosu[c("Id", "Caption", "Slen")])
```

```
## data frame with 0 columns and 0 rows
```

5. gives nucleotide or protein bases text query sequences, and then writes the sequences to a file in fasta format (show the beginning of the file);

```
s <- esearch(db = "protein", term = "crustacyanin AND Homarus americanus[orgn]")
f <- efetch(uid = s[1:10], db = "protein", rettype = "fasta", retmode = "text")
write(content(f), "Ham_crcn.fa")
fastaf <- readLines("Ham_crcn.fa")
head(fastaf)
```

```
## [1] ">XP_042236484.1 crustacyanin-C1 subunit-like [Homarus americanus]"
## [2] "MNSLSILLVFVASVAADKIPDFVVPVKCASVDRNKLWAEQTPNRNNYAGVWYQFALTNNPYQLIEKCVRN"
## [3] "EYSFDGEQFVITSTGIAYDGNLLKRNGKLYPNPFGEPHLSIDYENSFAAPLVILETDYSNYACLYSCIDY"
## [4] "NFGYHSDFSIFSRSANLAEQYVKKCEAAFKNINVDTRFVKTVQGSSCPYDTQKTL"
## [5] ""
## [6] ">XP_042236483.1 crustacyanin-A2 subunit-like [Homarus americanus]"
```

6. downloads a protein corresponding to a known nucleotide UID;

```
lnk1 <- elink(uid = "2065188392", dbFrom = "nucleotide", dbTo = "protein")
efetch(lnk1, rettype = "fasta", retmode = "text")
```

```
## Object of class 'efetch'
## >XP_042223242.1 crustacyanin-A2 subunit-like [Homarus americanus]
## MGVWYEIQAQPNIFQSIKSLASSYKRVKTEIHVLSEGLDSSGASTTTKSILKIVDPQNP AHMVTDFVPG
## VEPPFDIVDTDYKTFSCAHSCLSIVGIKTEFVFIYSRNRTLRSNSTQHCLSIFEVSIIGIISFYTNANNY
##
##
## ...
## EFetch query using the 'protein' database.
## Query url: 'https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?efe...'
## Retrieval type: 'fasta', retrieval mode: 'text'
```

7. downloads all sequences from work with PMID ... (for example, from the first task) and writes them to the fasta file.

```
ms2 <- esearch(term = "lobster microsporidia", db = "pubmed")
lnk <- elink(ms2[4], dbFrom = "pubmed", dbTo = "nuccore")
f2 <- efetch(lnk, rettype = "fasta", retmode = "text")
write(content(f2), "lobster_microsporidia.fa")
```