

```
from Bio import Entrez
Entrez.email = 'annabalan267@gmail.com'
```

## ▼ 1. esearch searches articles in NCBI PubMed

```
handle = Entrez.esearch(db = "pubmed", term = "crustacyanin")
record = Entrez.read(handle)
print(record)
```

```
{'Count': '79', 'RetMax': '20', 'RetStart': '0', 'IdList': ['35247793', '35010161', '34436301', '33919403', '33465290',
```

## ▼ 2. efetch returns abstracts of 3 first articles

```
mshandle = Entrez.efetch(db="pubmed", id=record["IdList"][0:3], rettype="abstract", retmode="text")
print(mshandle.read())
```

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 Agriculture, Fisheries College, Jimei University, Xiamen 361021, China.

Electronic address: [llzhang@jmu.edu.cn](mailto:llzhang@jmu.edu.cn).

Aquaria species are characterized by their amazing colors and patterns. Research on the breeding molecular genetics of ornamental shrimps is surprisingly limited. We conducted a transcriptome analysis to investigate the expression of encoding genes in the integument of the strains Neocaridina denticulate sinensis. After assembled and filtered, 19,992 unigenes were annotated by aligning with public functional databases (NR, Swiss-Prot, KEGG, COG). 14,915 unigenes with significantly different expressions were found by comparing three strains integument transcriptomes. Ribosomal protein genes, ABC transporter families, calmodulin, carotenoid proteins and crustacyanin may play roles in the cytological process of pigment migration and chromatophore maintenance. Numerous color genes associated with multiple pathways including melanin, ommochrome and pteridines pathways were identified. The expression patterns of 25 candidate genes were analysis by qPCR in red, yellow, transparent and glass strains. The qPCR results in red, yellow and transparent were consistent with the level of RPKM values in the transcriptomes. The above results will advance our knowledge of integument color varieties in N. denticulate sinensis and help the genetic selection of crustaceans with consumer-favored colors. Furthermore, it also provides some candidate pigmentation genes to investigate the correlation between coloration and sympatric speciation in crustaceans.

Copyright © 2022. Published by Elsevier Inc.

DOI: 10.1016/j.cbd.2022.100977

PMID: 35247793 [Indexed for MEDLINE]

2. Foods. 2021 Dec 23;11(1):35. doi: 10.3390/foods11010035.

Purification and Characterisation of Two Novel Pigment Proteins from the Carapace of Red Swamp Crayfish (Procambarus clarkii).

Chen H(1)(2), Ji H(1)(3)(4)(5)(6), Pan C(6)(7), Zhang D(1)(3)(4)(5), Su W(1)(3)(4)(5), Liu S(1)(3)(4)(5)(6), Deng Y(1), Huang X(1).

Author information:

(1)Guangdong Provincial Key Laboratory of Aquatic Product Processing and Safety, College of Food Science and Technology, Guangdong Ocean University, Zhanjiang 524088, China.

(2)Hunan Provincial Key Laboratory of Soybean Products Processing and Safety Control, College of Food and Chemical Engineering, Shaoyang University, Shaoyang 422000, China.

(3)Guangdong Provincial Engineering Technology Research Center of Seafood, College of Food Science and Technology, Guangdong Ocean University, Zhanjiang

## ▼ 3. esearch searches in the base all the sequences for a certain gene and species, returning a list if IDs

```
handle = Entrez.esearch(db = "nucleotide", term = "crustacyanin AND Homarus[orgn]") #orgn=organism
record = Entrez.read(handle)
```

```
print(record)
Entrez.efetch(db = "nucleotide", id = record["IdList"])

{'Count': '19', 'RetMax': '19', 'RetStart': '0', 'IdList': ['2065206281', '2065193120', '2065190079', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392', '2065188392'], 'TranslationSet': [], 'TranslationStack': [{ 'Term': '2065188392', 'LinkSetDb': '1', 'Link': '1', 'Id': '2065188392'}]}
```

#### ▼ searches taxon ID with a given name

```
handle = Entrez.esearch(db = "taxonomy", term = "Homarus americanus")
record = Entrez.read(handle)
print(record)
print(record['IdList'])

{'Count': '1', 'RetMax': '1', 'RetStart': '0', 'IdList': ['6706'], 'TranslationSet': [], 'TranslationStack': [{ 'Term': '6706', 'LinkSetDb': '1', 'Link': '1', 'Id': '6706'}]}
```

#### ▼ 4. esearch+summary searches in database of proteins and nucl sequences with a name of a protein. Returns UID. actually its XML parcing.

```
handle = Entrez.esearch(db="protein", term="crustacyanin AND Homarus americanus[orgn]")
record = Entrez.read(handle)
for rec in record["IdList"]:
    temphandle = Entrez.read(Entrez.esummary(db="protein", id=rec, retmode="text"))
    print(temphandle[0]['Id']+"\t"+temphandle[0]['Caption']+"\t"+str(temphandle[0]['Length']))#+"\n")
##str(int(temphandle[0]['Length'])))
```

2068680993	XP_042236484	197
2068680990	XP_042236483	190
2068650119	XP_042225885	190
2068650116	XP_042225884	197
2068642615	XP_042227234	190
2068642613	XP_042227223	197
2068642611	XP_042227211	190
2068642608	XP_042227198	197
2068642605	XP_042227187	190
2068642602	XP_042227176	197
2056515232	KAG7177238	197
2056515231	KAG7177237	190
2056515230	KAG7177236	197
2056515229	KAG7177235	190
2056515228	KAG7177234	197
2056515227	KAG7177233	190
2056483091	KAG7166898	197
2056483090	KAG7166897	190
2056467580	KAG7160784	197
2056467579	KAG7160783	190

#### ▼ 5. returns fasta and writes it in a file

```
handle = Entrez.esearch(db="protein", term="crustacyanin AND Homarus americanus[orgn]")
record = Entrez.read(handle)

Entrez.efetch(db="protein", id=record["IdList"], retmode="text", rettype="fasta").read()
with open("crn.fasta", "w") as outf:
    for rec in record["IdList"]:
        lne = Entrez.efetch(db="protein", id=rec, retmode="text", rettype="fasta").read()
        outf.write(lne+"\n")
with open("crn.fasta", "r") as fastaf:
    snippet = [next(fastaf) for x in range(5)]
    print(snippet)

['>XP_042236484.1 crustacyanin-C1 subunit-like [Homarus americanus]\n', 'MNSLSILLVFVASVAADKIPDFVVPKGKASVDRNKLWAEQTPNRRNN']
```

#### ▼ 6. downloads a protein, takes a UID of a nucleotide

```
lhandle = Entrez.elink(dbfrom="nucleotide", db="protein", id="2065188392")
lrecord = Entrez.read(lhandle)
prothandle = lrecord[0]["LinkSetDb"][0]['Link'][0]['Id']
rrecord = Entrez.efetch(db="protein", id=prothandle, rettype="fasta", retmode="text")
with open("prot_from_nt.fasta", "w") as outf:
    outf.write(rrecord.read()+"\n")
```

## ▼ 7. Downloads fasta sequences from a work with PMID (e.g. from the first task)

```
lhandle = Entrez.elink(dbfrom="pubmed", db="nucleotide", id="20558169")
lrecord = Entrez.read(lhandle)
ids = []
for el in lrecord[0]["LinkSetDb"][0]["Link"]:
    ids.append(el['Id'])
rrecord = Entrez.efetch(db="nucleotide", id=ids[:4], rettype="fasta", retmode="text")
with open ("py_fasta_pmid.fasta", "w") as outf:
    outf.write(rrecord.read()+"\n")
```

## ▼ SSH

1. ищет в PubMed статьи по интересному для вас запросу и возвращает абстракты этих статей (можно N первых статей в списке) в простом текстовом формате (можно записать в файл);

```
esearch -email your@email.com -db pubmed -query "crustacyanin"
esearch -email your@email.com -db pubmed -query "crustacyanin AND lobster[orgn]" | efetch -mode text -format abstract
```

2. запрашивает в базе нуклеотидных последовательностей все последовательности по названию гена для организма по названию вида и возвращает список ID или информацию об их количестве;

```
esearch -email your@email.com -db nucleotide -query "crustacyanin AND Homarus americanus[orgn]" | esummary
```

3. ищет ID организма по названию в базе;

```
esearch -email your@email.com -db taxonomy -query "Homarus gammarus" | esummary | grep TaxId
```

4. запрашивает в базе белковых или нуклеотидных последовательностей по названию гена, после чего возвращает таблицу с UID (в XML это поле называется Id), accession number (в XML это поле называется Caption), длиной последовательности (Slen);

```
esearch -email your@email.com -db protein -query "crustacyanin AND Homarus americanus[orgn]" | esummary -mode xml -format doc
```

5. даёт в базу нуклеотидных или белковых последовательностей текстовый запрос, а затем пишет последовательности в файл в формате fasta (покажите начало файла);

```
esearch -email your@email.com -db protein -query "crustacyanin AND Homarus americanus[orgn]" | efetch -format fasta -mode text > lobster_msp.fasta
```

```
XP_042236484.1 crustacyanin-C1 subunit-like [Homarus americanus]
MNSLSILLVVASVAADKIPDFVVPKGKASVDRNKLWAEQTPNRNNYAGVWYQFALTNNPYQLIEKCVRN
EYSFDGEQFVITSTGIAYDGNLLKRNGKLYPNPFGEPHLSIDYENSFAAPLVILETDYSNYACLYSCIDY
NFGYHSDFSFIFSRANLAEQYVKCEAAFKNINVDTTFRVKTQVQSSCPYDTQKTL XP_042236483.1 crustacyanin-A2 subunit-like
[Homarus americanus] MFRTVIVAALVACVAADGIPSVFTAGKASVANQDNFDLRRYAGRWTTHIENAYQPVTRCINSNYEYS
GNDYGFVKVTTAGFNPNDYKIDFKVYPTKEFPAAHMLIDAPSVFAAPYEVIEDYDTYSCVYSCITTDN
YKSEFAFVFSRTPQTSGPAVEKCAAVFNKNGVEFSKFVPVSHAEVCYRA XP_042225885.1 crustacyanin-A2 subunit-like [Homarus
americanus] MFRTVIVAALVACVAADGIPSVFTAGKASVANQDNFDLPRYAGRWTTHIENAYQPVTRCINSNYEYS
```

6. скачивает белок, соответствующий известному UID нуклеотида;

```
link -id 2065188392 -db nuccore -target protein | efetch -db protein -format fasta -mode text
```

7. скачивает все последовательности из работы с PMID ... (например, из первого задания) и пишет их в файл fasta.

```
elink -db pubmed -target nucleotide -id 20558169 | efetch -format fasta -mode text > lobster_msp.fasta
```

