Ecological ubiquity and phylogeny drive nestedness in phages–bacteria networks and shape the bacterial defensome

Chloé Feltin [1]*, Sylvain Piry [1], Benoit Moury [1], Lola Chateau [1], Karine Berthier [1], Jonathan M. Jacobs [2,3], Jules Butchacas [2,3], Linda Fenske [4], Lillian Ebeling-Koning [2,3], Theo H. M. Smits [5], Cindy E. Morris [1], Clara Torres-Barceló [6]*

[1] INRAE, Pathologie Végétale, 84140 Montfavet, France

[2] Department of Plant Pathology, The Ohio State University, Columbus, OH 43210, USA

[3] Infectious Disease Institute, The Ohio State University, Columbus, OH 43210, USA

[4] Bioinformatics and Systems Biology, Justus Liebig University Giessen, Giessen, Germany

[5] Environmental Genomics and Systems Biology Research Group, Institute of Natural Resource Sciences (IUNR), Zurich University of Applied Sciences ZHAW, Wädenswil, Switzerland

[6] PHIM Plant Health Institute, University of Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France

* Corresponding authors

E-mails: clara.torresbarcelo@inrae.fr and chloe.feltin@inrae.fr

Abstract

Identifying the ecological and evolutionary factors that shape phage–bacterial interactions is key to understanding their dynamics in microbial communities. Yet, such interactions remain poorly characterised in plant agroecosystems. Here, we investigate the ecological determinants of the interaction between a highly diverse set of 23 phages isolated from diseased apricot trees and 44 bacterial strains from the *Pseudomonas syringae* species complex collected either from diseased apricot trees, healthy plants or non-agricultural environment. Based on their ecological origin, we expected phages to preferentially infect bacterial strains from the same ecological context, forming modular host-range patterns. Contrary to these expectations, we discovered a significantly nested structure, suggesting generalised infection dynamics rather than local adaptation, primarily driven by the broad ecological dynamics of this pathosystem. Analysis of the bacterial genomes showed that both the profiles of anti-phage defence systems and the distribution profiles of prophages are strongly shaped by bacterial phylogeny. Furthermore, while the number of defence systems showed limited correlation with the breadth of bacterial sensitivity to phages, prophage abundance exhibited a strong, non-linear link with phage virulence. Together, these findings provide an ecological and evolutionary perspective on phage–bacterium infection networks and new insights into a better understanding of the role of phages in agricultural ecosystems.

Author Summary

Viruses that infect bacteria, known as phages, are part of microbial communities and influence the abundance, diversity, and traits of their hosts. In an agriculture-related context, they are commonly considered as potential biocontrol agents, but studying the

bases of fundamental phage–bacterial interactions may help us better understand the plant microbiome and its applications. Many factors influence these interactions, and identifying which ones matter most remains a challenge. In our study, we investigated how phages from diseased plants interact with bacteria collected from diseased and healthy plants, as well as from surrounding environments. We expected phages to mainly infect bacteria from similar environments, but instead observed that they often infected bacteria regardless of their source. This suggests that phage activity in this system has few barriers, reflecting the wide ecological distribution of their bacterial hosts. We further investigated how bacteria defend against phages by identifying both defence systems and prophages within their genomes and using this information to explore their contribution to bacterial resistance or sensitivity to phages. Together, our findings offer new insights into how phage–bacterium relationships evolve and function in plant ecosystems.

Introduction

Phages profoundly influence bacterial diversity and population dynamics (1). The impact of phages on bacterial ecology and evolution has been extensively examined in ecosystems like oceans and the human gut (2–7), research in agriculture (8). A deeper understanding of the ecological roles, population dynamics, and interactions of phages with bacterial communities in plant-associated systems, could provide fundamental insights to guide the development of phage-based biocontrol strategies and the sustainable management of microbial communities in agriculture.

Phage interactions with bacteria depend on a multitude of factors. The phage must recognise and infect a host, integrate or replicate, while avoiding or blocking bacterial defence systems (9,10). A range of approaches have been employed to unravel the

67 determinants of phage-bacterial interactions, including ecological and evolutionary analyses
68 of natural populations and experimental evolution (11,12). Mechanistic studies have
69 identified bacterial receptors (13,14), with more recent work focusing on defence systems
70 associated with susceptibility to phage (14–17). For example, *Vibrio lentus* strains with
71 identical phage receptors display varying levels of susceptibility depending on their phage
72 defence elements (18). However, research on prophages and defence systems in
73 phytopathogenic bacteria remains limited and have yet to connect these elements to host
74 range data (19–25).

75 The complexities of phage-bacterial interactions can be visualised as ecological networks,
76 which exhibit distinct structural patterns, commonly modularity or nestedness (26).
77 Nestedness, a structure in which phages exhibit a gradient of host range breadth and
78 bacteria varying levels of susceptibility, is commonly observed with limited genetic diversity
79 of phages or bacteria (26–28). In contrast, modularity, where phages and bacteria interact
80 within modules but rarely across them, tends to emerge at broader evolutionary or
81 ecological scales, such as when interactions involve hosts, or are examined over extended
82 time periods (11,29). Nonetheless, modules can also evolve rapidly and emerge within
83 species (*e.g.* (12)), although such events may be rare. A recent study introduced statistical
84 methods that incorporate quantitative host–parasite interaction data into ecological
85 networks (30). By considering interaction strength rather than simple infection presence or
86 absence, these approaches provide a more nuanced view of infection patterns, host
87 susceptibility, and phage infectivity, capturing dynamics overlooked by binary analyses.

88 The *Pseudomonas syringae* species complex is particularly interesting for studying phage-
89 bacterial interactions in agriculture, due to its high genetic diversity and broad ecological

90 distribution across cultivated plants, wild environments and aquatic habitats (31–33). *P.*
91 *syringae* strains responsible for bacterial canker of apricot trees (*Prunus armeniaca*) are
92 particularly diverse, spanning multiple phylogroups (PG01, PG02, PG03, PG07) (34). It is
93 hypothesized that the pathogenic success of these apricot strains may be driven primarily by
94 general ecological fitness rather than specific virulence factors. These lineages are not
95 apricot-specific but thrive across varied environments, raising questions about the ecological
96 contexts of their phage interactions, whether within the host or external reservoirs (34).
97 Exploring phage infectivity patterns on these strains can reveal how phylogenetic and
98 ecological diversity shapes phage-bacterium dynamics.

99 Historically, phage host range profiling was used to differentiate *P. syringae* strains (35–39).
100 More recent studies have emphasised the isolation and characterisation of novel *P. syringae*
101 phages and their potential as biocontrol agents (40,41). Most phage host range assays were
102 confined to the isolation host and the intended target, reflecting an applied focus. Others
103 included a broader set of strains, though typically limited to agricultural isolates. For
104 example, phages infecting *P. syringae* pv. *porri* exhibited strict specificity to this variant,
105 having been tested mainly against *porri* strains and a few non-*porri* isolates, and *P. syringae*
106 pv. *actinidiae* phages were largely pathovar-specific, with some extending activity to closely
107 related lineages (42).

108 We focused on the causal agents of bacterial canker of apricot trees to investigate the
109 structure and ecological drivers of phage–bacterium infection networks. We hypothesised
110 that phages would be locally adapted to their bacterial hosts, leading to a modular infection
111 network structure that could be explained by the presence of prophages and diverse
112 bacterial defence systems. To test this, we addressed the following questions: i) Are phages

113 isolated from diseased apricot trees specifically adapted to resident *P. syringae* compared to
114 closely-related strains from other environmental reservoirs? ii) Do closely-related bacteria
115 isolated from different environments differ in their adaptation to phages? iii) What bacterial
116 genomic traits underlie the observed infection network structure? We conducted an
117 integrative approach utilizing a quantitative host-range assay involving 23 genetically-diverse
118 *P. syringae* phages and 44 *P. syringae* strains. This was coupled with whole-genome
119 sequencing of all bacterial strains for detailed annotation of prophages and defence systems.
120 This comprehensive methodology provides mechanistic insights into how ecological origin
121 and bacterial genomic content shape phage–bacterial interactions.

122 Results

123 *Host range characterization of a diverse panel of phages on P. syringae bacteria*

124 We characterised the phage–bacterial infection network by testing 23 phages against 44
125 strains from the *P. syringae* species complex (hereafter referred to as *P. syringae* strains) in
126 liquid culture over 48 hours of bacterial growth monitoring. Unlike traditional plaque assays,
127 which primarily assess phage replication on solid media, this quantitative approach
128 measures the extent of bacterial growth suppression in liquid conditions, thereby capturing
129 a complementary aspect of phage–bacterium interactions (9,43). Each interaction was
130 quantified using two complementary metrics: 1/ a continuous variable, corresponding to the
131 percentage of bacterial growth inhibition by phages (0–100%) (S1A Table), and 2/ a binary
132 variable, indicating whether the inhibition was statistically significant (1) or not (0), based on
133 triplicate measurements (S1B Table). Because these values can be interpreted from either
134 the phage or bacterial perspective, we provide clear terminology to distinguish them. From
135 the bacterial perspective, these values represent susceptibility (percentage inhibition) and

136 sensitivity (binary significance of inhibition). Conversely, from the phage perspective, these
137 values reflect virulence (percentage inhibition) and infectivity (binary success of infection).

138 Overall, the tested bacterial strains showed continuous distributions of susceptibility and
139 sensitivity values, reflecting the diversity of interactions of the panel with phages. On
140 average, bacterial susceptibility to phages was moderate, with a 24% reduction in growth,
141 and strains were sensitive to 12.3 out of the 23 tested phages (53.5%) (S1A Fig), indicating a
142 broader sensitivity than suggested by inhibition levels alone. The most susceptible bacterial
143 strains were DG1a.7, GAW0089, and DG12.6 (all from PG02), with average inhibition levels
144 of 46%, 50%, and 42%, respectively (S1A Table). The most broadly sensitive strains, DG1a.7,
145 7C, and CCE0067 (all from PG02), were significantly inhibited by 20, 19, and 19 phages,
146 respectively (87%, 83%, and 83% of tested phages respectively) (S1B Table). In contrast,
147 bacterial strains CC1583, DG8.15, and DGA9CS5-8 (from PG10, PG07 and PG10 respectively)
148 showed the highest resistance, with average inhibition percentages of only 3%, 6%, and 6%,
149 respectively (S1A Table). DG8.15 (from PG07) resisted to 17 phages (74% of tested phages)
150 and seven other strains (19B, 41D, CST0086, DGA9CS5-15, FMU-107, and USA0088 from
151 PG01, PG02, PG02, PG10, PG07 and PG02) were resistant to 16 phages (70%) (S1B Table).

152 From the phage perspective, average bacterial growth inhibition across the strain panel was
153 23%, reflecting moderate virulence. However, each phage was able to infect an average of
154 22.4 out of 42 tested *P. syringae* strains (53.3%) (S1B Fig), indicating a relatively broad
155 infectivity despite moderate inhibition levels. The most virulent phages were Ghual01 (65%
156 inhibition), Arace01 (57%), and Draal03 (44%) (S1A Table). The most infecting phages were
157 Nican01, Arace01, and Ghual01 infecting 39 (93% of tested bacteria), 37 (88%), and 34 (81%)
158 strains, respectively (S1B Table). Moreover, Arace01 and Ghual01 were the only phages

159 capable of infecting strains from other species in the *Pseudomonas* genus. Arace01 infected
160 a *P. rhizosphaerae* strain at a low efficiency of plating (EOP) of 0.005, whereas Ghual01
161 infected a *P. graminis* strain at a moderate efficiency (EOP = 0.017). In contrast, Orimi01,
162 Cygsa01, and Pavpe01 displayed the lowest virulence, with an average bacterial growth
163 inhibition of 0%, 3%, and 4% respectively (S1A Table), while Orimi01, Cygsa01, and Touem01
164 had the lowest infectivity, infecting only 2 (5% of tested bacteria), 5 (12%), and 11 (26%)
165 strains, respectively (S1B Table).

166 The relationship between the breadth of infection and the strength of inhibition can reveals
167 underlying evolutionary dynamics, including trade-offs in both phages and bacteria (44,45).
168 When considering only significant inhibition interactions, we detected a positive correlation
169 between the phage virulence and the phage infectivity (Spearman rho = 0.56; p = 0.01; S9C
170 Fig), indicating that phages with broader host ranges tend to exhibit higher average
171 virulence not supporting the presence of a trade-off. However, this interpretation should be
172 viewed with caution, as some degree of metric dependency may persist. In contrast, no
173 correlation was observed between bacterial susceptibility and bacterial sensitivity
174 (Spearman rho = 0.24; p = 0.12; S9D Fig), reflecting greater variability among bacterial strains
175 in their response to phage infections and suggesting the potential for trade-offs in bacterial
176 defensive strategies.

177 These findings reveal substantial variability in phage virulence and infectivity, as well as
178 bacterial susceptibility and sensitivity, with over than half (53%) of combinations resulting in
179 significant growth inhibition. While some phages exhibit a broad host range, extending even
180 beyond *P. syringae*, others are highly specialised, only infecting a few strains, reflecting a
181 complex interplay between phage infectivity and bacterial resistance.

182 *Nested structure of the phage-bacterial interaction network and association with*
183 *bacterial traits*

184 The structure of the phage-bacterial infection network was analysed using multiple
185 algorithms to assess its nestedness and modularity. Given the diversity of both phages and
186 bacterial hosts, spanning 13 phage genera and four phylogroups of *P. syringae*, and the
187 ecological origins of phages (diseased apricot trees) and bacteria (diseased apricot trees,
188 healthy plants and other environmental reservoirs), we hypothesised a modular structure
189 type. However, the network representing bacterial growth inhibition percentage by phages
190 displayed a strong nested structure, as indicated by scores of 0.48 and 0.44, on a scale
191 varying from 0 to 1, with the *WINE* and *wNODF* algorithms, respectively (Table 1). To
192 determine the statistical significance of the nested pattern, the network was compared to
193 null models, which are simulations differing in the constraints applied during randomization,
194 ranging from unconstrained (S, N) to fully constrained (B). Nestedness was highly significant
195 ($p < 0.01$) against nearly all null models, except null model B, which was shown to be
196 excessively conservative (30).

197 Modularity analyses differed depending on the algorithm (Table 1). With the *spinglass*
198 algorithm, selected for its good performance in terms of false positive rate (30), the
199 modularity score was low (0.01, on a scale varying from -0.5 to 1) and was not significant,
200 whatever the null model. The four alternative modularity algorithms produced slightly
201 higher scores ($0.09 < \text{modularity} < 0.25$) and showed significant modularity ($p < 0.01$) with
202 both null models C1 and R1. These four algorithms show lower performances than *spinglass*
203 in terms of false positive rates, but combined with null models C1 and R1, they show
204 generally acceptable performances in terms of false positive and false negative rates (30).

205 These results suggest globally a slight degree of modular organisation within the network.
206 The strong and significant nestedness and the weak modularity were further supported by
207 the analysis of the binary data representing the significance of bacterial growth inhibition (S2
208 Table).

209

210 **Table 1. Analysis of nestedness and modularity of the bacterial growth inhibition network**
211 **combining 44 strains of the *P. syringae* species complex and 23 phages**.

Analysis	Algorithm	Score (max. = 1.0)	Null model						
			B	N	C1	R1	S	C2	R2
Nestedness	wNODF	0.44	0.44	**0**	**0**	**0**	**0**	**0**	**0**
	WINE	0.48	1	**0**	**0.01**	**0**	**0**	**0**	**0**
Modularity	spinglass	0.01	1	1	1	1	0.07	0.29	0.18
	edge betweenness	0.09	**0**	**0.02**	**0**	**0**	1	1	1
	fast greedy	0.24	**0**	**0**	**0**	**0**	1	0.81	**0.02**
	leading eigenvector	0.14	**0**	**0**	**0**	**0**	1	1	0.94
	beckett	0.25	**0**	**0**	**0**	**0**	1	**0.01**	1

212 Null models are described in Moury *et al.* (30). For each null model and each algorithm, the
213 frequency of simulations, among 100, showing a strictly higher nestedness (or modularity)
214 score than that of the actual matrix is indicated. Significantly nested or modular networks (p
215 ≤ 0.05) are indicated in red.

216

217 To assess whether the nested structure of phage–bacterial networks is shaped by bacterial
218 evolutionary history, we first tested for a phylogenetic signal in bacterial sensitivity summed
219 by bacteria (Fig 1). The analysis revealed a strong phylogenetic signal ($\lambda = 1.0$; LR ($\lambda = 0$) =
10

220 48.26; $p < 0.01$), indicating that closely-related bacterial strains tend to be infected by a
221 similar number of phages. To explore this pattern further, we evaluated the effect of
222 phylogroup identity using both a Generalised Linear Model (GLM) with the bacterial
223 sensitivity summed by bacteria (Fig 1C) and a PErmutational Multivariate ANalysis Of
224 VAriance (PERMANOVA) on bacterial susceptibility profiles (S2A Fig). Both analyses
225 confirmed a significant influence of phylogroup on bacterial susceptibility, explaining almost
226 10% of the variance (GLM: $p < 0.01$; PERMANOVA: $R^2 = 0.09$; $p < 0.01$). In contrast, ecological
227 origin had no significant effect in either test (GLM: $p = 0.87$; PERMANOVA: $R^2 = 0.05$; $p =$
228 0.29). Specifically, PG02 strains emerged as the most sensitive to phages, being infected on
229 average by 14.6 phages and showing the highest susceptibility, with 20.7% growth inhibition.
230 PG01 strains showed intermediate sensitivity being infected by 10.4 phages and an average
231 of 14.6% inhibition, followed closely by PG10 strains with 15.3% inhibition. In contrast, PG07
232 strains were the most resistant overall, infected on average by only 8.4 phages and
233 displaying the lowest susceptibility with 8.73% inhibition.

234

235 **Fig 1. Networks of percentages of growth inhibition of 44 *P. syringae* species complex**
236 **strains (rows) by 23 phages (columns).**

237 (A) Taxonomic organisation of the network presented on a maximum likelihood phylogenetic
238 tree based on bacterial core-genomes and phage genera. A node with a maximum likelihood
239 support value below 0.90 is marked with a red dot. (B) The same network after row and
240 column permutations showing its nested pattern, with phylogroups and ecological origins of
241 *P. syringae* strains displayed on the right side of panels A and B. (C) Bacterial sensitivity to
242 phages by bacterial phylogroup and ecological origins. Letters indicate Tukey post-hoc test

243 results from a generalised linear model (GLM) with a Poisson distribution; ns indicates non-
244 significant differences ($p > 0.05$). (D) Phylogenetic Host Range Index (PHRI) across phage
245 genera.

246

247 The broad taxonomic diversity of the phages studied here, covering 13 genera, includes both
248 specialists and generalists. To characterise phage host range, we calculated the Phylogenetic
249 Host Range Index (PHRI), which incorporates the number, the evenness, and the
250 phylogenetic diversity of host bacteria (46). The PHRI values ranged from 0 (Orimi01, a
251 specialist phage) to 0.38 (Arace01, a generalist phage), forming a gradient that reflects the
252 nested structure of the phage-bacteria interaction network (Fig 1D). Based on PHRI
253 classification, 9% of phages were categorised as specialists, 65% as generalists, and 26% as
254 having an intermediate host range (Fig 1D). At the genetic level, significant differences in
255 virulence were observed between phage genera (GLMM; $p < 0.01$). The generalist genus
256 *Aravirus* exhibited the highest virulence (52.8%), while the specialist genera *Cygnusvirus* and
257 *Orionvirus* displayed markedly lower virulence levels, averaging 1.1% and 4.2%, respectively
258 (S2B Fig). Additionally, phages with similar host infection profiles clustered by taxonomy,
259 with those sharing >70% nucleotide identity (same genus) explaining 63.5% of the variation
260 in virulence (PERMANOVA: $R^2 = 0.64$, $p < 0.01$), indicating genus-level host specificity (Fig
261 1A). For example, the two phages from the *Lyravirus sulafat* species and the *Uliginvirus*
262 genus exhibited highly similar infection host profiles.

263 These results demonstrate that the phage–bacterial interaction networks are predominantly
264 nested within the investigated panel of strains, with minimal modularity suggesting broad
265 host range overlap among phages rather than local adaptation based on ecological origin.

266 Bacterial susceptibility is phylogenetically constrained, with PG02 strains being highly
267 sensitive and PG07 strains resistant. Additionally, although more than half of the phages
268 were generalists, genus-level similarity drives host-range patterns.

269 *Defence systems: profile and association with ecology and phylogeny*

270 The 44 strains of *P. syringae* analysed in this study were examined for their accessory
271 genome content, with a particular focus on their repertoire of defence systems to
272 counteract mobile genetic elements such as phages. Across these strains, a total of 48
273 defence system types and 86 subtypes were identified (Fig 2A). The most abundant defence
274 subtypes included Septu (n = 81), type IV Restriction-Modification (RM) (n = 37), and Gabija
275 (n = 31) (Fig 2A). In contrast, 31 subtypes were sporadically detected across the bacterial
276 genomes, including rare defence systems such as three Abi subtypes (AbiH, AbiJ, and AbiO),
277 two Asv subtypes (AsvI and AsvII), and three DTR subtypes (DTR6, DTR8, and DTR3) (Fig 2A).
278 Defence systems are present at an average rate of 9.93 per strain, with two strains (AF0015
279 and P3.01.09C09, both from PG01) harbouring the highest number (15 systems), and four
280 strains (DG1A_7 from PG02, CSZ0293, SZ0119, and SZB0008 from PG10) the lowest number
281 (5 systems) (Fig 2A). Each defence system can be classified into different functional
282 categories to better characterise its role. Specifically, 38.4% of defence system subtypes are
283 associated with abortive infection, 24.4% with nucleic acid degradation, and 37.2% remain
284 unclassified (Fig 2A).

285

286 **Fig 2. Defence system distribution in bacterial genomes and relationship with bacterial**
287 **sensitivity to phages.**

288 (A) Matrix showing the number of defence systems present in the 44 *P. syringae* strains,
289 organised on a maximum likelihood phylogenetic tree based on their core-genomes. A node
290 with a maximum likelihood support value below 0.90 is marked with a red dot. The
291 phylogroup affiliation of each bacterial strain is indicated on the left using colours, and
292 defence systems are represented by subtypes with colours on the right indicating their
293 respective functional category. (B) Number of defence systems per bacterial strain according
294 to their phylogroups and ecological origins. ns: not significant ($p > 0.05$), according to post-
295 hoc Tukey tests from a Generalised Linear Model with a Poisson distribution. (C) Bacterial
296 sensitivity to phages as a function of the number of defence systems, with linear regressions
297 performed within each phylogroup.

298

299 To assess whether the distribution of defence systems across bacterial genomes was
300 influenced by phylogeny, we tested for a phylogenetic signal in the total number of encoded
301 systems per genome. A strong signal indicated that closely related strains tend to harbour
302 similar numbers of defence systems ($\lambda = 1.0$; LR test ($\lambda = 0$): 44.18; $p < 0.01$) (S3 Fig), as
303 described in a previous study (24). Phylogroup significantly explained variation in defence
304 system profiles (PERMANOVA, $R^2 = 0.14$, $p < 0.01$), with a weaker effect from ecological
305 origin ($R^2 = 0.06$, $p < 0.01$), while growth rate of the bacteria had no significant impact ($R^2 =$
306 0.02, $p = 0.51$) (Fig 2A). Post-hoc comparisons based on this analysis revealed that strains
307 from PG10 displayed significantly distinct defence system profiles compared to those from
308 PG02 and PG07, while strains isolated from healthy plants exhibited distinct defence system
309 profiles compared to other ecological origins. However, none of these factors significantly
310 affected the total number of defence systems, although the effect of phylogroup

311 approached significance (GLM: $p = 0.06$) (Fig 2B). This phylogroup trend is supported by a
312 larger dataset of 590 *P. syringae* genome assemblies, where PG01 strains encode the highest
313 number of defence systems, while PG07 and PG10 harbour some of the lowest counts (24).

314 While closely-related strains tend to possess similar numbers of defence systems, we further
315 investigated whether they also share the same subtypes of defence systems. To this end, we
316 calculated the percentage of shared defence system subtypes between each pair of bacterial
317 strains, where 100% indicates complete identity in the sets of subtypes. A weak but
318 significant positive correlation was observed between the percentage of shared defence
319 system subtypes and the genome-wide Average Nucleotide Identity (ANI) among bacterial
320 strains (GLMM; estimate = 0.015; $p < 0.01$) (S4 Fig). This relationship suggests that some
321 defence systems are partly inherited through vertical transmission, reflecting the influence
322 of phylogeny on their distribution. Furthermore, only 1.14% of the identified defence
323 systems were linked to prophages within bacterial genomes, indicating that the vast majority
324 are located outside prophage regions. This low prevalence indicates that while defence
325 systems may spread horizontally among bacterial populations, their dissemination is likely
326 mediated by mobile genetic elements other than prophages.

327 Defence systems may impose varying energetic costs on bacterial cells, potentially
328 influencing their allocation of resources (47,48). In strains of *P. syringae*, defence systems
329 related to nucleic acid degradation comprised, on average, 54% of the total defence
330 systems, while abortive infection systems, which are considered more metabolically costly
331 (49,50), constituted an average of 25% (S3 Fig). We then revealed that the proportion of
332 defence system categories were significantly shaped by phylogroup, explaining 16% of the
333 variance (PERMANOVA; $R^2 = 0.16$; $p = 0.03$), and growth rate ($R^2 = 0.07$; $p = 0.04$), but not

334 ecological origin (R^2 = 0.08; p = 0.11). Pairwise comparisons based on this analysis showed
335 that strains from PG07 had significantly different defence category profiles compared to
336 those from PG01 and PG02.

337 Our results revealed that the defence arsenal profiles in *P. syringae* are shaped by phylogeny
338 and slightly by ecological origin, consistent with partial vertical inheritance of defence
339 system profiles. Overall, these findings underscore the complexity of bacterial defence
340 strategies and indicate that multiple evolutionary and ecological factors influence phage
341 resistance.

342 *Prophages: profile and association with ecology and phylogeny*

343 We identified 88 prophages among the 44 genomes of *P. syringae*. On average, each strain
344 carried two prophages, with one strain (FMU-107) harbouring five prophages and 18 strains
345 (41%) containing a single cryptic prophage (Fig 3A). Among these prophages, 36% were
346 classified as high quality (complete or nearly complete according to CheckV software), 39%
347 as medium quality, and 25% as low quality. Prophage sizes ranged from 12.0 kb to 89.8 kb,
348 with an average size of 32.2 kb. The size distribution exhibited three peaks at approximately
349 18 kb, 44 kb, and 57 kb. To classify these prophages taxonomically, we used VIRIDIC, which
350 identified 65 novel phage species and 42 novel genera (S3 Table), while proteomic tree
351 analysis with Viptree revealed six distinct prophage clusters, corresponding approximately to
352 family-level taxonomic divisions (Fig 3A) (51). One of these, the CL4 family-like prophage,
353 was found in all 44 *P. syringae* strains, which span four distinct phylogroups (Fig 3A). CL4
354 family-like prophage is a tailocin, specifically an R-type syringacin, a bacteriocin derived from
355 a bacteriophage (52,53). These tailocins are widely distributed across the *P. syringae* species
356 complex and are thought to play a key role in inter-bacterial competition by enabling strains

357 to kill closely-related competitors (54,55). Moreover, we found that the taxonomy of the CL4
358 family-like prophage mirrors the phylogeny of their bacterial hosts across 68 strains from 10
359 phylogroups (including the 44 strains analysed in this study), with a significant
360 cophylogenetic congruence signal (Procrustes Approach to Cophylogeny (PACo): residual
361 sum of squares = 0.15; $p < 0.01$) (S5 Fig).

362

363 **Fig 3. Diversity and distribution of prophages in bacterial genomes and relationship with**
364 **bacterial sensitivity to phages.**

365 (A) Distribution of prophages in the genomes of the 44 *P. syringae* strains, categorised by
366 genus and family-like cluster (VipTree). Bacterial strains are organised based on a maximum
367 likelihood phylogenetic tree based on their core-genomes, with their phylogroups and
368 ecological origins indicated on the right. A node with a maximum likelihood support value
369 below 0.90 is marked with a red dot. (B) Abundance of prophages per bacterial strain
370 according to their phylogroups and ecological origins. ns: not significant ($p > 0.05$), according
371 to post-hoc Tukey tests from a Generalised Linear Model with a Poisson distribution. (C)
372 Bacterial sensitivity to phages as a function of the number of prophages, coloured by
373 phylogroups.

374

375 The abundance of prophages (*i.e.* the number of prophages per bacterial genome) and the
376 prophage distribution profiles in bacterial genomes are strongly shaped by phylogeny. A
377 significant phylogenetic signal was detected in prophage abundance ($\lambda = 1.0$; LR ($\lambda = 0$) =
378 46.48; $p < 0.01$), indicating that closely related strains tend to harbour similar numbers of
379 prophages (Fig 3B). Phylogroup significantly explained 16% of variation in prophage

380 distribution profiles (PERMANOVA, $R^2 = 0.16$; $p < 0.01$), with a weaker effect from ecological
381 origin ($R^2 = 0.07$; $p < 0.01$), while growth rate had no significant impact ($R^2 = 0.03$; $p = 0.09$)
382 (Fig 3A). All distinct phylogroups, as well as ecological origins, exhibited significantly different
383 prophage distribution profiles, supporting the idea that lineage and ecological context
384 influence the genomic landscape of prophages. However, none of these factors significantly
385 affected the abundance of prophages (GLM: phylogroup, $p = 0.27$; ecological origin, $p = 0.92$;
386 growth rate, $p = 0.30$), despite some observed variation. Indeed, PG07 strains showed the
387 highest prophage abundance, nearly double that of PG10 strains (Fig 3B). However, this
388 pattern contrasts with findings from broader dataset (590 *P. syringae* genome assemblies),
389 in which PG02 strains encoded among the highest numbers of prophages, PG07 among the
390 lowest, and PG01 and PG10 showed intermediate levels (24). In contrast, differences across
391 ecological origins were minimal (Fig 3B).

392 In conclusion, the comprehensive analysis of prophages within the *P. syringae* species
393 complex revealed a diverse array of complete and incomplete prophages integrated into the
394 genomes. The prophage distribution profiles are strongly explained by bacterial phylogeny
395 and ecological origins of the host strains.

396 *Contribution of bacterial defence systems and prophages to bacterial sensitivity to*
397 *phages*

398 Since bacterial defence systems target phages (50), a negative correlation is expected
399 between their number and bacterial sensitivity to phages. To test this, we examined the
400 correlation between bacterial sensitivity and the number of encoded defence systems, while
401 accounting for phylogenetic background. When considering all bacterial strains together, no
402 overall correlation was identified ($p = 0.36$) with a phylogroup effect ($p < 0.01$) and no effect

403 of ecological origins ($p = 0.88$). For both PG01 and PG02, no significant correlation was
404 detected (linear regression for PG01: adjusted $R^2 = 0.12$; $p = 0.18$; for PG02: adjusted $R^2 = -$
405 0.07; $p = 0.94$) (Fig 2C). In contrast, a negative correlation was detected in PG10, supporting
406 our expectation (linear regression for PG10: adjusted $R^2 = 0.34$; $p = 0.05$) (Fig 2C). In contrast,
407 PG07 exhibited an unexpected positive correlation, suggesting that strains with a higher
408 number of defence systems were more sensitive to phages (linear regression for PG07:
409 adjusted $R^2 = 0.92$; $p < 0.01$) (Fig 2C).

410 Although prophages can contribute to bacterial immunity against phage infection (56), we
411 found no significant association between prophage abundance and bacterial sensitivity to
412 phages ($p = 0.82$), and neither phylogroup ($p = 0.31$) nor ecological origin ($p = 0.80$) had a
413 detectable effect (Fig. 3C). Moreover, no correlation was observed between prophage
414 abundance and the number of defence systems encoded in bacterial genomes, suggesting
415 that these systems do not provide protection against prophage integration or maintenance
416 (linear regression; adjusted $R^2 = -0.02$; $p = 0.64$).

417 *Influence of multiple bacterial and phage traits on bacterial growth inhibition*

418 Having identified several bacterial and phage factors that individually influenced the
419 bacterial growth inhibition, we next sought to integrate these explanatory variables into a
420 unified framework to gain a more comprehensive understanding of phage–bacteria
421 interaction dynamics. The primary model included prophage abundance and the total
422 number of defence systems per strain as bacterial traits associated with phage resistance,
423 alongside phage- and bacterium-related traits examined in this study. We also tested an
424 alternative model incorporating PCA-derived variables related to bacterial defensome, but it
425 did not improve model performance (Akaike Information Criterion (AIC) = 16.70) compared

426 to the initial model (AIC = 16.69). Therefore, we retained the first model for its greater
427 interpretability.

428 Consistent with the above findings, bacterial phylogeny and phage taxonomy emerged as
429 the strongest explanatory variables for inhibition percentage, with an importance weight of
430 1.0 across the 256 models (Table 2). In addition, a variable indicating whether the phage had
431 been isolated using one of the tested bacterial strains was significantly associated with
432 inhibition level. Specifically, phages isolated from strains within the tested host range
433 exhibited a slightly lower inhibition capacity (mean = 15.1%) compared to those isolated
434 from external strains (mean = 17.7%).

435

436 **Table 2. Relative importance of explanatory variables influencing bacterial growth**
437 **inhibition by phages.**

Explanatory variable	*Sum of AIC weights*	*N Models*
Phage genera	**1.00**	54
Bacterial phylogroups	**1.00**	53
Number of prophages in bacterial genomes	**0.72**	29
Phages isolated with bacteria from the data set	**0.67**	31
Number of defence systems in bacterial genomes	0.38	26
Bacteria from the data set having isolated phages	0.37	24
Bacterial growth rate	0.29	23

Sum of AIC weights was computed from model averaging across the top-ranked models whose cumulative weights reached ≥ 0.95, selected from a full set of 256 candidate models generated by multi-model inference. The relative importance of each variable corresponds to the sum of AIC weights across all retained models in which it appears. Bold values indicate variables with high relative importance (sum of weights ≥ 0.60), selected as significant predictors in the analysis. N Models: Number of retained models in which each variable appears. Variables related to phages are marked with the symbol ⍰; bacterial variables are unmarked.

The model identified the total number of prophages in bacterial genomes as a significant predictor of inhibition of bacterial growth (sum of weights = 0.72; Table 2). Inhibition levels varied nonlinearly with prophage abundance: inhibition levels were 15.6%, 14.7%, 17.5%, 28.4%, and 10.9% for strains carrying 1 to 5 prophages, respectively. While prophage abundance did not predict bacterial sensitivity, it appeared to influence bacterial susceptibility, suggesting that prophage-rich bacteria may be more susceptible to highly-effective phages. In contrast, the total number of defence systems did not significantly explain bacterial growth inhibition (sum of weights = 0.38; Table 2), despite its correlation with the bacterial sensitivity for some phylogroups.

Overall, these findings highlight that the primary determinants of the bacterial growth inhibition by phage infection within the *P. syringae* species complex are bacterial phylogeny and phage taxonomy rather than bacterial ecological origins. The integration of multiple explanatory variables within a unified modelling framework further revealed that prophage

460 abundance, but not the total number of defence systems, significantly contributes to
461 variations in inhibition levels. This suggests that while prophages may not act as strict
462 barriers to infection, their presence can modulate the efficacy of phage infection. Together,
463 these results emphasize the complex and multifactorial nature of bacterial susceptibility to
464 phages, shaped by both evolutionary history and prophage-mediated genomic landscapes.

465

466 Discussion

467 Our study uncovered a strong nested structure in the phage–bacterial infection network
468 involving genetically-diverse *P. syringae* strains and phages. This discovery is in strong
469 contrast with the modular structure frequently reported (11,29). This pronounced
470 nestedness likely reflects the ubiquitous nature of the highly diverse *P. syringae* complex,
471 mirrored by a phage community ranging from specialists to generalists in host range. While
472 bacterial phylogeny had a marked influence on the nestedness of the infection network, we
473 found no evidence of local phages-bacteria co-adaptation. Defence systems and prophages
474 contributed to phage resistance, but their distribution was largely shaped by bacterial
475 phylogeny and ecological origin, rather than phage pressure alone. Together, these findings
476 highlight how evolutionary and ecological forces shape phage–bacterial infection networks
477 in natural environments.

478 *Ecological origin and phylogenetic influences on bacterial sensitivity to phages*

479 The observed nested infection network is unexpected when we consider the high genetic
480 diversity of the phages (13 genera, 17 species) and bacteria (4 phylogroups corresponding to
481 species levels and 7 clades), as well as the varied ecological origins of the bacterial strains

482 included in this study. Such a diversity of organisms could be expected to reveal local
483 adaptation, where phages perform better on sympatric than allopatric bacterial hosts, in line
484 with what has been demonstrated previously (57,11). However, our findings suggest that
485 phage–host interactions in this pathosystem transcend ecological boundaries. Phages
486 isolated from apricot tree soil infected bacterial strains from both diseased orchards and
487 aquatic environments with comparable efficiency, and bacterial origin did not broadly
488 influence bacterial sensitivity/susceptibility to phages. This absence of modularity, combined
489 with a lack of ecological signal in host range, points to broader ecological connectivity and
490 population flow within the *P. syringae* species complex. These results align with previous
491 work (58) showing that *P. syringae* population structures are shaped more by fluctuating
492 ecological dynamics than by geographic isolation, likely due to documented exchanges
493 between agricultural and aquatic environments (59). Overall, the nested structure observed
494 here implies that interaction dynamics in this system are maintained by widespread
495 generalism and ongoing ecological mixing, rather than niche-specific adaptation.

496 Bacterial phylogeny plays a central role in shaping both the composition of defence systems,
497 the distribution of prophages, and the sensitivity of *P. syringae* strains to phages. In our
498 study, PG02 strains, known for their ubiquitous presence across habitats and for containing
499 strains capable of infecting a broad range of plant species (60,33), were also the most
500 sensitive to phages, carried a moderate number of defence systems, and encoded one of the
501 highest prophage loads (24). In contrast, PG07 strains, which are predominantly associated
502 with diverse environmental reservoirs (60), were the most resistant to phages despite
503 encoding relatively few defence systems and prophages. PG01 strains, mainly isolated from
504 diseased plants (60), showed intermediate bacterial sensitivity to phages, possessed the
505 greatest number of defence systems, and encoded a moderate prophage load (24). Finally,

506 PG10 strains, exclusively environmental, showed moderate bacterial sensitivity to phages,
507 low numbers of defence systems, and a moderate prophage count (24). These findings
508 suggest a potential trade-off between ecological ubiquity and bacterial sensitivity to phages,
509 a hypothesis that requires direct experimental investigation. In a similar vein, a recent study
510 reported that effector gene counts in *P. syringae*, which are key to pathogenicity, positively
511 correlate with phage defences and show partial conservation within phylogroups (24).

512 Phylogroups show consistent patterns in defence system profiles, with closely-related strains
513 sharing portions of their defence arsenals, while ecological origins accounted for only minor
514 similarities. Strains within the same phylogroup also tend to retain consistent features in
515 their defensome, including the number and functional categories of defence systems.
516 Collectively, these findings suggest that each phylogroup may adopt a common anti-phage
517 strategy, implemented through distinct combinations of defence systems, mainly acquired
518 by horizontal gene transfers (50,61). For strains belonging to phylogroup PG10, we observed
519 a negative correlation between the number of defence systems and sensitivity to phages,
520 suggesting a protective effect from accumulating diverse systems in their genome (17).
521 Conversely, strains in PG07, although generally the most resistant to phages, showed a
522 positive correlation between bacterial sensitivity to phages and defence systems. This
523 counter-intuitive pattern could reflect historical predation by phages, where strains
524 subjected to intense phage attack would have accumulated defence systems (62).
525 Alternatively, the apparent lack of protection may arise because these systems target phages
526 or mobile genetic elements (MGEs) other than those tested here. However, it is important to
527 interpret these trends cautiously, as they are based on relatively small sample sizes (n=6 for
528 PG07; n=10 for PG10), and our understanding remains partial given the continual discovery
529 of novel defence systems (63). Although the sensitivity of specific phylogroups to phages

530 may be correlated to their defence systems, the overall contribution of these systems to
531 phage resistance in *P. syringae* strains appears limited and may depend on the activity of
532 particular defence mechanisms (64).

533 *Limitations and alternative approaches for understanding phage-bacterium*
534 *interactions*

535 Liquid culture assays offer the advantage of measuring a proxy for phage virulence by
536 continuously monitoring bacterial growth kinetics in a high-throughput format. This
537 automated approach enhances reproducibility, allows the simultaneous processing of a large
538 number of samples, and enables the use of diverse metrics to quantify phage-induced
539 inhibition of bacterial growth (65), including host-range, and potential indicators of bacterial
540 resistance to phages (9,43). In our dataset, certain phage–bacterium combinations exhibited
541 signs of evolutionary rescue (when a bacterial population initially suppressed by phages
542 recovers through the growth of resistant mutants) (66) (S11 Fig). Future studies could build
543 on these findings by examining their genetic co-evolution through temporal sampling and
544 genomic analyses. Nevertheless, we acknowledge that the phenomenon of lysis from
545 without (referring to bacterial lysis without phage replication, typically resulting from high
546 phage adsorption rates or abortive infections) cannot be entirely ruled out in our dataset
547 (67). Follow-up experiments measuring phage titres, bacterial population dynamics, or
548 adsorption efficiency would be valuable to confirm the underlying mechanisms of bacterial
549 suppression. Such analyses would also refine our understanding of the role of defence
550 systems during phage infection (14,17). In particular, quantifying phage adsorption rates
551 could help discriminate between pre- and post-adsorption resistance: successful adsorption
552 without subsequent lysis would suggest the involvement of intracellular defence

553 mechanisms, whereas failed adsorption would point to surface receptor incompatibility as
554 the main barrier to infection.

555 Analysis of bacterial membrane receptors recognised by phages is therefore essential in
556 interpreting host range patterns (68,12). Such factors have enabled accurate prediction of
557 phage-bacteria interactions at the strain level in *E. coli* (14). However, receptors for phages
558 infecting *P. syringae* remain poorly characterised compared to other bacterial models, with
559 only three distinct receptors identified in the *P. syringae* phage receptor database compared
560 to over 50 in *E. coli* (69). Increasing research on phage–phytopathogen interactions will help
561 to close this knowledge gap and improve our understanding of the molecular basis
562 underlying the observed interaction patterns.

563 A complementary strategy to uncover bacterial genetic determinants of sensitivity to phages
564 and resistance phenotypes involves genotype–phenotype association analyses, including
565 approaches based on gene presence–absence matrices, phylogenetic trait correlation
566 methods and genome-wide association studies (GWAS). This latter method has recently
567 been applied to 80 *P. aeruginosa* strains, identifying host genes that influence susceptibility
568 to two phages, particularly those related to flagella and LPS/O-antigens associated with
569 lipopolysaccharide biosynthesis (70). While our dataset of 44 *P. syringae* genomes may
570 provide a sufficient basis for GWAS, challenges remain. The high genomic diversity of the *P.*
571 *syringae* species complex and the large number of phages tested (n = 23), each potentially
572 interacting with different receptors or resistance pathways, add considerable complexity.
573 Future studies may address this by focusing on specific phage-host combinations or by
574 applying comparative genomics approaches that can help disentangle the genetic basis of
575 resistance across diverse strains.

576 *Applied perspectives and future directions*

577 From an applied perspective, we identified several generalist phages capable of infecting a
578 wide range of *P. syringae* species complex strains. Understanding the molecular basis of
579 their generalism could provide important phage evolutionary clues, *e.g.,* whether generalism
580 involves flexible receptor-binding proteins, the ability to evade diverse bacterial defence
581 systems, or other mechanisms. Many *P. syringae* strains are non-pathogenic members of the
582 plant microbiome, highlighting the need for targeted phage cocktails that selectively control
583 pathogenic strains while preserving beneficial microbes. Such specificity offers a more
584 sustainable and ecologically-balanced approach to biocontrol compared to using broad-
585 spectrum generalist phages (71). This exemplifies how fundamental ecological
586 understanding of phage–bacterium infection networks can directly inform and enhance the
587 development of effective biocontrol strategies.

588 To further test our hypothesis that ecological connectivity shapes the nested structure of
589 phage–*P. syringae* interactions, future studies should isolate phages from diverse non-
590 agricultural environmental, such as water and healthy plants to include cases of true
591 sympatry. These newly isolated phages should then be tested against the same bacterial
592 strains used in our study, or against new strains from corresponding habitats. If similar
593 nested network structures emerge, it would strengthen the idea that phage and *P. syringae*
594 populations continuously migrate between agricultural and non-agricultural environments,
595 forming an ecologically-connected meta-community. Moving forward, several key research
596 priorities should guide future investigations: verification of the role of specific defence
597 systems through genetic manipulation like gene deletion and heterologous expression;
598 systematic investigation of receptor usage by our studied phages to reveal molecular

599 determinants of host range; and strategic expansion of the bacterial strain collection to
600 improve phylogenetic and ecological coverage, with particular focus on understudied
601 phylogroups. These complementary approaches would collectively strengthen our
602 understanding of the molecular and ecological factors driving the phage-bacteria
603 interactions observed in this system. Extending such analyses to other ecologically and
604 phylogenetically-distinct plant-associated bacteria, such as *Xanthomonas*, *Ralstonia*, or
605 *Agrobacterium*, could reveal whether the patterns observed in *P. syringae*-phages infection
606 network are generalisable or reflect unique evolutionary dynamics of this highly diverse
607 species complex.

608

609 **Materials and Methods**

610 *Phages panel*

611 The 23 phages infecting strains of the *P. syringae* species complex used in this study were
612 isolated from soil samples collected in apricot orchards affected by bacterial canker and fully
613 described in a previous study (72). Briefly, this phage collection exhibits high genetic
614 diversity, encompassing 13 genera and 17 species, representing the three most common
615 morphological types (myovirus, podovirus, and siphovirus) and comprising 20 virulent and 3
616 temperate phages. The collection included two characterised DNA packaging mechanisms
617 (direct terminal repeat (DTR) and headful) alongside two newly identified but unverified
618 mechanisms, while the remaining packaging strategies remain unknown. The genomic GC
619 content of the phages ranged from 46% to 58%. All phage stocks were standardised to a

620 concentration of 10^7 PFU/mL and stored in SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50
621 mM Tris-HCl, pH 7.4) at 4°C.

622 *Bacterial strains panel*

623 The 44 *P. syringae* species complex strains were obtained from the Plant Pathology Unit
624 collection at INRAE (Montfavet, France) and were not isolated as part of this project (S4
625 Table). The strains belonged to seven clades within four phylogroups (PG01, PG02, PG07,
626 PG10) corresponding to the species levels, according to Berge *et al.* (60). The strains were
627 sampled from three different environments: diseased apricot trees (exhibiting bacterial
628 canker symptoms) (34), healthy plant tissues, and non-agricultural environments (S4 Table).
629 Bacterial strains were revived from -80°C storage and cultured on King's B agar (73). Cultures
630 were maintained at 4°C for a maximum of one week.

631 *Sequencing and analysis of whole genomes of bacterial strains*

632 The bacterial cultures were grown on King's B (KB) agar medium for 48 hours at 25°C. A
633 bacterial suspension was then prepared to achieve an approximate concentration of 2×10^9
634 CFU in 4 mL of demineralised water. The suspension was subsequently centrifuged at 3,215
635 *g* for 10 minutes at 20°C to pellet the bacterial cells, after which the supernatant was
636 carefully discarded. Genomic DNA was extracted using the Qiagen Genomic-tips kit (Qiagen,
637 Valencia, CA, USA) following the manufacturer's protocol.

638 Of the 44 *P. syringae* strains, 39 were sequenced using 20 µL of genomic DNA, with
639 sequencing performed by Plasmidsaurus Inc (San Francisco, USA) using Oxford Nanopore
640 Technology (ONT) and a custom analysis and annotation pipeline
641 (https://plasmidsaurus.com/). The long-read sequencing library was prepared without
642 amplification using v14 library preparation chemistry, and sequencing was conducted with

643 R10.4.1 flow cells. The genomes of five additional strains were obtained from the RefSeq
644 library at NCBI: CC1559 (GCF_000452685.1), USA0007 (GCF_000452545.1), 41A
645 (GCF_000935775.1), MAFF302273^{PT} (also known as JD08) (GCF_000145925.1), and CC1583
646 (GCF_000452665.1).

647 The core genome analysis was performed using a private database within the EDGAR 3.5
648 platform (74). The core genome was defined as the set of orthologous genes present in all
649 332 genomes of the *P. syringae* species complex, including the 44 strains analysed in this
650 study. A total of 1,346 core genes were identified per genome, corresponding to 446,872
651 aligned genes across all strains. Each set of orthologous genes was individually aligned at the
652 protein level using MUSCLE (74). The resulting alignment were then concatenated into a
653 single alignment, yielding approximatly 512,223 amino acid residues per genome and
654 170,058,036 in total across all genomes. Phylogenetic relationships were inferred using the
655 FastTree algorithm within EDGAR, which constructs maximum-likelihood phylogenetic trees
656 based on the core genome alignment (75). All phylogenetic trees of bacterial strains shown
657 in the figures were rooted at the midpoint of the longest path and the nodes with a
658 maximum likelihood support value below 0.90 were marked with red dots.

659 *Identification of defence systems and prophages in bacterial genomes*

660 The defence systems of the 44 *P. syringae* strains were identified using *DefenseFinder*
661 version 1.2.2 (76) based on assembled nucleotide sequences from the bacterial genomes.
662 Each defence system type was categorised into one of three functional groups: nucleic acid
663 degradation, abortive infection (Abi), or undetermined (ND).

664 The prophages within the 44 *P. syringae* strains were identified using *geNomad* version 1.5.0
665 (77) based on assembled nucleotide sequences of bacterial genomes. As prophage

666 prediction tools are still imperfect and may miss or misclassify certain elements, these
667 results should be interpreted with caution. Among the 97 identified prophage sequences,
668 only those classified as viral by *geNomad* and *CheckV* (78) were retained. Specifically,
669 sequences labelled as "No Terminal Repeats" in the topology column and those with a
670 "viral_genes" value of 0 were excluded. This filtering resulted in a final dataset of 88 verified
671 viral prophages, accounting for 91% of the total identified prophages across the 44 strains.
672 Additionally, a genetic analysis of the prophages was performed using *VIRIDIC* (79) to classify
673 them into genera and species (S3 Table). *VipTree* was used to generate prophage clusters
674 based on a proteomic tree analysis using distance-based methods (BIONJ from tBLASTx
675 scores) (80). *geNomad* and *VipTree* were then employed to identify prophages belonging to
676 the CL4 cluster and to construct a proteomic tree encompassing 68 CL4 prophages (24
677 reference strains + 44 analysed strains) (81).

678 To assess the congruence between the taxonomy of CL4 prophages and the phylogeny of
679 their bacterial hosts, we performed a PACo (Procrustes Approach to Cophylogeny) analysis.
680 Phylogenetic distance matrices were generated from the host and prophage trees using the
681 *cophenetic* function (82). These matrices were then centred and corrected for negative
682 eigenvalues using the "Cailliez" correction method. Congruence was tested using the *PACo*
683 function from the *paco* R package with 1,000 permutations to assess statistical significance.
684 The co-phylogeny visualisation was performed with *phytools* R package and *cophylo* function
685 (83).

686 *Phage host range analysis*

687 *a. Experimental setup*

688 To analyse the host range of phages infecting *P. syringae*, bacterial growth kinetics in the
689 presence of phages were measured using 96-well plates (65). This liquid culture approach
690 has been compared to other methods, including qPCR, spot assays, and plaque assays,
691 demonstrating its relevance for estimating phage host range across diverse bacterial–phage
692 combinations (65,84–86). Given the broad panel of phages and bacterial strains analysed, we
693 selected this method as it provides an optimal balance between throughput, reproducibility,
694 and comparability, while avoiding the more labour-intensive phage progeny quantification
695 assays.

696 The dataset comprised 44 bacterial strains tested in biological triplicate without phage as a
697 positive control, and the same 44 strains in biological triplicate exposed individually to each
698 of the 23 phages (S6 Fig). Each phage was tested against each biological triplicate of bacterial
699 strain, resulting in 3,036 growth kinetic measurements (23 phages × 44 bacterial strains × 3
700 replicates) (S6 Fig). Negative controls (culture medium) were used to check for
701 contamination, with plates containing more than two contaminated wells out of four
702 discarded (S6 Fig and S7A Fig). Two internal controls were included on each plate: the *P.*
703 *syringae* strain P1.01.01B03 (PG01a, isolated from apricot buds in France in 2014) and phage
704 Ghuch01, chosen because it completely inhibited the growth of strain P1.01.01B03 under
705 the experimental conditions used in this study (34,72) (S6 Fig and S7B-C Fig). A plate was
706 considered valid only when all three controls were visually confirmed: the internal bacterial
707 control exhibited visible turbidity, while both the internal phage control and the media
708 controls showed clear wells (S6 Fig).

709 Bacteria were incubated in 5 mL of liquid KB medium at 25°C and 200 rpm for 4 hours. Each
710 replicate was standardised to an OD_{600nm} = 0.01 (~10^7 CFU/mL) in phosphate buffer and
711 diluted in liquid KB medium to a final concentration of 10^5 CFU/mL. Each 96-well
712 spectrophotometric plate (NUNC, Thermo Scientific 267334) contained most phages
713 confronted with a single bacterial strain, with positive control wells containing 180 µL of
714 bacterial suspension and 20 µL of SM buffer. Experimental wells comprised 180 µL of
715 bacterial suspension and 20 µL of phage suspension at 10^7 PFU/mL, corresponding to a
716 multiplicity of infection (MOI) of 10, as determined from preliminary optimization tests.
717 Positive control measurements included between two and seven technical replicates per
718 biological replicate (except for bacterial strains 41A and 93D, which had only one technical
719 replicate per biological replicate). Measurements were taken using the SPECTROstar Omega
720 spectrophotometer (BMG Labtech, France) every 5 minutes for 48 hours at 25°C, with
721 intermittent orbital shaking at 200 rpm (200 seconds "ON" followed by 400 seconds "OFF").

722 To assess the host specificity of the 23 phages, their infectivity was tested on four
723 *Pseudomonas* spp. strains outside the *P. syringae* species complex using the double agar
724 layer method on solid medium. The selected strains were *P. aeruginosa* CFBP 2466, *P.*
725 *fluorescens* CFBP 2102, *P. graminis* 38B9, and *P. rhizosphaerae* 6B4. Strains were grown on
726 King's B (KB) agar for 48 h at 25°C (*P. fluorescens*, *P. graminis*) or 28°C (*P. aeruginosa*, *P.*
727 *rhizosphaerae*). Biological triplicates were inoculated into 5 mL of KB broth and incubated
728 overnight at 25°C with shaking at 200 rpm. Then, 500 µL of each culture was transferred to 5
729 mL fresh KB and incubated under the same conditions for 4 h. 300 µL of the culture in
730 exponential phase (OD between 0.2 and 0.8 at 600 nm) was mixed with 6 mL of 0.7% KB
731 semi-solid agar for the overlay. Each phage at an initial concentration of 10^7 PFU/mL was
732 serially diluted, and 10 µL of each dilution were spotted in duplicate onto the bacterial lawn.

733 Plates were allowed to dry and incubated overnight at 25°C before evaluation of plaque
734 lysis. The phage concentrations on the tested strains were divided by the concentration
735 obtained on their respective reference strains to calculate the efficiency of plating (EOP), a
736 standard metric.

737 *b. Calculation of bacterial growth inhibition percentages and probability of*
738 *significant inhibitory effects of phages on bacteria*

739 Bacterial growth curve data from each plate were analysed using R software (version 4.4.0)
740 with the *growthcurver* package (87). The area under the empirical growth curve (auc*e*) was
741 the primary parameter analysed. For each biological replicate of the bacterial control
742 condition (without phage), the mean and standard deviation of the auc*e* were calculated.
743 The percentage of bacterial growth inhibition caused by each phage was then determined
744 for each biological replicate using the following formula:

$$Bacterial\ growth\ inhibition(\%) = \frac{auce\ bacteria - auce\ phage}{auce\ bacteria} \times 100$$

745 where auc*e* bacteria represent the mean auc*e* of each bacterial control biological replicates,
746 and auc*e* phage corresponds to the auc*e* of the same bacterial strain when exposed to a
747 given phage. This normalisation accounts for potential differences in bacterial growth rates,
748 allowing for direct comparison of phage effects across bacterial strains (88). To evaluate
749 potential bias in this calculation, particularly when bacteria exhibit different maximal
750 densities (88), we assessed the coefficient of variation (CV) of the *k* parameter among
751 bacteria control condition. The CV was low (9.22%), supporting the validity of this approach
752 and indicating minimal bias associated with differences in growth maxima. Additionally, the
753 CV of the auc*e* values under bacteria control condition was moderate (11.34%), further

754 supporting the robustness of the percentage inhibition metric for cross-strain comparisons
755 while limiting potential biases (88).

756 Phage impact on bacterial growth was assessed using a probability value derived from a
757 normalised distribution, calibrated based on the mean (μ in the formula) and standard
758 deviation (σ in the formula) of the auc*e* for each biological replicate of each bacterium
759 exposed to phages. The probability ranges from 0 to 1 and is calculated using:

$$auce\ phage \sim N(\mu(auce\ bacteria), \sigma(auce\ bacteria))$$

760 A probability value below 0.05 indicates a significant growth reduction in presence of phage,
761 while a value exceeding 0.95 indicates a significant growth stimulation of bacteria in
762 presence of phage. The distribution of inhibition percentages within the 0.05 to 0.95
763 probability range is approximately symmetric and unimodal (S8 Fig), resembling a Gaussian
764 distribution. However, the Shapiro-Wilk test indicated a significant deviation from strict
765 normality ($p < 0.01$), likely due to the large sample size. Despite this, the distribution shape
766 supports the use of this probability window to robustly summarise phage effects on bacterial
767 growth. Bacterial strains 41A and 93D had only one technical replicate per biological
768 replicate, precluding probability calculation.

769 A phage was considered capable of inhibiting a bacterial strain if at least one of the three
770 bacterial replicates showed a significant inhibition (probability < 0.05); in this case, the
771 phage–bacterium combination was assigned a value of 1, in order to capture the maximal
772 host range potential of each phage. If no replicates showed significant inhibition, the
773 combination was assigned a value of 0. This threshold was verified to yield consistent results
774 with a more stringent criterion (requiring at least two out of three replicates to be
775 significant), in order to confirm the robustness of our analysis.

776 The relationship between the breadth of infection and the strength of inhibition was first
777 examined using the full dataset. For the phage perspective, we plotted virulence (mean
778 percentage of inhibition per phage) against infectivity (the number of bacterial strains each
779 phage significantly infected) (S9A Fig). For the bacterial perspective, we plotted susceptibility
780 (mean percentage of inhibition per bacterial strain) against sensitivity (the number of phages
781 significantly infecting each strain) (S9B Fig). We then considered only combinations showing
782 significant inhibition, for both phages (S9C Fig) and bacteria (S9D Fig), in order to reduce the
783 dependency between variables. Spearman correlation tests were calculated exclusively on
784 this subset of significant interactions, as this approach removes part of the inherent
785 dependence generated by the shared underlying metric.

786 Generally, bacterial populations decreased in the presence of phages. In 17% of observed
787 phage-bacteria combinations, a "stimulatory" effect was identified, with inhibition values
788 reaching up to -40%. Given the lack of reproducibility of the stimulatory effect (S10 Fig), only
789 positive inhibition values were retained for all analyses.

790 *Statistical analysis*

791 *a. Analysis of the structure of phage-bacteria interaction networks*

792 The analysis of networks representing bacterial growth inhibition percentages of 44 bacterial
793 strains combined with 23 phages was computed using the mean of the three biological
794 replicates for each combination. The analyses were performed using the R software (version
795 4.4.0) with the *bipartite* (for nestedness and modularity analysis) and *igraph* (for modularity
796 analysis) packages, following Moury *et al.* (30) and Torres-Barceló *et al.* (46).

797 The nestedness of the inhibition percentage network was evaluated using two algorithms:
798 *wNODF* (a weighted nestedness metric based on overlap and decreasing fill) (89) and *WINE*

799 (weighted-interaction nestedness estimator) (90). Modularity was assessed using six
800 algorithms: *spinglass* (91–93), *edge betweenness* (91), *fast greedy* (94), *leading eigenvector*
801 (95), *Beckett* (96), and *louvain* (97).

802 Prior to nestedness and modularity analyses, the continuous inhibitory percentage data of
803 the matrix were transformed into discrete integer values spanning from 0 to 100.

804 In addition, we analysed a network of binary data (0 and 1 values) representing the
805 statistical significance of bacterial growth inhibition by phages. For this network, three
806 algorithms were used for nestedness: *WINE* as applied to quantitative data, *Binmatnest* (98),
807 and *NODF2* (89) and three algorithms were used for modularity (*spinglass, leading*
808 *eigenvector,* and *Beckett)*.

809 For each algorithm and each network, the calculated nestedness and modularity scores were
810 compared with the scores obtained with seven (network of continuous inhibition data) or six
811 (network of binary inhibition significance data) different null models (i.e. simulated networks
812 for which no particular nestedness or modularity patterns are expected) in order to estimate
813 the statistical significance of nestedness and modularity of the actual networks (99,30,100).
814 Briefly, each null model differs by the constraints applied during randomization. The N
815 model preserves the total sum and number of zeros but not row or column totals. C1 and R1
816 apply the same principle column- or row-wise, respectively. The B model maintains both row
817 and column sums. The S model fully randomizes cell values, while C2 and R2 randomize
818 values within columns or rows, respectively. The PD (Probabilistic Degree) model takes the
819 original matrix and reshuffles positive infections while maintaining the same 'degree
820 distribution', on average, of both phages and hosts while reshuffling. Together, these models

821 span a gradient of constraint strength, from the least constrained (S, N) to the most
822 constrained (B).

823 *b. Statistical analyses to explain bacterial sensitivity and susceptibility to phages*

824 All statistical analyses were conducted in R version 4.4.2. The phylogenetic signal of the
825 number of phages infecting each bacterial strain was calculated using the *phylosig* function
826 from the *phytools* package, based on a phylogenetic tree reconstructed from bacterial core-
827 genome analysis (83).

828 The bacterial susceptibility profiles were tested with a PERmutational Multivariate ANalysis
829 Of VAriance (PERMANOVA), a non-parametric method that assesses the influence of
830 categorical variables on multivariate data by partitioning variation within a distance matrix.
831 Unlike traditional ANOVA, it makes no assumptions about data distribution and relies on
832 permutations to evaluate the statistical significance of group differences (101). Here, Bray–
833 Curtis dissimilarity was used to construct a distance matrix based on the mean percentage of
834 bacterial growth inhibition (from three replicates per combination). This matrix was then
835 analysed using the *adonis2* function from the *vegan* R package, with 9,999 permutations to
836 ensure robust p-value estimation (102). "*Phylogroup", "ecological origin",* and "*whether the*
837 *strain was an isolation bacterial strain for the phages*" were included as explanatory
838 variables. Multiple comparisons were conducted using the function *pairwise.adonis2* from
839 the *pairwiseAdonis* package (103), which is specifically adapted for PERMANOVA post-hoc
840 testing.

841 Additionally, the summed sensitivity, determined from the significant inhibition network
842 (see above), was used as a response variable in a Generalised Linear Model (GLM) with a
843 Poisson distribution with the function *glmmTMB* from the package *glmmTMB* (104). The

844 model included *"phylogroup"*, *"ecological origin"*, and their interaction as fixed effects. As no
845 significant interaction was detected, the simplified additive model (*phylogroup + ecological*
846 *origin*) was retained. No overdispersion was observed in the data (deviance/degrees of
847 freedom = 1.01). The effects of each explanatory variable were assessed using Type III
848 ANOVA with the *Anova* function from the *car* package. Post-hoc pairwise comparisons were
849 performed using Tukey's test implemented via the *multcomp* and *multcompView* packages,
850 with significance groups assigned accordingly.

851 *c. Statistical analysis to explain infectivity and virulence of phages*

852 The Phylogenetic Host Range Index (PHRI) (46), derived from the phylogenetic diversity
853 index *Hp* (105), was calculated using the *hilldiv* package . For this analysis, the bacterial
854 phylogenetic tree, constructed from core-genome analysis (see above), was transformed
855 into an ultrametric tree, and the PHRI index was determined using the matrix of percentage
856 inhibition.

857 The ranking of phage genera based on inhibition percentage (offset by +0.1 to avoid zeros)
858 was assessed using a GLMM with a Gamma distribution with a log-link, suitable for strictly
859 positive, continuous, and skewed data. The model included a random effect accounting for
860 the *"phage"* tested, the *"bacterial"* strain tested, and *"replicates"* nested within each
861 *"phage*:*bacterium"* combination. No overdispersion was detected (χ^2/ddl = 0.67). Post-hoc
862 pairwise comparisons between genera were performed using Tukey's test.

863 The phage-host infection profiles were tested with a PERMANOVA test based on the
864 distance matrix based on the mean percentage of bacterial growth inhibition compared with
865 *"phage genus"* and *"whether the phage had been isolated using one of the bacterial strains*
866 *included in the experimental panel*" explanatory variables (Bray-Curtis distance matrix; 9,999

867 permutations). Pairwise comparisons of phage infection patterns were performed using the
868 *pairwise.adonis2* function (103).

869 *d. Statistical analysis of defence systems and prophages*

870 The phylogenetic signal of the total number of defence systems and the abundance of
871 prophages was calculated using the *phylosig* function from the *phytools* package, based on a
872 phylogenetic tree reconstructed from bacterial core-genome analysis (83).

873 The total number of bacterial defence systems per bacterial genome and the abundance of
874 prophages were tested against the explanatory variables *"phylogroup"*, *"ecological origin",*
875 and *"growth rate"* in a GLM with a Poisson distribution with an additive model of the three
876 explanatory variables. No overdispersion was observed in the data (deviance/degrees of
877 freedom for defence systems = 0.61; for abundance of prophages = 0.47), and the effects of
878 each explanatory variable were assessed using Type III ANOVA. Finally, the post-hoc pairwise
879 comparisons were performed using Tukey's test and significance groups were assigned
880 accordingly.

881 The relationship between ANI of bacterial pairs and their percentage similarity in their
882 defence systems (offset by +0.1 to avoid zeros) was assessed using a GLMM with a Gamma
883 distribution and a log-link function. Random effects for each bacterium in the pair were
884 included to account for the non-independence of observations. No overdispersion was
885 detected (χ^2/ddl = 0.24).

886 The defence system profiles, the proportions of the defence system categories, and
887 prophages distribution profile across bacterial strains were analysed using a PERMANOVA
888 test based on a Bray-Curtis dissimilarity matrix with 9,999 permutations. *"Phylogroup"*,
889 *"ecological origin"*, and *"growth rate"* were included as explanatory variables. Pairwise

890 comparisons of the defence system and prophage composition between groups were
891 conducted using the *pairwise.adonis2* function (103).

892 Linear regressions were conducted to explain the number of phages infecting each bacterial
893 strain, using the *total number of defence system subtypes per bacterium*, the *phylogroups*,
894 and *ecological origins* as explanatory variables in a multiple linear regression model
895 (adjusted $R^2 = 0.21$; $p = 0.02$ for the overall model). The same analysis was applied using the
896 *number of prophages per bacterium* instead of defence systems, with *phylogroups* and
897 *ecological origins* as predictors (adjusted $R^2 = -0.06$; $p = 0.81$ for the overall model). Before
898 conducting these analyses, all assumptions of linear regression, including normality of
899 residuals and homoscedasticity, were verified.

900 *e. Statistical analysis of host range data with multimodel analysis*

901 To account for the widest range of explanatory variables influencing the percentage of
902 inhibition, we performed a multimodel analysis coupled with a generalised linear mixed
903 model (GLMM). To enhance model efficiency, the number of explanatory variables was
904 reduced through a multi-step process. Variables with more than 20% missing data were
905 excluded. To ensure independence among explanatory variables before multimodel analysis,
906 we assessed pairwise associations using Spearman's correlation for quantitative variables,
907 chi-squared tests for categorical variables, and Kruskal–Wallis tests for associations between
908 categorical and quantitative variables. Variables showing significant correlations ($p < 0.05$)
909 were considered dependent. In such cases, one of the correlated variables was retained,
910 while the others were excluded. To further refine the selection, Principal Component
911 Analyses (PCA) were conducted separately on the matrices representing the distributions of

912 defence systems and prophages across bacterial genomes. The most informative variables
913 contributing to the first principal components were retained for subsequent modelling.

914 The PCA on defence systems was performed without centring or scaling the data. Two
915 dimensions accounted for 32.5% of the variance and led to the selection of the variables
916 *"Septu"*, *"Gabija"*, and *"RloC"*, which were among the most frequent defence systems in the
917 genomes. Similarly, the PCA on the distribution of prophage species in bacterial genomes,
918 also conducted without centring or scaling (already binary), captured 14% of the variance in
919 the first two dimensions and identified two perpendicular variable groups composed
920 respectively of prophage species *"17"*, *"54"*, *"26",* and *"29"* on one side, and species *"7"*,
921 *"47"*, *"57"*, *"22"*, and *"31"* on the other. One representative species from each group was
922 selected for the multimodel analysis (species *"17"* and *"22"*).

923 To identify factors explaining variations in bacterial growth inhibition percentage, we fitted a
924 GLMM. The response variable (offset by +0.1 to avoid zeros) was modelled using a Gamma
925 distribution with a log-link function. Random effects included *"phage"*, *"bacterium"*, and
926 their *"imbricated replicates"*.

927 To compare two approaches for modelling the effects of bacterial defence systems and
928 prophages on the percentage of inhibition, two separate GLMMs were constructed: one
929 using total counts of defence systems and prophages, and another incorporating PCA-
930 derived principal components of the distribution matrices (of prophages and defence
931 systems). No overdispersion was detected in either model (total counts: $\chi^2/df = 0.65$; PCA-
932 derived: $\chi^2/df = 0.66$), supporting the appropriateness of the Gamma distribution for
933 modelling the response variable.

934 To identify the most informative predictors explaining variation in phage replication
935 efficiency, we used a multimodel analysis. Model selection was performed using the *dredge*
936 function from the *MuMIn* package, ranking models based on corrected Akaike information
937 criterion (AIC). All possible sub-models were generated and evaluated. Models whose
938 cumulative AIC weights accounted for 95% of the total were retained for model averaging
939 using the *model.avg* function. The relative importance of each explanatory variable was then
940 estimated using the *sw* function, which computes the sum of the AIC weights across all
941 models in which the variable appears.

942

943 **Acknowledgments**

944 We thank Elisabeth Villard, Cécile Monteil, and Charlotte Chandeysson for their assistance
945 with experimental procedures. We are also grateful to Dominique Holtappels for valuable
946 discussions, and to Hannah Toth for her assistance in depositing bacterial genomes into
947 NCBI.

948

949

950 **References**

951 1. Brown TL, Charity OJ, Adriaenssens EM. Ecological and functional roles of bacteriophages
952 in contrasting environments: marine, terrestrial and human gut. Curr Opin Microbiol.
953 2022 Dec 1;70:102229.

954 2. Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea: viruses play critical roles in
955 the structure and function of aquatic food webs. BioScience. 1999 Oct 1;49(10):781–8.

956 3. Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol
957 Biol Rev. 2000 Mar;64(1):69–114.

958 4. Suttle CA. Marine viruses — major players in the global ecosystem. Nat Rev Microbiol.
959 2007 Oct;5(10):801–12.

960 5. Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, Heldal M, et al.
961 Global-scale processes with a nanoscale drive: the role of marine viruses. ISME J. 2008
962 June;2(6):575–8.

963 6. Mirzaei MK, Maurice CF. Ménage à trois in the human gut: interactions between host,
964 bacteria and phages. Nat Rev Microbiol. 2017 July;15(7):397–408.

965 7. De Sordi L, Lourenço M, Debarbieux L. The battle within: interactions of bacteriophages
966 and bacteria in the gastrointestinal tract. Cell Host Microbe. 2019 Feb;25(2):210–8.

967 8. Svircev A, Roach D, Castle A. Framing the future with bacteriophages in agriculture.
968 Viruses. 2018 May;10(5):218.

969 9. Hyman P, Abedon ST. Bacteriophage host range and bacterial resistance. In: Advances in
970 Applied Microbiology [Internet]. Elsevier; 2010 [cited 2024 Aug 12]. p. 217–48. Available
971 from: https://linkinghub.elsevier.com/retrieve/pii/S0065216410700071

972 10. Georjon H, Bernheim A. The highly diverse antiphage defence systems of bacteria. Nat
973 Rev Microbiol. 2023 Oct;21(10):686–700.

974 11. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, et al. Phage–host
975 coevolution in natural populations. Nat Microbiol. 2022 July;7(7):1075–86.

976 12. Borin JM, Lee JJ, Lucia-Sanz A, Gerbino KR, Weitz JS, Meyer JR. Rapid bacteria-phage
977 coevolution drives the emergence of multiscale networks. Science. 2023 Nov
978 10;382(6671):674–8.

979 13. Beamud B, García-González N, Gómez-Ortega M, González-Candelas F, Domingo-Calap P,
980 Sanjuan R. Genetic determinants of host tropism in *Klebsiella* phages. Cell Rep. 2023 Feb
981 6;42(2):112048.

982 14. Gaborieau B, Vaysset H, Tesson F, Charachon I, Dib N, Bernier J, et al. Prediction of strain
983 level phage–host interactions across the *Escherichia* genus using only genomic
984 information. Nat Microbiol. 2024 Nov;9(11):2847–61.

985 15. Johnson MC, Laderman E, Huiting E, Zhang C, Davidson A, Bondy-Denomy J. Core
986 defense hotspots within *Pseudomonas aeruginosa* are a consistent and rich source of
987 anti-phage defense systems. Nucleic Acids Res. 2023 June 9;51(10):4995–5005.

988 16. Burke KA, Urick CD, Mzhavia N, Nikolich MP, Filippov AA. Correlation of *Pseudomonas*
989 *aeruginosa* phage resistance with the numbers and types of antiphage systems. Int J Mol
990 Sci. 2024 Jan;25(3):1424.

991 17. Costa AR, van den Berg DF, Esser JQ, Muralidharan A, van den Bossche H, Bonilla BE, et
992 al. Accumulation of defense systems in phage-resistant strains of *Pseudomonas*
993 *aeruginosa*. Sci Adv. 2024 Feb 23;10(8):eadj0341.

994 18. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, et al. Rapid
995 evolutionary turnover of mobile genetic elements drives bacterial resistance to phages.
996 Science. 2021 Oct 22;374(6566):488–92.

997 19. Czajkowski R. May the phage be with you? Prophage-like elements in the genomes of
998 soft rot Pectobacteriaceae: *Pectobacterium* spp. and *Dickeya* spp. Front Microbiol
999 [Internet]. 2019 Feb 14 [cited 2025 Apr 28];10. Available from:
1000 https://www.frontiersin.orghttps://www.frontiersin.org/journals/microbiology/articles/
1001 10.3389/fmicb.2019.00138/full

1002 20. Dominguez-Mirazo M, Jin R, Weitz JS. Functional and comparative genomic analysis of
1003 integrated prophage-like sequences in "*Candidatus Liberibacter asiaticus*". mSphere.
1004 2019 Nov 13;4(6):10.1128/msphere.00409-19.

1005 21. Castillo JA, Secaira-Morocho H, Maldonado S, Sarmiento KN. Diversity and evolutionary
1006 dynamics of antiphage defense systems in *Ralstonia solanacearum* species complex.
1007 Front Microbiol [Internet]. 2020 May 20 [cited 2024 Sept 24];11. Available from:
1008 https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.00961/f
1009 ull

1010 22. Greenrod STE, Stoycheva M, Elphinstone J, Friman VP. Global diversity and distribution
1011 of prophages are lineage-specific within the *Ralstonia solanacearum* species complex.
1012 BMC Genomics. 2022 Oct 6;23(1):689.

1013 23. Dougherty PE, Nielsen TK, Riber L, Lading HH, Forero-Junco LM, Kot W, et al. Widespread
1014 and largely unknown prophage activity, diversity, and function in two genera of wheat
1015 phyllosphere bacteria. ISME J. 2023 Dec;17(12):2415–25.

1016 24. Holtappels D, Rickus GEJ, Morgan T, Rezende RR de, Koskella B, Alfenas-Zerbini P.
1017 Evidence for the key roles of the *Pseudomonas syringae* mobilome in shaping biotic
1018 interactions [Internet]. bioRxiv; 2024 [cited 2024 Aug 20]. p. 2024.03.19.585818.
1019 Available from: https://www.biorxiv.org/content/10.1101/2024.03.19.585818v2

1020 25. Jimenez D, Beltran D, Castillo JA. Mobile genetic elements of *Xylella fastidiosa* and their
1021 contribution to pathogenicity. Plant Pathol. 2024 Dec;73(9):2490–9.

1022 26. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host-phage
1023 interactions. Proc Natl Acad Sci U S A. 2011 July 12;108(28):E288–97.

1024 27. Gupta A, Peng S, Leung CY, Borin JM, Medina SJ, Weitz JS, et al. Leapfrog dynamics in
1025 phage-bacteria coevolution revealed by joint analysis of cross-infection phenotypes and
1026 whole genome sequencing. Ecol Lett. 2022;25(4):876–88.

1027 28. Molina F, Menor-Flores M, Fernández L, Vega-Rodríguez MA, García P. Systematic
1028 analysis of putative phage-phage interactions on minimum-sized phage cocktails. Sci
1029 Rep. 2022 Feb 14;12(1):2458.

1030 29. Kauffman KM, Chang WK, Brown JM, Hussain FA, Yang J, Polz MF, et al. Resolving the
1031 structure of phage–bacteria interactions in the context of natural diversity. Nat
1032 Commun. 2022 Jan 18;13(1):372.

1033 30. Moury B, Audergon JM, Baudracco-Arnas S, Ben Krima S, Bertrand F, Boissot N, et al. The
1034 quasi-universality of nestedness in the structure of quantitative plant-parasite
1035 interactions. Peer Community J. 2021 Nov 30;1:e44.

1036 31. Morris CE, Monteil CL, Berge O. The life history of *Pseudomonas syringae*: linking
1037 agriculture to earth system processes. Annu Rev Phytopathol. 2013;51:85–104.

1038 32. Morris CE, Sands DC, Vinatzer BA, Glaux C, Guilbaud C, Buffière A, et al. The life history of
1039 the plant pathogen *Pseudomonas syringae* is linked to the water cycle. ISME J. 2008
1040 Mar;2(3):321–34.

1041 33. Morris CE, Lamichhane JR, Nikolić I, Stanković S, Moury B. The overlapping continuum of
1042 host range among strains in the *Pseudomonas syringae* complex. Phytopathol Res. 2019
1043 Jan 16;1(1):4.

1044 34. Parisi L, Morgaint B, Blanco-Garcia J, Guilbaud C, Chandeysson C, Bourgeay JF, et al.
1045 Bacteria from four phylogroups of the *Pseudomonas syringae* complex can cause
1046 bacterial canker of apricot. Plant Pathol. 2019 Sept;68(7):1249–58.

1047 35. Thomas MD, Leary JV. Bacteriophages from sewage specific for fluorescent phytogenic
1048 Pseudomonads. Phytopathology. 1983 Sept;(73):403–6.

1049 36. Cuppels DA. Isolation and characterization of phages useful for identifying *Pseudomonas*
1050 *syringae pv. tomato*. Phytopathology. 1983 Apr;(73):1376–81.

1051 37. Gross DC, Cody YS, Proebsting EL, Radamaker GK, Spotts RA. Ecotypes and pathogenicity
1052 of ice-nucleation-active *Pseudomonas syringae* isolated from deciduous fruit tree
1053 orchards. 1983 Sept;(74):241–8.

1054 38. Cuppels DA. The use of pathovar-indicative bacteriophages for rapidly detecting
1055 *Pseudomonas syringae pv. tomato* in tomato leaf and fruit lesions. 1984 Feb;(74):891–4.

1056 39. Taylor JD, Teverson DM, Allen DJ, Pastor-Corrales MA. Identification and origin of races
1057 of *Pseudomonas syringae pv. phaseolicola* from Africa and other bean growing areas.
1058 Plant Pathol. 1996;45(3):469–78.

1059 40. Frampton RA, Taylor C, Holguín Moreno AV, Visnovsky SB, Petty NK, Pitman AR, et al.
1060 Identification of bacteriophages for biocontrol of the kiwifruit canker phytopathogen
1061 *Pseudomonas syringae pv. actinidiae*. Appl Environ Microbiol. 2014 Apr;80(7):2216–28.

1062 41. James SL, Rabiey M, Neuman BW, Percival G, Jackson RW. Isolation, characterisation and
1063 experimental evolution of phage that infect the horse chestnut tree pathogen,
1064 *Pseudomonas syringae pv. aesculi*. Curr Microbiol. 2020 Aug 1;77(8):1438–47.

1065 42. Rombouts S, Volckaert A, Venneman S, Declercq B, Vandenheuvel D, Allonsius CN, et al.
1066 Characterization of novel bacteriophages for biocontrol of bacterial blight in leek caused
1067 by *Pseudomonas syringae pv. porri*. Front Microbiol [Internet]. 2016 Mar 15 [cited 2024
1068 Oct 8];7. Available from:
1069 https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2016.00279/f
1070 ull

1071 43. Geng Y, Nguyen TVP, Homaee E, Golding I. Using bacterial population dynamics to count
1072 phages and their lysogens. Nat Commun. 2024 Sept 6;15(1):7814.

1073 44. Wright RCT, Friman VP, Smith MCM, Brockhurst MA. Cross-resistance is modular in
1074 bacteria–phage interactions. PLOS Biol. 2018 Oct 3;16(10):e2006057.

1075 45. Wright RCT, Friman VP, Smith MCM, Brockhurst MA. Resistance evolution against phage
1076 combinations depends on the timing and order of exposure. mBio. 2019 Sept
1077 24;10(5):10.1128/mbio.01652-19.

1078 46. Torres-Barceló C, Boyer C, Garneau JR, Poussier S, Robène I, Moury B. A Phylogenetic
1079 Host-Range Index reveals ecological constraints in phage specialisation and virulence.
1080 Mol Ecol. 2025 July 28;n/a(n/a):e70052.

1081 47. Seidel R, Bloom JGP, Dekker C, Szczelkun MD. Motor step size and ATP coupling
1082 efficiency of the dsDNA translocase EcoR124I. EMBO J. 2008 May 7;27(9):1388–98.

1083 48. Pleška M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, et al. Bacterial
1084 autoimmunity due to a Restriction-Modification system. Curr Biol CB. 2016 Feb
1085 8;26(3):404–9.

1086 49. Berngruber TW, Lion S, Gandon S. Evolution of suicide as a defence strategy against
1087 pathogens in a spatially structured environment. Ecol Lett. 2013 Apr;16(4):446–53.

1088 50. van Houte S, Buckling A, Westra ER. Evolutionary ecology of prokaryotic immune
1089 mechanisms. Microbiol Mol Biol Rev. 2016 July 13;80(3):745–63.

1090 51. Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage
1091 taxonomy. Viruses. 2021 Mar;13(3):506.

1092 52. Mavrodi DV, Loper JE, Paulsen IT, Thomashow LS. Mobile genetic elements in the
1093 genome of the beneficial rhizobacterium *Pseudomonas fluorescens* Pf-5. BMC Microbiol.
1094 2009 Jan 13;9(1):8.

1095 53. Hockett KL, Renner T, Baltrus DA. Independent co-option of a tailed bacteriophage into a
1096 killing complex in *Pseudomonas*. mBio. 2015 Aug 11;6(4):10.1128/mbio.00452-15.

1097 54. Backman T, Latorre SM, Symeonidi E, Muszyński A, Bleak E, Eads L, et al. A phage tail–like
1098 bacteriocin suppresses competitors in metapopulations of pathogenic bacteria. Science.
1099 2024 June 14;384(6701):eado0713.

1100 55. Fautt C, Hockett K, Delattre S, Baltrus D, Couradeau E. Tailocin tail fiber diversity
1101 correlates with rfbD variation in the *Pseudomonas syringae* species complex. ISME
1102 Commun. 2025 Jan 1;5(1):ycaf099.

1103 56. Hunter M, Fusco D. Superinfection exclusion: A viral strategy with short-term benefits
1104 and long-term drawbacks. PLOS Comput Biol. 2022 May 10;18(5):e1010125.

1105 57. Koskella B, Lin DM, Buckling A, Thompson JN. The costs of evolving resistance in
1106 heterogeneous parasite environments. Proc R Soc B Biol Sci. 2011 Dec
1107 14;279(1735):1896–903.

1108 58. Morris CE, Sands DC, Vanneste JL, Montarry J, Oakley B, Guilbaud C, et al. Inferring the
1109 evolutionary history of the plant pathogen *Pseudomonas syringae* from its biogeography
1110 in headwaters of rivers in North America, Europe, and New Zealand. mBio. 2010 June
1111 29;1(3):e00107-10.

1112 59. Monteil CL, Yahara K, Studholme DJ, Mageiros L, Méric G, Swingle B, et al. Population-
1113 genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas*
1114 *syringae* pathogens. Microb Genomics [Internet]. 2016 Oct 1 [cited 2021 Sept 15];2(10).
1115 Available from:
1116 https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000089

1117 60. Berge O, Monteil CL, Bartoli C, Chandeysson C, Guilbaud C, Sands DC, et al. A user's guide
1118 to a data base of the diversity of *Pseudomonas syringae* and its application to classifying
1119 strains in this phylogenetic complex. PLOS ONE. 2014 Sept 3;9(9):e105547.

1120 61. Koonin EV, Makarova KS, Wolf YI. Evolutionary genomics of defense systems in archaea
1121 and bacteria. Annu Rev Microbiol. 2017 Sept 8;71(Volume 71, 2017):233–61.

1122 62. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a
1123 community resource. Nat Rev Microbiol. 2020 Feb;18(2):113–9.

1124 63. Makarova KS, Wolf YI, Koonin EV. Comparative genomics of defense systems in archaea
1125 and bacteria. Nucleic Acids Res. 2013 Apr 1;41(8):4360–77.

1126 64. Cooney-Nutley K, Chakravorty S, Nix I, Zeng Z, Greer SF, Rabiey M. Understanding
1127 bacterial antiviral defence systems and phage receptors to better inform rational phage
1128 cocktail design to treat bacterial canker. Microb Biotechnol. 2025 Sept 19;18(9):e70232.

1129 65. Xie Y, Wahab L, Gill JJ. Development and validation of a microtiter plate-based assay for
1130 determination of bacteriophage host range and virulence. Viruses. 2018 Apr;10(4):189.

1131 66. Bell G. Evolutionary Rescue. Annu Rev Ecol Evol Syst. 2017 Nov 2;48(Volume 48,
1132 2017):605–27.

1133 67. Abedon S. Lysis from without. Bacteriophage. 2011 Jan 1;1:46–9.

1134 68. Nobrega FL, Vlot M, de Jonge PA, Dreesens LL, Beaumont HJE, Lavigne R, et al. Targeting
1135 mechanisms of tailed bacteriophages. Nat Rev Microbiol. 2018 Dec;16(12):760–73.

1136 69. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption.
1137 FEMS Microbiol Lett. 2016 Feb 1;363(4):fnw002.

1138 70. Müller DM, Pourtois JD, Kim MK, Targ B, Burgener EB, Milla C, et al. Bacterial receptors
1139 but not anti-phage defence mechanisms determine host range for a pair of
1140 *Pseudomonas aeruginosa* lytic phages [Internet]. bioRxiv; 2024 [cited 2024 Sept 23]. p.
1141 2024.04.30.591980. Available from:
1142 https://www.biorxiv.org/content/10.1101/2024.04.30.591980v2

1143 71. Kering KK, Kibii BJ, Wei H. Biocontrol of phytobacteria with bacteriophage cocktails. Pest
1144 Manag Sci. 2019;75(7):1775–81.

1145 72. Feltin C, Garneau JR, Morris CE, Bérard A, Torres-Barceló C. Novel phages of
1146 *Pseudomonas syringae* unveil numerous potential auxiliary metabolic genes. J Gen Virol.
1147 2024;105(6):001990.

1148 73. King EO, Ward MK, Raney DE. Two simple media for the demonstration of pyocyanin and
1149 fluorescin. J Lab Clin Med. 1954 Aug 1;44(2):301–7.

1150 74. Dieckmann MA, Beyvers S, Nkouamedjo-Fankep RC, Hanel PHG, Jelonek L, Blom J, et al.
1151 EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure.
1152 Nucleic Acids Res. 2021 July 2;49(W1):W185–92.

1153 75. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for
1154 large alignments. PLOS ONE. 2010 Mar 10;5(3):e9490.

1155 76. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, et al. Systematic and
1156 quantitative view of the antiviral arsenal of prokaryotes. Nat Commun. 2022 May
1157 10;13(1):2561.

1158 77. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile
1159 genetic elements with geNomad. Nat Biotechnol. 2023 Sept 21;1–10.

1160 78. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses
1161 the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol.
1162 2021 May;39(5):578–85.

1163 79. Moraru C, Varsani A, Kropinski AM. VIRIDIC—a novel tool to calculate the intergenomic
1164 similarities of prokaryote-infecting viruses. Viruses. 2020 Nov 6;12(11):1268.

1165 80. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral
1166 proteomic tree server. Bioinforma Oxf Engl. 2017 Aug 1;33(15):2379–80.

1167 81. Orfei B, Pothier JF, Fenske L, Blom J, Moretti C, Buonaurio R, et al. Race-specific
1168 genotypes of *Pseudomonas syringae pv. tomato* are defined by the presence of mobile
1169 DNA elements within the genome. Front Plant Sci [Internet]. 2023 July 5 [cited 2025 June
1170 7];14. Available from: https://www.frontiersin.org/journals/plant-
1171 science/articles/10.3389/fpls.2023.1197706/full

1172 82. Balbuena JA, Míguez-Lozano R, Blasco-Costa I. PACo: A Novel Procrustes Application to
1173 Cophylogenetic Analysis. PLOS ONE. 2013 Apr 8;8(4):e61048.

1174 83. Revell LJ. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods
1175 (and other things). PeerJ. 2024 Jan 5;12:e16505.

1176 84. Walsh SK, Imrie RM, Matuszewska M, Paterson GK, Weinert LA, Hadfield JD, et al. The
1177 host phylogeny determines viral infectivity and replication across *Staphylococcus* host
1178 species. PLOS Pathog. 2023 June 8;19(6):e1011433.

1179 85. Holtappels D, Alfenas-Zerbini P, Koskella B. Drivers and consequences of bacteriophage
1180 host range. FEMS Microbiol Rev. 2023 July 1;47(4):fuad038.

1181 86. Alseth EO, Roush C, Irby I, Kopylov M, Bobe D, Diggs MW, et al. Mystique, a broad host
1182 range *Acinetobacter* phage, reveals the impact of culturing conditions on phage isolation
1183 and infectivity. PLOS Pathog. 2025 Apr 10;21(4):e1012986.

1184 87. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics
1185 from microbial growth curves. BMC Bioinformatics. 2016 Apr 19;17:172.

1186 88. Blazanin M, Vasen E, Vilaró Jolis C, An W, Turner PE. Quantifying phage infectivity from
1187 characteristics of bacterial population dynamics. Proc Natl Acad Sci. 2025 Sept
1188 16;122(37):e2513377122.

1189 89. Almeida-Neto M, Guimarães P, Guimarães Jr PR, Loyola RD, Ulrich W. A consistent metric
1190 for nestedness analysis in ecological systems: reconciling concept and measurement.
1191 Oikos. 2008;117(8):1227–39.

1192 90. Galeano J, Pastor JM, Iriondo JM. Weighted-Interaction Nestedness Estimator (WINE): A
1193 new estimator to calculate over frequency matrices. Environ Model Softw. 2009 Nov
1194 1;24(11):1342–6.

1195 91. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys
1196 Rev E. 2004 Feb 26;69(2):026113.

1197 92. Reichardt J, Bornholdt S. Statistical mechanics of community detection. Phys Rev E. 2006
1198 July 18;74(1):016110.

1199 93. Traag VA, Bruggeman J. Community detection in networks with positive and negative
1200 links. Phys Rev E. 2009 Sept 21;80(3):036115.

1201 94. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks.
1202 Phys Rev E. 2004 Dec 6;70(6):066111.

95. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E. 2006 Sept 11;74(3):036104.

96. Beckett SJ. Improved community detection in weighted bipartite networks. R Soc Open Sci. 2016 Jan;3(1):140536.

97. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008 Oct;2008(10):P10008.

98. Rodríguez-Gironés MA, Santamaría L. A new algorithm to calculate the nestedness temperature of presence–absence matrices. J Biogeogr. 2006;33(5):924–35.

99. Moury B, Fabre F, Hébrard E, Froissart R. Determinants of host species range in plant viruses. J Gen Virol. 2017 Apr;98(4):862–73.

100. Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, et al. Phage-bacteria infection networks. Trends Microbiol. 2013 Feb;21(2):82–91.

101. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). In: Wiley StatsRef: Statistics Reference Online [Internet]. John Wiley & Sons, Ltd; 2017 [cited 2025 June 4]. p. 1–15. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07841

102. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14(6):927–30.

103. Arbizu PM. pmartinezarbizu/pairwiseAdonis [Internet]. 2025 [cited 2025 Apr 14]. Available from: https://github.com/pmartinezarbizu/pairwiseAdonis

104. Brooks M E, Kristensen K, Benthem K J ,van, Magnusson A, Berg C W, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 2017;9(2):378.

105. Allen B, Kon M, Bar-Yam Y. A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. Am Nat. 2009 Aug;174(2):236–43.

Supporting information

S1 Fig. Bacterial resistance to phages and host range breadth of the tested phages.

(A) Number of phages to which bacteria are sensitive or resistant, with the phylogeny of the strains indicated at the bottom of the histogram. (B) Number of bacteria infected or not by

1234 the phages, with the taxonomy of the phages listed at the bottom of the histogram. Values
1235 are derived from the binary inhibition matrix.

1236

1237 **S2 Fig. Taxonomic and phylogenetic characteristics, and ecological origins of phages and**
1238 **bacteria explaining host range.**

1239 (A) Bacterial percentage of inhibition by phages across bacterial phylogroups and ecological
1240 origins. Significance letters indicate results from the PERMANOVA test and associated
1241 pairwise comparisons. (B) Bacterial inhibition percentage of different phage genera.
1242 Significance letters represent results from a GLMM with a Gamma regression distribution
1243 and random effect (phage, bacteria, and the replicate imbricated in each of the phage-
1244 bacteria interactions), and post hoc pairwise comparisons performed using Tukey's test.

1245

1246 **S3 Fig. Relative proportions of defence systems by functional category (%) and total**
1247 **number of defence systems across bacterial genomes.**

1248 On the left, a phylogenetic tree based on the core genomes of the bacterial strains is shown,
1249 with corresponding phylogroups indicated in the adjacent column.

1250

1251 **S4 Fig. Percentage of similarity of defence system subtypes among bacterial strains in**
1252 **relation to their genome-wide average nucleotide identity (ANI).**

1253 Transparent grey dots represent individual observations (darker areas indicate overlap). The
1254 overall trend is shown by the grey line, which represents the mean prediction from the

1255 GLMM accounting for random effects associated with bacterial strains. The shaded ribbon
1256 indicates the 95 % confidence interval of the model's marginal predictions.

1257

1258 **S5 Fig. Co-phylogeny of the CL4 prophage family-like (R-type syringacin).**

1259 Co-phylogenetic analysis between 68 strains of the *P. syringae* species complex (left),
1260 represented by a maximum likelihood phylogenetic tree based on their core-genomes, and
1261 their associated 68 CL4 family-like prophages (right), represented by a proteomic tree. The
1262 44 bacterial strains included in this study are marked with dots at the tips of the bacterial
1263 tree. Phylogroups and prophage-carrying strains are colour-coded accordingly. Two nodes
1264 with a maximum likelihood support value below 0.90 are marked with red dots.

1265

1266 **S6 Fig. Experimental setup illustrating bacterial growth inhibition by phages.**

1267 Each biological replicate per bacterial strain is represented by a different colour, and each
1268 row corresponds to the technical replicates within that biological replicate. For clarity, only 3
1269 of the 23 tested phages are shown in this schematic, which does not depict the
1270 randomisation of phages across plates. Three types of internal controls are included: positive
1271 controls with bacteria only, negative controls with medium only, and internal reference
1272 controls using the bacterial strain P1.01.01B03 (growth shown in S7B Fig) and the internal
1273 phage Ghuch01, which consistently inhibits this strain (see S7C Fig). Icons were obtained
1274 from open-source resources (microplate: https://commons.wikimedia.org/wiki/File:96-
1275 Well_plate.svg; phage: https://openclipart.org/detail/62785/virus; bacteria:

1276 https://openclipart.org/detail/221334/lamarque-disease; bottle of media:
1277 https://openclipart.org/detail/128449/reagent-bottle).

1278

1279 **S7 Fig. Analysis of the different controls used in the phage host range *in vitro* tests.**

1280 (A) Analysis of culture medium controls by plate (labelled by bacteria). (B) Area under the
1281 empirical curve (auc*e*) of the internal bacterial control, *P. syringae* strain P1.01.01B03. (C)
1282 Inhibition of the internal phage control Ghuch01 on the internal bacterial control (%).

1283

1284 **S8 Fig. Histogram showing the distribution of inhibition percentages within the 0.05 to**
1285 **0.95 probability range, displaying an approximately symmetric and unimodal shape.**

1286

1287 **S9 Fig. Relationship between infection breadth of infection and inhibition strength.**

1288 (A, C) Phage perspective: mean percentage of inhibition per phage (y-axis) is plotted against
1289 the number of bacterial strains each phage significantly infected (x-axis). Colours indicate
1290 phage genera. (B, D) Bacterial perspective: mean percentage of inhibition per bacterial strain
1291 (y-axis) is plotted against the number of phages significantly infecting each strain (x-axis).
1292 Colours represent bacterial phylogroups, and shapes indicate ecological niches. (A, B)
1293 Analyses based on overall inhibition values. (C, D) Analyses restricted to significant phage–
1294 bacterium interactions. Spearman's rho and associated p-values are reported for panels (C)
1295 and (D).

1296

1297 **S10 Fig. Results of additional experiments on five potentially stimulating phages and five**
1298 **potentially stimulated bacteria**.

1299 Generally, bacterial populations decreased in the presence of phages. In 17% of observed
1300 phage-bacteria combinations, a "stimulatory" effect was identified, with inhibition values
1301 reaching up to -40%. Five phages showing strong stimulatory effect and five bacteria
1302 showing significant stimulation were chosen to test the reproducibility of the stimulation
1303 effect in subsequent experiments. (A) Matrix of mean inhibition percentages for the five *P.*
1304 *syringae* strains (rows) and five phages (columns), with their phylogenetic and taxonomic
1305 affiliations indicated by colour. The number of stars indicates the number of replicates
1306 showing a significantly inhibiting or stimulating percentage. Of ten initially identified
1307 significant stimulating interactions (delimited by red lines), only three were confirmed in
1308 subsequent experiments. (B) Size of the five bacterial populations in response to the phages.
1309 Stars correspond to significant differences from the control (no phage) based on the results
1310 of the GLM model, family = Gamma. Arrows indicate interactions with at least one of the
1311 replicates showing a significant inhibition or stimulation percentage. Out of the three
1312 significant interactions of stimulatory phenotype, only 2 phage-bacteria interactions resulted
1313 in slightly higher bacterial densities in the presence of phages compared to phage-free
1314 conditions, though this difference was not statistically significant. (C) Spearman correlation
1315 between bacterial concentration and inhibition percentage, by bacterium and phage. No
1316 correlation was significant and bacterial population size only poorly explains the inhibition
1317 percentages.

1318

1319 **S11 Fig. Kinetics of bacterial growth inhibition across the full dataset.**

1320 Optical density (OD) is plotted over time (hours) for each phage–bacteria interaction, with
1321 phages (n=23) in columns and bacterial strains (n=44) in rows. Grey curves represent the
1322 average growth of bacteria alone for each phage–bacteria pair. Blue curves show individual
1323 replicates for each interaction (n = 3), illustrating the variability in bacterial response to
1324 phage infection.

1325

1326 **S1 Table. Phage–bacterium matrices showing (A) the percentage of bacterial growth**
1327 **inhibition and (B) the number of biological replicates (out of three) showing significant**
1328 **inhibition.**

1329 Rows correspond to bacterial strains, while columns represent phage clones, each annotated
1330 with their respective phylogroups and genera.

1331

1332 **S2 Table: Analysis of nestedness and modularity of the network of binary data**
1333 **representing the significance or non-significance of bacterial growth inhibition by phages**
1334 **(S1B Table).**

1335 Null models are described in Moury *et al.* (30) for B, N, C2, R2 and S, and in Moury *et al.* (99)
1336 for PD. For each null model and each algorithm, the frequency of simulations, among 100,
1337 showing a strictly higher nestedness (or modularity) score than that of the actual matrix is
1338 indicated. Significantly nested or modular networks ($p \leq 0.05$) are indicated in red.

1339

1340 **S3 Table. Genetic analysis of the 88 prophages identified in the 44 *P. syringae* strains using**
1341 **VIRIDIC.**

1342 The 27 reference phages are highlighted in yellow (25 from the collection described in Feltin
1343 *et al.* (72) and 3 from NCBI). Species and genus-level clusters of the prophages are indicated
1344 on the right in colour. High genetic similarities are highlighted in green.

1345

1346 **S4 Table. Phylogenetic characteristics, ecological origins, countries, and dates of sampling**
1347 **of the 44 *P. syringae* strains used for the host range study.**

1348 The four *P. syringae* strains highlighted in bold are strains that were used to isolate some of
1349 the phages tested in the host range assay. Four strains are in the CIRM-CFBP, the French
1350 collection for plant-associated bacteria (https://cirm-cfbp.fr/ ; June 2025): USA0007(CFBP
1351 8574), CC1559 (CFBP 8551), MAFF302273PT (CFBP 2339), and CC1583 (CFBP 8564).

1352

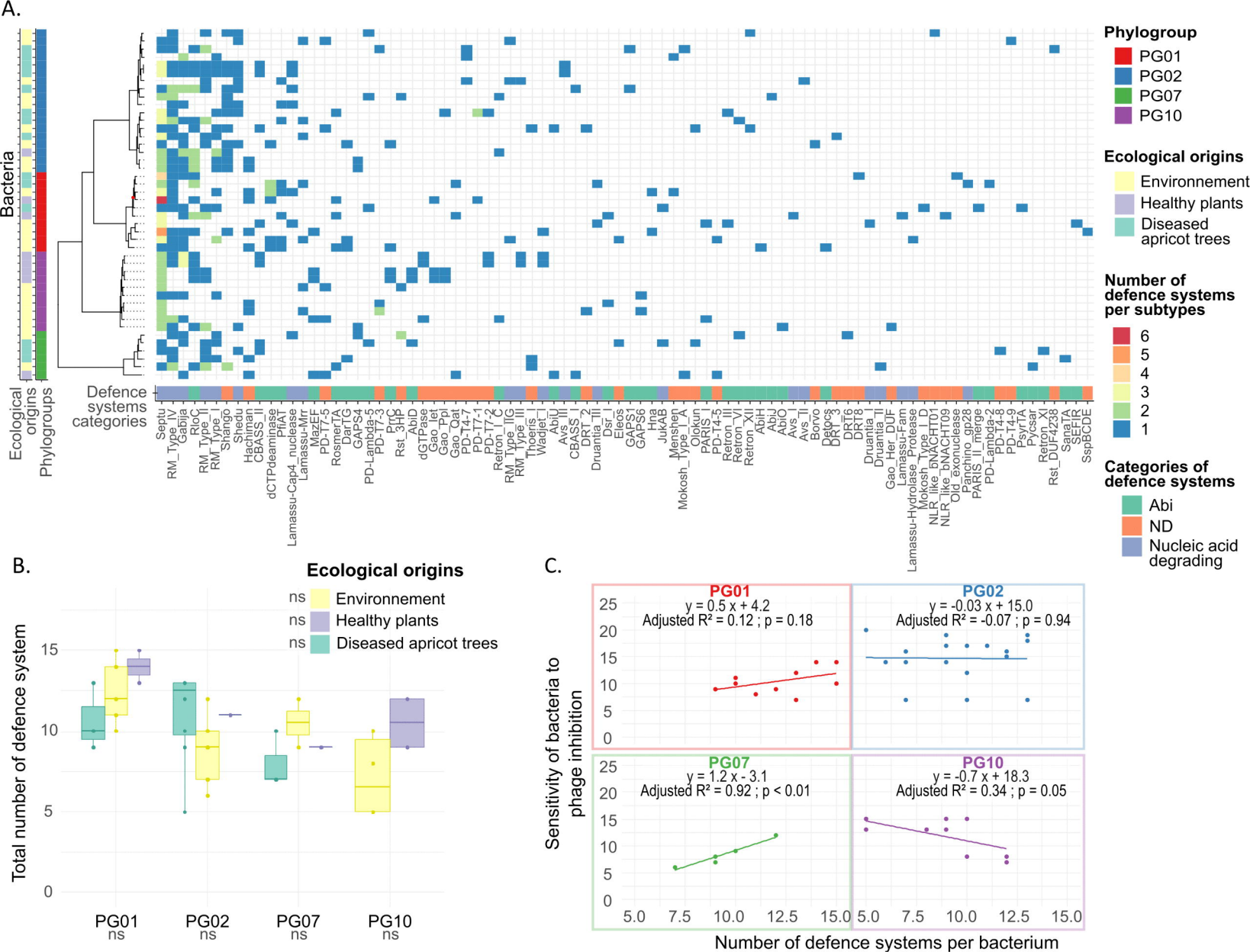
1353 **Data Availability Statement**

1354 The phage–bacterium infection network, based on average values, has been deposited in the
1355 Viral Host Range Database hosted by Institut Pasteur
1356 (https://viralhostrangedb.pasteur.cloud/data-source/287/). The raw optical density (OD)
1357 kinetics data and the processed datasets used for analysis (e.g., AUC, inhibition values,
1358 phage effect probability) have been deposited in Zenodo (DOI: 10.5281/zenodo.17735985).
1359 The 39 bacterial genomes sequenced in this study have been deposited under the BioProject
1360 accession number PRJNA1292126.

1361

1362 **Financial Disclosure Statement**

The work described in our manuscript was founded by the Plant Health and Environment (SPE) Department of the French National Research Institute for Agriculture Food and the Environment (INRAE) (to C-TB) and Avignon University (AU) (to C-TB). CF received PhD funding from both the SPE Department of INRAE and AU. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.


A.
Bacteria
Ecological origins
Phylogroups
Defence systems categories
Phylogroup
PG01
PG02
PG07
PG10
Ecological origins
Environnement
Healthy plants
Diseased apricot trees
Number of defence systems per subtypes
6
5
4
3
2
1
Categories of defence systems
Abi
ND
Nucleic acid degrading
B.
Ecological origins
ns Environnement
ns Healthy plants
ns Diseased apricot trees
Total number of defence system
PG01 ns
PG02 ns
PG07 ns
PG10 ns
C.
PG01
y = 0.5 x + 4.2
Adjusted R² = 0.12 ; p = 0.18
PG02
y = -0.03 x + 15.0
Adjusted R² = -0.07 ; p = 0.94
PG07
y = 1.2 x - 3.1
Adjusted R² = 0.92 ; p < 0.01
PG10
y = -0.7 x + 18.3
Adjusted R² = 0.34 ; p = 0.05
Sensitivity of bacteria to phage inhibition
Number of defence systems per bacterium

A.

Bacteria

Ecological origins

Phylogroups

Prophage genera

Prophage clusters

5 2 35 40 42 10 12 27 37 38 41 8 14 23 26 34 36 1 3 4 6 7 9 11 13 15 17 18 19 20 21 22 24 28 29 30 31 32 33 39 16 25

Phylogroup

- PG01
- PG02
- PG07
- PG10

Ecological origins

- Environnement
- Healthy plants
- Diseased apricot trees

Number of prophages

- 1
- 0

Cluster of prophages

- CL1
- CL2
- CL3
- CL4
- CL5
- CL6

B.

Ecological origins

- ns Environnement
- ns Healthy plants
- ns Diseased apricot trees

Number of prophages

0 2 4 6

PG01 ns PG02 ns PG07 ns PG10 ns

C.

Phylogroup

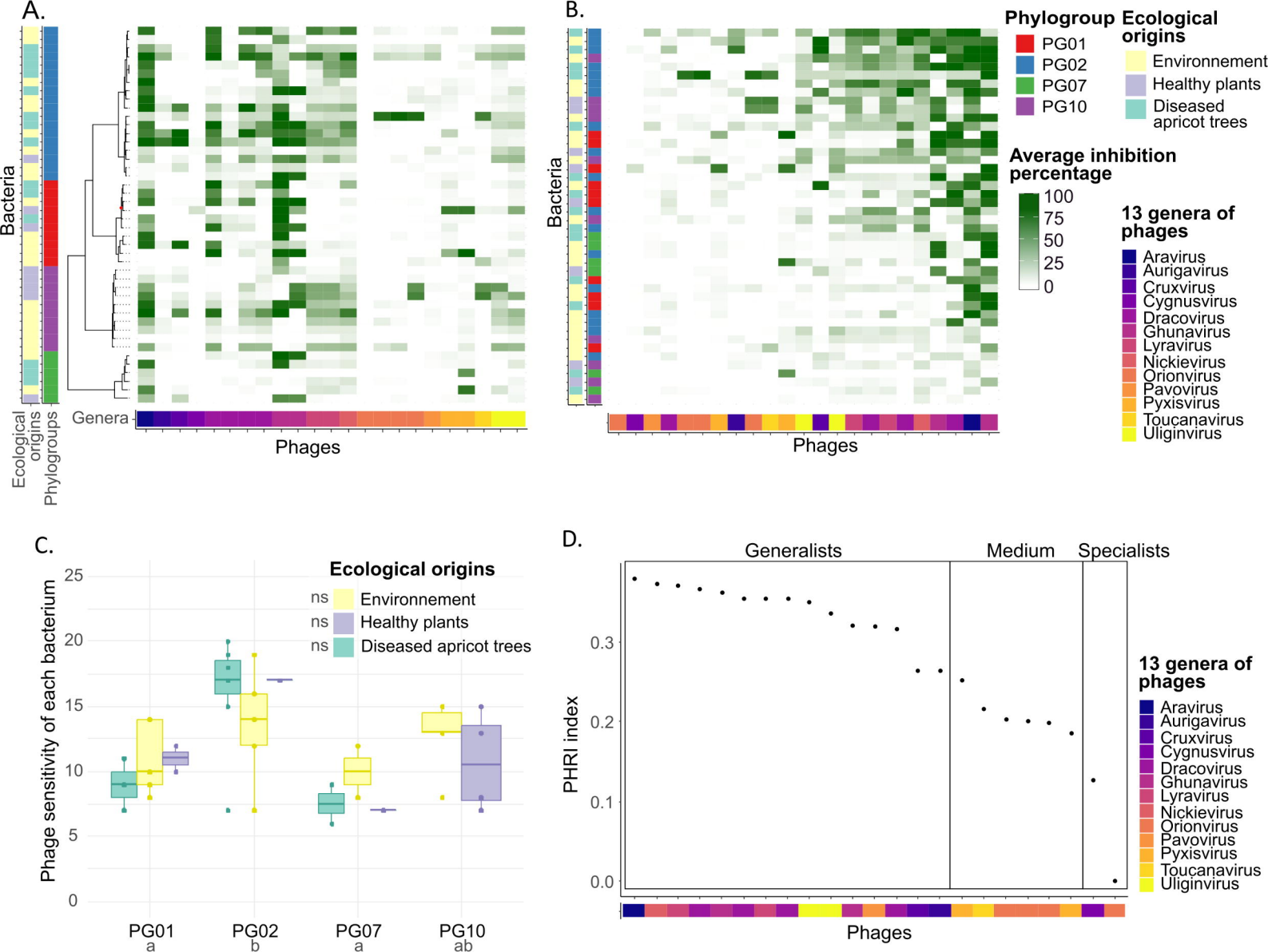
- PG01
- PG02
- PG07
- PG10

Sensitivity of bacteria to phage inhibition

0 5 10 15 20 25

1 2 3 4 5

Number of prophages


A.
Bacteria
Genera
Phages
Ecological origins
Phylogroups
B.
Bacteria
Phages
Phylogroup
PG01
PG02
PG07
PG10
Ecological origins
Environnement
Healthy plants
Diseased apricot trees
Average inhibition percentage
100
75
50
25
0
13 genera of phages
Aravirus
Aurigavirus
Cruxvirus
Cygnusvirus
Dracovirus
Ghunavirus
Lyravirus
Nickievirus
Orionvirus
Pavovirus
Pyxisvirus
Toucanavirus
Uliginvirus
C.
Phage sensitivity of each bacterium
25
20
15
10
5
0
Ecological origins
ns Environnement
ns Healthy plants
ns Diseased apricot trees
PG01
a
PG02
b
PG07
a
PG10
ab
D.
Generalists
Medium
Specialists
PHRI index
0.3
0.2
0.1
0.0
Phages
13 genera of phages
Aravirus
Aurigavirus
Cruxvirus
Cygnusvirus
Dracovirus
Ghunavirus
Lyravirus
Nickievirus
Orionvirus
Pavovirus
Pyxisvirus
Toucanavirus
Uliginvirus