A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases.

Tesson F.[1,4], Planel R[3]., Egorov A[2]., Georjon H.[1], Vaysset H.[1,5,6], Brancotte B[3]., Néron B[3]., Mordret E.[1], Atkinson G[2]., Bernheim A[1], Cury J.[1]*.*

[1]Institut Pasteur, Université Paris Cité, INSERM U1284, Molecular Diversity of Microbes lab, 75015, Paris, France
[2]Department of Experimental Medical Science, and Lund University Virus Centre, Lund University, 221 00 Lund, Sweden
[3]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France
[4]Université Paris Cité, INSERM, IAME, 75018 Paris, France
[5]Institut Curie, Université PSL, Sorbonne Université, CNRS, Laboratoire Physico Chimie Curie, UMR 168, 75005 Paris, France
[6]Agroparistech, Paris, France

Abstract

In recent years, a vast number of novel antiphage defense mechanisms were uncovered. To facilitate the exploration of mechanistic, ecological, and evolutionary aspects related to antiphage defense systems, we released DefenseFinder in 2021 (Tesson et al., 2022). DefenseFinder is a bioinformatic program designed for the systematic identification of known antiphage defense mechanisms. The initial release of DefenseFinder v1.0.0 included 60 systems. Over the past three years, the number of antiphage systems incorporated into DefenseFinder has grown to 152. The increasing number of known systems makes it a challenge to enter the field and makes the interpretation of detections of antiphage systems difficult. Moreover, the rapid development of sequence-based predictions of structures offers novel possibilities of analysis and should be easily available. To overcome these challenges, we present a hub of resources on defense systems, including: 1) an updated version of DefenseFinder with a web-service search function, 2) a community-curated repository of knowledge on the systems, and 3) precomputed databases, which include annotations done on RefSeq genomes and structure predictions generated by AlphaFold. These pages can be freely accessed for users as a starting point on their journey to better understand a given system. We anticipate that these resources will foster the use of bioinformatics in the study of antiphage systems and will serve the community of researchers who study antiphage systems. This resource is available at: https://defensefinder.mdmlab.fr.

Introduction

In the last few years, a considerable number of newly discovered antiphage defense mechanisms have come to light [1]. The most widespread mechanisms, Restriction-Modification and CRISPR-Cas systems, target foreign nucleic acids[2,3]. However, recent discoveries have revealed an important diversity of molecular modalities by which bacteria defend themselves against phages. This diversity of mechanisms includes nucleotide depletion 4–10, membrane disruption11–14, and production of antiviral molecules[15]. Importantly, many defense mechanisms remain unknown.

The discovery of diverse antiphage systems not only provides new insights into bacterial immunity mechanisms but also transforms the exploration of interactions between phages and bacteria and microbial evolution. Evaluating the impact of different defense systems in naturally occurring *Vibrio* isolates has revealed that a rapid turnover of a few mobile genetic elements encoding defense systems can completely alter their susceptibility to phages[16,17]. Ongoing research is delving into the broader phylogenetic scale to explore the role of defense systems and their potential implications for phage therapy strategies[18,19]. Furthermore, examples indicate that bacterial defense systems can be co-regulated and operate synergistically in multi-layered defense strategies[20]. Beyond experimental approaches, there is a growing interest in identifying the antiphage systems encoded in diverse species or environments to understand the diversity of antiviral strategies employed by bacteria.

To investigate mechanistic, ecological, and evolutionary questions related to antiphage defense systems, we created in 2021 DefenseFinder[21], a tool to detect known antiphage defense systems systematically. At the time of publication of DefenseFinder v1.0.0, the tool encompassed 60 systems. In the last 3 years, the number of antiviral systems included in DefenseFinder has grown to 152. The increasing number of systems has been accompanied by emerging challenges in comprehending DefenseFinder results biologically. For non-specialist users, the analysis can be quite intricate, demanding a high level of knowledge in this rapidly evolving field.

To bridge this gap, we decided to create a website dedicated to defense systems, encompassing an improved version of a web service to run DefenseFinder and diverse databases to illuminate and increase the understandability of bioinformatic detection of antiphage defense systems. Here, we provide the release of 92 new defense systems models and a software update as well as 3 databases. 1/ A collaborative knowledge base (wiki) summarizing information on known defense systems. 2/ A structure database, with experimentally determined and AlphaFold2 predicted structures. 3/ A precomputed database of DefenseFinder results in over 20,000 complete genomes. Besides, we designed the website to keep it as up-to-date as possible. Wiki pages are easily editable by anyone and reviewed by experts in the field before publication on the website. They can also be edited automatically to generate sections that aggregate data, such as the phylogenetic distribution of a system. All those website components are also integrated within the DefenseFinder webserver output to easily find information on a system found in a genome.

Results

Updates of DefenseFinder program

To improve detections performed by DefenseFinder, we updated both the program and the models. Since the release of DefenseFinder models v1.0.0, we have constantly updated the defense system models included in DefenseFinder by adding a total of 92 defense systems and 132 subsystems (**Figure 1A**). Among those systems, 63 were discovered and described after releasing the first version of DefenseFinder models. Other systems were missed in the first version and are now added after a deeper literature review of the field. Those systems encompass the "Abi" group discovered between 1990 and 2006. DefenseFinder models v1.0.0 included 3 of these systems, while the latest version v1.2.3 detects 22 Abi systems[22]. Beyond Abi systems, we also added SanaTA [23], MazEF [24], antiphage defense systems identified in mycobacterium prophages [25,26] and pAgo [27]. The defense system named "Rst_DprA-PPRT " [28] was renamed "ShosTA" as it was initially discovered in 2013 [23] but without phage activity testing. Finally, for Lamassu-Fam[29], we separated the systems into different subsystems according to the evolving literature on the topic. Models were also adapted to the latest version of MacSyFinder[30].

To evaluate the proportion of newly detected systems, we ran DefenseFinder v1.2.3 on the RefSeq complete genome database of prokaryotes (Bacteria N = 22,422; Archaea N = 381). We found a total of 152,386 different systems, among which 33% correspond to systems present in v1.2.3 but absent from v1.0.0 **(Figure 1B**). Without restriction-modification and CRISPR-Cas systems, which are by far the most abundant systems, newly added systems now represent 54% of the detected systems (41-73% across phyla) (**Figure 1B**).

The previous version of DefenseFinder used a "profile coverage" parameter of 0.1. This value was chosen to detect defense systems when the protein was split. This low coverage threshold was balanced with a high score threshold for single-gene systems and the necessity to find the different components of the system colocalizing in the genome for multi-gene systems. We changed that threshold to 0.4 as a default value in the latest version. This threshold was set to maintain the possibility of finding incomplete genes or a gene with a domain replacement while removing very low coverage off-target hits. Users can still manually change the coverage threshold (--coverage-profile) to decide on how conservative detection should be depending on the biological problem at hand. For ease of use, we also added the possibility to directly input a nucleic acid fasta file. We use pyrodigal v3.0.1[31] to identify and translate the coding regions, which are then processed by DefenseFinder. Finally, we added non-regressive tests in a continuous integration pipeline to make the development of future features more robust against introduction of bugs.

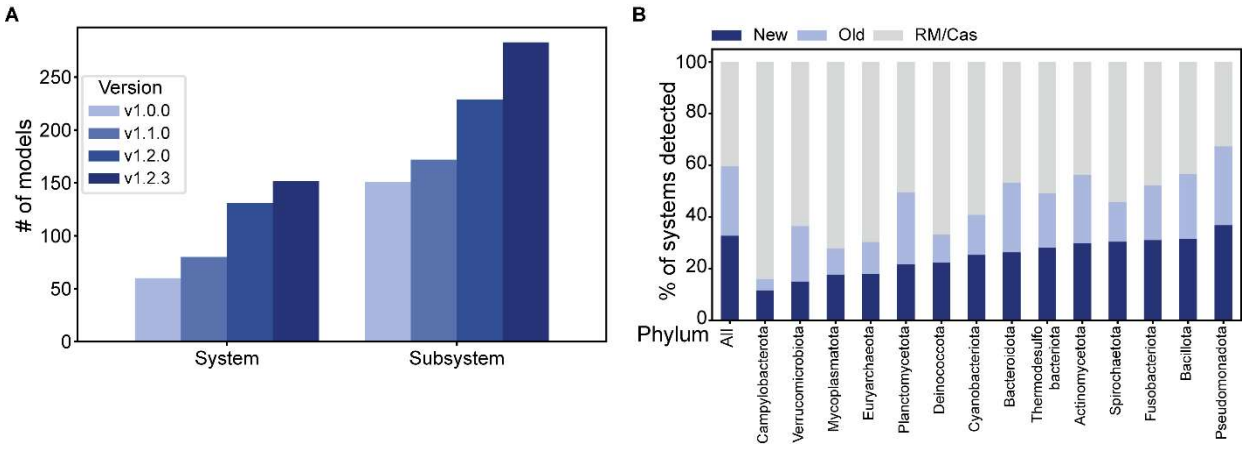

Figure 1. Evolution of DefenseFinder models. A. Evolution of the number of defense systems across the different versions. **B.** Proportion of the currently detected systems which were added in v1.0.0 (Old and RM/Cas) or in more recent versions (New) of DefenseFinder models across different phyla of prokaryotes.

An improved web service

DefenseFinder is a Python program running in the command line. To enhance the software's usability, DefenseFinder has been available as a web service since its launch. Some useful features were missing, such as the possibility of accessing previous analyses. We updated the web service to add new features: a simple interface for depositing fasta nucleic acid or fasta amino acid entries on the "Home" page (**Figure 2A**). Results of previous analyses are now displayed in the “Analysis” page. On this page, jobs can be accessed and renamed at convenience for 1 month without any action on the web service. All the different DefenseFinder outputs are displayed in dynamic tables that can be easily downloaded. The results are displayed in a genome browser to better visualize the hits in their context (**Figure 2B**). All results are displayed, including orphan HMMs, which do not form a system. Those orphan HMMs should be used cautiously and analyzed using their score and profile coverage. From the result table, getting more information on a given system is now easier via a link to the collaborative knowledge base. Importantly for the community, the latest version of DefenseFinder has been packaged in Galaxy [32] and can be run on any Galaxy instance.

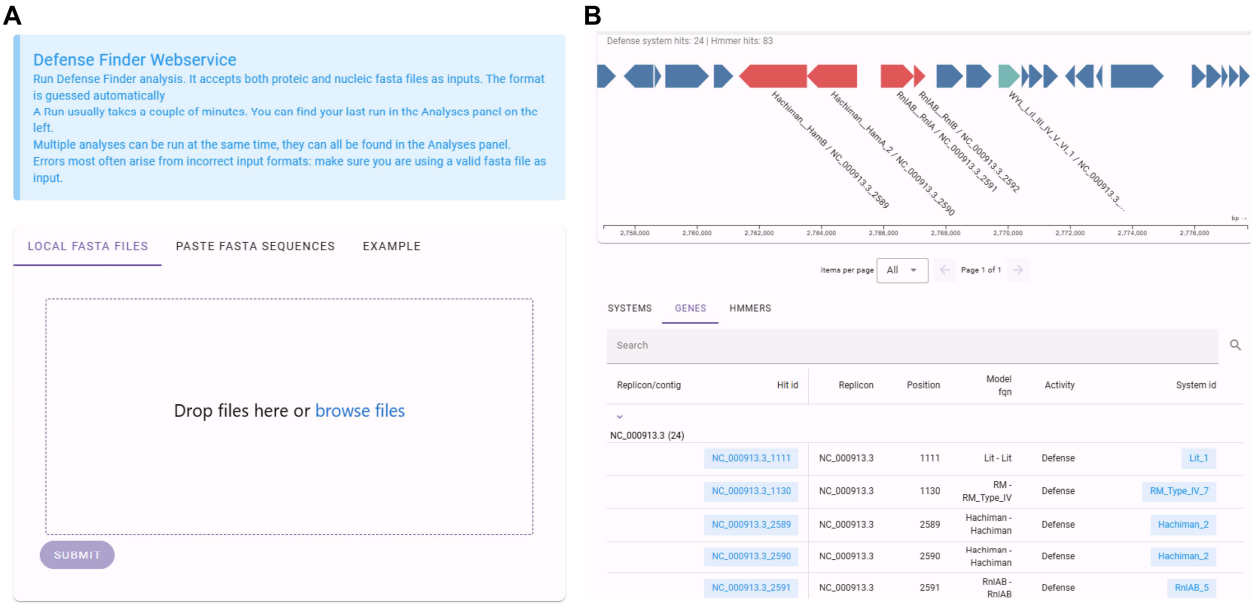

Figure 2. Architecture of DefenseFinder Webservice. **A.** Homepage of the DefenseFinder webservice. **B.** Results page of the webservice. Three different output tables and a visualization of the chromosome with the different defense system hits are available.

A collaborative knowledge base for Defense systems

DefenseFinder output constitutes 3 tables (defense_finder_systems, defense_finder_genes, and defense_finder_hmmer) containing respectively one line per detected system, detected genes inside a system, and defensive HMM profiles hits. Given the large (and ever-growing) number of defense systems, this output can be hard to analyze. Thus, we decided to create a participative knowledge base of defense systems available at https://defensefinder.mdmlab.fr/wiki/. This website aims to provide information to understand better bioinformatic predictions provided by DefenseFinder and to allow a dynamic, up-to-date sharing of knowledge on antiphage systems. The wiki provides a few pages on general concepts of the field and a page per defense system detected by DefenseFinder. The different pages are accessible directly from the results of the DefenseFinder web service or can be explored individually and are all summarized in a table containing the main references, part of their mechanism (sensor, activator, effector) when this information is available in the literature and the Pfam domain present in the system (**Figure 3A**).

Each defense system page is organized into distinct sections (**Figure 3B**). 1/ **Description**: summarizing simple information such as the system's discovery, protein names, and known domains. 2/ **The molecular mechanism**: summarizing known mechanisms and highlighting instances where the mechanism remains unknown. 3/ **Genomic architecture**: a detailed breakdown of system components and associated domains is provided, accompanied by examples of genomic loci organization. 4/ **Distribution among prokaryotes**: the system distribution across phylogenetic phyla using DefenseFinder results on the RefSeq complete genome database. 5/ **3D Structure of the system**: showcases both experimentally validated 3D structures (if available) and predicted structures using AlphaFold2 [33] for validated systems. Foldseek [34] results starting from the given predicted structure are also precomputed and available.

6/ **Experimental validation**: shows the tested system, the expression organism and against which phage(s) it was shown to be effective. 7/ **References**: relevant publications (denoted with a star) are included to facilitate a deeper understanding of each system.

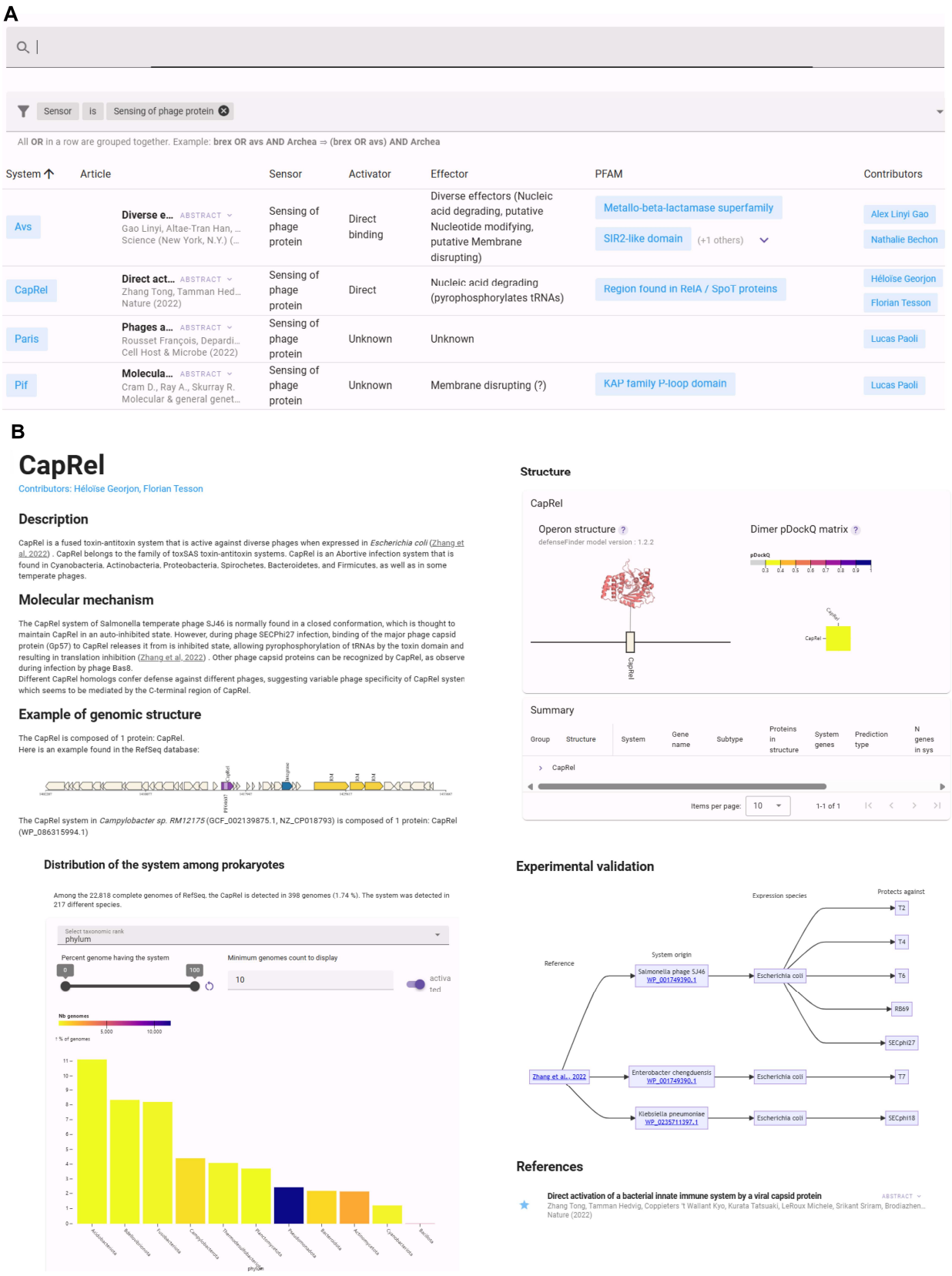

Figure 3. Architecture of the DefenseFinderWiki collaborative knowledge base. A. Table presenting all the different defense systems. The table can be filtered by system, type of

sensor/activator or effector and Pfam inside the system. **B.** An example of one page of the knowledge base covering the CapRel defense system.

Each system featured in DefenseFinder models v1.2.3 has a dedicated page on the knowledge base. These pages were populated through collaborative efforts, including a hackathon organized internally and contributions from other researchers from the community. Recognizing the rapid pace of defense system discovery and mechanism characterization, we opted for a collaborative approach by making the website open for updates. To foster a collaborative framework, we based the website on a Gitlab repository, with pages written in Markdown and designed to be easy to use by people that are not Git experts. To modify and add information to wiki pages, contributors can click on “Edit a page” (found at the bottom of every page), add some content (and add themselves as contributors) and create an automatic merge request that will be verified by at least one expert in the field before being merged. When describing a new system, the discoverers are encouraged to request the creation of a new page to describe it with all the necessary information.

Precomputed results of DefenseFinder on 22,738 complete genomes.

Running DefenseFinder on a large database is time and resource-consuming; thus, we created a precomputed database. DefenseFinder v1.1.0 with models v1.2.3 was run on the complete genomes database from RefSeq (From July 2022, Bacteria N = 22,422, Archaea N = 381). A total of 152,386 defense systems were detected from 152 different defense systems (264 subtypes). Those precomputed results can be visualized on the “RefSeq DB” tab of the DefenseFinder website: https://defensefinder.mdmlab.fr/wiki/refseq/. Results are displayed in an interactive table, allowing research by system, accession, or taxonomy (**Figure 4A**). This table is linked to interactive graphs displaying the relative abundances of the system in the phylogeny and distribution of systems (**Figure 4B and 4C**). Results (tables and graphs) can be downloaded from the website either as a whole or only a subset using different filters. Sequences of detected proteins are accessible through a link to their NCBI protein page.

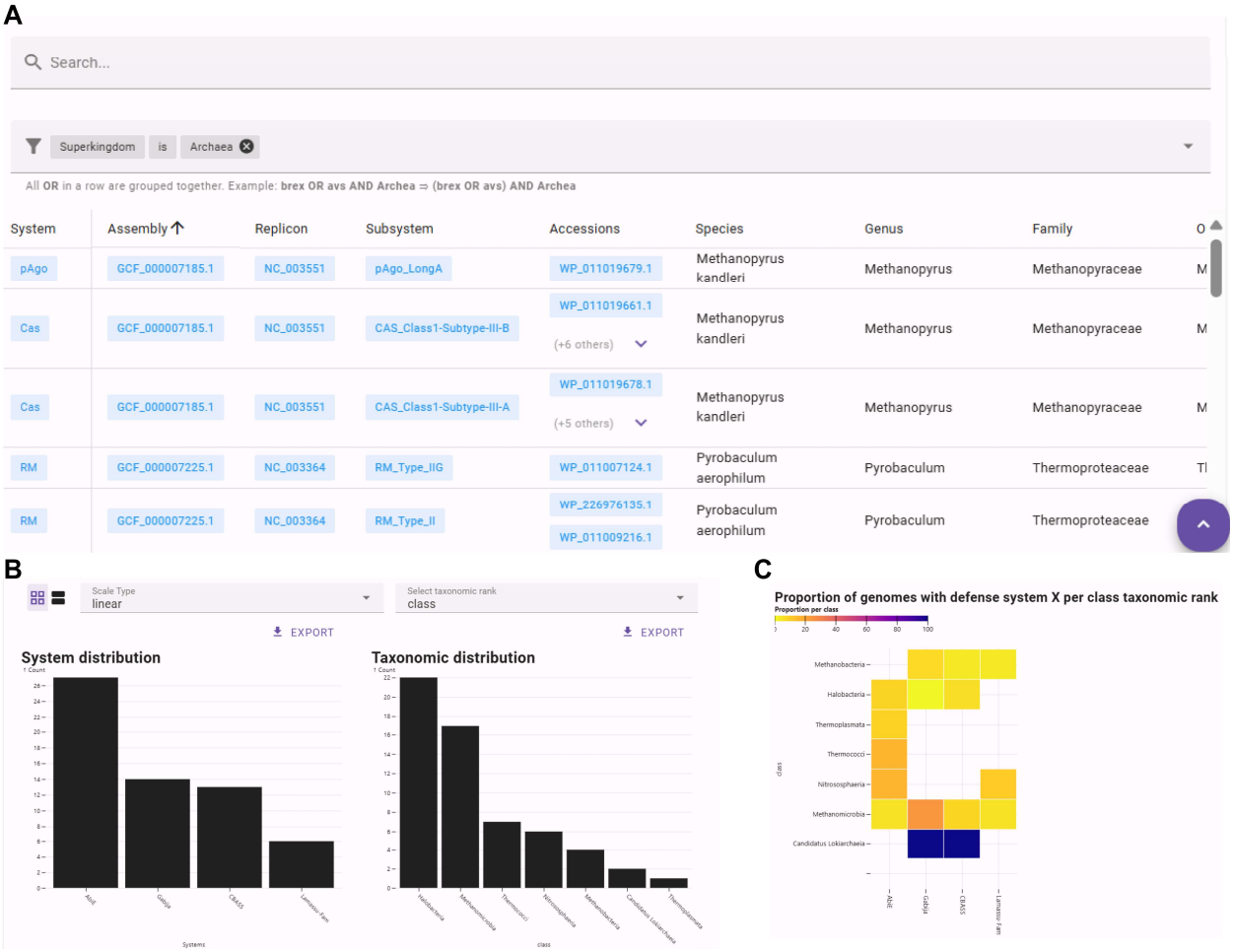

Figure 4. DefenseFinder results database A. Main table used to search through the DefenseFinder results. Searching through this table can be done using only keywords or by using a combination of conditions on the system, the phylogeny, or the name of the replicon. **B.** Interactive bar charts of the count of each system on the left and the count of each taxonomic order reactive to the filter in the main table. **C.** Interactive heatmap displaying the count of each system in each taxonomic order reactive to the main table. The taxonomic level can be modified from superkingdom to the species level for both the barchart and heatmap.

Precomputed 3D structure predictions of defense systems

Many mechanisms of the newly described systems remain to be elucidated. Multiple studies were conducted to elucidate such mechanisms[5,9,13,35–38]. Often, the 3D structure of the system was an important step towards understanding its mechanism[4,8,10,39–41]. Recent developments in structure predictions, such as the development of AlphaFold2 or ESMFold[33,42], allow good-quality predictions of proteins and protein complexes. We thus created a database of predicted structures for known defense systems.

Experimentally validated proteins were retrieved for each system, and the 3D predicted structure was computed using AlphaFold2. For some systems, we could not find the original protein sequence or accession of the experimentally validated system. Some subsystems (CBASS[43],

227 Retron[44,45], Lamassu[29], etc.) were not experimentally validated and are included in
228 DefenseFinder. In such cases, we selected a different representative from DefenseFinder for
229 structural prediction (See Methods). Recent studies demonstrated that many systems function as
230 complexes [4,8,10,39,40]. To provide insight on the possible complexes, all possible dimers (homo and
231 heterodimers) were computed for each system. For each predicted complex, pDockQ scores [46]
232 were computed to assess the probability of the protein-protein interaction. For several systems
233 with three or more proteins, complexes with up to four proteins were also computed in 1:1
234 stoichiometry. To discover similar folds in structural databases, we ran FoldSeek[34] using the
235 predicted 3D structure of the monomers against the PDB [47] and AlphaFold Uniprot[48,49] databases,
236 and provided the precomputed results in the structure table.
237
238 In total, we provide more than 1,500 predictions of monomers, homodimers, or heterodimers.
239 Results are available on the DefenseFinder structure database at:
240 https://defensefinder.mdmlab.fr/wiki/structure/ as a table with all predicted structures present in
241 the database (**Figure 5A**). Users can search for specific proteins or systems as well as filter all
242 the structure by monomers, dimers, or quality statistics (pLDDT, iptm+ptm or pDockQ for
243 multimers). Predicted structures can be visualized on the website using Mol* [50] with the predicted
244 structure's confidence score (pLDDT) (**Figure 5B**). Results can also be downloaded individually
245 or in bulk directly on the website.
246
247
248
249

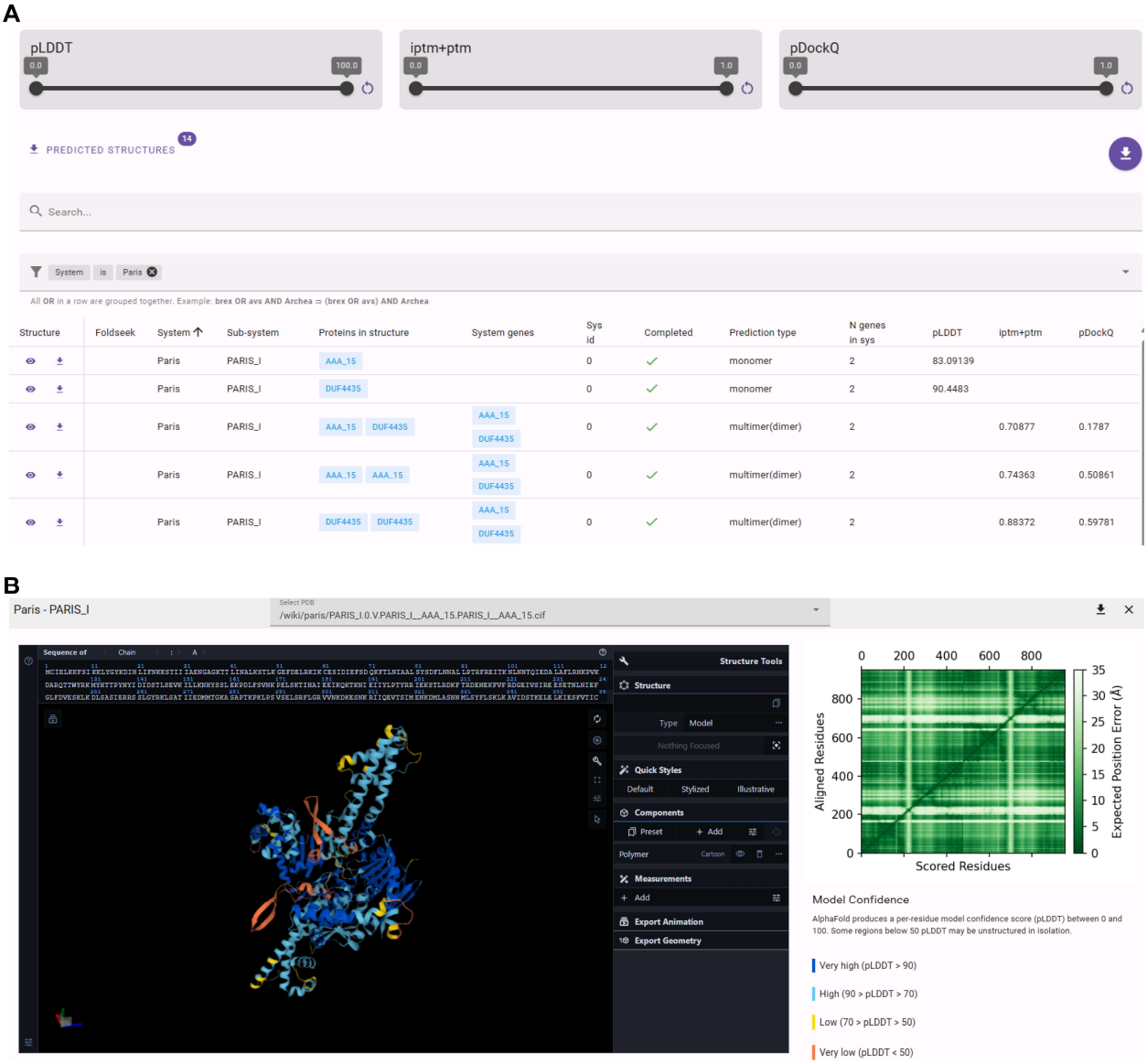

Figure 5. DefenseFinder predicted structures' database: Organization of the structure database. **A.** The dynamic table to search across the 1,543 predicted structures. **B.** Example of protein structure visualization with a dimer of AriA_AAA15 from the Paris defense system.

Conclusion and perspectives

The field of antiphage systems is growing rapidly, which can be difficult to follow for many researchers and students whose primary focus is not antiphage systems. Typically, literature reviews can become quickly outdated. At the same time, bioinformatics predictions guide many studies on antiphage systems. We offer a comprehensive resource for studying antiphage systems to lower the barrier of bioinformatics predictions and to have flexible and fast-evolving access to knowledge on antiphage systems.

The integration of DefenseFinderWiki enables the outputs produced by DefenseFinder to be linked to wiki-style pages, consolidating information on various defense systems thereby improving the understanding of the defense arsenal. Identifying defense systems can pose computational challenges, particularly when applied to large datasets. Here, we ran DefenseFinder on the RefSeq database. The resulting data is showcased through a dynamic table, complemented by visualizations depicting their distribution across the prokaryotic taxonomy. The final section of the website is devoted to protein structures, showcasing visual representations of predicted structures for all identified defense systems.

The ongoing commitment to regular updates of DefenseFinder ensures the incorporation of newly discovered defense systems, thereby enhancing the platform's utility and adaptability within the scientific community. To match the speed of discovery, we built the platform as a collaborative database. This will allow researchers from the community to add information about previously described systems and to create new pages for undescribed defense systems. This database is easy to modify via Gitlab, even for non-git users, while taking advantage of a version-controlled system.

We will continue to develop the community aspect of the knowledge base by providing tutorials and organizing workshops to encourage people to contribute to the project. New updates will be made to increase the information on the website (new predicted structure, alphafoldDB[49], increase in the number of genomes, sequence availability). We plan to add in a future release, a new section where users can test whether a system is related to a known one or not. If the system is new, we will provide a form to add the new system both for DefenseFinder and the website.

Overall, we are providing a new hub, gathering many resources that we hope will be useful for those exploring antiphage defense systems. This website aims to support newcomers in the field, including students, seasoned researchers, and enthusiasts of defense systems, by providing a platform to initiate work or deepen their understanding and knowledge in this domain.

Materials and Methods

Website development

This website comprises two distinct components: the webservice and the knowledge base (wiki and databases) each functioning separately. On the frontend, both utilize Vue.js and its ecosystem, particularly Nuxt.

The Webservice: The backend is built on top of Django and Django Ninja, designed for user-friendly and intuitive API development. It establishes communication with the Pasteur Galaxy [51] instance to execute the DefenseFinder tool and retrieve results. Subsequently, outcomes are stored in a PostgreSQL database accessible through anonymous sessions.

The knowledge base: Devoid of a backend, this is a fully static website implemented using Nuxt and Nuxt Content. Nuxt Content, a Git-based Headless CMS, facilitates the creation and management of static pages via Markdown or JSON files stored in a GitLab repository. This architectural approach alleviates the need for database and backend maintenance, delegating these responsibilities to GitLab, which seamlessly handles authentication, permissions, modification history, and a rich web editor. Therefore, the content is easily editable and accessible. The wiki also provides an interface for searching through large datasets (RefSeq and structure predictions) using Meilisearch (self-hosted) with filters and complex queries.
The plots are generated using Observable Plot and D3.
Both websites undergo automated deployment to a Kubernetes cluster through a GitLab CI workflow. Additionally, a custom linter can be executed against the markdown content via the GitLab web interface.

Protein selection for structure prediction

Protein accessions of experimentally validated systems were retrieved and used for structure prediction for each system and subsystem. For several subsystems, it was not possible to retrieve experimentally validated sequences for two reasons: no protein sequences or accessions in the original paper or, it's a subsystem with no experimental validation. For those systems, one of the best system hits from DefenseFinder was randomly selected and used for the protein structure prediction. Best hits were selected based on their hit scores and profile coverage (fourth quantile of hit score for each gene of the system and more than 75% of profile coverage).

AlphaFold2 (v2.3.1) [33] was used to predict protein structures, using the AlphaFold-Multimer [52] protocol for complexes. Monomer and complex models were sorted based on pLDDT (mean of predicted Local Distance Difference Test) and iptm+ptm (weighted sum of interface and all residues template-modeling scores), respectively. The models with the highest scores were taken for subsequent analysis. Additionally, for dimer structures, pDockQ scores (predicted DockQ) [46] were calculated to assess interface quality; (models with acceptable quality are considered if they have DockQ ≥ 0.23). For all monomeric structures we used FoldSeek (v5) [34] to perform structure similarity searches against the PDB [47] and AlphaFold Uniprot [48,49] databases.

Precomputed DefenseFinder results

For the precomputed database, we used all the RefSeq complete genomes [53] of both Bacteria (N = 22,422) and Archaea (N = 381) from July 2022. DefenseFinder v1.2.0 [21] using models v1.2.3 was run on all genomes separately using standard settings (coverage 0.4).

Pfam annotation

We ran hmmsearch HMMER 3.3.2 [54] on proteins that were detected by DefenseFinder with a coverage above 75%, to make sure they are complete, against Pfam-A database v33.1 [55]. For each protein family of a given defense system, PFAM that hits at least 50% of the members of the family are assigned to the protein family and therefore to the system.

Pfam annotation of the example of genomic locus was done on each protein of the system using hmmsearch "with --ga_cut" argument. If two PFAM hits were overlapping in a single protein sequence, only the best hit (hit_score) was kept.

New profile modeling

New HMM profiles were built using the same method as in the first version of DefenseFinder (see methods in [21]. For protein with a single representative available, the first profile was made either by blasting one homolog (if only one is available) on BLAST RefSeq (nr) [56] database (filter with 30% identity and 70% coverage). If a first detection is available in supplementary material, the first profile is done using those proteins.

Addition of new models

Using the architecture of MacSyFinder v2 [57], we added new definitions and profiles inside the previous MacSyFinder models. Inconsistencies between models were then checked and reduced: overlapping systems and systems blocking the detection of the other.

Funding

F. T., H. G., H. V., E. M., A. B. and J. C. are supported by the CRI Research Fellowship to A.B. from the Bettencourt Schueller Foundation, the ATIP-Avenir program from INSERM (R21042KS/RSE22002KSA), the Emergence program from the University of Paris-Cité (RSFVJ21IDXB6_DANA) ERC Starting Grant (PECAN 101040529) and the core funding of Institut Pasteur. H.G. PhD is funded by Generare Bioscience. G. C. A. and A. E. acknowledge support from the Knut and Alice Wallenberg Foundation (project grant 2020-0037) and the Swedish Research Council (Vetenskapsrådet) (grants 2019-01085, 2022-01603 and 2023-02353). The AlphaFold2 computations were enabled by the supercomputing resources Berzelius provided by the National Supercomputer Centre (NSC) at Linköping University and the Knut and Alice Wallenberg foundation. Additional computational resources for structural modeling were provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing at NSC, Chalmers University Centre for Computational Science and Engineering (C3SE), and PDC Centre for High Performance Computing, KTH Royal Institute of Technology, partially funded by the Swedish Research Council through grants 2018-05973 and 2022-06725.

Data availability

DefenseFinder website is available at: https://defensefinder.mdmlab.fr/. This website is free and open to all users and there is no login requirement. The source code of the wiki pages is hosted on the GitLab instance of the Pasteur Institute (https://gitlab.pasteur.fr/mdm-lab/wiki) and is under GPLv3 license.

Acknowledgement

We are grateful to all the members of MDMLab for their help in filling most of the knowledge base pages with relevant information. We thank Nathalie Béchon, Alex L. Gao, Adi Millman, François Rousset, Avigail Stokar-Avihail and Daan Swarts who accepted to write pages on their favorite subjects. We thank the IT Department of Institut Pasteur, including Thomas Menard, in particular, for providing access to the Kubernetes cluster and initial training.

393 **References**
394

395 1. Georjon, H., and Bernheim, A. (2023). The highly diverse antiphage defence systems of
396 bacteria. Nat Rev Microbiol *21*, 686–700. https://doi.org/10.1038/s41579-023-00934-x.

397 2. Nussenzweig, P.M., and Marraffini, L.A. (2020). Molecular Mechanisms of CRISPR-Cas
398 Immunity in Bacteria. Annu. Rev. Genet. *54*, 93–120. https://doi.org/10.1146/annurev-genet-
399 022120-112523.

400 3. Tock, M.R., and Dryden, D.T. (2005). The biology of restriction and anti-restriction. Current
401 Opinion in Microbiology *8*, 466–472. https://doi.org/10.1016/j.mib.2005.06.003.

402 4. Duncan-Lowey, B., Tal, N., Johnson, A.G., Rawson, S., Mayer, M.L., Doron, S., Millman, A.,
403 Melamed, S., Fedorenko, T., Kacen, A., et al. (2023). Cryo-EM structure of the RADAR
404 supramolecular anti-phage defense complex. Cell *186*, 987-998.e15.
405 https://doi.org/10.1016/j.cell.2023.01.012.

406 5. Hsueh, B.Y., Severin, G.B., Elg, C.A., Waldron, E.J., Kant, A., Wessel, A.J., Dover, J.A.,
407 Rhoades, C.R., Ridenhour, B.J., Parent, K.N., et al. (2022). Phage defence by deaminase-
408 mediated depletion of deoxynucleotides in bacteria. Nat Microbiol *7*, 1210–1220.
409 https://doi.org/10.1038/s41564-022-01162-4.

410 6. Ka, D., Oh, H., Park, E., Kim, J.-H., and Bae, E. (2020). Structural and functional evidence
411 of bacterial antiphage protection by Thoeris defense system via NAD+ degradation. Nat
412 Commun *11*, 2816. https://doi.org/10.1038/s41467-020-16703-w.

413 7. Rousset, F., Yirmiya, E., Nesher, S., Brandis, A., Mehlman, T., Itkin, M., Malitsky, S.,
414 Millman, A., Melamed, S., and Sorek, R. (2023). A conserved family of immune effectors
415 cleaves cellular ATP upon viral infection. Cell *186*, 3619-3631.e13.
416 https://doi.org/10.1016/j.cell.2023.07.020.

417 8. Shen, Z., Lin, Q., Yang, X.-Y., Fosuah, E., and Fu, T.-M. (2023). Assembly-mediated
418 activation of the SIR2-HerA supramolecular complex for anti-phage defense. Molecular Cell
419 *83*, 4586-4599.e5. https://doi.org/10.1016/j.molcel.2023.11.007.

420 9. Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E.,
421 Avraham, C., Brandis, A., Mehlman, T., et al. (2022). Bacteria deplete deoxynucleotides to
422 defend against bacteriophage infection. Nat Microbiol *7*, 1200–1209.
423 https://doi.org/10.1038/s41564-022-01158-0.

424 10. Tang, D., Chen, Y., Chen, H., Jia, T., Chen, Q., and Yu, Y. (2023). Multiple enzymatic
425 activities of a Sir2-HerA system cooperate for anti-phage defense. Molecular Cell *83*, 4600-
426 4613.e6. https://doi.org/10.1016/j.molcel.2023.11.010.

427 11. Duncan-Lowey, B., McNamara-Bordewick, N.K., Tal, N., Sorek, R., and Kranzusch, P.J.
428 (2021). Effector-mediated membrane disruption controls cell death in CBASS antiphage
429 defense. Molecular Cell *81*, 5039-5051.e5. https://doi.org/10.1016/j.molcel.2021.10.020.

430 12. Durmaz, E., and Klaenhammer, T.R. (2007). Abortive Phage Resistance Mechanism AbiZ
431 Speeds the Lysis Clock To Cause Premature Lysis of Phage-Infected Lactococcus lactis. J
432 Bacteriol *189*, 1417–1425. https://doi.org/10.1128/JB.00904-06.

433 13. Johnson, A.G., Wein, T., Mayer, M.L., Duncan-Lowey, B., Yirmiya, E., Oppenheimer-
434 Shaanan, Y., Amitai, G., Sorek, R., and Kranzusch, P.J. (2022). Bacterial gasdermins reveal
435 an ancient mechanism of cell death. Science *375*, 221–225.
436 https://doi.org/10.1126/science.abj8432.

437 14. Zeng, Z., Chen, Y., Pinilla-Redondo, R., Shah, S.A., Zhao, F., Wang, C., Hu, Z., Wu, C.,
438 Zhang, C., Whitaker, R.J., et al. (2022). A short prokaryotic Argonaute activates membrane
439 effector to confer antiviral defense. Cell Host & Microbe *30*, 930-943.e6.
440 https://doi.org/10.1016/j.chom.2022.04.015.

441 15. Bernheim, A., Millman, A., Ofir, G., Meitav, G., Avraham, C., Shomar, H., Rosenberg, M.M.,
442 Tal, N., Melamed, S., Amitai, G., et al. (2021). Prokaryotic viperins produce diverse antiviral
443 molecules. Nature *589*, 120–124. https://doi.org/10.1038/s41586-020-2762-2.

444 16. Hussain, F.A., Dubert, J., Elsherbini, J., Murphy, M., VanInsberghe, D., Arevalo, P.,
445 Kauffman, K., Rodino-Janeiro, B.K., Gavin, H., Gomez, A., et al. (2021). Rapid evolutionary
446 turnover of mobile genetic elements drives bacterial resistance to phages. Science *374*,
447 488–492. https://doi.org/10.1126/science.abb1083.

448 17. Piel, D., Bruto, M., Labreuche, Y., Blanquart, F., Goudenège, D., Barcia-Cruz, R.,
449 Chenivesse, S., Le Panse, S., James, A., Dubert, J., et al. (2022). Phage–host coevolution
450 in natural populations. Nat Microbiol *7*, 1075–1086. https://doi.org/10.1038/s41564-022-
451 01157-1.

452 18. Costa, A.R., Berg, D.F. van den, Esser, J.Q., Muralidharan, A., Bossche, H. van den,
453 Bonilla, B.E., Steen, B.A. van der, Haagsma, A.C., Fluit, A.C., Nobrega, F.L., et al. (2023).
454 Accumulation of defense systems in phage resistant strains of Pseudomonas aeruginosa.
455 Preprint at bioRxiv, https://doi.org/10.1101/2022.08.12.503731
456 https://doi.org/10.1101/2022.08.12.503731.

457 19. Gaborieau, B., Vaysset, H., Tesson, F., Charachon, I., Dib, N., Bernier, J., Dequidt, T.,
458 Georjon, H., Clermont, O., Hersen, P., et al. (2023). Predicting phage-bacteria interactions
459 at the strain level from genomes. Preprint at bioRxiv,
460 https://doi.org/10.1101/2023.11.22.567924 https://doi.org/10.1101/2023.11.22.567924.

461 20. Tesson, F., and Bernheim, A. (2023). Synergy and regulation of antiphage systems: toward
462 the existence of a bacterial immune system? Current Opinion in Microbiology *71*, 102238.
463 https://doi.org/10.1016/j.mib.2022.102238.

464 21. Tesson, F., Hervé, A., Mordret, E., Touchon, M., d'Humières, C., Cury, J., and Bernheim, A.
465 (2022). Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat
466 Commun *13*, 2561. https://doi.org/10.1038/s41467-022-30269-9.

467 22. Chopin, M.-C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci:
468 variations on a theme. Curr Opin Microbiol *8*, 473–479.
469 https://doi.org/10.1016/j.mib.2005.06.006.

470 23. Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013).
471 Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. Mol Cell *50*,
472 136–148. https://doi.org/10.1016/j.molcel.2013.02.002.

24. Hazan, R., and Engelberg-Kulka, H. (2004). Escherichia coli mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol Genet Genomics *272*, 227–234. https://doi.org/10.1007/s00438-004-1048-y.

25. Dedrick, R.M., Jacobs-Sera, D., Bustamante, C.A.G., Garlena, R.A., Mavrich, T.N., Pope, W.H., Reyes, J.C.C., Russell, D.A., Adair, T., Alvey, R., et al. (2017). Prophage-mediated defence against viral attack and viral counter-defence. Nat Microbiol *2*, 1–13. https://doi.org/10.1038/nmicrobiol.2016.251.

26. Mageeney, C.M., Mohammed, H.T., Dies, M., Anbari, S., Cudkevich, N., Chen, Y., Buceta, J., and Ware, V.C. (2020). Mycobacterium Phage Butters-Encoded Proteins Contribute to Host Defense against Viral Attack. mSystems *5*, e00534-20. https://doi.org/10.1128/mSystems.00534-20.

27. Bobadilla Ugarte, P., Barendse, P., and Swarts, D.C. (2023). Argonaute proteins confer immunity in all domains of life. Current Opinion in Microbiology *74*, 102313. https://doi.org/10.1016/j.mib.2023.102313.

28. Rousset, F., Depardieu, F., Miele, S., Dowding, J., Laval, A.-L., Lieberman, E., Garry, D., Rocha, E.P.C., Bernheim, A., and Bikard, D. (2022). Phages and their satellites encode hotspots of antiviral systems. Cell Host & Microbe *30*, 740-753.e5. https://doi.org/10.1016/j.chom.2022.02.018.

29. Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hör, J., Garb, J., Bechon, N., Brandis, A., Lopatina, A., et al. (2022). An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe *30*, 1556-1569.e5. https://doi.org/10.1016/j.chom.2022.09.017.

30. Néron, B., Denise, R., Coluzzi, C., Touchon, M., Rocha, E.P.C., and Abby, S.S. (2023). MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. Peer Community Journal *3*. https://doi.org/10.24072/pcjournal.250.

31. Larralde, M. (2022). Pyrodigal: Python bindings and interface to Prodigal, an efficient method for gene prediction in prokaryotes. Journal of Open Source Software *7*, 4296. https://doi.org/10.21105/joss.04296.

32. The Galaxy Community (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Research *50*, W345–W351. https://doi.org/10.1093/nar/gkac247.

33. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

34. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. Nat Biotechnol, 1–4. https://doi.org/10.1038/s41587-023-01773-0.

511 35. Koga, M., Otsuka, Y., Lemire, S., and Yonesaki, T. (2011). Escherichia coli rnlA and rnlB
512 Compose a Novel Toxin–Antitoxin System. Genetics *187*, 123–130.
513 https://doi.org/10.1534/genetics.110.121798.

514 36. LeRoux, M., Srikant, S., Littlehale, M.H., Teodoro, G., Doron, S., Badiee, M., Leung, A.K.L.,
515 Sorek, R., and Laub, M.T. (2021). The DarTG toxin-antitoxin system provides phage
516 defense by ADP-ribosylating viral DNA. bioRxiv, 2021.09.27.462013.
517 https://doi.org/10.1101/2021.09.27.462013.

518 37. Wein, T., Johnson, A.G., Millman, A., Lange, K., Yirmiya, E., Hadary, R., Garb, J.,
519 Steinruecke, F., Hill, A.B., Kranzusch, P.J., et al. (2023). CARD-like domains mediate anti-
520 phage defense in bacterial gasdermin systems. bioRxiv, 2023.05.28.542683.
521 https://doi.org/10.1101/2023.05.28.542683.

522 38. Zhang, T., Tamman, H., Coppieters ’t Wallant, K., Kurata, T., LeRoux, M., Srikant, S.,
523 Brodiazhenko, T., Cepauskas, A., Talavera, A., Martens, C., et al. (2022). Direct activation
524 of a bacterial innate immune system by a viral capsid protein. Nature *612*, 132–140.
525 https://doi.org/10.1038/s41586-022-05444-z.

526 39. Antine, S.P., Johnson, A.G., Mooney, S.E., Leavitt, A., Mayer, M.L., Yirmiya, E., Amitai, G.,
527 Sorek, R., and Kranzusch, P.J. (2023). Structural basis of Gabija anti-phage defence and
528 viral immune evasion. Nature. https://doi.org/10.1038/s41586-023-06855-2.

529 40. Deep, A., Gu, Y., Gao, Y.-Q., Ego, K.M., Herzik, M.A., Zhou, H., and Corbett, K.D. (2022).
530 The SMC-family Wadjet complex protects bacteria from plasmid transformation by
531 recognition and cleavage of closed-circular DNA. Molecular Cell *82*, 4145-4159.e7.
532 https://doi.org/10.1016/j.molcel.2022.09.008.

533 41. Li, Y., Shen, Z., Zhang, M., Yang, X.-Y., Cleary, S.P., Xie, J., Marathe, I.A., Kostelic, M.,
534 Greenwald, J., Rish, A.D., et al. (2024). PtuA and PtuB assemble into an inflammasome-like
535 oligomer for anti-phage defense. Nat Struct Mol Biol, 1–11. https://doi.org/10.1038/s41594-
536 023-01172-8.

537 42. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O.,
538 Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with
539 a language model. Science *379*, 1123–1130. https://doi.org/10.1126/science.ade2574.

540 43. Millman, A., Melamed, S., Amitai, G., and Sorek, R. (2020). Diversity and classification of
541 cyclic-oligonucleotide-based anti-phage signalling systems. Nat Microbiol *5*, 1608–1615.
542 https://doi.org/10.1038/s41564-020-0777-y.

543 44. Mestre, M.R., González-Delgado, A., Gutiérrez-Rus, L.I., Martínez-Abarca, F., and Toro, N.
544 (2020). Systematic prediction of genes functionally associated with bacterial retrons and
545 classification of the encoded tripartite systems. Nucleic Acids Research *48*, 12632–12647.
546 https://doi.org/10.1093/nar/gkaa1149.

547 45. Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A.,
548 Oppenheimer-Shaanan, Y., and Sorek, R. (2020). Bacterial Retrons Function In Anti-Phage
549 Defense. Cell *183*, 1551-1561.e12. https://doi.org/10.1016/j.cell.2020.09.065.

46. Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun *13*, 1265. https://doi.org/10.1038/s41467-022-28865-w.

47. Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, P.A., Crichlow, G.V., Dalenberg, K., Duarte, J.M., et al. (2023). RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. Nucleic Acids Res *51*, D488–D508. https://doi.org/10.1093/nar/gkac1077.

48. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

49. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

50. Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S.K., Koča, J., and Rose, A.S. (2021). Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Research *49*, W431–W437. https://doi.org/10.1093/nar/gkab314.

51. Mareuil, F., Doppelt-Azeroual, O., and Ménager, H. (2017). A public Galaxy platform at Pasteur used as an execution engine for web services . F1000Research *6*. https://doi.org/10.7490/f1000research.1114334.1.

52. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer (Bioinformatics) https://doi.org/10.1101/2021.10.04.463034.

53. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res *44*, D733-745. https://doi.org/10.1093/nar/gkv1189.

54. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLOS Computational Biology *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

55. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

56. 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. https://doi.org/10.1186/1471-2105-10-421.

57. Néron, B., Denise, R., Coluzzi, C., Touchon, M., Rocha, E.P.C., and Abby, S.S. (2022). MacSyFinder v2: Improved modelling and search engine to identify molecular systems in

589 genomes. Preprint at bioRxiv, https://doi.org/10.1101/2022.09.02.506364
590 https://doi.org/10.1101/2022.09.02.506364.

591