

API tutorial

To use the proposed scripts you need a linux environment.

I. Two ways to submit an API request

There are two possibilities to submit a request. The first type is to submit a sequence and the second a Pfam identifier.

A Pfam identifier (Pfam ID) uses the following format PFXXXXX (e.g. PF00111 for Ferredoxin) and can be downloaded directly from the Pfam website using OGA interface. Pfam IDs and entry annotation details can be found on <https://pfam.xfam.org>.

A. With a sequence using the [APIpipeline_seq.sh](#) script.

Three fields will need to be modified in the script :

1. The « requestName » which is a free text used to name downloadable files.
2. The « dbName » to choose the dataset to query such as OM- RGC_v1, OM- RGC_v2_metaG, OM-RGC_v2_metaT, MATOU_v1_metaG, MATOU_v1_metaT, SMAGs, 530MagProkArc_metaG, BacArcMag, MGT or 530MagProkArc_metaT.

```
#choose a name for the query
requestName="APITest"

#choose a dataset
dbName="OM-RGC_v1"
```

3. Add the sequence between the quote marks.

```
seq=">EAQ46983.1 Metallo-phosphoesterase [Roseobacter sp. MED193]\n\
MKYRTIFLSDIHLGTPGCQADLLLGLNSHEADTYLVDIVDAWRIRRKGFLLWPQAHNDVVQTLLAKAH\n\
DGARIFLIPGNHDEFLLRSYYGTHFGGIEVVATADVFASDGKRYLVTHGDQFDVAVVTNAKWLHLGDQAYE\n\
FMLWLNTRINRLRHLWGGQYWSLSKWAKHQVKQAVNFISEYENVLTAEARRGGYDGVICGHIHSAAIRDL\n\
EGMTYVNTGDWVESCTAIVERDNGSLALIDWERSARRSRHRARRARQKDKVLENV"

tool="blastp"
```

The sequence must be in FASTA format with a header of less than 100 characters (n is the line break). If you put a nucleotide sequence change tool="blastp" to tool="blastn".

B. With a Pfam_id usig the [APIpipeline_HMM.sh](#) script.

As previously, three fields will need to be modified in the script :

1. The « requestName » which is a free text used to name downloadable files.
2. The « dbName » to choose the dataset to query such as OM- RGC_v1, OM-RGC_v2_metaG, OM-RGC_v2_metaT, MATOU_v1_metaG, MATOU_v1_metaT, SMAGs, 530MagProkArc_metaG, BacArcMag, MGT or 530MagProkArc_metaT.

```
#choose a name for the query  
requestName="APitest"  
  
#choose a dataset  
dbName="OM-RGC_v1"
```

3. Add the Pfam_id between the quote marks.

```
#enter your Pfam_id  
Pfam_id="PF01228"
```

II. Submit your job

1. First, save the modified script.
2. In a linux terminal, go to the location of the script.
3. Make it executable with this command: *chmod +x name_of_script*
4. Then run the script with: *./name_of_script*

In the terminal, you have an estimate of the time with the "eta" comment.

III. Results

A directory will be created with the name corresponding to the "requestName" you have defined. Inside are the files of alignment result, homolog sequences, abundances and environmental data. In the terminal, you will get the URL of the result page.

Please do not submit or retrieve more than one request every 30 seconds.

A limit is set at 200 jobs per 24 hours and queries launched on the web interface have priority.