ABC-finder: A containerized web server for the identification and topology prediction of ABC proteins

Poonam Vishwakarma[1#], Naveen Kumar Meena[1#], Rajendra Prasad[2]*, Andrew M. Lynn[1]* and Atanu Banerjee[2]*

[1]*School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India*

[2]*Amity Institute of Biotechnology and Amity Institute of Integrative Sciences and Health, Amity University Haryana, Gurugram, India*

 # contributed equally

**corresponding authors:*

Atanu Banerjee; abanerjee1@ggn.amity.edu

Andrew M. Lynn; andrew@jnu.ac.in

Rajendra Prasad; rprasad@ggn.amity.edu

ABSTRACT

In view of the multiple clinical and physiological implications of ABC transporter proteins, there is a considerable interest among researchers to characterize them functionally. However, such characterizations are based on the premise that ABC proteins are accurately identified in the proteome of an organism, and their topology is correctly predicted. With this objective, we have developed ABC-finder, i.e., a Docker-based package for the identification of ABC proteins in all organisms, and visualization of the topology of ABC proteins using a web browser. ABC-finder is built and deployed in a Linux container, making it scalable for many concurrent users on our servers and enabling users to download and run it locally. Overall, ABC-finder is a convenient, portable, and platform-independent tool for the identification and topology prediction of ABC proteins.

ABC-finder is accessible at http://abc-finder.osdd.jnu.ac.in

KEYWORDS

ABC proteins, Nucleotide-binding domains, Transmembrane domains, Docker, Reproducibility

RUNNING TITLE

ABC-finder web server

INTRODUCTION

ABC proteins comprise one of the largest and most important protein families, most proteins being involved in active transport [1]. A typical ABC transporter is composed of a pair of nucleotide-binding domains (NBDs) and transmembrane domains (TMDs) [2]. While the NBDs are involved in fueling the transport process by means of ATP hydrolysis, the TMDs are responsible for substrate recognition and form the translocation channel [3]. NBDs have several conserved motifs, namely, Walker A, Walker B, Signature sequence (C-motif), H-loop, D-loop, and Q-loop [4]. Contrarily, the TMDs display poor conservation across different subfamilies. HUGO Gene Nomenclature Committee divided the ABC superfamily into 7 subfamilies from ABCA to ABCG. This classification was later extended for non-mammalian proteins with the addition of the ABCH subfamily found in insects [5] and fishes [6,7], and the ABCI subfamily found in plants [8]. ABC transporters can function both as importers as well as exporters, however, the former is restricted only to bacteria and plants and mediate the uptake of various nutrients, micronutrients, and phytohormones, etc. [9]. On the other hand, the ABC exporters facilitate the extrusion of a wide spectrum of substrates including ions, lipids, peptides, toxic xenobiotics, etc. [10]. Numerous studies have suggested the role of ABC transporters in various human diseases and chemoresistance [1]. Besides their role as transporters, some ABC proteins harbor just the NBDs, for instance, the ABCE and ABCF family representatives, and have implications in ribosome biogenesis, translation control, etc. [11,12], further adding to the immense biological relevance of this superfamily. The primary requirement for in-depth investigations pertaining to ABC proteins is their accurate identification from the genome/proteome and analysis of its topology. Even though programs performing certain steps in isolation exist, there is a need for a unified package to do the job with lesser hassles and more emphasis on reproducibility, keeping in view that reproducibility of research is the key element in modern science [13]. In an effort to make a simpler program available to the biology researcher community, we herein present “ABC-finder” as a simple and fast tool for identification and topology prediction of ABC proteins based on our previously established pipeline which led to the inventorization of ABC proteins in a number of yeast species [14–16]. ABC-finder combines stand-out methodologies, namely profile-HMM and TOPCONS for homology detection, and prediction of the topology of membrane proteins, respectively in a seamless manner whereby the users need to provide only the organism’s name or the proteome

file as an input. Our analysis with reference organisms shows a strong correlation with the ABC protein inventories available in the literature.

METHODOLOGIES AND WORKFLOW

Data submission

Protein sequence data can be submitted in the following manner to the ABC-finder platform. The user can either specify the organism's name (*Homo sapiens*, *Arabidopsis thaliana*, or *Candida glabrata*) or can upload a raw FASTA file (uncompressed) of the proteome. The file size is, however, limited to 95 MB. In the last step, the user needs to provide an email address to get the result files directly delivered.

Method

Prediction of potential ABC proteins based on NBD HMM

To extract the putative ABC proteins from the sequence input, we utilized a protocol that was developed previously by us for inventorization of ABC proteins in yeasts [14–16]. Briefly, the HMM profile of the ABC-tran model (accession PF00005) obtained from the Pfam database [17] is used as a query to match against the proteome of an organism with the help of the "hmmsearch" function within the HMMER package [18] using the default settings. Positive hits above the default threshold are then further filtered based on a cutoff defined from the plot of domain-score and E-value. The default cut-off is system-generated from the E-value and requires manual filtering to remove proteins incorrectly classified above the cut-off. To automate this step, we use the observation that there is a small difference between the scores and E-values for proteins that belong to the same fold, but a large difference when scoring proteins from unrelated folds. This sudden drop in the sorted scores is exploited to mine proteins belonging to the defined class with minimal false positives.

Clustering of sequences using Cd-hit

Next, ABC-finder uses the Cluster Database at High Identity with Tolerance (Cd-hit) [19]. This clustering algorithm allows the grouping of all the sequences based on their sequence

similarities. Given that Cd-hit clusters the highly similar (redundant) sequences, it allows the grouping of sequences resulting from different transcript variants of each ABC protein that usually populate the database in case of higher eukaryotes. The user also has the option to optimize the input parameters for the program from the "advanced options" menu.

Prediction of the TMD and protein topology

Since most of the families under the ABC superfamily include membrane transporters, it is relevant to detect their presence among putative proteins. Following the Cd-hit run, ABC-finder utilizes the stand-out program TOPCONS [20], which is a widely used program for consensus prediction of membrane protein topology. Once the TM helices are defined using TOPCONS, ABC-finder uses a set of in-house written Python scripts to demarcate the NBDs and TMDs in each of the predicted ABC proteins.

Visualization

E-value and domain score values of all the proteins above the inclusion threshold, as obtained from the HMM analysis are presented in the form of a plot. This plot allows the further filtering of sequences as defined in the "Prediction of potential ABC proteins based on NBD HMM" section**.** The arrangement of TMDs and NBDs as obtained with the help of TOPCONS is used to generate the final topology diagram of the proteins, respectively. Dynamic and Static plots are generated with Plotly (https://plot.ly/) and Orca (https://plot.ly/python/orca-management/). All the results files and several other log files are generated and zipped as the output data for downloading.

The overall workflow implemented in the ABC-finder server is shown in **Figure 1**. The help page on the web server provides a step-by-step guide to ABC-finder. All results are kept on the server for ten days.

IMPLEMENTATION

ABC-finder server is implemented in Python 3.7.6 (http://www.python.org/) using the Web framework Django 3.0.3 (http://www.djangoproject.com/). The system is containerized using Docker 19.03.3 [21] and docker-compose 1.25.0. Job queuing is carried out by the asynchronous

task queue Celery 3.0.19 that uses the distributed message passing system Redis 3.3.11.

Users can also install the ABC-finder web server directly from the source code or build and run it from within Docker container using the instructions provided in the **Supplementary information**. The **Supplementary information** also contains FAQs and some useful links.

RESULTS AND DISCUSSION

Once the input proteome file is uploaded or the organism is specified to ABC-finder, the server provides a zipped result folder that includes the following files in different formats (indicated within the parenthesis):

1. **search(.faa)** → it is the proteome file that is submitted to ABC-finder for analysis. It is either uploaded by the user or retrieved by the web server depending upon the organism specified by the user.
2. **profile(.hmm)** → it is the profile HMM derived from the PF00005 seed alignment.
3. **hmm_output** → it is the output obtained after running the "hmmsearch" function with the user-specified proteome file using profile.hmm as the query.
4. **E-value_Domain_Score(.html/.pdf/.PNG/.SVG/.JPEG)** → it is the plot of E-value and domain score for the sequences above the inclusion threshold as obtained from the hmm_output file. A representative plot for the analysis of *S. cerevisiae* proteome is provided in **Figure 2**.
5. **threshold(.txt)** → this file lists the IDs of the putative ABC protein candidates that are identified in the query proteome. These are predicted based on a cutoff defined from the plot of domain-score and E-value.
6. **search_faa0(.clstr)** → it is the output file obtained after running Cd-hit. It contains the different protein clusters based on sequence similarity. Users will need to refer to this file in special cases where multiple variants are reported for each protein in the database.
7. **search_faa1(.faa)** → FASTA file utilized as the input for TOPCONS.
8. **query.result(.txt)** → it is the final output file obtained from TOPCONS.
9. **Topology Plot(.html/.pdf/.PNG/.SVG/.JPEG)** → this plot represents the topology of the identified ABC proteins. TOPCONS is utilized for the prediction of the transmembrane domains. A representative topology plot for the predicted ABC proteins

in the *S. cerevisiae* S288C strain is provided in **Figure 3**. The topology plot should be referred alongside the Cd-hit output file "search_faa0.clstr" to detect the highly similar sequence clusters, especially in the case of higher eukaryotes, wherein sequences from all the transcript variants of each protein are present in the database.

To evaluate the performance of ABC-finder, we utilized already reported inventories of ABC proteins from prokaryotic as well as eukaryotic systems. The results are summarized in **Table 1**. Overall, the results show a very good correlation with the number of ABC proteins that are reported in the literature for the reference organisms.

Since the ABC-finder protocol uses relatively stringent parameters, the results obtained should reflect the most likely ABC candidate proteins in the user-specified proteome. To test this further, we analyzed the results for *S. cerevisiae*, most of whose ABC proteins are experimentally characterized. As per earlier studies, the organism comprises 30 ABC proteins [22], 27 of which come with some degree of experimental characterization. As summarized in **Table 2**, ABC-finder could identify all 30 of the reported ABC proteins. In addition, it also identifies one uncharacterized protein YKR104W (NP_013030.1). Further sequence and structure-function analysis will help confirm its status as an ABC protein. Since, ABC-finder utilizes TOPCONS, which combines an arbitrary number of topology predictions into one consensus prediction [20] and combines it with the results of HMM-based NBD-mapping. Hence, it presents a highly plausible domain(s) arrangement (topology) in ABC proteins **(Figure 3)**. Furthermore, the Cd-hit aided clustering, which allows the grouping of the candidate proteins according to sequence similarity, can be exploited to classify the ABC members into various subfamilies.

Altogether, the web server caters to the ABC research community by providing a very simple tool that meets all the preliminary needs pertaining to sequence/domain analysis of the ABC proteins. Considering the need for analysis of larger datasets by some users, we have also made it feasible for the users to run ABC-finder locally. Given the numerous physiological and clinical implications of ABC proteins across the living world, the web server is an extremely useful tool that will be further improvised in the subsequent versions.

SOFTWARE AVAILABILITY

The scripts we used to deploy the containers in our study are publicly available on GitHub and

the Docker images are available on DockerHub.

GitHub URL (project home page for source code): https://github.com/lynngroup/abcfinder

DockerHub URL: https://hub.docker.com/r/lynngroup/abcfinder

FUNDING

AB, AML, and RP are supported by funding from the Department of Biotechnology, Government of India, grant no. BT/PR32349/MED/29/1456/2019.

ACKNOWLEDGEMENTS

We are grateful to all our colleagues involved in testing the ABC-finder web server. AB acknowledges funding support from SERB (SRG/2019/000514). PV and NKM are thankful to funding agencies ICMR and Department of Biotechnology (DBT/JRF/15/AL, DBT/2016/JNU/695), Govt. of India and JNU for financial support and fellowships. The authors acknowledge the High-Performance Computational Facility (HPCF) at SCIS, JNU.

CONFLICT OF INTEREST

None

REFERENCES

[1] C.F. Higgins, ABC transporters: physiology, structure and mechanism--an overview, Res. Microbiol. 152 (2001) 205–210. https://www.sciencedirect.com/science/article/pii/S0923250801011937.
[2] K.P. Locher, Mechanistic diversity in ATP-binding cassette (ABC) transporters, Nat. Struct. Mol. Biol. 23 (2016) 487–493. https://doi.org/10.1038/nsmb.3216.
[3] R. Prasad, A. Banerjee, N.K. Khandelwal, S. Dhamgaye, The ABCs of Candida albicans Multidrug Transporter Cdr1, Eukaryot. Cell. 14 (2015) 1154–1164. https://doi.org/10.1128/EC.00137-15.
[4] A. Moreno, A. Banerjee, R. Prasad, P. Falson, PDR-like ABC systems in pathogenic fungi, Res. Microbiol. 170 (2019) 417–425. https://doi.org/10.1016/j.resmic.2019.09.002.
[5] M. Dean, T. Annilo, Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates, Annu. Rev. Genomics Hum. Genet. 6 (2005) 123–142. https://doi.org/10.1146/annurev.genom.6.080604.162122.
[6] T. Luckenbach, S. Fischer, A. Sturm, Current advances on ABC drug transporters in fish, Comp. Biochem. Physiol. C. Toxicol. Pharmacol. 165 (2014) 28–52.

https://doi.org/10.1016/j.cbpc.2014.05.002.
[7] M. Popovic, R. Zaja, J. Loncar, T. Smital, A novel ABC transporter: the first insight into zebrafish (Danio rerio) ABCH1, Mar. Environ. Res. 69 Suppl (2010) S11–3. https://doi.org/10.1016/j.marenvres.2009.10.016.
[8] P.J. Verrier, D. Bird, B. Burla, E. Dassa, C. Forestier, M. Geisler, M. Klein, Ü. Kolukisaoglu, Y. Lee, E. Martinoia, Others, Plant ABC proteins--a unified nomenclature and updated inventory, Trends Plant Sci. 13 (2008) 151–159. https://www.sciencedirect.com/science/article/pii/S1360138508000368.
[9] O. Lewinson, N. Livnat-Levanon, Mechanism of Action of ABC Importers: Conservation, Divergence, and Physiological Adaptations, J. Mol. Biol. 429 (2017) 606–619. https://doi.org/10.1016/j.jmb.2017.01.010.
[10] K. Beis, Structural basis for the mechanism of ABC transporters, Biochem. Soc. Trans. 43 (2015) 889–893. https://doi.org/10.1042/BST20150047.
[11] V. Murina, M. Kasari, H. Takada, M. Hinnu, C.K. Saha, J.W. Grimshaw, T. Seki, M. Reith, M. Putrinš, T. Tenson, H. Strahl, V. Hauryliuk, G.C. Atkinson, ABCF ATPases Involved in Protein Synthesis, Ribosome Assembly and Antibiotic Resistance: Structural and Functional Diversification across the Tree of Life, J. Mol. Biol. 431 (2019) 3568–3590. https://doi.org/10.1016/j.jmb.2018.12.013.
[12] V. Vasiliou, K. Vasiliou, D.W. Nebert, Human ATP-binding cassette (ABC) transporter family, Hum. Genomics. 3 (2009) 281–290. https://doi.org/10.1186/1479-7364-3-3-281.
[13] M. Baker, 1,500 scientists lift the lid on reproducibility, Nature. 533 (2016) 452–454. https://doi.org/10.1038/533452a.
[14] S. Kumari, M. Kumar, N.K. Khandelwal, P. Kumari, M. Varma, P. Vishwakarma, G. Shahi, S. Sharma, A.M. Lynn, R. Prasad, N.A. Gaur, ABC transportome inventory of human pathogenic yeast Candida glabrata: Phylogenetic and expression analysis, PLoS One. 13 (2018) e0202993. https://doi.org/10.1371/journal.pone.0202993.
[15] M. Wasi, N.K. Khandelwal, P. Vishwakarma, A.M. Lynn, A.K. Mondal, R. Prasad, Inventory of ABC proteins and their putative role in salt and drug tolerance in Debaryomyces hansenii, Gene. 676 (2018) 227–242. https://doi.org/10.1016/j.gene.2018.07.029.
[16] M. Wasi, N.K. Khandelwal, A.J. Moorhouse, R. Nair, P. Vishwakarma, G. Bravo Ruiz, Z.K. Ross, A. Lorenz, S.M. Rudramurthy, A. Chakrabarti, A.M. Lynn, A.K. Mondal, N.A.R. Gow, R. Prasad, ABC Transporter Genes Show Upregulated Expression in Drug-Resistant Clinical Isolates of Candida auris: A Genome-Wide Characterization of ATP-Binding Cassette (ABC) Transporter Genes, Front. Microbiol. 10 (2019) 1445. https://doi.org/10.3389/fmicb.2019.01445.
[17] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279–85. https://doi.org/10.1093/nar/gkv1344.
[18] S.R. Eddy, A new generation of homology search tools based on probabilistic inference, Genome Inform. 23 (2009) 205–211. https://www.ncbi.nlm.nih.gov/pubmed/20180275.
[19] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics. 22 (2006) 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
[20] K.D. Tsirigos, C. Peters, N. Shu, L. Käll, A. Elofsson, The TOPCONS web server for

consensus prediction of membrane protein topology and signal peptides, Nucleic Acids Res. 43 (2015) W401–7. https://doi.org/10.1093/nar/gkv485.
[21] C. Boettiger, An introduction to Docker for reproducible research, Oper. Syst. Rev. 49 (2015) 71–79. https://doi.org/10.1145/2723872.2723882.
[22] C.M. Paumi, M. Chuk, J. Snider, I. Stagljar, S. Michaelis, ABC transporters in Saccharomyces cerevisiae and their interactors: new technology advances the biology of the ABCC (MRP) subfamily, Microbiol. Mol. Biol. Rev. 73 (2009) 577–593. https://doi.org/10.1128/MMBR.00020-09.
[23] M. Dean, A. Rzhetsky, R. Allikmets, The human ATP-binding cassette (ABC) transporter superfamily, Genome Res. 11 (2001) 1156–1166. https://doi.org/10.1101/gr.184901.
[24] W. Dermauw, T. Van Leeuwen, The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance, Insect Biochem. Mol. Biol. 45 (2014) 89–110. https://doi.org/10.1016/j.ibmb.2013.11.001.
[25] Z. Feng, D. Liu, L. Wang, Y. Wang, Z. Zang, Z. Liu, B. Song, L. Gu, Z. Fan, S. Yang, J. Chen, Y. Cui, A Putative Efflux Transporter of the ABC Family, YbhFSR, in Escherichia coli Functions in Tetracycline Efflux and Na+(Li+)/H+ Transport, Front. Microbiol. 11 (2020) 556. https://doi.org/10.3389/fmicb.2020.00556.
[26] Y. Kim, S.-Y. Park, D. Kim, J. Choi, Y.-H. Lee, J.-H. Lee, W. Choi, Genome-scale analysis of ABC transporter genes and characterization of the ABCC type transporter genes in Magnaporthe oryzae, Genomics. 101 (2013) 354–361. https://doi.org/10.1016/j.ygeno.2013.04.003.
[27] J.A. Sheps, S. Ralph, Z. Zhao, D.L. Baillie, V. Ling, The ABC transporter gene family of Caenorhabditis elegans has implications for the evolutionary dynamics of multidrug resistance in eukaryotes, Genome Biol. 5 (2004) R15. https://doi.org/10.1186/gb-2004-5-3-r15.
[28] P. Leprohon, D. Légaré, I. Girard, B. Papadopoulou, M. Ouellette, Modulation of Leishmania ABC protein gene expression through life stages and among drug-resistant parasites, Eukaryot. Cell. 5 (2006) 1713–1725. https://doi.org/10.1128/EC.00152-06.
[29] O. Garcia, P. Bouige, C. Forestier, E. Dassa, Inventory and comparative analysis of rice and Arabidopsis ATP-binding cassette (ABC) systems, J. Mol. Biol. 343 (2004) 249–265. https://doi.org/10.1016/j.jmb.2004.07.093.
[30] K. Gulshan, W.S. Moye-Rowley, Vacuolar import of phosphatidylcholine requires the ATP-binding cassette transporter Ybt1, Traffic. 12 (2011) 1257–1268. https://doi.org/10.1111/j.1600-0854.2011.01228.x.
[31] D. Wawrzycka, I. Sobczak, G. Bartosz, T. Bocer, S. Ułaszewski, A. Goffeau, Vmr 1p is a novel vacuolar multidrug resistance ABC transporter in Saccharomyces cerevisiae, FEMS Yeast Res. 10 (2010) 828–838. https://doi.org/10.1111/j.1567-1364.2010.00673.x.
[32] N.C. Rockwell, H. Wolfger, K. Kuchler, J. Thorner, ABC transporter Pdr10 regulates the membrane microenvironment of Pdr12 in Saccharomyces cerevisiae, J. Membr. Biol. 229 (2009) 27–52. https://doi.org/10.1007/s00232-009-9173-5.
[33] S. Petrovic, L. Pascolo, R. Gallo, F. Cupelli, J.D. Ostrow, A. Goffeau, C. Tiribelli, C.V. Bruschi, The products of YCF1 and YLL015w (BPT1) cooperate for the ATP-dependent vacuolar transport of unconjugated bilirubin in Saccharomyces cerevisiae, Yeast. 16 (2000) 561–571. https://doi.org/3.0.CO;2-L.">10.1002/(SICI)1097-0061(200004)16:6<561::AID-YEA551>3.0.CO;2-L.
[34] H. Wolfger, Y.M. Mamnun, K. Kuchler, The Yeast Pdr15p ATP-binding Cassette (ABC)

Protein Is a General Stress Response Factor Implicated in Cellular Detoxification, Journal of Biological Chemistry. 279 (2004) 11593–11599. https://doi.org/10.1074/jbc.m311282200.
[35] J. Golin, S.V. Ambudkar, M.M. Gottesman, A.D. Habib, J. Sczepanski, W. Ziccardi, L. May, Studies with novel Pdr5p substrates demonstrate a strong size dependence for xenobiotic efflux, J. Biol. Chem. 278 (2003) 5963–5969. https://doi.org/10.1074/jbc.M210908200.
[36] P. Piper, Y. Mahé, S. Thompson, R. Pandjaitan, C. Holyoak, R. Egner, M. Mühlbauer, P. Coote, K. Kuchler, The pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast, EMBO J. 17 (1998) 4257–4265. https://doi.org/10.1093/emboj/17.15.4257.
[37] J. Servos, E. Haase, M. Brendel, Gene SNQ2 of Saccharomyces cerevislae, which confers resistance to 4-nitroquinoline-N-oxide and other chemicals, encodes a 169 kDa protein homologous to ATP-dependent permeases, Molecular and General Genetics MGG. 236-236 (1993) 214–218. https://doi.org/10.1007/bf00277115.
[38] T.R. Cabrito, M.C. Teixeira, A. Singh, R. Prasad, I. Sá-Correia, The yeast ABC transporter Pdr18 (ORF YNR070w) controls plasma membrane sterol composition, playing a role in multidrug resistance, Biochem. J. 440 (2011) 195–202. https://doi.org/10.1042/BJ20110876.
[39] A. Decottignies, A.M. Grant, J.W. Nichols, H. de Wet, D.B. McIntosh, A. Goffeau, ATPase and multidrug transport activities of the overexpressed yeast ABC protein Yor1p, J. Biol. Chem. 273 (1998) 12612–12622. https://doi.org/10.1074/jbc.273.20.12612.
[40] M. Papay, C. Klein, I. Hapala, L. Petriskova, K. Kuchler, M. Valachovic, Mutations in the nucleotide-binding domain of putative sterol importers Aus1 and Pdr11 selectively affect utilization of exogenous sterol species in yeast, Yeast. 37 (2020) 5–14. https://doi.org/10.1002/yea.3456.
[41] S. Michaelis, STE6, the yeast a-factor transporter, Semin. Cell Biol. 4 (1993) 17–27. https://doi.org/10.1006/scel.1993.1003.
[42] V. Kasari, A.A. Pochopien, T. Margus, V. Murina, K. Turnbull, Y. Zhou, T. Nissan, M. Graf, J. Nováček, G.C. Atkinson, M.J.O. Johansson, D.N. Wilson, V. Hauryliuk, A role for the Saccharomyces cerevisiae ABCF protein New1 in translation termination/recycling, Nucleic Acids Res. 47 (2019) 8807–8820. https://doi.org/10.1093/nar/gkz600.
[43] K. Kiriyama, K.Y. Hara, A. Kondo, Extracellular glutathione fermentation using engineered Saccharomyces cerevisiae expressing a novel glutathione exporter, Appl. Microbiol. Biotechnol. 96 (2012) 1021–1027. https://doi.org/10.1007/s00253-012-4075-3.
[44] C.B.F. Andersen, T. Becker, M. Blau, M. Anand, M. Halic, B. Balar, T. Mielke, T. Boesen, J.S. Pedersen, C.M.T. Spahn, T.G. Kinzy, G.R. Andersen, R. Beckmann, Structure of eEF3 and the mechanism of transfer RNA release from the E-site, Nature. 443 (2006) 663–668. https://doi.org/10.1038/nature05126.
[45] K. Gościńska, S. Shahmoradi Ghahe, S. Domogała, U. Topf, Eukaryotic Elongation Factor 3 Protects Saccharomyces cerevisiae Yeast from Oxidative Stress, Genes . 11 (2020). https://doi.org/10.3390/genes11121432.
[46] C.W.T. van Roermund, L. IJlst, A. Baker, R.J.A. Wanders, F.L. Theodoulou, H.R. Waterham, The Saccharomyces cerevisiae ABC subfamily D transporter Pxa1/Pxa2p co-imports CoASH into the peroxisome, FEBS Lett. (2020). https://febs.onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.13974.

[47] A. Zutz, S. Gompf, H. Schägger, R. Tampé, Mitochondrial ABC proteins in health and disease, Biochim. Biophys. Acta. 1787 (2009) 681–690. https://doi.org/10.1016/j.bbabio.2009.02.009.
[48] D. Lockshon, L.E. Surface, E.O. Kerr, M. Kaeberlein, B.K. Kennedy, The sensitivity of yeast mutants to oleic acid implicates the peroxisome and other processes in membrane function, Genetics. 175 (2007) 77–91. https://doi.org/10.1534/genetics.106.064428.
[49] C.R.V. de Aldana, C.R.V. de Aldana, M.J. Marton, A.G. Hinnebusch, GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF-2 alpha kinase GCN2 in amino acid-starved cells, The EMBO Journal. 14 (1995) 3184–3199. https://doi.org/10.1002/j.1460-2075.1995.tb07321.x.
[50] L. Young, K. Leonhard, T. Tatsuta, J. Trowsdale, T. Langer, Role of the ABC transporter Mdl1 in peptide export from mitochondria, Science. 291 (2001) 2135–2138. https://doi.org/10.1126/science.1056957.
[51] V. Srinivasan, A.J. Pierik, R. Lill, Crystal structures of nucleotide-free and glutathione-bound mitochondrial ABC transporter Atm1, Science. 343 (2014) 1137–1140. https://doi.org/10.1126/science.1246729.
[52] J. Dong, R. Lai, J.L. Jennings, A.J. Link, A.G. Hinnebusch, The novel ATP-binding cassette protein ARB1 is a shuttling factor that stimulates 40S and 60S ribosome biogenesis, Mol. Cell. Biol. 25 (2005) 9859–9873. https://doi.org/10.1128/MCB.25.22.9859-9873.2005.
[53] A. Yarunin, V.G. Panse, E. Petfalski, C. Dez, D. Tollervey, E.C. Hurt, Functional link between ribosome formation and biogenesis of iron–sulfur proteins, EMBO J. 24 (2005) 580–588. https://doi.org/10.1038/sj.emboj.7600540.
[54] H.-Y. Liu, Y.-C. Chiang, J. Pan, J. Chen, C. Salvadore, D.C. Audino, V. Badarinarayana, V. Palaniswamy, B. Anderson, C.L. Denis, Characterization of CAF4 and CAF16 Reveals a Functional Connection between the CCR4-NOT Complex and a Subset of SRB Proteins of the RNA Polymerase II Holoenzyme, Journal of Biological Chemistry. 276 (2001) 7541–7548. https://doi.org/10.1074/jbc.m009112200.

TABLES

Table 1: A comparison between the number of ABC proteins reported in the literature v/s predicted by ABC-finder. cd-hit parameter used: **-c** = sequence identity threshold; **-n** = word length; **Default** = (-c = 0.9; -n = 5). Result files for all the below mentioned organisms are available at https://zenodo.org/record/4603195.

S. no.	**Organism name**	**ABC proteins reported [literature reference]**	**ABC proteins predicted by ABC-finder**	**Cd-hit parameter used**	**Match percent (between the literature reported and ABC-finder reported numbers)**
1	*Homo sapiens*	48-49 [12,23]	48	-c 0.6 -n 4	100
2	*Drosophila melanogaster*	56 [24]	55	Default	98
3	*Escherichia coli*	79 [25]	77	Default	97
4	*Saccharomyces cerevisiae*	30 [22]	31	Default	97
5	*Candida glabrata*	25 [14]	25	Default	100
6	*Candida auris*	28 [16]	28	Default	100

7	*Debaryomyces hansenii*	30 [15]	30	Default	100
8	*Magnaporthe oryzae*	50 [26]	50	-c 1 -n 5	100
9	*Caenorhabditis elegans*	60 [27]	56	Default	93
10	*Leishmania major*	42 [28]	41	-c 1 -n 5	98
11	*Bombyx mori*	53-55 [24]	53	Default	100
12	*Oryza sativa*	121 [29]	111	-c 1 -n 5	92
13	*Arabidopsis thaliana*	120 [29]	121	Default	99

Table 2: List of ABC proteins predicted by ABC-finder in the *S. cerevisiae* S288C proteome. NP_013030.1 has not been reported in previous inventories of *S. cerevisiae* ABC proteins.

S.no.	**Sequence ID**	**Protein details**	**Literature confirming status as ABC protein**
1	NP_013052.1	bile acid-transporting ATPase YBT1	[30]

2	NP_011828.1	putative ATP-binding cassette multidrug transporter VMR1	[31]
3	NP_014973.1	ATP-binding cassette multidrug transporter PDR10	[32]
4	NP_013086.1	ATP-binding cassette bilirubin transporter BPT1	[33]
5	NP_010694.1	ATP-binding cassette multidrug transporter PDR15	[34]
6	NP_014796.3	ATP-binding cassette multidrug transporter PDR5	[35]
7	NP_010419.3	ATP-binding cassette glutathione S-conjugate transporter YCF1	[33]
8	NP_015267.1	ATP-binding cassette multidrug transporter PDR12	[36]
9	NP_010294.1	ATP-binding cassette transporter SNQ2	[37]
10	NP_014468.3	ATP-binding cassette multidrug transporter PDR18	[38]
11	NP_011797.3	ATP-binding cassette transporter YOR1	[39]
12	NP_012252.1	ATP-binding cassette multidrug transporter PDR11	[40]
13	NP_014654.1	ATP-binding cassette sterol transporter AUS1	[40]
14	NP_014567.2	uncharacterized protein YOL075C	-

15	NP_012713.1	ATP-binding cassette a-factor transporter STE6	[41]
16	NP_013029	putative multidrug transporter NFT1	-
17	NP_015098.1	New1	[42]
18	NP_009937.2	putative ATP-dependent permease ADP1	[43]
19	NP_013350.1	translation elongation factor EF-3 (YEF3)	[11,44]
20	NP_014384.3	translation elongation factor EF-3 (HEF3)	[11,45]
21	NP_015178.1	ATP-binding cassette long-chain fatty acid transporter PXA1	[46]
22	NP_012733.1	ATP-binding cassette long-chain fatty acid transporter PXA2	[46]
23	NP_015053.2	ATP-binding cassette permease MDL2	[47,48]
24	NP_116664.1	putative AAA family ATPase GCN20	[49]
25	NP_013289.1	ATP-binding cassette permease MDL1	[50]
26	NP_014030.1	ATP-binding cassette Fe/S cluster precursor transporter ATM1	[51]
27	NP_010953.1	ATP-binding cassette family ATPase ARB1	[52]

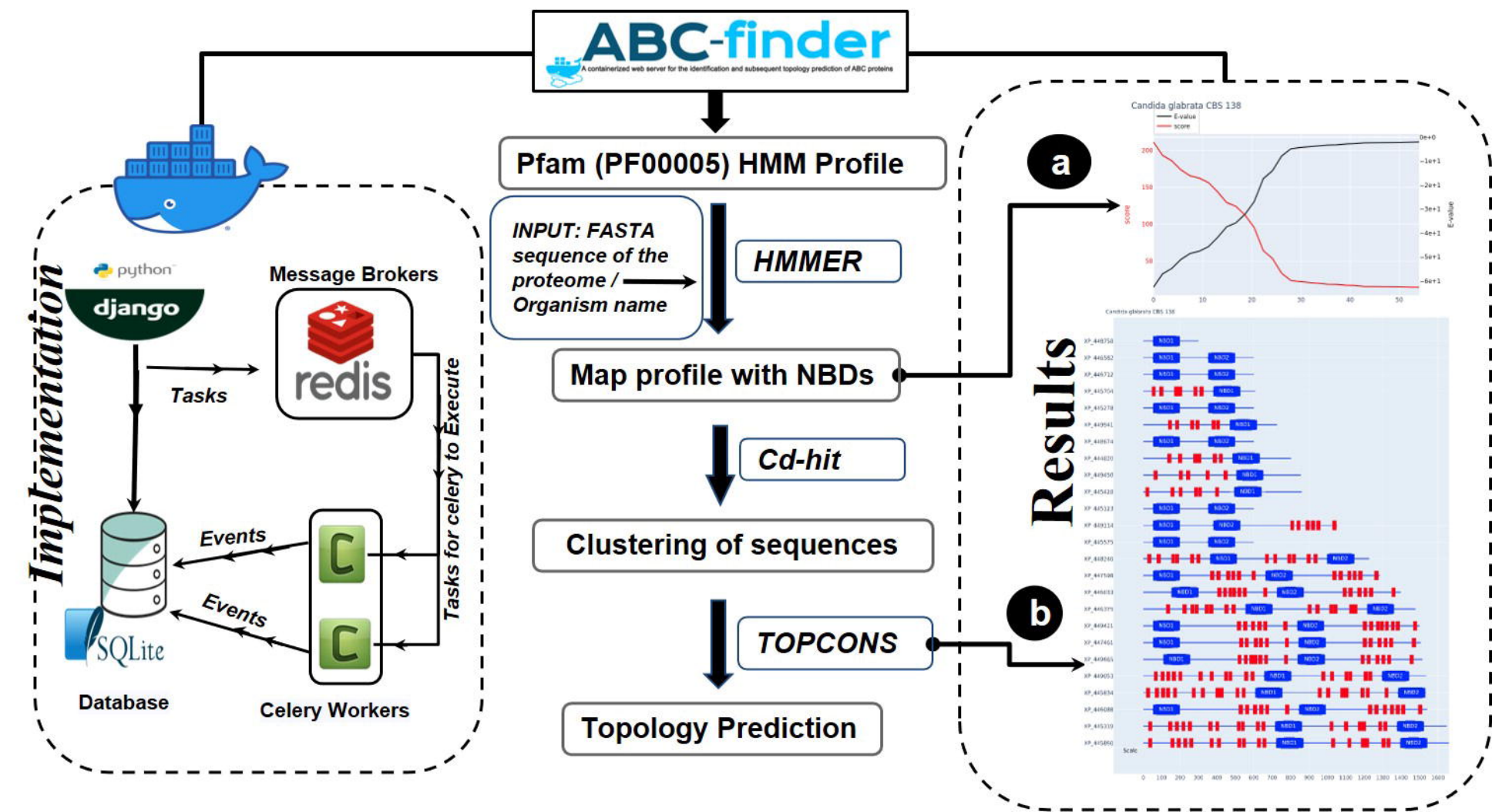
28	NP_010376.3	Fe-S cluster-binding ribosome biosynthesis protein RL1	[53]
29	NP_010346.3	uncharacterized protein YDR061W	-
30	**NP_013030.1**	**uncharacterized protein YKR104W**	**-**
31	NP_116625.1	putative ATP-binding cassette family ATPase CAF16	[54]

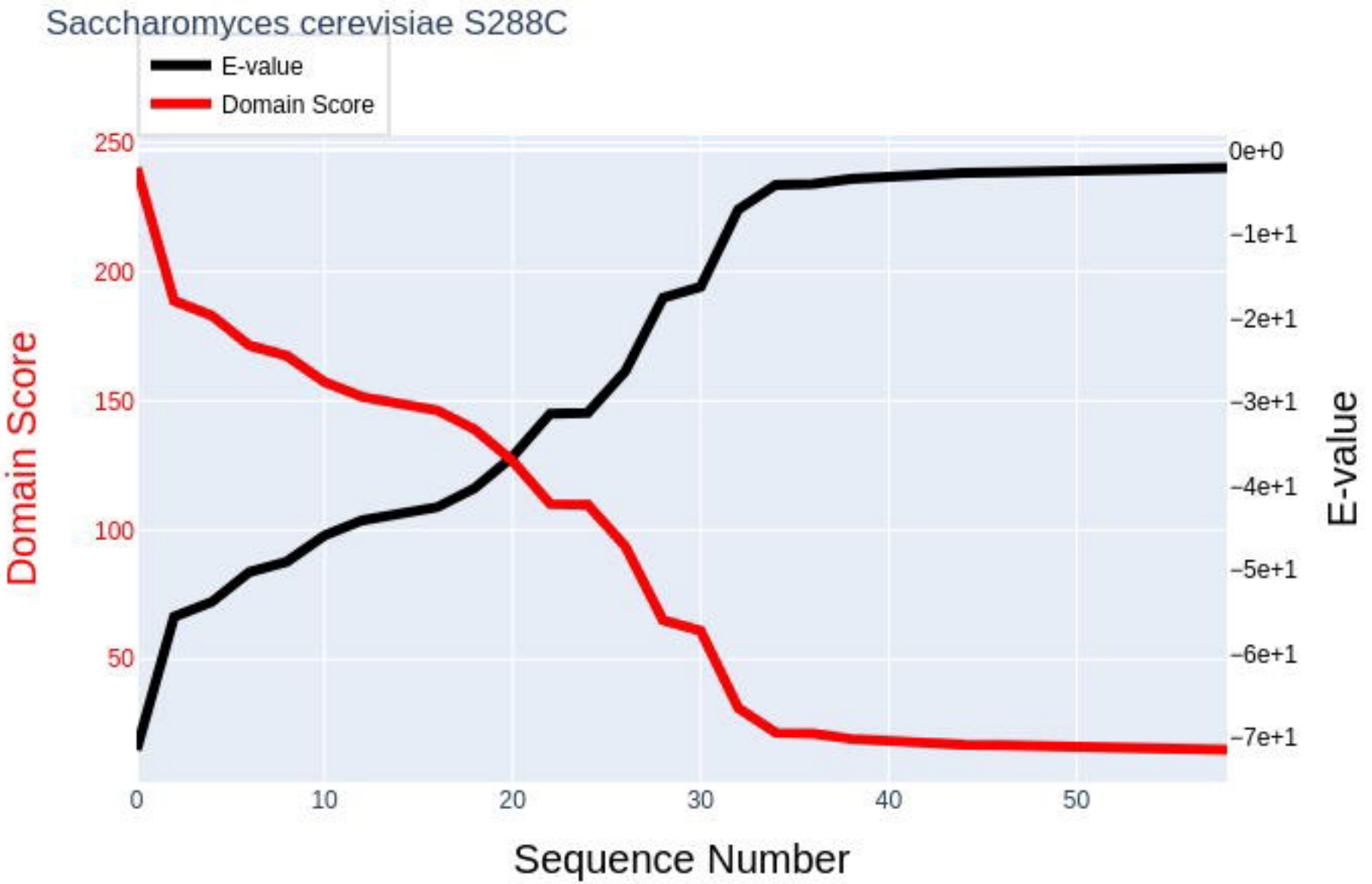
FIGURE LEGENDS

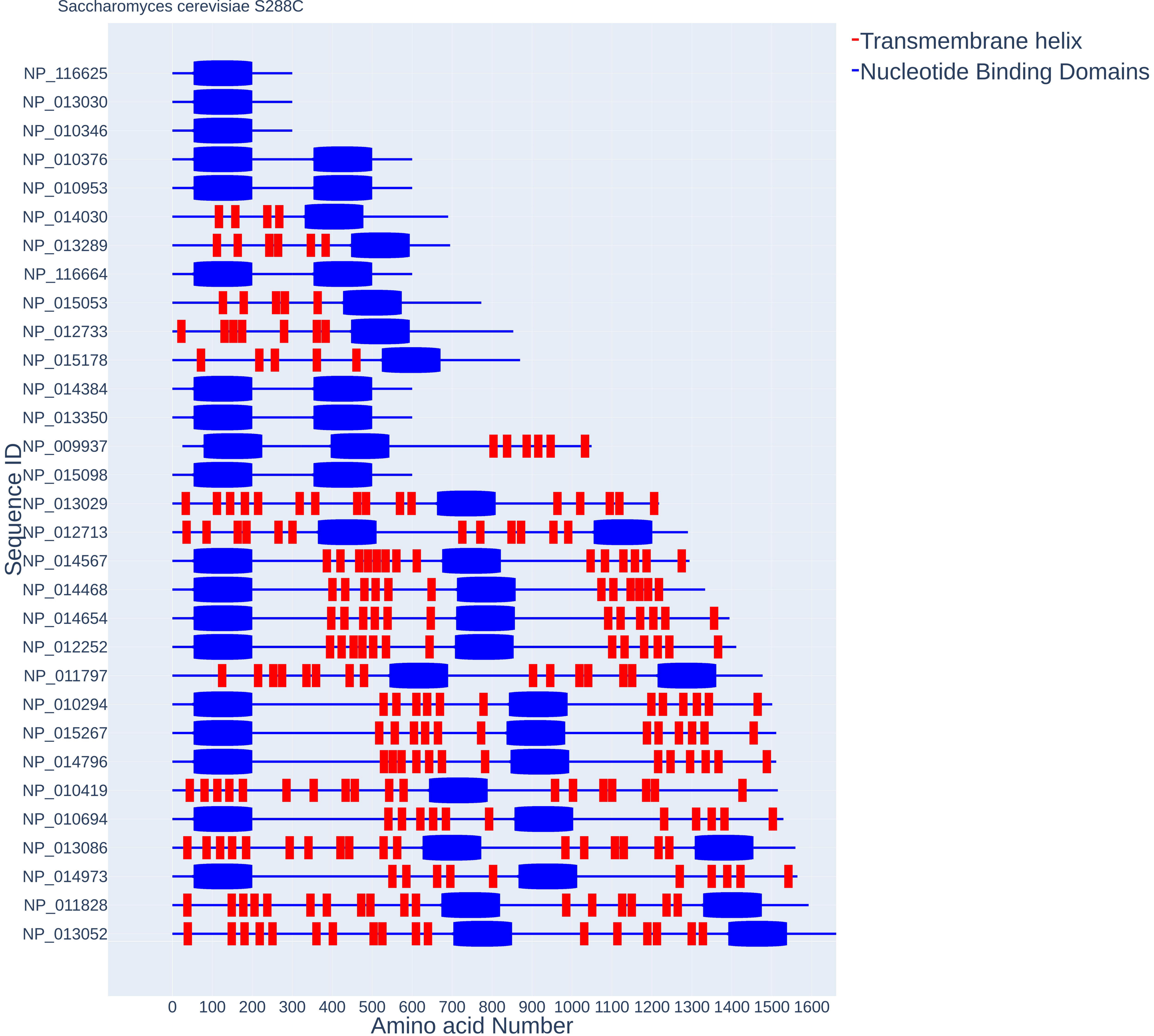
Figure 1: Generalized workflow of ABC-finder.

Figure 2: Plot of E-value and domain score of the sequences above the inclusion threshold value obtained after running the "hmmsearch" function with the *S. cerevisiae* S288C proteome.

Figure 3: Plot representing the topology of the predicted ABC proteins in the *S. cerevisiae* S288C strain. While the x-axis denotes the amino acid number, the y-axis denotes the sequence IDs. TOPCONS is utilized for the prediction of the transmembrane domains. Red columns represent transmembrane domains and blue spheres represent the nucleotide binding domains.


ABC-finder
A containerized web server for the identification and subsequent topology prediction of ABC proteins
Pfam (PF00005) HMM Profile
INPUT: FASTA sequence of the proteome / Organism name
HMMER
Map profile with NBDs
Cd-hit
Clustering of sequences
TOPCONS
Topology Prediction
Implementation
python
django
Message Brokers
redis
Tasks
Tasks for celery to Execute
Events
Events
SQLite
Database
Celery Workers
Results
a
b
Candida glabrata CBS 138
E-value
score


Saccharomyces cerevisiae S288C
E-value
Domain Score
250
200
150
100
50
Domain Score
0e+0
−1e+1
−2e+1
−3e+1
−4e+1
−5e+1
−6e+1
−7e+1
E-value
0
10
20
30
40
50
Sequence Number


Saccharomyces cerevisiae S288C
Transmembrane helix
Nucleotide Binding Domains
Sequence ID
NP_116625
NP_013030
NP_010346
NP_010376
NP_010953
NP_014030
NP_013289
NP_116664
NP_015053
NP_012733
NP_015178
NP_014384
NP_013350
NP_009937
NP_015098
NP_013029
NP_012713
NP_014567
NP_014468
NP_014654
NP_012252
NP_011797
NP_010294
NP_015267
NP_014796
NP_010419
NP_010694
NP_013086
NP_014973
NP_011828
NP_013052
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600
Amino acid Number