



HAL
open science

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

F. Tesson, R. Planel, A. Egorov, H. Georjon, H. Vaysset, B. Brancotte, B. Néron, E. Mordret, G. Atkinson, A Bernheim, et al.

► To cite this version:

F. Tesson, R. Planel, A. Egorov, H. Georjon, H. Vaysset, et al.. A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases. 2024. hal-04600530

HAL Id: hal-04600530

<https://hal.science/hal-04600530v1>

Preprint submitted on 4 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

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

3 *Tesson F.^{1,4}, Planel R³, Egorov A², Georjon H.¹, Vaysset H.^{1,5,6}, Brancotte B³, Néron B³, Mordret E.¹,*
4 *Atkinson G², Bernheim A^{*1}, Cury J.^{1*}*

5
6 ¹Institut Pasteur, Université Paris Cité, INSERM U1284, Molecular Diversity of Microbes lab, 75015, Paris, France

7 ²Department of Experimental Medical Science, and Lund University Virus Centre, Lund University, 221 00 Lund,
8 Sweden

9 ³Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France

10 ⁴Université Paris Cité, INSERM, IAME, 75018 Paris, France

11 ⁵Institut Curie, Université PSL, Sorbonne Université, CNRS, Laboratoire Physico Chimie Curie, UMR 168, 75005 Paris,
12 France

13 ⁶Agroparistech, Paris, France

14

15 **Abstract**

16

17 In recent years, a vast number of novel antiphage defense mechanisms were uncovered. To
18 facilitate the exploration of mechanistic, ecological, and evolutionary aspects related to antiphage
19 defense systems, we released DefenseFinder in 2021 (Tesson et al., 2022). DefenseFinder is a
20 bioinformatic program designed for the systematic identification of all known antiphage defense
21 mechanisms. The initial release of DefenseFinder v1.0.0 included 60 systems. Over the past three
22 years, the number of antiphage systems incorporated into DefenseFinder has grown to 152. The
23 increasing number of known systems makes it a challenge to enter the field and makes the
24 interpretation of detections of antiphage systems difficult. Moreover, the rapid development of
25 sequence-based predictions of structures offers novel possibilities of analysis and should be easily
26 available. To overcome these challenges, we present a hub of resources on defense systems
27 including 1) a search tool (DefenseFinder as a web service) 2) a knowledge base, and 3)
28 precomputed databases (results of DefenseFinder ran on RefSeq database, predicted structure
29 database computed with AlphaFold). These pages can be freely accessed for users as a starting
30 point on their journey to better understand a given system. We anticipate that these resources
31 will foster the use of bioinformatics in the study of antiphage systems and will serve the community
32 of researchers who study antiphage systems. This resource is available at:
33 <https://defensefinder.mdmlab.fr>.

34

35 Introduction

36
37 In the last few years, a considerable number of newly discovered antiphage defense mechanisms
38 have come to light ¹. The most widespread mechanisms, Restriction-Modification and CRISPR-
39 Cas systems, target foreign nucleic acids^{2,3}. However, recent discoveries have revealed an
40 important diversity of molecular modalities by which bacteria defend themselves against phages
41 among which nucleotide depletion ⁴⁻¹⁰, membrane disruption¹¹⁻¹⁴, production of antiphage
42 molecules ¹⁵. Importantly, many antiphage defense mechanisms remain unknown.

43
44 The discovery of diverse antiphage systems not only provides new insights into bacterial immunity
45 mechanisms but also transforms the exploration of interactions between phages and bacteria, as
46 well as microbial evolution. Evaluating the impact of different defense systems in naturally
47 occurring *Vibrio* isolates has revealed that a rapid turnover of a few mobile genetic elements
48 encoding defense systems can completely alter their susceptibility to phages^{16,17}. Ongoing
49 research is delving into the broader phylogenetic scale to explore the role of antiphage systems
50 and their potential implications for phage therapy strategies^{18,19}. Furthermore, examples indicate
51 that antiphage systems can be co-regulated and operate synergistically in multi-layered defense
52 strategies²⁰. Beyond experimental approaches, there is a growing interest in identifying the
53 antiphage systems encoded in diverse species or environments, to understand the diversity of
54 antiviral strategies employed by bacteria.

55
56 To investigate mechanistic, ecological and evolutionary questions related to antiphage defense
57 systems, we created in 2021 DefenseFinder²¹, a tool to systematically detect all known anti phage
58 defense systems. At the time of publication of DefenseFinder v1.0.0, the tool encompassed 60
59 systems. In the last 3 years, the number of antiphage systems included in DefenseFinder has
60 grown to reach 152. The increasing number of systems has been accompanied by emerging
61 challenges in comprehending DefenseFinder results biologically. For non-specialist users, the
62 analysis can be quite intricate, demanding a high level of knowledge in this rapidly evolving field.

63
64 To bridge this gap, we decided to create a website dedicated to defense systems, encompassing
65 both an improved version of a web service to run DefenseFinder and diverse databases to
66 illuminate and increase the understandability of bioinformatic detection of antiphage defense
67 systems. Here, we provide the release of 92 new defense systems models and a software update
68 as well as 3 databases. 1/ A collaborative knowledge base (wiki) summarizing information on
69 known defense systems. 2/ A structure database, with experimentally determined as well as
70 AlphaFold2 predicted structures. 3/ A precomputed database of DefenseFinder results over more
71 than 20 000 complete genomes. Besides, we designed the website to keep it as up to date as
72 possible. Wiki pages are easily editable by anyone and reviewed by experts in the field before
73 publication on the website. They can also be edited automatically to generate sections that
74 aggregate data such as the phylogenetic distribution of a system. The information of the web
75 server is integrated into the rest of the website.

76
77
78

79 Results

80

81 Updates of DefenseFinder program

82 To improve detections performed by DefenseFinder, we updated both the program and the
83 models.

84

85 Since the release of DefenseFinder models v1.0.0, we have constantly updated the defense
86 system models included in DefenseFinder by adding a total of 92 defense systems and 132
87 subsystems (**Figure 1A**). Among those systems, 63 were discovered and described after the
88 release of the first version of DefenseFinder models. Other systems were missed in the first
89 version and are now added after a deeper literature review of the field. Those systems encompass
90 the “Abi” group discovered between 1990 and 2006. DefenseFinder models v1.0.0 included 3 of
91 these systems, while the latest version v1.2.3 detects the 22 different ones²². Beyond Abi
92 systems, we also added: SanaTA²³, MazEF²⁴, antiphage defense systems identified in
93 mycobacterium prophages^{25,26} and pAgo²⁷. The defense system named “Rst_DprA-PPRT”²⁸
94 was renamed “ShosTA” as it was initially discovered in 2013²³ but without phage activity testing.
95 Finally, for Lamassu-Fam²⁹, we separated the systems into different subsystems according to
96 the evolving literature on the topic. Models were also adapted to the latest version of MacSyFinder
97³⁰.

98

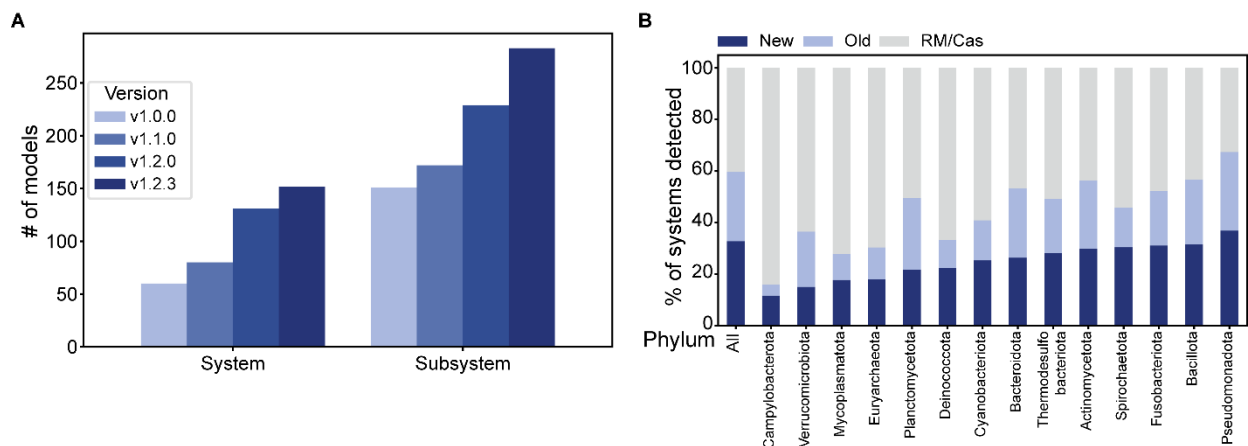
99 To evaluate the proportion of newly detected systems, we ran DefenseFinder v1.2.3 on the
100 RefSeq complete genome database of prokaryotes (Bacteria N = 22,422; Archaea N = 381). We
101 found a total of 152,386 different systems, among which 33% correspond to systems present in
102 v1.2.3 but absent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

105

106 The previous version of DefenseFinder used a “profile coverage” parameter of 0.1. This value
107 was chosen to detect defense systems when the protein was split. This low coverage threshold
108 was balanced with a high score threshold for single gene systems and the necessity to find the
109 different components of the system colocalizing in the genome for multi-genes systems. In the
110 latest version, we changed that threshold to 0.4 as a default value. This threshold was set to
111 maintain the possibility of finding incomplete genes or with a domain replacement while removing
112 very low coverage off target hits. Users can still manually change the coverage threshold (--
113 coverage-profile) to decide on how conservative detection should be depending on the biological
114 problem at hand. For ease of use, we also added the possibility to directly input a nucleic acid
115 fasta file. We use pyrodigal v3.0.1³¹ to identify and annotate the coding regions, which are then
116 processed by DefenseFinder. Finally, we added non-regressive tests in a continuous integration
117 pipeline to make the development of future features more robust.

118

119



120
121 **Figure 1. Evolution of DefenseFinder models.** **A.** Evolution of the number of defense systems
122 across the different versions. **B.** Proportion of the currently detected systems which were added
123 in v1.0.0 (Old and RM/Cas) or in more recent versions (New) of DefenseFinder models across
124 different phyla of prokaryotes.

125 126 **An improved web service**

127 DefenseFinder is a python program running in the command line. To enhance the software's
128 usability, DefenseFinder has been available as a web service since its launch. Some useful
129 features were however missing such as the possibility to access previous analyses. We updated
130 the web service to add new features: a simple interface for depositing fasta nucleic acid or fasta
131 amino acid entries on the "Home" page (**Figure 2A**). Results of previous analyses are now
132 displayed in the "Analysis" page. On this page, jobs can be accessed and renamed at
133 convenience for 1 month without any action on the web service. All the different DefenseFinder
134 outputs are displayed in dynamic tables that can be easily downloaded. The results are displayed
135 in a genome browser to better visualize the hits in their context (**Figure 2B**). From the result table,
136 it is now easier to get more information on a given system via a link to the collaborative knowledge
137 base.

138 Importantly for the community, the latest version of DefenseFinder has been packaged in Galaxy
139 ³² and can be run on any Galaxy instance.

A

Defense Finder

Welcome on DefenseFinder's web page. On this site, you can freely use (without any login) the DefenseFinder webservice (see below) and get help to navigate the ever expanding world of defense systems. There is a collaborative knowledge base of defense systems that provide essential information on any given defense systems. You can further explore defense systems across all fully sequenced bacteria and get insight on the structure of each protein of those systems thanks to protein structure prediction using alphafold for monomer and dimers. We encourage you to contribute to this knowledge database, and if you see any non up-to-date information, let us know, or contribute yourself.

Defense Finder Webservice

Run Defense Finder analysis. It accepts both proteic and nucleic fasta files as inputs. The format is guessed automatically. A Run usually takes a couple of minutes. You can find your last run in the Analyses panel on the left. Multiple analyses can be run at the same time, they can all be found in the Analyses panel. Errors most often arise from incorrect input formats: make sure you are using a valid fasta file as input.

LOCAL FASTA FILES PASTE FASTA SEQUENCES EXAMPLE

Drop files here or browse files

If you are using DefenseFinder please cite:

Systematic and quantitative view of the antiviral arsenal of prokaryotes

Tesson F, Hervé A, Touchon M, d'Humières C, Cury J, Benneheim A, Nature Communication, 2022

MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes

Néron, B, Denise, R, Clakatz, C, Touchon, M, Rocha, E.P.C., Abby, S.S., Peer Community Journal, 2023

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

Tesson, F., Planel, R., Egorov, A., Georjon, H., Vaysse, H., Brancotte, B., Néron, B., Mondret, E., Benneheim, A., Atkinson, G., & Cury, J., Preprint, 2024

List defense systems

The list of all systems describes in the knowledge database can be found here.

DefenseFinder issues

Any issues concerning DefenseFinder, the software, can be raised here.

Wiki issues

Any issues concerning the knowledge database website can be raised here.

B

Home / Analyses / BE.fasta



141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

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

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. This output can be hard to analyze, given the large (and ever-growing) number of defense systems. 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 better understand 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 phylums using DefenseFinder results on the RefSeq complete

167 genome database. 5/ **3D Structure of the system**: showcases both experimentally validated 3D
168 structures (if available) and predicted structures using AlphaFold2 ³³ for validated systems.
169 Foldseek ³⁴ results starting from the given predicted structure are also precomputed and available.
170 6/ **Experimental validation**: shows the tested system, the expression organism and against
171 which phage(s) it was shown to be effective. 7/ **References**: relevant publications (denoted with
172 a star) are included to facilitate a deeper understanding of each system.

A

Knowledge database of all known defense systems

WEBSERVICE WIKI REFSEQ DB STRUCTURES DB HELP

DefenseFinder Wiki

Help

General Conc...

Defense Syst...

List of defense systems

Abi2

AbiA

AbiB

AbiC

AbiD

AbiE

AbiG

AbiH

AbiI

AbiJ

AbiK

AbiL

List of defense systems

The knowledge of anti-phage defense systems is ever expanding. The spectacular increase of the number of known systems in the past years suggests that many of them are still to be discovered. As of April 2022, 63 defense systems have been described. To our knowledge, there are currently in the following table, 151 systems. This table summarizes some info for each systems (that can be search in the **Search...** field below), and contains link toward corresponding wiki page.

LIST SYSTEMS 5

▼ Sensor is Sensing of phage protein

All OR in a row are grouped together. Example: *brex* OR *avs* AND *Archea* ⇒ (*brex* OR *avs*) AND *Archea*

System ↑	Article	Sensor	Activator	Effector	PFAM	Contributors
Avs	ITRAC Ga... Sci...	Sensing of phage protein	Direct binding	Diverse effectors (Nucleic acid degrading, putative Nucleotide modifying, putative Membrane disrupting)	Metallo-beta-lactamase superfamily SIR2-like domain (+1 others)	Alex Linyi Gao Nathalie Bechon
CapRel	ITRAC Zh... Nat...	Sensing of phage protein	Direct	Nucleic acid degrading (pyrophosphorylates tRNAs)	Region found in RelA / SpoT proteins	H�eloise Georjon Florian Tesson

B

CapRel

Contributors: H eloise Georjon, Florian Tesson

Description

CapRel is a fused toxin-antitoxin system that is active against diverse phages when expressed in *Escherichia coli* (Zhang et al., 2022). CapRel belongs to the family of toxSAS toxin-antitoxin systems. CapRel is an Abortive infection system which is found in Cyanobacteria, Actinobacteria, and Proteobacteria, Spirochetes, Bacteroidetes, and Firmicutes, as well as in some temperate phages.

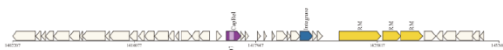
Molecular mechanism

The CapRel system of Salmonella temperate phage S146 is normally found in a closed conformation, which is thought to maintain CapRel in an auto-inhibited state. However during phage SECPH27 infection, binding of the major phage capsid protein (Gp57) to CapRel releases it from its inhibited state, allowing pyrophosphorylation of tRNAs by the toxin domain and resulting in translation inhibition (Zhang et al., 2022). Other phage capsid proteins can be recognized by CapRel, as observed during infection by phage Das8.

Different CapRel homologues confer defense against different phages, suggesting variable phage specificity of CapRel system which seems to be mediated by the C-terminal region of CapRel.

Example of genomic structure

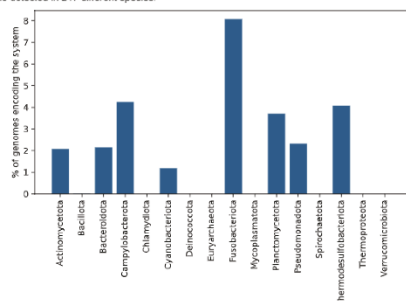
The CapRel is composed of 1 protein: CapRel. Here is an example found in the RefSeq database:



The CapRel system in *Campylobacter* sp. RM12175 (GCF_002139875.1, NZ_CP016793) is composed of 1 protein: CapRel (WP_086315994.1)

Distribution of the system among prokaryotes

Among the 22,003 complete genomes of RefSeq, the CapRel is detected in 398 genomes (1.75%). The system was detected in 217 different species.



Proportion of genome encoding the CapRel system for the 14 phyla with more than 50 genomes in the RefSeq database.

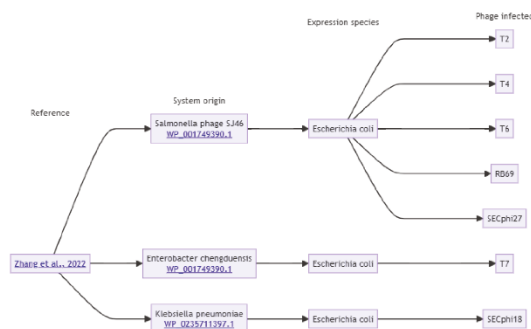
Structure

CapRel

Example 1

Select PDB

Experimental validation



References

★ Direct activation of a bacterial innate immune system by a viral capsid protein ABSTRACT
 Zhang Tong, Tamman Hedvig, Coppieters T Wallant Kyo, Kurita Tatsuki, LeRoux Michele, Srikant Sriram, Brodziszko Tetiana, Ce... Nature (2022)

EDIT ON GITLAB

← BUTTERS_GP57R

CARD_NLR →

173
174
175
176
177
178

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.

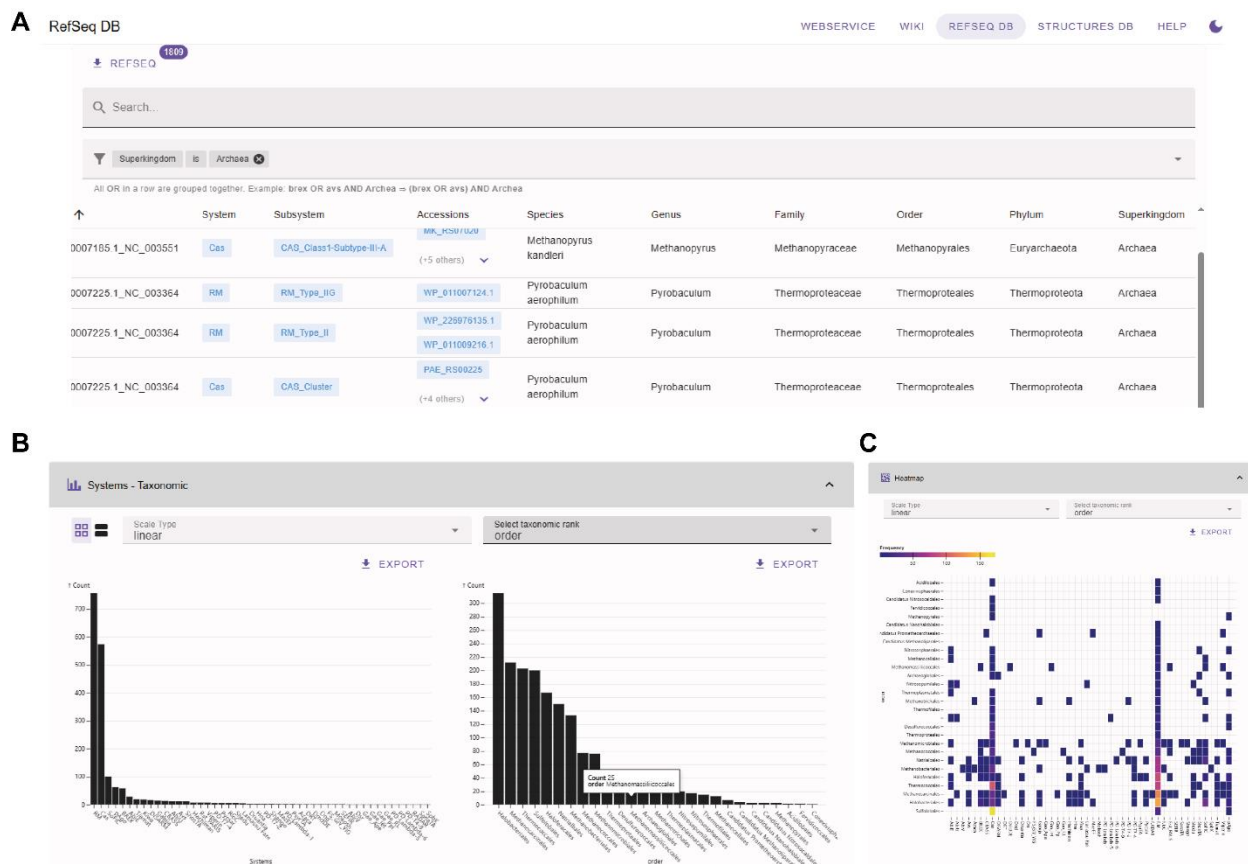
179 Each system featured in DefenseFinder models v1.2.3 has a dedicated page on the knowledge
180 base. These pages were populated through collaborative efforts, including a hackathon organized
181 internally and contributions from other researchers from the community. Recognizing the rapid
182 pace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

190

191 **Precomputed results of DefenseFinder on 22,738 complete genomes.**

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

203



204
205 **Figure 4. DefenseFinder results database A.** Main table used to search through the
206 DefenseFinder results. Searching through this table can be done using only keywords or by using
207 a combination of conditions on the system, the phylogeny, or the name of the replicon. **B.**
208 Interactive bar charts of the count of each system on the left and the count of each taxonomic
209 order reactive to the filter in the main table. **C.** Interactive heatmap displaying the count of each
210 system in each taxonomic order reactive to the main table. The taxonomic level can be modified
211 from superkingdom to the species level for both the bar chart and heatmap.

212

213 Precomputed 3D structures predictions of defense systems

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

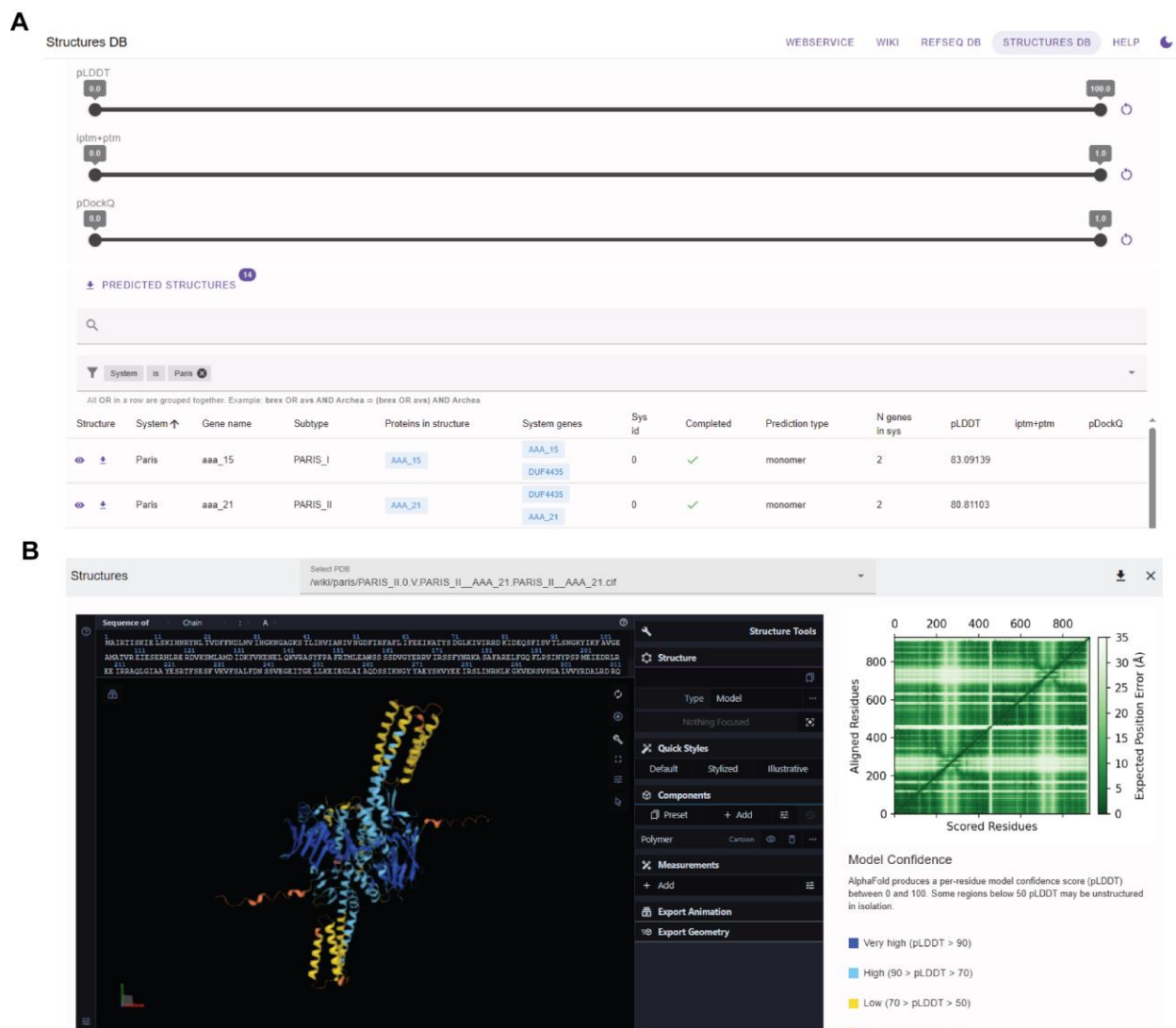
220

221 For each system, experimentally validated proteins were retrieved, and the 3D predicted structure
222 was computed using AlphaFold2. For systems and subsystems where protein accessions were
223 impossible to retrieve, we selected another representative. Recent studies demonstrated that
224 many systems function as complexes^{4,8,10,39,40}. To provide insight on the possible complexes, all
225 possible dimers (homo and heterodimers) were computed for each system. For each predicted
226 complex, pDockQ scores⁴³ were computed to assess the probability of the protein-protein

227 interaction. For several systems with more than 2 proteins, complexes with up to 4 proteins were
228 also computed in 1:1 stoichiometry.
229 To discover similar folds in structural databases, we ran FoldSeek³⁴ using the predicted 3D
230 structure of the monomers against the PDB⁴⁴ and AlphaFold Uniprot^{45,46} databases, and provided
231 the precomputed results in the structure table.
232

233 In total, we provide more than 1,500 predictions of homodimers or dimers. Results are available
234 on the DefenseFinder structure database at: <https://defensefinder.mdmlab.fr/wiki/structure/> as a
235 table with all predicted structures present in the database (**Figure 5A**). Users can search for
236 specific proteins or systems as well as filter all the structure by monomers, dimers, or quality
237 statistics (pLDDT, iptm+ptm or pDockQ for multimers). Predicted structures can be visualized on
238 the website using Mol*⁴⁷ with the confidence score (pLDDT) of the predicted structure (**Figure**
239 **5B**). Results can also be downloaded one by one or in bulk directly on the website.
240

241
242
243



244
245 **Figure 5. DefenseFinder predicted structures' database:** Organization of the structure
246 database. **A.** The dynamic table to search across the 1,543 predicted structures. **B.** Example of
247 protein structure visualization with a dimer of AriA_AAA15 from the Paris defense system.
248

249
250 **Conclusion and perspectives**
251

252 The field of antiphage systems is growing at a rapid pace which can be difficult to follow for many
253 researchers and students whose primary focus of studies is not antiphage systems. Typically,
254 reviews of literature can become quickly outdated. At the same time, bioinformatics predictions
255 allow guiding many studies on antiphage systems. We offer here a comprehensive resource for
256 the study of antiphage systems with the aim of lowering the barrier of bioinformatics predictions
257 and to have flexible and fast evolving access to knowledge on antiphage systems.

258 The integration of DefenseFinderWiki enables the outputs produced by DefenseFinder to be
259 linked to wiki-style pages consolidating information on various defense systems, thereby

260 improving the understanding of the defense arsenal. Identifying defense systems can pose
261 computational challenges, particularly when applied to large datasets. Here, we ran
262 DefenseFinder on the RefSeq database. The resulting data is showcased through a dynamic
263 table, complemented by visualizations depicting their distribution across the prokaryotic
264 taxonomy. The final section of the website is devoted to protein structures, showcasing visual
265 representations of predicted structures for all identified defense systems.

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

273
274 Overall, we are providing a new hub, gathering many resources that we hope will be useful for
275 those embarking on the exploration of antiphage defense systems. This website aims to support
276 newcomers in the field, including students, seasoned researchers, and enthusiasts of defense
277 systems, by providing a platform to initiate work or deepen their understanding and knowledge in
278 this domain.

279 **Materials and Methods**

280 *Website development*

281
282 This website comprises two distinct components: the webservice and the knowledge base (wiki
283 and databases) each functioning separately. On the frontend, both utilize Vue.js and its
284 ecosystem, particularly Nuxt.

285
286 The Webservice: The backend is built on top of Django and Django Ninja, designed for user-
287 friendly and intuitive API development. It establishes communication with the Pasteur Galaxy ⁴⁸
288 instance to execute the DefenseFinder tool and retrieve results. Subsequently, outcomes are
289 stored in a Postgresql database accessible through anonymous sessions.

290
291 The knowledge base: Devoid of a backend, this is a fully static website implemented using Nuxt
292 and Nuxt Content. Nuxt Content, a Git-based Headless CMS, facilitates the creation and
293 management of static pages via Markdown or JSON files stored in a GitLab repository. This
294 architectural approach alleviates the need for database and backend maintenance, delegating
295 these responsibilities to GitLab, which seamlessly handles authentication, permissions,
296 modification history, and a rich web editor. Therefore, the content is easily editable and
297 accessible. The wiki also provides an interface for searching through large datasets (Refseq and
298 structure predictions) using Meilisearch (self-hosted) with filters and complex queries.

299 The plots are generated using Observable Plot and D3.

300 Both websites undergo automated deployment to a Kubernetes cluster through a GitLab CI
301 workflow. Additionally, a custom linter can be executed against the markdown content via the
302 GitLab web interface.

303

304 *Protein selection for structure prediction*

305 For each system and subsystem, protein accessions of experimentally validated systems were
306 retrieved and used for structure prediction. For systems with no accession available, one of the
307 best systems hits (hit_score and coverage) was randomly selected and used for the protein
308 structure prediction.

309

310 *Structure prediction*

311 AlphaFold2 (v2.3.1) ³³ was used to predict protein structures, using the AlphaFold-Multimer ⁴⁹
312 protocol for complexes. Monomer and complex models were sorted based on pLDDT (mean of
313 predicted Local Distance Difference Test) and iptm+ptm (weighted sum of interface and all
314 residues template-modeling scores), respectively. The models with the highest scores were
315 taken for subsequent analysis. Additionally, for dimer structures, pDockQ scores (predicted
316 DockQ) ⁴³ were calculated to assess interface quality; (models with acceptable quality are
317 considered if they have DockQ ≥ 0.23). For all monomeric structures we used FoldSeek (v5) ³⁴
318 to perform structure similarity searches against the PDB ⁴⁴ and AlphaFold Uniprot ^{45,46}
319 databases.

320

321 *Precomputed DefenseFinder results*

322 For the precomputed database, we used all the RefSeq complete genomes⁵⁰ of both Archaea (N
323 = 22,422) and Bacteria (N = 381) from July 2022. DefenseFinder v1.2.0²¹ using models v1.2.3
324 was run on all genomes separately using standard settings (coverage 0.4).

325

326 *Pfam annotation*

327 We ran hmmsearch HMMER 3.3.2⁵¹ on proteins that were detected by DefenseFinder with a
328 coverage above 75%, to make sure they are complete, against Pfam-A database v33.1⁵². For
329 each protein family of a given defense system, PFAM that hits at least 50% of the members of
330 the family are assigned to the protein family and therefore to the system.

331 Pfam annotation of the example of genomic locus was done on each protein of the system using
332 hmmsearch with --ga_cut argument. If two PFAM hits were superposed only the best hit
333 (hit_score) was kept.

334

335 *New profile modeling*

336 New HMM profiles were built using the same method as in the first version of DefenseFinder (see
337 methods in²¹. For protein with a single representative available, the first profile was made either
338 by blasting one homolog (if only one is available) on BLAST RefSeq (nr)⁵³ database (filter with
339 30% identity and 70% coverage). If a first detection is available in supplementary material, the
340 first profile is done using those proteins.

341

342 *Addition of new models*

343 Using the architecture of MacSyFinder v2⁵⁴, we added new definitions and profiles inside the
344 previous MacSyFinder models. Inconsistencies between models were then checked and reduced:
345 overlapping systems and systems blocking the detection of the other.

346

347 **Funding**

348 F. T., H. G., H. V., E. M., A. B. and J. C. are supported by the CRI Research Fellowship to A.B.
349 from the Bettencourt Schueller Foundation, the ATIP-Avenir program from INSERM
350 (R21042KS/RSE22002KSA), the Emergence program from the University of Paris-Cité
351 (RSFVJ21IDX6_DANA) ERC Starting Grant (PECAN 101040529) and the core funding of
352 Institut Pasteur. H.G. PhD is funded by Generare Bioscience. G. C. A. and A. E. acknowledge
353 support from the Knut and Alice Wallenberg Foundation (project grant 2020-0037) and the
354 Swedish Research Council (Vetenskapsrådet) (grants 2019-01085, 2022-01603 and 2023-
355 02353). The AlphaFold2 computations were enabled by the supercomputing resources Berzelius
356 provided by the National Supercomputer Centre (NSC) at Linköping University and the Knut and
357 Alice Wallenberg foundation. Additional computational resources for structural modeling were
358 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the
359 Swedish National Infrastructure for Computing at NSC, Chalmers University Centre for
360 Computational Science and Engineering (C3SE), and PDC Centre for High Performance
361 Computing, KTH Royal Institute of Technology, partially funded by the Swedish Research Council
362 through grants 2018-05973 and 2022-06725.

363

364 **Data availability**

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

369

370 **Acknowledgement**

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

376

377

378

379 References

380

- 381 1. Georjon, H., and Bernheim, A. (2023). The highly diverse antiphage defence systems of
382 bacteria. *Nat Rev Microbiol* 21, 686–700. 10.1038/s41579-023-00934-x.
- 383 2. Nussenzweig, P.M., and Marraffini, L.A. (2020). Molecular Mechanisms of CRISPR-Cas
384 Immunity in Bacteria. *Annu. Rev. Genet.* 54, 93–120. 10.1146/annurev-genet-022120-
385 112523.
- 386 3. Tock, M.R., and Dryden, D.T. (2005). The biology of restriction and anti-restriction. *Current*
387 *Opinion in Microbiology* 8, 466–472. 10.1016/j.mib.2005.06.003.
- 388 4. Duncan-Lowey, B., Tal, N., Johnson, A.G., Rawson, S., Mayer, M.L., Doron, S., Millman, A.,
389 Melamed, S., Fedorenko, T., Kacen, A., et al. (2023). Cryo-EM structure of the RADAR
390 supramolecular anti-phage defense complex. *Cell* 186, 987-998.e15.
391 10.1016/j.cell.2023.01.012.
- 392 5. Hsueh, B.Y., Severin, G.B., Elg, C.A., Waldron, E.J., Kant, A., Wessel, A.J., Dover, J.A.,
393 Rhoades, C.R., Ridenhour, B.J., Parent, K.N., et al. (2022). Phage defence by deaminase-
394 mediated depletion of deoxynucleotides in bacteria. *Nat Microbiol* 7, 1210–1220.
395 10.1038/s41564-022-01162-4.
- 396 6. Ka, D., Oh, H., Park, E., Kim, J.-H., and Bae, E. (2020). Structural and functional evidence
397 of bacterial antiphage protection by Thoeris defense system via NAD⁺ degradation. *Nat*
398 *Commun* 11, 2816. 10.1038/s41467-020-16703-w.
- 399 7. Rousset, F., Yirmiya, E., Neshet, S., Brandis, A., Mehlman, T., Itkin, M., Malitsky, S.,
400 Millman, A., Melamed, S., and Sorek, R. (2023). A conserved family of immune effectors
401 cleaves cellular ATP upon viral infection. *Cell* 186, 3619-3631.e13.
402 10.1016/j.cell.2023.07.020.
- 403 8. Shen, Z., Lin, Q., Yang, X.-Y., Fosuah, E., and Fu, T.-M. (2023). Assembly-mediated
404 activation of the SIR2-HerA supramolecular complex for anti-phage defense. *Molecular Cell*
405 83, 4586-4599.e5. 10.1016/j.molcel.2023.11.007.
- 406 9. Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E.,
407 Avraham, C., Brandis, A., Mehlman, T., et al. (2022). Bacteria deplete deoxynucleotides to
408 defend against bacteriophage infection. *Nat Microbiol* 7, 1200–1209. 10.1038/s41564-022-
409 01158-0.
- 410 10. Tang, D., Chen, Y., Chen, H., Jia, T., Chen, Q., and Yu, Y. (2023). Multiple enzymatic
411 activities of a Sir2-HerA system cooperate for anti-phage defense. *Molecular Cell* 83, 4600-
412 4613.e6. 10.1016/j.molcel.2023.11.010.
- 413 11. Duncan-Lowey, B., McNamara-Bordewick, N.K., Tal, N., Sorek, R., and Kranzusch, P.J.
414 (2021). Effector-mediated membrane disruption controls cell death in CBASS antiphage
415 defense. *Molecular Cell* 81, 5039-5051.e5. 10.1016/j.molcel.2021.10.020.
- 416 12. Durmaz, E., and Klaenhammer, T.R. (2007). Abortive Phage Resistance Mechanism AbiZ
417 Speeds the Lysis Clock To Cause Premature Lysis of Phage-Infected *Lactococcus lactis*. *J*
418 *Bacteriol* 189, 1417–1425. 10.1128/JB.00904-06.

- 419 13. Johnson, A.G., Wein, T., Mayer, M.L., Duncan-Lowey, B., Yirmiya, E., Oppenheimer-
420 Shaanan, Y., Amitai, G., Sorek, R., and Kranzusch, P.J. (2022). Bacterial gasdermins reveal
421 an ancient mechanism of cell death. *Science* 375, 221–225. 10.1126/science.abj8432.
- 422 14. Zeng, Z., Chen, Y., Pinilla-Redondo, R., Shah, S.A., Zhao, F., Wang, C., Hu, Z., Wu, C.,
423 Zhang, C., Whitaker, R.J., et al. (2022). A short prokaryotic Argonaute activates membrane
424 effector to confer antiviral defense. *Cell Host & Microbe* 30, 930-943.e6.
425 10.1016/j.chom.2022.04.015.
- 426 15. Bernheim, A., Millman, A., Ofir, G., Meitav, G., Avraham, C., Shomar, H., Rosenberg, M.M.,
427 Tal, N., Melamed, S., Amitai, G., et al. (2021). Prokaryotic viperins produce diverse antiviral
428 molecules. *Nature* 589, 120–124. 10.1038/s41586-020-2762-2.
- 429 16. Hussain, F.A., Dubert, J., Elsherbini, J., Murphy, M., VanInsberghe, D., Arevalo, P.,
430 Kauffman, K., Rodino-Janeiro, B.K., Gavin, H., Gomez, A., et al. (2021). Rapid evolutionary
431 turnover of mobile genetic elements drives bacterial resistance to phages. *Science* 374,
432 488–492. 10.1126/science.abb1083.
- 433 17. Piel, D., Bruto, M., Labreuche, Y., Blanquart, F., Goudenège, D., Barcia-Cruz, R.,
434 Chenivresse, S., Le Panse, S., James, A., Dubert, J., et al. (2022). Phage–host coevolution
435 in natural populations. *Nat Microbiol* 7, 1075–1086. 10.1038/s41564-022-01157-1.
- 436 18. Costa, A.R., Berg, D.F. van den, Esser, J.Q., Muralidharan, A., Bossche, H. van den,
437 Bonilla, B.E., Steen, B.A. van der, Haagsma, A.C., Fluit, A.C., Nobrega, F.L., et al. (2023).
438 Accumulation of defense systems in phage resistant strains of *Pseudomonas aeruginosa*.
439 Preprint at bioRxiv, 10.1101/2022.08.12.503731 10.1101/2022.08.12.503731.
- 440 19. Gaborieau, B., Vaysset, H., Tesson, F., Charachon, I., Dib, N., Bernier, J., Dequidt, T.,
441 Georjon, H., Clermont, O., Hersen, P., et al. (2023). Predicting phage-bacteria interactions
442 at the strain level from genomes. Preprint at bioRxiv, 10.1101/2023.11.22.567924
443 10.1101/2023.11.22.567924.
- 444 20. Tesson, F., and Bernheim, A. (2023). Synergy and regulation of antiphage systems: toward
445 the existence of a bacterial immune system? *Current Opinion in Microbiology* 71, 102238.
446 10.1016/j.mib.2022.102238.
- 447 21. Tesson, F., Hervé, A., Mordret, E., Touchon, M., d’Humières, C., Cury, J., and Bernheim, A.
448 (2022). Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat*
449 *Commun* 13, 2561. 10.1038/s41467-022-30269-9.
- 450 22. Chopin, M.-C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci:
451 variations on a theme. *Curr Opin Microbiol* 8, 473–479. 10.1016/j.mib.2005.06.006.
- 452 23. Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013).
453 Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Mol Cell* 50,
454 136–148. 10.1016/j.molcel.2013.02.002.
- 455 24. Hazan, R., and Engelberg-Kulka, H. (2004). *Escherichia coli* mazEF-mediated cell death as
456 a defense mechanism that inhibits the spread of phage P1. *Mol Genet Genomics* 272, 227–
457 234. 10.1007/s00438-004-1048-y.

- 458 25. Dedrick, R.M., Jacobs-Sera, D., Bustamante, C.A.G., Garlena, R.A., Mavrich, T.N., Pope,
459 W.H., Reyes, J.C.C., Russell, D.A., Adair, T., Alvey, R., et al. (2017). Prophage-mediated
460 defence against viral attack and viral counter-defence. *Nat Microbiol* 2, 1–13.
461 10.1038/nmicrobiol.2016.251.
- 462 26. Mageeney, C.M., Mohammed, H.T., Dies, M., Anbari, S., Cudkevich, N., Chen, Y., Buceta,
463 J., and Ware, V.C. (2020). Mycobacterium Phage Butters-Encoded Proteins Contribute to
464 Host Defense against Viral Attack. *mSystems* 5, e00534-20. 10.1128/mSystems.00534-20.
- 465 27. Bobadilla Ugarte, P., Barendse, P., and Swarts, D.C. (2023). Argonaute proteins confer
466 immunity in all domains of life. *Current Opinion in Microbiology* 74, 102313.
467 10.1016/j.mib.2023.102313.
- 468 28. Rousset, F., Depardieu, F., Miele, S., Dowding, J., Laval, A.-L., Lieberman, E., Garry, D.,
469 Rocha, E.P.C., Bernheim, A., and Bikard, D. (2022). Phages and their satellites encode
470 hotspots of antiviral systems. *Cell Host & Microbe* 30, 740-753.e5.
471 10.1016/j.chom.2022.02.018.
- 472 29. Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hör, J., Garb, J., Bechon, N.,
473 Brandis, A., Lopatina, A., et al. (2022). An expanded arsenal of immune systems that
474 protect bacteria from phages. *Cell Host Microbe* 30, 1556-1569.e5.
475 10.1016/j.chom.2022.09.017.
- 476 30. Néron, B., Denise, R., Coluzzi, C., Touchon, M., Rocha, E.P.C., and Abby, S.S. (2023).
477 MacSyFinder v2: Improved modelling and search engine to identify molecular systems in
478 genomes. *Peer Community Journal* 3. 10.24072/pcjournal.250.
- 479 31. Larralde, M. (2022). Pyrodigal: Python bindings and interface to Prodigal, an efficient
480 method for gene prediction in prokaryotes. *Journal of Open Source Software* 7, 4296.
481 10.21105/joss.04296.
- 482 32. The Galaxy Community (2022). The Galaxy platform for accessible, reproducible and
483 collaborative biomedical analyses: 2022 update. *Nucleic Acids Research* 50, W345–W351.
484 10.1093/nar/gkac247.
- 485 33. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
486 Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate
487 protein structure prediction with AlphaFold. *Nature* 596, 583–589. 10.1038/s41586-021-
488 03819-2.
- 489 34. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding,
490 J., and Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. *Nat*
491 *Biotechnol*, 1–4. 10.1038/s41587-023-01773-0.
- 492 35. Koga, M., Otsuka, Y., Lemire, S., and Yonesaki, T. (2011). *Escherichia coli* rnlA and rnlB
493 Compose a Novel Toxin–Antitoxin System. *Genetics* 187, 123–130.
494 10.1534/genetics.110.121798.
- 495 36. LeRoux, M., Srikant, S., Littlehale, M.H., Teodoro, G., Doron, S., Badiee, M., Leung, A.K.L.,
496 Sorek, R., and Laub, M.T. (2021). The DarTG toxin-antitoxin system provides phage

- 497 defense by ADP-ribosylating viral DNA. bioRxiv, 2021.09.27.462013.
498 10.1101/2021.09.27.462013.
- 499 37. Wein, T., Johnson, A.G., Millman, A., Lange, K., Yirmiya, E., Hadary, R., Garb, J.,
500 Steinruecke, F., Hill, A.B., Kranzusch, P.J., et al. (2023). CARD-like domains mediate anti-
501 phage defense in bacterial gasdermin systems. bioRxiv, 2023.05.28.542683.
502 10.1101/2023.05.28.542683.
- 503 38. Zhang, T., Tamman, H., Coppieters 't Wallant, K., Kurata, T., LeRoux, M., Srikant, S.,
504 Brodiazhenko, T., Cepauskas, A., Talavera, A., Martens, C., et al. (2022). Direct activation
505 of a bacterial innate immune system by a viral capsid protein. *Nature* 612, 132–140.
506 10.1038/s41586-022-05444-z.
- 507 39. Antine, S.P., Johnson, A.G., Mooney, S.E., Leavitt, A., Mayer, M.L., Yirmiya, E., Amitai, G.,
508 Sorek, R., and Kranzusch, P.J. (2023). Structural basis of Gabija anti-phage defence and
509 viral immune evasion. *Nature*. 10.1038/s41586-023-06855-2.
- 510 40. Deep, A., Gu, Y., Gao, Y.-Q., Ego, K.M., Herzik, M.A., Zhou, H., and Corbett, K.D. (2022).
511 The SMC-family Wadjet complex protects bacteria from plasmid transformation by
512 recognition and cleavage of closed-circular DNA. *Molecular Cell* 82, 4145-4159.e7.
513 10.1016/j.molcel.2022.09.008.
- 514 41. Li, Y., Shen, Z., Zhang, M., Yang, X.-Y., Cleary, S.P., Xie, J., Marathe, I.A., Kostelic, M.,
515 Greenwald, J., Rish, A.D., et al. (2024). PtuA and PtuB assemble into an inflammasome-like
516 oligomer for anti-phage defense. *Nat Struct Mol Biol*, 1–11. 10.1038/s41594-023-01172-8.
- 517 42. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O.,
518 Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with
519 a language model. *Science* 379, 1123–1130. 10.1126/science.ade2574.
- 520 43. Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein
521 interactions using AlphaFold2. *Nat Commun* 13, 1265. 10.1038/s41467-022-28865-w.
- 522 44. Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, P.A., Crichlow,
523 G.V., Dalenberg, K., Duarte, J.M., et al. (2023). RCSB Protein Data Bank (RCSB.org):
524 delivery of experimentally-determined PDB structures alongside one million computed
525 structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res*
526 51, D488–D508. 10.1093/nar/gkac1077.
- 527 45. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic*
528 *Acids Res* 51, D523–D531. 10.1093/nar/gkac1052.
- 529 46. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D.,
530 Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database:
531 massively expanding the structural coverage of protein-sequence space with high-accuracy
532 models. *Nucleic Acids Res* 50, D439–D444. 10.1093/nar/gkab1061.
- 533 47. Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S.,
534 Burley, S.K., Koča, J., and Rose, A.S. (2021). Mol* Viewer: modern web app for 3D
535 visualization and analysis of large biomolecular structures. *Nucleic Acids Research* 49,
536 W431–W437. 10.1093/nar/gkab314.

- 537 48. Mareuil, F., Doppelt-Azeroual, O., and Ménager, H. (2017). <p>A public Galaxy platform at
538 Pasteur used as an execution engine for web services</p>. F1000Research 6.
539 10.7490/f1000research.1114334.1.
- 540 49. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R.,
541 Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer
542 (Bioinformatics) 10.1101/2021.10.04.463034.
- 543 50. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B.,
544 Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq)
545 database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic
546 Acids Res 44, D733-745. 10.1093/nar/gkv1189.
- 547 51. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLOS Computational Biology 7,
548 e1002195. 10.1371/journal.pcbi.1002195.
- 549 52. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L.,
550 Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein
551 families database in 2021. Nucleic Acids Research 49, D412–D419. 10.1093/nar/gkaa913.
- 552 53. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and
553 Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10, 421.
554 10.1186/1471-2105-10-421.
- 555 54. Néron, B., Denise, R., Coluzzi, C., Touchon, M., Rocha, E.P.C., and Abby, S.S. (2022).
556 MacSyFinder v2: Improved modelling and search engine to identify molecular systems in
557 genomes. Preprint at bioRxiv, 10.1101/2022.09.02.506364 10.1101/2022.09.02.506364.
- 558