

AcrHub: an integrative hub for investigating, predicting and mapping anti-CRISPR proteins

Jiawei Wang^{1,*}, Wei Dai^{1,2}, Jiahui Li², Qi Li², Ruopeng Xie², Yanju Zhang², Christopher Stubenrauch¹ and Trevor Lithgow^{1,*}

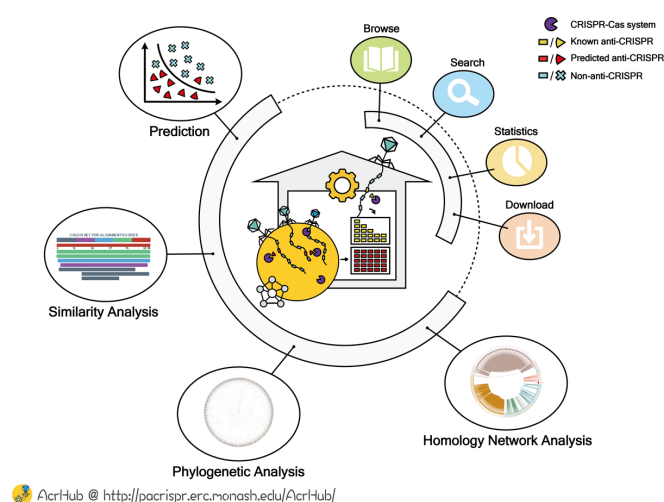
¹Infection and Immunity Program, Biomedicine Discovery Institute and Department of Microbiology, Monash University, VIC 3800, Australia and ²School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China

Received August 15, 2020; Revised September 29, 2020; Editorial Decision October 06, 2020; Accepted October 08, 2020

ABSTRACT

Anti-CRISPR (Acr) proteins naturally inhibit CRISPR-Cas adaptive immune systems across bacterial and archaeal domains of life. This emerging field has caused a paradigm shift in the way we think about the CRISPR-Cas system, and promises a number of useful applications from gene editing to phage therapy. As the number of verified and predicted Acrs rapidly expands, few online resources have been developed to deal with this wealth of information. To overcome this shortcoming, we developed AcrHub, an integrative database to provide an all-in-one solution for investigating, predicting and mapping Acr proteins. AcrHub catalogs 339 non-redundant experimentally validated Acrs and over 70 000 predicted Acrs extracted from genome sequence data from a diverse range of prokaryotic organisms and their viruses. It integrates state-of-the-art predictors to predict potential Acrs, and incorporates three analytical modules: similarity analysis, phylogenetic analysis and homology network analysis, to analyze their relationships with known Acrs. By interconnecting all modules as a platform, AcrHub presents enriched and in-depth analysis of known and potential Acrs and therefore provides new and exciting insights into the future of Acr discovery and validation. AcrHub is freely available at <http://pacrispr.erc.monash.edu/AcrHub/>.

GRAPHICAL ABSTRACT



INTRODUCTION

Bacteria and archaea have evolved a wide variety of CRISPR-Cas systems to protect themselves from harmful mobile genetic elements (MGEs), such as phages and plasmids (1–4). In response, MGEs evolved a series of potent inhibitors, known as anti-CRISPRs (Acrs), to counteract host CRISPR-Cas defence systems (5–11). Except for their universally short sequences, Acrs have little in common with each other, including very low sequence and structural similarity. At least 50 distinct Acr families have been identified across both bacterial and archaeal domains of life where they each use different molecular mechanisms to inhibit CRISPR-Cas systems (8–10,12–13). Outside the confined environment of a microbial cell, Acrs have inspired a number of downstream applications, from gene editing technologies and protein engineering to phage therapy (10,14–17), applications that are only limited by the relatively small number of known Acr systems compared to the thousands hidden in sequenced genomes.

*To whom correspondence should be addressed. Tel: +61 3 9902 9217; Email: trevor.lithgow@monash.edu
Correspondence may also be addressed to Jiawei Wang. Tel: +61 3 9905 1031; Email: jiawei.wang@monash.edu

To overcome this shortfall, several resources were developed either to help categorize Acrs or predict new ones. The three main resources involved in categorizing Acrs are Anti-CRISPRdb (18), which is a compendium of experimentally verified Acrs, the Bondy-Denomy *et al.* online spreadsheet (19) that consolidates Acr records under a uniform naming convention, and AcrCatalog (20), which is a list of predicted Acrs generated using a random forest based model. Alternatively, to identify putative Acrs from within user-defined sequences, there are three online toolkits available. Firstly, AcrRanker (21), which ranks phage proteomes for potential Acrs using an eXtreme Gradient Boosting based model. Secondly, AcrFinder, which integrates homology-based, guilt-by-association-based and self-targeting approaches to identify potential Acrs from prokaryotes and phage genomes (22). Thirdly, PaCRISPR (23), which incorporates evolutionary information within an ensemble model to predict Acr proteins from genomic and metagenomic sequences. However, these toolkits are prediction-oriented and have limited downstream capabilities for further exploring those predicted Acrs.

Here, we present an integrative platform, AcrHub, to offer an all-in-one, user-friendly solution for Acr protein prediction, characterization and relationship analysis (Figure 1). AcrHub catalogs 339 non-redundant, experimentally validated Acrs and 71 728 putative Acrs, each annotated with their relationships to known Acrs. Additionally, three predictors were integrated into AcrHub for prediction of Acr homologues and novel Acrs: a Hidden Markov Model (HMM) based predictor, AcrRanker (21) and PaCRISPR (23). To facilitate relationship analysis between known and potential Acrs, we developed three analytical modules: a BLAST-based similarity analysis, a multiple sequence alignment-based phylogenetic analysis, and a homology-based network analysis. These tools can either work independently or within the AcrHub pipeline to facilitate downstream relationship analysis of a newly predicted Acr and thereby shorten the gap between prediction, functional characterization and eventual experimental validation.

MATERIALS AND METHODS

Here we present the overall workflow of AcrHub in terms of how we curated and annotated the data, and how we designed and implemented the platform. The architecture of AcrHub incorporates different kinds of modules, including the Acr investigation modules and Acr functional modules (Figure 1).

Data curation and annotation

We extracted 574 experimentally verified Acrs along with their basic annotations from current literature and available resources, including the Anti-CRISPRdb (18) and unified online Acr spreadsheet (19). After removing redundant sequences, we obtained 339 Acr proteins. Where required, we further manually annotated their attributes and functions from the literature (10,12,14–15,24), including inhibited stages, molecular mechanisms, targeted proteins and the cellular contexts of Acrs (i.e. the species or cell types

each Acr has been validated in). Acrs were further annotated with their tertiary structure, derived from the protein data bank (PDB) (25), and their secondary structures using the PSIPRED 4.0 server (26). For each Acr, the disorder area was predicted using the IUPred2A server (27), and visualized by ECharts (<https://echarts.apache.org/>).

We then retrieved 232 616 non-redundant putative Acrs from AcrCatalog (20). These putative Acrs were extracted from 10 938 430 proteins from the genomes of prokaryotic species and their viruses. Their potential inhibitor types were also annotated according to the CRISPR-Cas types that existed within the original host genomes. Using our PaCRISPR predictor (23) that had been updated on the latest Acr data (described in the following section), we refined our list of 232 616 putative Acrs to 71 728. These putative Acrs were then incorporated into the AcrHub catalogue and were further annotated with their PaCRISPR prediction scores, HMM based prediction results, AcrRanker prediction scores and ranks, inhibited CRISPR types, sequences, lengths, accession links to track their original information from the NCBI resource (28) and their predicted disorder areas.

For both experimentally verified and predicted Acrs, we additionally provided multiple pre-calculated relationship analyses to link them to known Acrs. For each entry, the homologous sequences were generated using the BLAST tool (version 2.8.1+) (29) by searching against known Acrs and then were visualized by BlasterJS (30). The R library msa (31) was additionally used to generate and visualize their multiple sequence alignments. For each entry, a phylogenetic tree was generated using FastTree (version 2.1.8) (32) based on the multiple sequence alignments of this entry and known Acrs, which were generated using the MAFFT tool (version v7.271) (33). The open source phylogram.d3 (<https://github.com/ConstantinoSchillebeekx/phylogram.d3>) was used to visualize the tree. For each entry, a homology network was generated using the all-against-all BLAST (version blast-2.2.26) (34) on this entry and known Acrs, and then visualized by ECharts. For the linked nodes within the network, their pairwise sequence alignments were generated using EMBOSS Stretcher through its web service (35).

Functional module implementation

The functional modules of AcrHub are comprised of two types of modules: one predictive module and three analytical modules. The analytical modules were designed to complement the predictive modules and have been implemented in the same manner as detailed within the 'Data curation and annotation' section. For the predictive module, we incorporated existing prominent predictors for users to choose from. Designed using different computational techniques, currently existing tools have their own strengths and drawbacks. Implementing a bioinformatic workflow based on well-accepted biological discoveries, AcrFinder (22) accepts nucleotide sequences as input and can provide users with the genomic context of Acrs. However, AcrFinder has limited functionality because it cannot predict Acrs that lack nearby anti-CRISPR-associated genes nor can it predict Acrs in genomes without complete CRISPR-Cas sys-

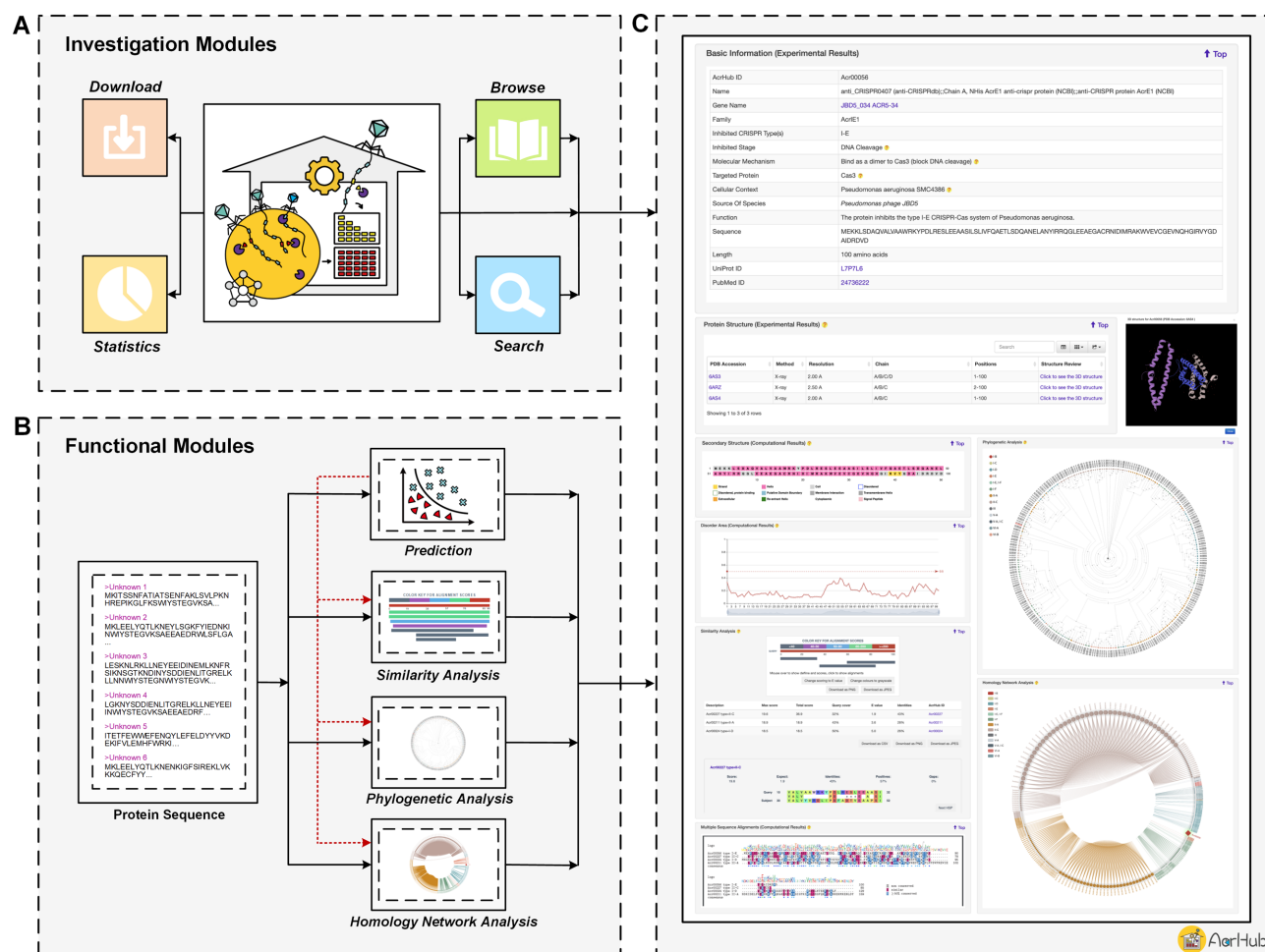


Figure 1. Overview of AcrHub. (A) The investigation modules that can be used to examine catalogued Acrs; (B) the functional modules and their interactions; (C) the 'Detailed information' page, using an experimentally validated Acr as example.

tems. In contrast, machine learning based tools are not restricted in this way, and are usually more accurate in predicting novel Acrs. Among them, AcRanker (21) and the AcrCatalog database method (20) mainly use sequence associated features to predict novel Acrs with similar characteristics to known Acrs. Compared to AcRanker that accepts protein sequences as input, the AcrCatalog method requires a complex matrix input before it will predict Acrs.

An alternative machine learning based predictor is PaCRISPR (23), which was instead trained using position-specific scoring matrix (PSSM) based features. Those high-level features enable PaCRISPR to track the evolutionary history of Acrs, promising a greater capability to predict more distantly related Acrs. However, this potential increase in accuracy comes at the expense of time, due to the requirement to generate PSSMs. We therefore decided to incorporate the more universal and user-friendly prediction tools, AcRanker and PaCRISPR, into the predictive module. AcRanker within AcrHub was implemented using its open-source code and model (21), whereas PaCRISPR within AcrHub was retrained using the same algorithm as the PaCRISPR predictor (23), but with a more up-to-date list of known Acrs. In this way, the evolutionary informa-

tion derived from the new Acrs was also incorporated into PaCRISPR to expand its detection scope to discover future novel Acrs. Additionally, we developed a lightweight HMM based predictor using HMMER (36) to allow users to rapidly detect homologues of known Acrs from their protein sequences of interest.

Website architecture

In general, the web interface of AcrHub was displayed by JSP, CSS, jQuery (<https://jquery.com/>), Bootstrap (<https://bootstrapdocs.com/>) and their extension packages. The logic actions were controlled by the JAVA (<https://www.java.com/>) server development suite, including Struts 2 (<https://struts.apache.org/>) and Hibernate (<https://hibernate.org/>). All experimentally validated and predicted Acr data was stored in a MySQL database (<https://www.mysql.com/>).

Specifically, the functional modules introduce an additional Perl CGI (<https://metacpan.org/pod/CGI>) based server backend to handle those time-consuming operations. Each functional module (implemented as above) was stringed together by a Perl thread within the server end. A Gearman framework (<http://gearman.org/>) based queueing

system was used to decouple the prompt-response-required web interface and the time-consuming server backend for better user experience.

PROFILING ACRHUB

AcrHub contains a browsable list of the known, experimentally validated Acrs and a vast number of pre-calculated potential Acrs, making it possible to conduct systematic, comparative analyses to enhance current understanding and facilitate future discovery. A majority of experimentally validated Acrs are fewer than 150 amino acids in length (Figure 2A), while the largest is comprised of 322 amino acids: AcrVA2 (AcrHub ID: Acr00311), identified from a prophage and inhibiting the V-A CRISPR-Cas system. Likewise, predicted Acrs are small molecules, where the majority of Acrs were smaller than 150 amino acids (Figure 2A). To date (as of September 2020), 11 inhibited CRISPR types have been confirmed in Acrs (Figure 2C), which has been identified from multiple origins, including prokaryotes, phages and prophages (Figure 2B). Even from our list of predicted Acrs catalogued in AcrHub, there are novel types of CRISPR-Cas systems that are possibly inhibited (Figure 2D). In most cases, homologs exist within the same inhibited CRISPR types due to additional related homolog retrieval applied after a novel Acr was discovered (Figure 2E). Few links exist across different inhibited CRISPR types, highlighting their low sequence similarity. Looking into the predicted Acrs (Figure 2E), we can see some of them have links to known Acrs, which indicates that they might have similar inhibited CRISPR types or functions. On the other hand, most of the predicted Acrs have no links to each other and only very few links within the predicted Acrs (Figure 2E). This promises the future discovery of novel Acrs with distinct functions.

USING ACRHUB

The AcrHub platform includes *Home*, *Anti-CRISPR investigation*, *Prediction*, *Relationship analysis*, *Help* and *Contact* modules (Figures 1, 3 and 4). Within *Anti-CRISPR investigation*, AcrHub provides its investigation modules to explore its list of 339 experimentally verified proteins and 71 728 predicted proteins with *Browse*, *Search*, *Statistics* and *Download* functions included (Figures 1 and 3). The *Prediction* module provides users with state-of-the-art predictors to better assist users in predicting new Acrs using their own query sequences. After successful prediction of a new Acr, users can select from three visualization tools to explore their data, or they can access those modules independently under *Relationship analysis: Similarity analysis*, *Phylogenetic analysis* and *Homology network analysis* (Figures 1 and 4). All of these tools and options are described within the *Help* module, which includes a detailed overview of AcrHub.

Using investigation modules to explore known and predicted Acrs

The Acr investigation modules allow users to browse, search, download and explore statistics and details of 339

known and 71 728 predicted Acrs stored within AcrHub (Figures 1 and 3), and is found within the 'Anti-CRISPR investigation' tab.

Browsing. The 'Browse' page presents the list of experimentally validated and predicted Acrs. Experimentally validated Acrs are categorized by the CRISPR type that is inhibited, and summarized with their AcrHub IDs, families, inhibited CRISPR types, inhibited stages, targeted proteins, molecular mechanisms, cellular contexts, sources of species, lengths, PDB accessions and PubMed IDs (28). Predicted Acrs are sorted by their PaCRISPR prediction scores, and summarized with their AcrHub IDs, inhibited CRISPR types, lengths, NCBI resource accession numbers and prediction results from PaCRISPR, AcRanker and the HMM based predictor.

Searching. The 'Search' page allows users to search for known or predicted Acrs using AcrHub ID, UniProt ID, possible protein name (smart partial match supported), inhibited CRISPR type and specified length. For users who have a particular interest in Acrs with experimental attributes and functions, we provide a search function to allow users to specify an option, including protein structure, inhibited stage, molecular mechanism, targeted protein and cellular context. The 'Search results' page has an output format similar to that described for the 'Browse' page, but filtered according to the search options.

Detailed information. The 'Detailed information' page opens when a user selects one of the listed Acrs or predicted Acrs. The page provides all of the relevant annotations for that entry. For each of the known Acrs (Figure 1C), experimental annotations (where available) are: its name (as complete as possible), gene name, family, inhibited CRISPR type(s), inhibited stage, molecular mechanism, targeted protein, cellular context, source of species, function, sequence, length, UniProt ID to trace the original sequence information and PubMed ID to trace the original reference. Experimentally confirmed protein structures are provided where available, including PDB accession numbers to track more comprehensive annotations therein and a link to trigger a popup for an interactive three-dimensional structure. A predicted visualization of secondary structure is also presented. For each of the 71 728 predicted Acrs within AcrHub (Figure 3C), annotations include possible inhibited CRISPR types, amino acid sequence and length, and its prediction results, which indicate the possibility of being an Acr. A link to its entry within the NCBI resource is also provided for more comprehensive annotations. For both experimentally verified and predicted Acrs (Figures 1C and 3C), AcrHub presents predicted disorder areas and pre-calculated relationship analyses against known Acrs in AcrHub, including sequence similarity, multiple sequence alignments, phylogenetic relationships and homology networks. For each Acr, this allows users to identify the closest related Acr and provides links to access those known Acrs for more information.

Statistics. The 'Statistics' page provides interactive visualizations on the statistics of AcrHub catalogued Acrs (validated and putative) (Figure 3E). These include phylogenetic tree and homology network analyses, as well as pro-

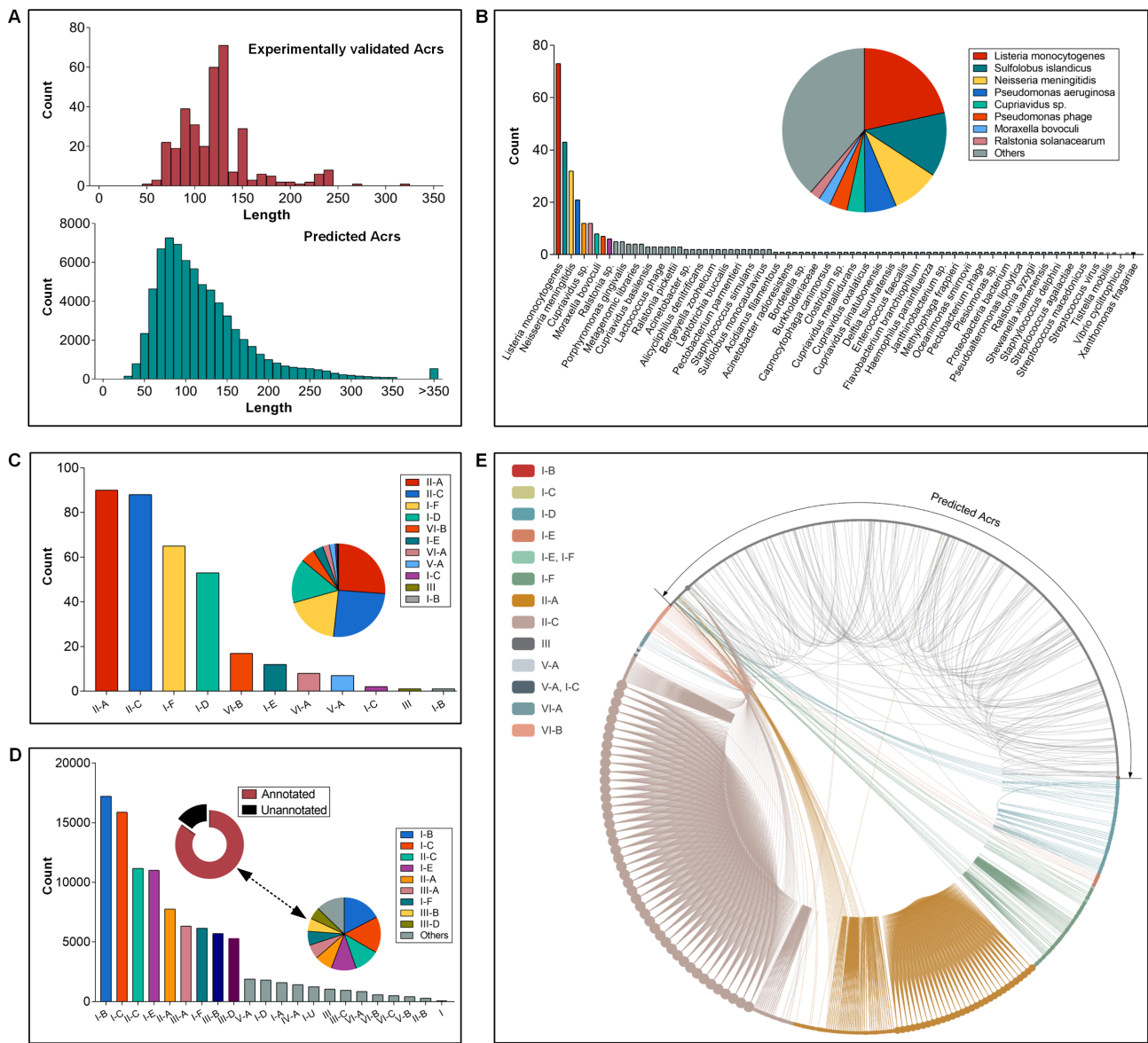


Figure 2. Acr Statistics. (A) Distribution of 339 experimentally verified Acrs and 71 728 predicted Acrs by their sequence lengths. (B) Distribution of 339 experimentally verified Acrs by their origin. The histogram chart counts the exact numbers, while the pie chart shows the percentage. This applies to panels C and D. (C) Distribution of 339 experimentally verified Acrs in terms of their inhibited CRISPR types. The Acrs with two CRISPR types inhibited were counted twice, once for each inhibited CRISPR type. (D) Distribution of 71 728 predicted Acrs in terms of their possible inhibited CRISPR types. As there are 10 902 predicted Acrs not yet annotated (denoted by black circle segment), only 60 826 with possible inhibited CRISPR types were counted. The predicted Acrs with more than one possible CRISPR types inhibited are counted for each of their inhibited CRISPR types. (E) Homology network of 339 experimentally verified Acrs and 71 728 predicted Acrs. The experimentally verified Acrs are grouped and colored by their inhibited CRISPR types, while the predicted Acrs are indicated by an arc. Each node represents an Acr or a predicted Acr protein. A link between two nodes represents they have a homology relationship. The more links a node has, the larger the node will be. This leads to the unequal distribution across the circle, with the much smaller number of validated Acrs accounting for a larger proportion of the circle. All statistics are reflected in the interactive visualizations in the ‘Statistics’ page of AcrHub.

tein distributions according to their (i) inhibited CRISPR types, (ii) sequence lengths and (iii) species of origin. Users can hover over any node of interest in the phylogenetic tree or homology network to browse the summary information for that Acr. Alternatively, users can click the node of interest to investigate the detailed information for that Acr. Users can also click any link in the homology network to view the pairwise sequence alignment between the two

linked Acrs. Similarly, users can click on a bar or part of a pie chart to obtain the corresponding list of Acrs, which is displayed similarly to the ‘Search results’ page. **Downloading the data.** The ‘Download’ page provides multiple options for users to download the database items, including sequences and annotations of experimentally validated and predicted Acrs (Figure 3D). For experimentally validated Acrs, we additionally provide predicted sec-

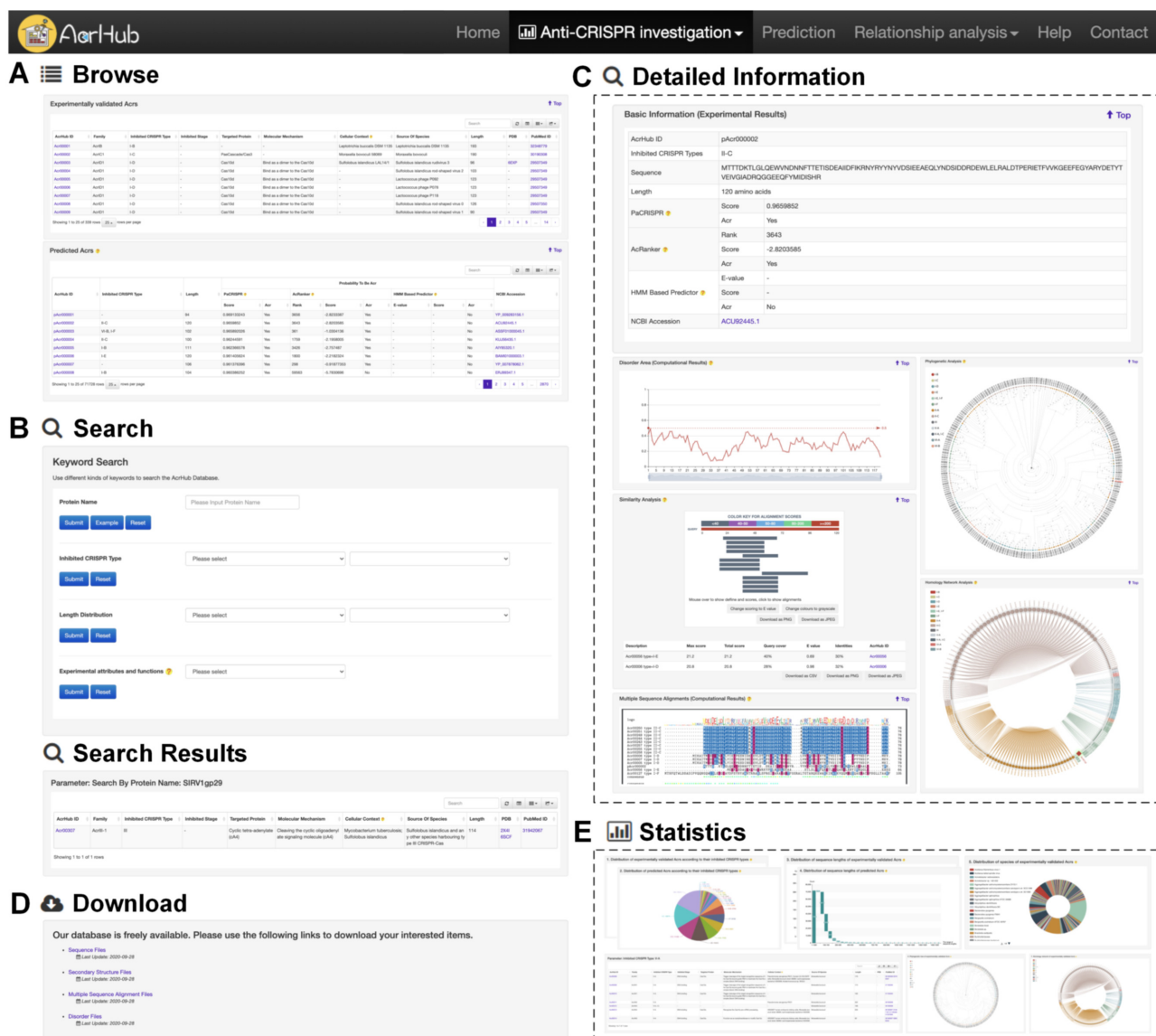


Figure 3. Graphical illustration of the investigation modules in AcrHub: (A) the 'Browse' page, (B) the 'Search' and 'Search results' pages, (C) the 'Detailed information' page to describe a predicted Acr protein, (D) the 'Download' page and (E) the 'Statistics' page.

ondary structures, predicted disorder files and predicted multiple sequence alignments.

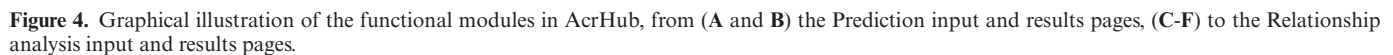
Using functional modules to explore potential Acrs

The functional modules allow users to predict novel Acrs and analyze their relationship to known Acrs listed in AcrHub (Figures 1 and 4). These modules can run independently and allow a maximum of 5 000 sequences per submission.

Prediction. AcrHub provides novel Acr predictive functions by incorporating PaCRISPR, AcRanker and an HMM based predictor. Accepting protein sequences as input, PaCRISPR and AcRanker can predict novel Acrs whereas the HMM based predictor can identify very close homologs to known Acrs. Users are provided with a checkbox so they can select one or more of the three predic-

tors to customize different prediction scenarios (Figure 4A). Submitted sequences will first be checked against the experimentally validated Acrs, and those with successful hits will be marked with 'Exp.' (i.e. Experimentally verified) and links to their 'Detailed information' pages (Figure 4B). The rest of the sequences will be sent to the prediction engine and prediction results of the selected predictor(s) will be retrieved.

Similarity analysis. The 'Similarity analysis' module of AcrHub aims to find regions of similarity between the query proteins and experimentally validated Acrs (Figure 4C). For each of the query sequences, AcrHub will visualize its known Acr homologs in terms of alignment scores, and list the detailed alignment information, including accurate alignment scores, query cover, E-value, identities and AcrHub ID (Figure 4D). Clicking any of the experimentally validated homologs will turn to its pairwise sequence alignment against the query sequence. Clicking



browse the summary information for that Acr, or click it to be redirected to its ‘Detailed information’ page.

Homology network analysis. The ‘Homology network’ analysis module aims to recognize the closest homologous relationships between query proteins and experimentally validated Acrs (Figure 4C). As a result, AcrHub presents an interactive homology network (Figure 4F): the nodes are categorized by their inhibited CRISPR types, the inquiry proteins are highlighted by red diamonds and links between two nodes indicates their close relationship. Users

can hover over any node of interest to browse the summary information for that Acr, click it to be redirected to the 'Detailed information' page for that Acr, or click any link between two nodes (if available) to display their pairwise sequence alignment information.

Analyzing data within the pipeline

The functional modules can also run within a pipeline (Figures 1 and 4). AcrHub provides options for users to redirect prediction results to relationship analysis tools easily, as well as links to 'Detailed information' pages for known Acrs homologous to the query proteins within each functional module.

From prediction to relationship analysis. Following successful prediction of novel Acr proteins, users have the option to further analyze their relationship to known Acrs. This will provide insight into the possible CRISPR types inhibited or functions of the predicted Acrs. Accordingly, the 'Prediction results' page of AcrHub provides an option for users to select their proteins using a specified prediction threshold and redirects them to any of the three relationship analysis modules (Figure 4B and C). This batch operation largely saves time, especially when screening a subset of Acrs from a much larger query.

From computational results to known Acrs. The functional modules interact well with the investigation modules (Figure 1). Within the prediction or relationship analysis results (Figure 4B and D-F), links are available that will direct users to the 'Detailed information' pages of known Acrs (with homology to the query proteins). This takes full advantage of the well-annotated known Acrs listed in AcrHub.

DISCUSSION

Deeper understanding of Acrs promises a number of real-world applications and exciting insights into the mechanisms of CRISPR-Cas systems, phage biology, and host-pathogen interactions. We therefore developed AcrHub, an integrative platform that consolidates the most up-to-date list of known and putative Acrs, while providing access to prediction and relationship analysis tools. All modules work independently or within the AcrHub pipeline to better assist biologists with extracting the data they need to help design experiments and formulate interesting hypotheses.

AcrHub will be systematically updated to include newly identified Acrs and the most up-to-date information for existing Acrs. Keeping track of the community efforts in Acr prediction, we will continue to incorporate new Acr predictors and update existing predictors once they become available. Additionally, our PaCRISPR and HMM based tools will be periodically retrained with the rapidly accumulating Acr data. Newly predicted Acrs will also be incorporated into AcrHub by executing PaCRISPR on the ever-expanding volume of prokaryotic and phage genomes. One of the key features of AcrHub is its relationship analysis modules that allow users to visualize their data, so AcrHub will naturally incorporate additional analysis tools as new visualization programs become available or if there is a clear need as defined by users.

DATA AVAILABILITY

The AcrHub platform is freely available at <http://pacrispr.erc.monash.edu/AcrHub/>. All data indexed by AcrHub can be downloaded via <http://pacrispr.erc.monash.edu/AcrHub/download.jsp>. Detailed user instructions can be accessed via the 'Help' page at <http://pacrispr.erc.monash.edu/AcrHub/help.jsp>.

FUNDING

National Health and Medical Research Council of Australia [1092262 to TL]. Funding for open access charge: National Health and Medical Research Council of Australia [1092262].

Conflict of interest statement. None declared.

REFERENCES

- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
- Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.*, **82**, 237–266.
- Barrangou, R. and Marraffini, L.A. (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell*, **54**, 234–244.
- Amitai, G. and Sorek, R. (2016) CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.*, **14**, 67–76.
- Bondy-Denomy, J., Pawluk, A., Maxwell, K.L. and Davidson, A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
- Borges, A.L., Davidson, A.R. and Bondy-Denomy, J. (2017) The discovery, mechanisms, and evolutionary impact of anti-CRISPRs. *Annu. Rev. Virol.*, **4**, 37–59.
- Maxwell, K.L. (2017) The anti-CRISPR story: a battle for survival. *Mol. Cell*, **68**, 8–14.
- Pawluk, A., Davidson, A.R. and Maxwell, K.L. (2018) Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol.*, **16**, 12–17.
- Stanley, S.Y. and Maxwell, K.L. (2018) Phage-encoded anti-CRISPR defenses. *Annu. Rev. Genet.*, **52**, 445–464.
- Davidson, A.R., Lu, W.T., Stanley, S.Y., Wang, J., Mejdani, M., Trost, C.N., Hicks, B.T., Lee, J. and Sontheimer, E.J. (2020) Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. *Annu. Rev. Biochem.*, **89**, 309–332.
- Peng, X., Mayo-Munoz, D., Bhoobalan-Chitty, Y. and Martinez-Alvarez, L. (2020) Anti-CRISPR proteins in archaea. *Trends Microbiol.*, **28**, 913–921.
- Hardouin, P. and Goulet, A. (2020) Diversity of molecular mechanisms used by anti-CRISPR proteins: the tip of an iceberg? *Biochem. Soc. Trans.*, **48**, 507–516.
- Wiegand, T., Karambelkar, S., Bondy-Denomy, J. and Wiedenheft, B. (2020) Structures and strategies of anti-CRISPR-mediated immune suppression. *Annu. Rev. Microbiol.*, **74**, 21–37.
- Marino, N.D., Pinilla-Redondo, R., Csorgo, B. and Bondy-Denomy, J. (2020) Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nat. Methods*, **17**, 471–479.
- Liu, Q., Zhang, H. and Huang, X. (2020) Anti-CRISPR proteins targeting the CRISPR-Cas system enrich the toolkit for genetic engineering. *FEBS J.*, **287**, 626–644.
- Basgall, E.M., Goetting, S.C., Goeckel, M.E., Giersch, R.M., Roggenkamp, E., Schrock, M.N., Halloran, M. and Finnigan, G.C. (2018) Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in *Saccharomyces cerevisiae*. *Microbiology*, **164**, 464–474.
- Bubeck, F., Hoffmann, M.D., Harteveld, Z., Aschenbrenner, S., Bietz, A., Waldhauer, M.C., Borner, K., Fakhiri, J., Schmela, C., Dietz, L. *et al.* (2018) Engineered anti-CRISPR proteins for optogenetic control of CRISPR-Cas9. *Nat. Methods*, **15**, 924–927.

18. Dong,C., Hao,G.F., Hua,H.L., Liu,S., Labena,A.A., Chai,G., Huang,J., Rao,N. and Guo,F.B. (2018) Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins. *Nucleic Acids Res.*, **46**, D393–D398.
19. Bondy-Denomy,J., Davidson,A.R., Doudna,J.A., Fineran,P.C., Maxwell,K.L., Moineau,S., Peng,X., Sontheimer,E.J. and Wiedenheft,B. (2018) A unified resource for tracking anti-CRISPR names. *CRISPR J.*, **1**, 304–305.
20. Gussow,A.B., Park,A.E., Borges,A.L., Shmakov,S.A., Makarova,K.S., Wolf,Y.I., Bondy-Denomy,J. and Koonin,E.V. (2020) Machine-learning approach expands the repertoire of anti-CRISPR protein families. *Nat. Commun.*, **11**, 3784.
21. Eitzinger,S., Asif,A., Watters,K.E., Iavarone,A.T., Knott,G.J., Doudna,J.A. and Minhas,F. (2020) Machine learning predicts new anti-CRISPR proteins. *Nucleic Acids Res.*, **48**, 4698–4708.
22. Yi,H., Huang,L., Yang,B., Gomez,J., Zhang,H. and Yin,Y. (2020) AcrFinder: genome mining anti-CRISPR operons in prokaryotes and their viruses. *Nucleic Acids Res.*, **48**, W358–W365.
23. Wang,J., Dai,W., Li,J., Xie,R., Dunstan,R.A., Stubenrauch,C., Zhang,Y. and Lithgow,T. (2020) PaCRISPR: a server for predicting and visualizing anti-CRISPR proteins. *Nucleic Acids Res.*, **48**, W348–W357.
24. Trasanidou,D., Geros,A.S., Mohanraju,P., Nieuwenweg,A.C., Nobrega,F.L. and Staals,R.H.J. (2019) Keeping CRISPR in check: diverse mechanisms of phage-encoded anti-CRISPRs. *FEMS Microbiol. Lett.*, **366**, fnz098.
25. Burley,S.K., Berman,H.M., Bhikadiya,C., Bi,C., Chen,L., Di Costanzo,L., Christie,C., Dalenberg,K., Duarte,J.M., Dutta,S. *et al.* (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.*, **47**, D464–D474.
26. Buchan,D.W.A. and Jones,D.T. (2019) The PSIPRED protein analysis Workbench: 20 years on. *Nucleic Acids Res.*, **47**, W402–W407.
27. Meszaros,B., Erdos,G. and Dosztanyi,Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
28. Sayers,E.W., Beck,J., Brister,J.R., Bolton,E.E., Canese,K., Comeau,D.C., Funk,K., Ketter,A., Kim,S., Kimchi,A. *et al.* (2020) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **48**, D9–D16.
29. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
30. Blanco-Miguez,A., Fdez-Riverola,F., Sanchez,B. and Lourenco,A. (2018) BlasterJS: a novel interactive JavaScript visualisation component for BLAST alignment results. *PLoS One*, **13**, e0205286.
31. Bodenhofer,U., Bonatesta,E., Horejs-Kainrath,C. and Hochreiter,S. (2015) msa: an R package for multiple sequence alignment. *Bioinformatics*, **31**, 3997–3999.
32. Price,M.N., Dehal,P.S. and Arkin,A.P. (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
33. Katoh,K., Misawa,K., Kuma,K. and Miyata,T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
34. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
35. Madeira,F., Park,Y.M., Lee,J., Buso,N., Gur,T., Madhusoodanan,N., Basutkar,P., Tivey,A.R.N., Potter,S.C., Finn,R.D. *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**, W636–W641.
36. Potter,S.C., Luciani,A., Eddy,S.R., Park,Y., Lopez,R. and Finn,R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.*, **46**, W200–W204.