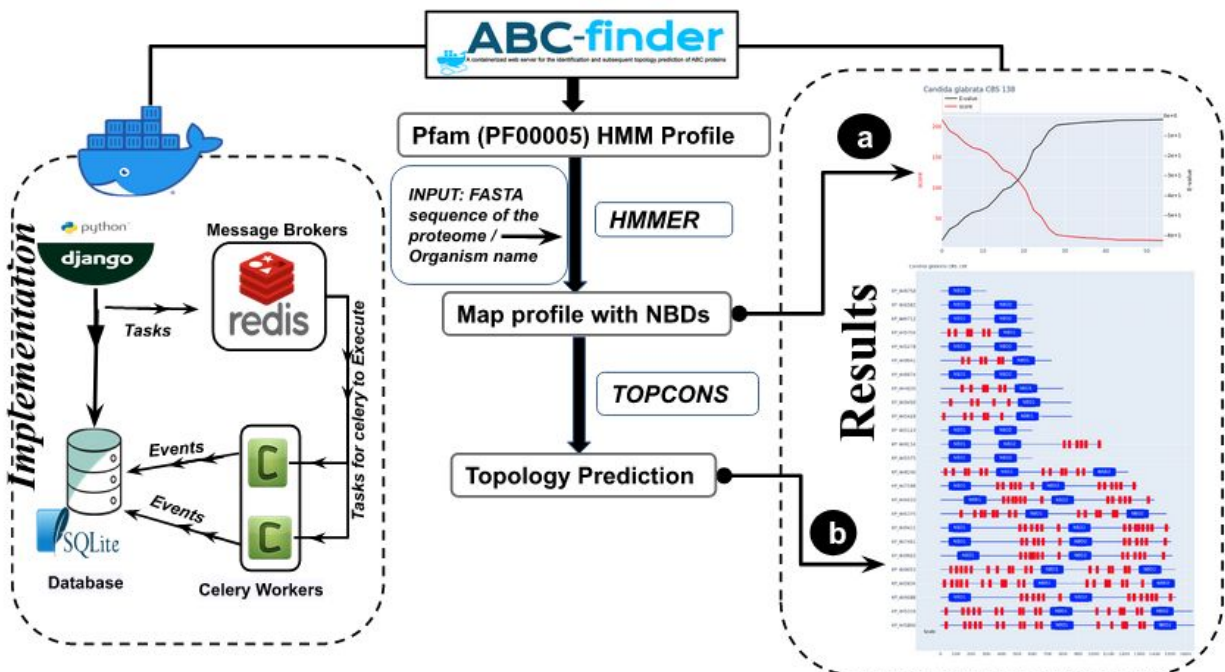


# Overview

ABC-Finder i.e., A Docker-based package for the identification of ABC proteins in all organisms, and downstream analysis and visualization of the topology of ABC proteins using an interactive web browser. ABC-Finder is built and deployed in a Linux container, making it scalable for many concurrent users on our servers and also enabling users to download and run ABC-Finder locally. Overall, ABC-Finder is an extremely convenient, portable and platform-independent package for the identification and subsequent domain prediction of ABC proteins.



## Installing

You can install ABC-finder-web server directly from the source code or build and run it from within Docker container.

### STEP 1: Install Docker Engine

Docker Engine is available on a variety of [Linux platforms](#), [macOS](#) and [Windows 10](#) through Docker Desktop, and as a [static binary installation](#). Find your preferred operating system below.

<https://docs.docker.com/engine/install/>

**NOTE:** We won't suggest to running ABC-finder on window or mac os , because ABC-finder is not a case of docker it's a more complicated it's a case of running a docker inside docker , so

until you have a advance skills in docker then please follow simple instructions on any linux based OS.

## STEP 2: Install Docker Compose

Docker Compose relies on Docker Engine for any meaningful work, so make sure you have Docker Engine installed either locally or remote, depending on your setup.

<https://docs.docker.com/compose/install/>

You can also install docker-compose using conda:

```
$ conda install -c conda-forge docker-compose
```

## STEP 3: Building the Docker container

First Clone ABC-finder Repository:

```
$ git clone https://github.com/lynngroup/abcfinder.git
```

Move into the repository directory:

```
$ cd abcfinder
```

## STEP 4: Building the docker images for ABC-finder

The file **Dockerfile** & **docker-compose.yaml** contains instructions for building a Docker container for ABC-finder-webserver.

If you have Docker & docker-compose installed on your computer, ABC-finder can be easily downloaded and installed using the following command:

```
$ docker-compose build
```

```
$ docker-compose up -d
```

## How to Access ABC-finder :

This will initiate ABC-finder at port 8000 of local server or localhost. The user may use another port to initiate another instance. [To manipulate Docker utilities refer to Docker Documentation] While the ABC-finder instance is running inside Docker container, ABC-finder User Interface (UI) can be accessed through a web browser at following URL:

<http://localhost:80> or

[http://IP\\_ADDRESS\\_OF\\_HOST\\_COMPUTER:80](http://IP_ADDRESS_OF_HOST_COMPUTER:80)

ABC-finder can now be used to upload your data using the browser.

**Average Accuracy Predicted by ABC-finder** for 14 Organisms listed below is **~95%**, **Except *Candida albicans* (~89% ) & *Trichophyton rubrum* ( ~70%)** for all other organisms, ABC-finder Accuracy is **>90%**. If you exclude ***Candida albicans* & *Trichophyton rubrum*** then Average accuracy will be ~98%.

***cd-hit parameter used:***

***-c = sequence identity threshold***

***-n = word\_length***

***Default = -c =.9 ; -n= 5***

**Table 1:** A comparison between Number of ABC proteins reported in Various Organisms VS Predicted by ABC-finder.

<b><i>S. no.</i></b>	<b><i>Organism Name</i></b>	<b><i>ABC proteins present</i></b>	<b><i>Predicted By ABC-finder</i></b>	<b><i>cd-hit parameter used</i></b>	<b><i>ABC-finder Accuracy %</i></b>
1	<i>Homo sapiens</i>	48	48	-c 0.6 -n 4	100
2	<i>Drosophila melanogaster</i>	56	55	Default	98
3	<i>Escherichia Coli</i>	79	77	Default	97

4	<i>Saccharomyces cerevisiae</i>	28	28	-c 0.7 -n 5	100
5	<i>Candida albicans</i>	28	25	Default	89
6	<i>Candida glabrata</i>	25	25	Default	100
7	<i>Candida auris</i>	28	28	Default	100
8	<i>Debaryomyces hansenii</i>	30	30	Default	100
9	<i>Magnaporthe oryzae</i> ( <i>Pyricularia oryzae</i> )	50	50	-c 1 -n 5	100
10	<i>Trichophyton rubrum</i>	26	37-40	Default & -c 0.4 -n 2	~70
11	<i>Caenorhabditis elegans</i>	60	56	Default	93

12	Leishmania major	42	41	-c 1 -n 5	98
13	Bombyx mori	53-55	53	Default	100
14	Oryza sativa	121	111	-c 1 -n 5	92
15	Arabidopsis thaliana	120	121	Default	99

## Frequently Asked Questions

### 1) How do I submit a sequence to ABC-finder?

<https://youtu.be/5ckaG4cXbRw>

### 2) How much time will it take to finish my sequence ?

It all depends on your sequence, if your sequence has a hit in the "preprocess pfam database" (If your sequence has a homologous family in Pfam) or not (see paper). **Without a hit it takes 5-10 minutes per sequence.**

For example *Debaryomyces hansenii* have 30 sequences so you can simply multiply  $5-10 \times 30 = 150-300 = 2.5-5$  hours i.e. to finish a 30 ABC-finder may take **2.5 to 5 hours**, you can calculate time according to your sequence, time will also vary according to your machine.

**Note:** Currently ABC-finder does not have caching, ABC-finder is a special case of running docker inside docker, ABC-finder itself is a docker container and it's using another docker container (topcons) inside it, so maybe later we try to enable caching on ABC-finder that will give you faster result if someone have already run your sequence.

### **3) Have you tested ABC-finder on some organism to validate your ABC-finder ?**

Yes we have tested ABC-finder on 15 organism, you can check result of all these 15 organism here: [TOPCONS Results for 15 organism](#)

### **4) How to install ABC-finder locally**

We have already explained all steps above in the Installing Section, please follow that.

### **5) Where can I contact, if my job is not finished or I have any doubt ?**

Please send your email to [abcfinder47@gmail.com](mailto:abcfinder47@gmail.com), you can also asks your doubts on github: <https://github.com/lynngroup/abcfinder/issues>

### **6) Can you please tell components of ABC-finder**

Please check ABC-finder related section

<http://abc-finder.osdd.jnu.ac.in/app/related>

### **7) How do I get notified about the status of my job on ABC-finder ?**

If you mention your email under the RUN section of ABC-finder then we will notify you once your job is started and we will also send you the result in a zip file once your job is finished.

### **8) Can you please give your docker hub link for docker images**

<https://hub.docker.com/r/lynngroup/abcfinder/tags> ( we are using 4 docker images tags )

1. lynngroup/abcfinder:db
2. lynngroup/abcfinder:redis
3. lynngroup/abcfinder:web
4. lynngroup/abcfinder:celery

### **9) Where to find the source code of ABC-finder ?**

<https://github.com/lynngroup/abcfinder>

## 10) What is the minimum hardware requirement for trying out ABC-finder locally ?

### Minimum

Linux 64 bits ( Ubuntu/centos7).

Any CPU (Intel i3/i5/ i7/ or Ryzen 3/5/7 , quad-core→ recommended )

4 GB RAM, 120 GB HDD Free Space.

### Recommended

Linux 64 bits ( Ubuntu/centos7).

**CPU quad-core or hexa-core Intel i7/Intel i9/ or Ryzen 5/7.**

8 GB RAM, 130 GB HDD Free Space. (ABC-finder require ~118 GB of TOPCONS database to be downloaded locally )

**Note:** ABC-finder uses CPU power and does not require GPU, you can easily change how many parallel jobs you want to run and what percentage of CPU you want to use, currently ABC finder uses max 60% of CPU. if you want to modify these setting, please send your email to [abcfinder47@gmail.com](mailto:abcfinder47@gmail.com)

## 11) What are these different types of files present inside my result folder ?

1. **Search.faa** → This is your fasta file that you have submitted to ABC-finder
2. **Search\_faa1.faa** → fasta file that will be your input for TOPCONS
3. **search\_faa0.clstr** → this file is a result of cd-hit.
4. **threshold .txt** → this file has no sequences that are taken from hmm\_output using threshold value to submit a sequence to ABC-finder.
5. **Hmm\_output** → it's a result of hmmer generated for your sequence to build a profile using hmmbuild
6. **Profile.hmm** → A profile HMM is a variant of an HMM relating specifically to biological sequences
7. **Plotv2.txt** → it's a final result file of TOPCONS

8. E-value\_Domain\_Score.pdf.pdf <http://abc-finder.osdd.jnu.ac.in/app/help>
9. topology\_plot.pdf→ this plot will show topology of ABC proteins for your sequence.  
<http://abc-finder.osdd.jnu.ac.in/app/help>

## Important Links:

1. ABC-finder is accessible at <http://abc-finder.osdd.jnu.ac.in>
2. *Project home page for source code:* <https://github.com/lynngroup/abcfinder>
3. *Docker Hub:* <https://hub.docker.com/r/lynngroup/abcfinder>
4. Zenodo : <https://zenodo.org/record/3733587#.X2YHHYbhUeZ>
5. ABC-finder Result for 14 Organisms Listed Below.  
[https://drive.google.com/drive/folders/ABC-finder\\_result](https://drive.google.com/drive/folders/ABC-finder_result)