Pseudomonas aeruginosa maintains an inducible array of novel and diverse prophages over lengthy persistence in CF lungs

Ifigeneia Kyrkou (✉ ifigeneia.kyrkou@gmail.com)

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark   https://orcid.org/0000-0002-0100-8245

Jennifer Bartell

Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark

Ana Lechuga

Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Heverlee, Belgium

Cédric Lood

Laboratory of Computational Systems Biology, Department of Microbial and Molecular Systems, KU Leuven, Heverlee, Belgium

Rasmus Lykke Marvig

Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark

Rob Lavigne

Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Heverlee, Belgium

Søren Molin

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

Helle Krogh Johansen (✉ hkj@biosustain.dtu.dk)

Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark

---

Research Article

1 **_Pseudomonas aeruginosa_ maintains an inducible**

2 **array of novel and diverse prophages over lengthy**

3 **persistence in CF lungs**

4 Ifigeneia Kyrkou[1,2,*], Jennifer Bartell[3], Ana Lechuga[4], Cédric Lood[4,5], Rasmus Lykke

5 Marvig[6], Rob Lavigne[4], Søren Molin[2], Helle Krogh Johansen[2,3,7,*]

6
7 [1]Department of Veterinary and Animal Sciences, Food Safety and Zoonosis, University
8 of Copenhagen, Copenhagen, Denmark
9 [2]The Novo Nordisk Foundation Center for Biosustainability, Technical University of
10 Denmark, Kgs. Lyngby, Denmark
11 [3]Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark
12 [4]Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Heverlee,
13 Belgium
14 [5]Laboratory of Computational Systems Biology, Department of Microbial and
15 Molecular Systems, KU Leuven, Heverlee, Belgium
16 [6]Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark
17 [7]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
18 *corresponding authors email address: Ifigeneia.kyrkou@sund.ku.dk,
19 hkj@biosustain.dtu.dk
20

Abstract

22 _Pseudomonas aeruginosa_ is a bacterium with increasing relevance in clinical

23 settings and among the most common bacteria occupying the cystic fibrosis

24 (CF) lung niche. Its ability to colonize and persist in diverse niches is attributed

25 to this bacterium's large accessory genome. In _P. aeruginosa_, prophages

26 represent a common feature of a strain's accessory genome. Hence, we

27 hypothesized that prophages play a role in the bacterium's fitness and

28 persistence in CF. We focused on the CF niche and used longitudinal isolates

29 of patients persistently infected by *P. aeruginosa*. Via *in silico* analysis we

30 predicted intact prophages in the genomes of each longitudinal isolate group

31 and scored their long-term persistence. We then confirmed whether they are

32 inducible and where they reside by induction experiments and lysate

33 sequencing. Lastly, we performed comparative genomics to evaluate

34 prophage diversity and confirm their predicted long-term persistence and level

35 of genomic maintenance. In concurrence with other studies, our findings

36 support that most *P. aeruginosa* harbour prophages, some of which can self-

37 induce. We also found ciprofloxacin, an antibiotic commonly used for *P.*

38 *aeruginosa* treatment in CF, to induce prophages. The induced prophage

39 genomes displayed a high degree of diversity and instances of genomic

40 novelty. Finally, we discovered that all induced prophages persisted long-term

41 with their genomes virtually unchanged, suggesting that they likely assist host

42 persistence. In addition to elucidating the role of prophages in *P. aeruginosa*,

43 we expect our findings to aid in developing novel diagnostics and phage-based

44 therapies for *P. aeruginosa* infections.

45 Keywords: prophages; genomics; persistence; ecology; *P. aeruginosa*;

46 temperate phages; cystic fibrosis; inducible

Introduction

48 Bacteriophages (phages) are viruses of bacteria that often show high infection

49 specificity. While phages with a strictly lytic lifestyle (virulent) rapidly kill their

50 bacterial host, phages with a lysogenic lifestyle (temperate) can also integrate

51 into the bacterial genome, in a form termed prophage. Temperate phage

52 integration incurs metabolic burden to the host bacterium (hereafter host).

53 This burden can be counterbalanced if the prophage (a) increases host fitness

54 via beneficial gene(-s) (aka morons), (b) offers immunity to infection by related

55 phages (aka superinfection exclusion), c) reverts to the lytic lifestyle (*i.e.*

56 induces) in part of the population and kills susceptible competitor strains,

57 and/or d) switches gene expression off or on upon its integration-induction,

58 essentially regulating the host phenotype [1, 2]. Prophages occur frequently

59 in the genomes of many human pathogenic bacteria, including *Acinetobacter*

60 *baumannii*, *Klebsiella pneumoniae*, *Escherichia coli* and *Staphylococcus*

61 *aureus* [3–5]. It is therefore theorized that prophages interfere with infection

62 processes of a pathogen either by controlling its population size or by

63 modifying its genomic content and/or phenotype [6].

64 Bacteria that establish persistent infections in human lungs show a broad

65 diversity of prophages [7]. An example of a bacterium that can persistently

66 infect the human lung is *Pseudomonas aeruginosa*. *P. aeruginosa* is associated

67 with severe morbidity and mortality in Cystic Fibrosis (CF) patients. Its

68 persistent infections cause chronic lung inflammation and can last for > 30

69 years [8], requiring continuous antibiotic treatment, and undermining life

70 quality due to impaired lung function [9]. While *P. aeruginosa* opportunistically

71 colonizes the human body, it also occupies a plethora of ecological niches,

72 from soil and water to plants, insects and animals [10]. The ability of this

73 bacterium to adapt to various niches may be attributed to mobile genetic

74 elements [10], especially considering the bacterium's mediocre capacity for

75 natural transformation [11]. Regardless of niche, *P. aeruginosa* is generally

76 estimated to be lysogenized by one or two temperate phages [12], which are

77 suggested to be important drivers of this bacterium's genomic plasticity [13].

78 In recent years, a renewed interest in temperate phages of *P. aeruginosa* has

79 mainly targeted their role in shaping host virulence, often overlooking other

80 impacts that these may have on host fitness and survival. Related studies

81 identified a number of prophage genes that contribute to host virulence [14].

82 The case of the Liverpool Epidemic Strain (LES) constitutes a notable example

83 of how prophages can influence host fitness; three of its five intact prophages,

84 LESφ2-3-5, were found essential for LES colonization in a rat lung infection

85 model [15]. Furthermore, PAO1 lysogens of prophages LESφ2-3-4 increased

86 competitiveness against non-lysogenic PAO1 in the same rat model [16].

87 Here, we investigated the abundance, activity, diversity and long-term

88 maintenance of intact prophages that reside in the genome of *P. aeruginosa*

89 isolates from the CF lung environment. With this study, we aim not only to

90 enrich current knowledge on *Pseudomonas* phage ecology, evolution and

91 genomics but also to interrogate the potential contributing role of prophages

92 for the persistence of *P. aeruginosa* in this specific niche.

93

94 **Materials and Methods**

Isolate collection and culture conditions

95

96 The study's collection comprises 201 longitudinal isolates from 12 CF patients

97 infected by a *P. aeruginosa* clone type (CT) for a continuous period of at least

98 four years. These isolates were routinely sampled from patients attending the

99 Copenhagen CF Center at the University Hospital, Rigshospitalet, Denmark.

100 Most were previously published [17], whereas 25 are presented here to extend

101 the timespan of this collection. Illumina reads of the published isolates were

102 extracted from https://www.ncbi.nlm.nih.gov/sra, cleaned with Cutadapt v4.4

103 [18] and assembled with SPAdes v3.14.0 [19] choosing BayesHammer

104 correction and careful mode. DNA from the new isolates was extracted with

105 the DNEasy Blood and Tissue kit (Qiagen) and libraries were built with

106 Nextera® XT and sequenced on an Illumina MiSeq (250-bp paired end) or

107 NextSeq (150-bp paired end). Clone typing of all isolates was conducted as

108 part of the Center's routine patient infection history surveillance, as described

109 earlier [17].

110 Additionally, isolate PaLo43 was sampled from a CF patient attending the

111 University Hospital of Leuven, Belgium, and strain PAO1 was purchased from

112 the DSMZ collection. PaLo43 and PAO1 served as indicator strains. For all

113 experiments, bacterial cultures were grown overnight in Lysogeny Broth (LB),

114 Lennox broth and agar (Sigma-Aldrich) at 37 °C and 200 rpm shaking.

115

Prophage predictions and longitudinal frequency counting

116

117 To predict active prophage-like elements, the longitudinal isolates of each

118 patient environment were scanned with Prophage Hunter's server [20],

119 choosing both default and "skip similarity matching" options. Results were

120 parsed with an in-house python script as follows: elements from the same CF

121 lung environment were merged, and "active" and "ambiguous" elements were

122 extracted and grouped under their corresponding "closest phage" hit. Next,

123 each group was listed in descending order of longitudinal frequency. To

124 confirm that elements under the same closest-phage group were closely

125 related, we additionally BLASTn-compared them using default settings. The

126 final curated results were used to count frequencies of the various prophage-

127 like elements to determine those likely significant to host long-term

128 persistence. Specifically, elements were considered significant when they

129 were often encountered in the "persistent" CT, *i.e.* the CT that was

130 longitudinally retraced for at least four years. Pf1-like prophage elements were

131 disregarded as these have already been extensively studied [21, 22]. For

132 subsequent experimentation, we selected one early isolate per CF lung

133 environment, provided its genome harboured all longitudinally frequent

134 prophage-like elements. These 12 early isolates were resequenced with

135 Oxford Nanopore for genome completion (see following section), Nanopore

136 assemblies were rescanned with Prophage Hunter and PHASTER [23] and

137 results were compared.

138

Prophage genome annotations

The genome of each longitudinally frequent prophage-like element was annotated to separate any likely intact prophages from other elements (e.g. pyocins). For that, we combined auto-annotations with manual annotations. Auto-annotations were conducted with RAST's annotation server v2.0 [24] using the RASTtk annotation scheme and GeneMark-Glimmer [25, 26] as gene callers. At this stage, if no or only tail-related structural genes were predicted, the element was deemed to not be an intact prophage and was excluded from further analysis. Predictions for proteins were verified with Blastp [27], HHpred [28] and InterProScan v5.62-94.0 [29], for tRNAs with tRNAscan-SE v2.0 [30] and Aragorn v1.2.41 [31], and a function was assigned when at least two predictions agreed. Prophage genomes were also scanned for genes encoding AMR and virulence factors with CARD (threshold of 80% identity over 40% coverage) [32] and PHIB-BLAST (PHI-Base v4.14; threshold of 50% identity over 50% coverage, e-value $<10^{-3}$) [33], respectively. Putative repressors were identified by gene cluster comparisons with known *Pseudomonas* prophages using Clinker v0.0.27 [34] and running PHMMER searches against the "Reference Proteomes" database [35]. Final annotation maps were designed with SnapGene (www.snapgene.com). Longitudinally non-frequent prophage elements were annotated via our auto-annotation method and deemed likely intact when a major capsid protein gene and other structural genes were predicted.

Nanopore whole-genome sequencing and de novo assembly

As additional corroborations of predictions and to fully resolve the genomic architecture of the strains, whole-genome sequencing of the 12 chosen isolates was expanded using the Oxford Nanopore long-read technology. This was done to prevent overlooking a prophage due to scanning low contiguity and low completeness assemblies [36]. High-molecular-weight gDNA was extracted from overnight cultures with Genomic-tip 100/G (Qiagen) following a published protocol and the Qiagen Genomic DNA Handbook [37]. Before quality control, DNA extracts were mildly sheared by 20x passage through a 25-G needle to encourage homogenization. Libraries were prepared with the SQK-RBK004 kit (Oxford Nanopore Technologies) for rapid barcoding, and sequenced on MinION R9.4.1 flow cells. High-accuracy basecalling was conducted with Guppy v4.2.2 (github.com/nanoporetech) by specifying "--min_score_mask 40" to reduce the number of false positives. Reads under 1,000 bp and scoring below Q10 were removed using SeqKit v0.13.2 [38]. Retained reads were assembled with Flye v2.9 [39] and assemblies were polished through four runs of Racon v1.4.21 [40] to remove random sequencing errors, then Medaka v1.2.0 (github.com/nanoporetech) and Homopolish v0.2.1 [41] to remove systematic Nanopore errors. Polished assembly accuracy and genome completeness were assessed using BUSCO

183 v5.1.3 [42] and CheckM v1.0.18 [43] against the Pseudomonadales database.

184 We evaluated coverage per 1,000 bp intervals with DepthOfCoverage of GATK

185 v 4.1.6.0 [44] and sequenced deeper any genomes with scores <20x. These

186 deep sequencing and post-assembly processing steps were conducted to

187 generate reference long-read assemblies with a final depth ≥65x for all

188 genomes.

189

Prophage induction and DNA extractions

191 For induction experiments, overnight cultures of the 12 isolates were diluted

192 to an optical density (OD) of 0.1 in 9 mL LB and incubated at 37 °C until early

193 exponential phase (hereafter t0), which corresponded to ODs of 0.2 – 0.3. At

194 t0 we harvested 700 uL per diluted culture, after adding either 2.5 µg/mL

195 mitomycin C (mitC) or approximately 0.5x the minimal inhibitory

196 concentration (MIC) of ciprofloxacin (Supplementary Table 1). Samples were

197 immediately centrifuged (17,000x g, 5 min, 25 °C) and supernatants were

198 passed through 0.45-µm cutoff cellulose acetate syringe filters (LABSOLUTE®)

199 and placed on ice until needed. Meanwhile, the diluted cultures were

200 reincubated and sampled again after 30 min, 1h, 2h, 3h, 4h, 19h. Similarly, we

201 sampled diluted cultures in the absence of antibiotics to check for self-induced

202 prophages.

203 Aliquots of the filtered samples were tested for induced prophages via double

204 agar overlay assays [45]. Briefly the overlays were produced using 4 mL of LB

205 broth supplemented with 0.4% w/v agarose (Fisher Scientific) and 0.1 mL

206 overnight of either PAO1 or PaLo43. Each sample was serially diluted tenfold

207 and 3x 10 uL per dilution were spotted against the indicator lawn. Following

208 24-h incubations, the resulting plates were inspected for individual plaques or

209 signs of cell lysis at the position of the spots. All positive samples were sorted

210 out to repeat double agar overlays, except that 0.1 mL of tenfold dilutions

211 were now blended with the overlay.

212 To capture all induced prophages, DNA extractions were performed for both

213 the positive samples and the 19-h-filtered samples of those inductions that

214 yielded no lysis. We extracted 2x 100 uL per sample and otherwise followed a

215 published protocol [46], with few modifications. Briefly, samples were passed

216 through 0.45-μm-cutoff ultrafiltration spin-columns (Millipore), then incubated

217 with 10 U of DNase I at 37 °C for 2h and for the next steps volumes were

218 doubled. The two sample copies were loaded to the same purification column

219 before the wash step and eluted with 25 uL of TE buffer (10 mM Tris–HCl, 0.1

220 mM EDTA, pH 7.5). Samples sequenced to contain high levels of bacterial-DNA-

221 read "noise" were re-extracted after pretreatment with 0.1 volumes

222 chloroform, according to PoT protocol [47].

223

224 ## Induced prophage DNA sequencing and read mapping

225 The gDNA libraries of all induced prophage DNA samples were prepared with

226 the Nextera Flex kit (Illumina) and sequenced with the Illumina MiniSeq using

227 a paired-end approach (2x 150 bp). Next, reads were quality-controlled with

228 FastQC [48] and scanned with Trimmomatic [49] to remove adapter

229 sequences, filter by length (>50 bp), and trim lower-quality regions (options:

230 ILLUMINACLIP:NexteraPE-PE.fa:2:30:10LEADING:3   TRAILING:3

231 SLIDINGWINDOW:4:15 MINLEN:50). Reads were mapped onto the

232 corresponding host genomes using bwa mem mapper [50] with default options

233 and resulting mapping files were visualized with WeeSAM

234 (https://github.com/centre-for-virus-research/weeSAM) and UGENE v42.0 [51].

235 Genomic regions corresponding to induced prophages were hereby localized

236 with high border accuracy. The graph of induced versus uninduced, intact

237 prophages per isolate was generated with GraphPad Prism v.10.0.0.

238

Comparative genomics, phylogenetics and CRISPR-Cas system predictions

241 The reference long-read genome assemblies of the 12 isolates were aligned

242 using Parsnp version 1.5 [52] and a tree was generated with RAxML [53] on

243 SNPs identified from the alignment. The tree was annotated with support

244 values from 500 bootstraps and visualized with iTOL v6.7 [54]. CRISPR-Cas

245 systems of the 12 isolates were predicted and classified with CRISPRCasTyper

246 v1.8.0 [55] using "Circular topology" and, otherwise, default settings. Self-

247 targeting against own intact prophages was investigated by first extracting

248 the list of spacers predicted per isolate genome and adding the protospacer

249 adjacent motif (5'-GG-'3 for I-F predictions and 5'-CAT-3' for I-E predictions)

250 upstream of each spacer sequence. Secondly, the edited spacers were aligned

251 against their own prophage genomes using the Megablast algorithm.

252 Prophage pairwise intergenomic similarities were computed via web-based

253 tools VIRIDIC [56] and ViPTree [57] using default settings. For that, all

254 prophages deemed intact were compared amongst them and against a

255 custom, literature-based database of all *P. aeruginosa* prophages proven to

256 exist as active particles (Pubmed search "Pseudomonas" AND

257 "lysogenic"/"prophage"/"excis?"/"temperate"; Supplementary Table 2).

258 Having sequenced the genome of all induced prophages, we verified their

259 presence in longitudinal isolates using the BLASTn algorithm and extracted

260 their genome sequence from the latest isolate they lysogenized with CLC

261 Genomic Workbench V8.0 (Qiagen). Using EasyFig v.2.2.5 [58], extracted

262 sequences were linearly BLASTn-compared to the corresponding sequenced

263 prophage genome. The phylogenetic tree based on concatenated amino acid

264 sequences of repressor/antirepressor was constructed via NGPhylogeny.fr [59]

265 with "PhyML+SMS/OneClick" for the tree inference.

266

Results and Discussion

268 Induction profiles of identified, intact prophages

269   Our bioinformatics analysis identified a sum of 29 intact prophages in the

270   genome of the 12 *P. aeruginosa* clinical isolates. Of these, 22 intact prophages

271   seemed to be longitudinally frequent when scanning the genomes of isolates

272   from the same CF lung. Prophage Hunter generally outperformed PHASTER in

273   predicting prophage completeness (Supplementary Table 3). Double agar

274   overlay assays were the first tests undertaken to identify free phages resulting

275   from our induction experiments. However, these assays alone would have

276   been insufficient to distinguish whether lawn lysis stems from phage rather

277   than bacteriocin killing. Such level of detail was possible thanks to the

278   sequencing analysis. By mapping lysate reads to their corresponding host

279   genome we could identify induced prophage locations with high border

280   accuracy. Looking back at the results of the double agar overlay assays (Fig.

281   1a), we noted that clearing zones and plaques did always result from prophage

282   induction, as confirmed from the sequencing (Fig. 1b).

283   In total, 15 predicted-as-intact prophages were found to exist as free particles

284   by lysate sequencing. While we did predict genes encoding a major capsid and

285   other structural proteins for the remaining 14 prophages (Supplementary

286   Table 4), no or scarce reads corresponding to them were captured (Fig.1b).

287   Given that lysate DNA was extracted from at least two different time points

288   per isolate and type of induction, it is unlikely that our method missed induced

289   dsDNA prophages. Explanations for why the remaining 14 prophages were not

290   induced can be attributed to various factors. For example, lysogenic to lytic

291   conversion for these prophages may be triggered by conditions different to

13

292   the ones tested [60], or they may belong to a non-inducible class of temperate

293   phages (such as Escherichia virus P2) [61]. Another improbable scenario is

294   that, while these prophages excised, our DNA filtration and Illumina

295   sequencing methods selected against them because of their ssDNA genome

296   [62].

297   Of the induced prophages some self-induced, while others excised due to one

298   or both antibiotics used (mitC or cipro; Fig. 1b). Under the tested *in vitro*

299   conditions, self-induction was common and occurred for almost 50% of the

300   prophages sequenced. This aligns with multiple studies on *P. aeruginosa*

301   clinical isolates from CF patients, such as [63, 64], which reported frequent *in*

302   *vitro* prophage self-induction. Cipro and mitC caused additional prophages to

303   excise from the 12 isolates (Fig.1b). In particular, cipro induction patterns

304   corroborated and occasionally expanded those of mitC (Fig.1a). Ciprofloxacin,

305   a fluoroquinolone commonly prescribed to CF patients including the

306   Copenhagen CF Center patients, was previously shown to trigger high rates of

307   *in vitro* prophage induction in clinical *P. aeruginosa* [65]. Considering these *in*

308   *vitro* observations, we can extrapolate that *in vivo P. aeruginosa*-infected CF

309   lungs often contain high titres of excised prophages, as already showcased for

310   isolate LESB58's prophages [66]. This extrapolation can be further supported

311   by the regularity of prophages in genomes of *P. aeruginosa* CF lung isolates.

312   Our results support findings in other studies that most clinical *P. aeruginosa*

313   are mono- or polylysogenic [63, 64], because except F004 all isolates

314   harboured prophages, with five of twelve being polylysogens harbouring two

315   to eight intact prophages (Fig.1b). It is tempting to hypothesize that in the CF

316   lung environment, lysogeny (especially polylysogeny) contributes to *P.*

317   *aeruginosa* persisting longer than its nonlysogenic or prophage-poor

318   counterparts. This would be due to prophage moron genes and/or

319   superinfection exclusion.

320   In the next sections, we present and discuss our results which support this

321   hypothesis, *i.e.* that intact prophages favour *P. aeruginosa* persistence in the

322   CF lung. More and more studies offer evidence that justifies this assumption.

323   For instance, *P. aeruginosa* prophages DMS3 and pp3 were shown to exclude

324   phages that require type IV pilus as a receptor and to assist in host adaptation

325   by promoting biofilm formation, respectively [67, 68]. In other studies, the

326   acquisition of prophages decreased antibiotic susceptibility and virulence of

327   clinical *P. aeruginosa* and increased biofilm formation [69]. Moreover, Burns *et*

328   *al.* [71] demonstrated that in mixed infections PAO1 polylysogens used phage

329   predation to prevail over their isogenic prophage-free competitors.

330   Intriguingly, monolysogeny here was indeed almost exclusively associated to

331   isolates from monoclonal infections, except for isolate F002 (Fig. 1b). Yet, the

332   CT of F002 (DK12) was reported to occur in multiple patients [72] likely

333   implying its direct acquisition from a polyclonal CF lung environment.

334   To explore possible correlations between polylysogeny of an isolate and a

335   higher permissiveness to phage invasion, we performed *in silico* predictions of

336   CRISPR-Cas systems. In line with others [73], our study revealed that isolate

337   genomes containing functional CRISPR-Cas systems harboured one or zero

338 prophages. (Table 1). Despite the reported strong association between type
339 I-F systems and self-targeting [73, 74], we detected no spacers targeting those
340 native intact prophages. Conclusive answers to the above would necessitate
341 experimental validation of the predicted CRISPR-Cas systems, and of any
342 phage countermeasures.

Isolate	Nr. of Cas Operons	Predicted Cas Subtypes	Nr. of Associated CRISPR Arrays	Nr. of Intact Prophages	Native Prophage (self-) Targeting?
37	1	I-F	2	1	No
135	1	I-E	2	1	No
199	1	I-F	2	1	No
F004	1	I-F	2	0	No
F023	1	I-F	2	1	No
F056	1	I-F	2	1	No
LRJ32	-	orphan	1	7	-

343 Table 1: List of studied isolates predicted to contain functional or orphan CRISPR-
344 Cas systems as compared to the number of intact prophages harboured by each
345 isolate.

346

347 # Genomic features and gene predictions of induced

348 # prophages

349 Fourteen of 29 intact prophage genomes could not be resolved with high-
350 border accuracy as they remained uninduced. For this reason, this section
351 focuses on the 15 induced prophages. In Table 2, we present the names and

16

352 key genomic characteristics of these 15 prophages discovered here. Marin,

353 Shamal, Haboob and Sirocco genomes were found to be terminally redundant

354 with terminal repeats of 61, 64, 60 and 46 bp, respectively (Supplementary

355 Table 4). Marin, Shamal and Haboob were also found to encode three, one and

356 four tRNAs, respectively, with Marin and Haboob displaying an almost identical

357 tRNA gene arrangement, despite their overall nucleotide dissimilarity (94%

358 identity over 46% query cover by BLASTn). This finding aligns with the concept

359 that prophages can have maximum four tRNAs and that phage genomes of

360 length longer than the average (here 42,061 bp) are more prone to harbouring

361 tRNA genes [75]. Notably, these same genomes had the highest %GC content

362 disparity as compared to *P. aeruginosa*'s genome, which averages 66.6% [76].

363 In general, it is hypothesized that phages carry selected tRNA genes either to

364 compensate for any codon usage differences with their host or to overcome

365 host tRNA-depleting anti-phage strategies [75–77]. Our observation regarding

366 the %GC content disparity would be better explained by the former theory.

367 Collectively considered, tRNA presence and %GC content disparity could imply

368 that Marin, Shamal and Haboob encountered various hosts throughout their

369 evolutionary history.

370 Functional annotations could averagely be assigned to one-third of the

371 predicted Open Reading Frames (ORFs; Table 2). Annotated ORFs were

372 typically associated with morphogenesis (portal, capsid and tail genes), DNA

373 packaging (terminase subunits) and lysogeny-lysis (repressor/antirepressor,

374 integrase, transposase, endolysin, holin) modules due to these proteins being

375 more conserved in phages. Transposition was predicted for prophages Bise,

376 Etesian, Gregale, Meltemi, Marin and Rashabar and agrees with early reports

377 on the wide distribution of transposable phages in *P. aeruginosa* [78]. By

378 combining functional annotations and synteny maps to related phages, we

379 predicted that the lysis-lysogeny switch was regulated by repressor CI and

380 either Cro- or Ner-like antirepressors for most prophages (Table 2).

381 Nevertheless, a combined repressor/antirepressor tree (Fig. 2b) failed to

382 cluster prophages according to their induction patterns, suggesting that

383 repressor/antirepressor genes are not alone in governing lysogeny-lysis

384 decisions.

385 Besides the repressors/antirepressors, we were particularly interested in

386 identifying morons. Moron proteins are recognised as important contributors

387 to the fitness of *P. aeruginosa* and other bacteria through, for instance,

388 antimicrobial resistance or virulence regulation [79]. Depending on their role,

389 some could hence support persistence of the prophage host in the CF lung.

390 Our screening against the CARD database located a single gene of prophage

391 Ostro (peg.66) encoding a protein with 92.5% identity to a bicyclomycin

392 resistance protein (NCBI: ALV80601.1). Our search against PHI-Base located a

393 considerable number of proteins with similarity to moron virulence factors, and

394 which were encoded by the genomes of Ostro, Alize, Haboob, Riah, Solano.

395 Ostro harboured two candidate virulence genes, pegs.69-70. Peg.69 encoded

396 a protein 100% identical to CcoN4 (PHI:7765), an orphan cbb3-type

397 cytochrome-c oxidase subunit, which was shown promote pathogenicity and

398   biofilm growth for strain PA14 in a *Caenorhabditis elegans* infection model

399   [80]. The peg.70-encoded protein had 50% identity to the GntR family protein

400   YdcR (PHI:7225), whose gene knock-outs led to dampened pathogenicity of

401   *Salmonella enterica* in a murine infection model, while the protein itself

402   seemed to directly regulate virulence factor SrfN [81]. Taken together the

403   results of CARD and PHI-Base strongly suggest that Ostro greatly assists the

404   fitness of its host, isolate 382. Solano and Alize harboured the homologous

405   candidate virulence genes peg.55 and peg.10, respectively, which encoded a

406   protein 50% identical to the Arc family DNA-binding protein AmrZ of *P.*

407   *syringae* (PHI:5372). AmrZ acts as a cellulose biosynthesis repressor, and

408   *amrZ* mutants displayed a hypovirulent phenotype implying that AmrZ may

409   regulate up to several virulence factors [82], like also proven for *P. aeruginosa*

410   [83]. Gene peg.67 of Haboob encoded a protein with 34% identity to AlgR, a

411   LytTR DNA-binding domain-containing protein of *P. aeruginosa* (PHI:8158).

412   AlgR activates the biosynthetic pathway of alginate, which leads to mucoidity

413   and immunoevasion and eventually persistence of *P. aeruginosa* in CF [84].

414   Moreover, AlgR was found to contribute to virulence via the regulation of type

415   IV pili, which are essential for *P. aeruginosa* mammalian cell colonization,

416   competence and pathogenesis [85]. Lastly, Riah's genome contained gene

417   peg.58 that encoded a protein with 43% identity to DprA from *Streptococcus*

418   *pneumoniae* (PHI:11059). Pneumococcal DprA's central role in virulence was

419   demonstrated through targeted deletion, where its deficiency led to

420 attenuated virulence in a bacteremia mouse model infection [86]. However, a

421 connection of DprA to *P. aeruginosa* virulence is yet to be established.

Prophage	Genome Size (bp)	Lysogenized Isolate	ORFs with Assigned Function	%GC Content	tRNA Genes	Repressor, Antirepressor
Rashabar	38,722	37	16/55	63.3		unclear, Ner-like
Riah	44,821	20	24/59	62.1		CI-like, Cro-like
Alize	38,819	F002	27/65	61.9		CI-like, Cro-like
Haboob	50,033	382	26/85	59.7	$tRNA^{Met}$ $tRNA^{Gly}$ $tRNA^{Asn}$ $tRNA^{Thr}$	CI-like, Cro-like
Marin	51,941	F002	29/84	59.9	$tRNA^{Gly}$ $tRNA^{Asn}$ $tRNA^{Thr}$	CI-like, Cro-like
Shamal	44,124	F002	19/76	57.3	$tRNA^{Ser}$	CI-like, Cro-like
Gregale	36,566	F002	15/53	64.2		CI-like, Ner-like
Solano	40,334	LRJ32	24/62	62		CI-like, Cro-like
Meltemi	36,609	LRJ32	15/54	64.5		CI-like, Ner-like
Lodos	44,613	135	24/55	63.9		CI-like, unclear
Bise	38,458	199	30/55	66		CI-like, Ner-like
Turba	42,092	F038	22/51	63.8		CI-like, unclear
Ostro	48,846	382	27/70	62.4		CI-like, Cro-like

Etesian	37,990	382	20/58	64.3		CI-like, Ner-like
Sirocco	36,945	188	35/52	62.5		CI-like, Cox-like

Table 2: Names and key genomic characteristics of the induced *P. aeruginosa* prophages. The ends of all genomes were defined based on their integration in the host chromosome.

The genomes of identified intact prophages are highly diverse

To explore how diverse the 29 prophages were as compared to an external - yet related- prophage population, we built a database of previously published dsDNA genomes belonging to 26 *P. aeruginosa* prophages proven to exist as active particles (Supplementary Table 2). Intergenomic similarity scoring for all 55 genomes was performed with VIRIDIC following the current International Committee on Taxonomy of Viruses thresholds for species (<95%) and genus (~70%) demarcation [56]. As anticipated by this threshold's stringency, only two species clusters were formed, grouping prophage Etesian together with database prophage JBD5. JDB5 possesses two genes encoding anti-CRISPR (Acr) proteins, AcrIF3 (NCBI:YP_007392740.1) and AcrIE1 (NCBI: YP_007392738.1) encoded by genes peg.25-6 in Etesian. These proteins should enable Etesian's spread to other *P. aeruginosa* that, unlike studied host 382, carry active type I-F and I-E CRISPR-Cas systems [87]. Along with Gregale and Meltemi, Etesian was also part of the genus cluster represented by database prophage D3112. Other formed genera clusters were represented

443   by database prophages JBD25 and JBD67. JBD25 cluster included the intact

444   but uninduced prophages Caju and Notus, while JBD67 cluster contained only

445   Rashabar. Lastly, phiCTX's cluster included prophage Sirocco and intact but

446   uninduced prophage Harmattan, but neither of them encoded a cytotoxin

447   related to phiCTX's (NCBI: NP_490598.1). Sirocco's clustering with prophages

448   of the P2 non-inducible class matches our observations on the inconsistency

449   of Sirocco's induction. Like with phiCTX, eventual lysis may occur by mutations

450   in a lysogeny-associated gene [88]. For even broader comparisons, this

451   study's prophage genomes were BLASTn-compared to the viral nucleotide

452   collection database (Supplementary Table 5).

453   The remaining intact prophages formed clusters with reduced or no similarity

454   to database prophages. Overall, our prophages were distributed across the

455   VIRIDIC heatmap. To assess the level of diversity among these prophages, we

456   compared results yielded by VIRIDIC with a ViPTree-built phylogeny. ViPtree

457   constructs trees based on tBlastX similarities, hence increasing resolution of

458   genomic relations. The derived tree (Fig. 3b) validated VIRIDIC and BLASTn

459   results, highlighting the clade of active prophages Lodos-Turba-Riah and the

460   monotypic taxon Bise for their very low similarity to any known prophages.

461   Other active prophages with reduced similarity to known prophages were

462   Haboob, Marin, Shamal, Alize, Ostro, and Solano. Notably, phylogenetically

463   related prophages (Fig. 3b) were not restricted to phylogenetically related

464   hosts (Fig. 3a). This, combined with the 29 prophages shared ecological niche,

465 corroborates prior findings on the remarkable versatility and diversity of *P.*

466 *aeruginosa* prophages [90].

467

The genomes of longitudinally frequent prophages remain

virtually unchanged over long evolutionary times

470 Each of the 12 *P. aeruginosa* isolates that harbour the 29 intact prophages

471 represents a patient persistently infected by *P. aeruginosa*. To gather more

472 evidence on whether any of these prophages favours host fitness and

473 persistence, we gauged how frequently they appear longitudinally. Per patient

474 environment, sequenced longitudinal isolates spanned a period of at least four

475 years (Fig. 4). All same-patient longitudinal isolates were BLASTn-compared to

476 the complete genomic sequence of each induced prophage to confirm earlier

477 bioinformatics-determined frequencies. Here we focus on the 15 intact and

478 induced prophages, as complete genomes of the 14 uninduced prophages

479 could not be verified with lysate sequencing. This approach is taken, because

480 sole reliance upon low-contiguity Illumina-based assemblies for prophage

481 discovery could lead to false-negative low prophage frequencies.

482 We found all 15 induced prophages to be present in at least 80% of the dates

483 when a persistent CT was isolated, including the most recent date. Hence, all

484 15 induced prophages were deemed longitudinally frequent, or else

485 persistent. Figure 4 presents the CT and longitudinal isolates scanned per

486 patient. Notably, we observed minimal alterations in all induced prophages'

487 genomes throughout the extensive evolutionary timeframe of four to nine

488 years that we considered. Figure 5 displays two cases of this genomic

489 conservation; a monoclonal infection exemplified by the synteny map of

490 prophage Alize and a polyclonal infection exemplified by that of Meltemi. The

491 genome of Alize matched by 93% overall nucleotide similarity a region

492 detected eight years later in isolate 32V99 of the same CT (Fig. 5a). Similarly,

493 the genome of Meltemi matched by 82% overall nucleotide similarity a region

494 detected nine years later in isolate 23V71 of the persistent CT DK67 (Fig. 5b).

495 Remaining synteny maps are available in Supplementary Figure 1.

496 Meltemi and its cohabiting prophage Solano constituted the only cases of a

497 persistent prophage likely moving from a transient (DK18) to the persistent CT

498 (DK67). We wanted to validate this hypothesis, versus the possibility that

499 Meltemi and Solano were inherent features of DK67's backbone. Indeed, the

500 two phages were not found in 36V42, an isolate of DK67 which only transiently

501 infected a patient. We extended this strategy and searched for additional CTs

502 that persisted in one patient but transiently infected another, and found

503 examples for DK12 and DK26. DK12 persisted in PID76609 and was transient

504 for patient PID08309 (Fig. 4). We screened the genomes of isolates F042, F043

505 and 93 from transient DK12 against the genomes of persistent prophages

506 Alize, Marin, Shamal, Gregale of isolate F002 (DK12). Again, we found no

507 homologous sequences, which indicated that these prophages do not belong

508 to DK12's backbone. Another CT, DK26, was persistent in patients PID42824

509 and PID08309, which were represented by isolates 382 and 188. We found

510 none of the four induced and two uninduced prophages from these isolates to

511 be shared except in one case; Sirocco had 76% overall nucleotide similarity to

512 a region in the genome of 382. The latter corresponded to the predicted as

513 intact but uninduced prophage Khazri (Fig. 3b), which was found to be a

514 longitudinally frequent element of 382 (Supplementary Table 4). To further

515 clarify the case of Sirocco, we looked at the genomes of isolates F011, F044

516 and 24V78. F011 -sampled in 2002- and 24V78 -sampled in 2021- originated

517 from the same CF patient and a region of 99% overall nucleotide similarity to

518 Sirocco was traced in both. The infection by DK26 in this patient seemed to

519 prevail for almost 19 years, but we did not have sequences of intermediate

520 isolates to validate its persistence. However, F044 from a CF patient

521 transiently infected by DK26, also harboured a region with 99% overall

522 nucleotide similarity to Sirocco. This identified ubiquity of Sirocco-type

523 prophages in DK26 isolates combined with the general non-inducibility of this

524 prophage lineage (Fig. 3b) suggests that Sirocco is an inherent part of DK26's

525 backbone.

526 Active prophages are bacterial parasites that burden the host in a dual

527 manner; they are metabolically costly and can revert to the lytic cycle

528 resulting in host death. Thus, the trend would be that they undergo mutational

529 degradation along with other parts of the bacterial genome and become

530 grounded over time [89], unless they confer ecological and evolutionary

531 benefits to their host [90]. Conversely, a prophage carrying a moron gene that

532 provides fitness benefits unmatched by the host genome is predicted to

533 persist within the host, even amidst conflicting selection pressures [91].

534 Stability in the host environment is estimated to further prolong maintenance

535 of such prophages [91]. Indeed, stability is a trait that characterises

536 *Pseudomonas*-dominated CF lungs [92], and likely creates a positive feedback

537 loop between beneficial prophage maintenance and host persistence.

538 Considering all these, we expect that, excluding Sirocco, all other induced

539 prophages enhance their hosts' fitness and persistence in the CF lung.

540 Suggestive to this role are their high genomic conservation within the

541 persistent CT and their long-term longitudinal persistence (in contrast to

542 transient examples of the same CT). The facil inducibility of all but Sirocco,

543 albeit supportive to host survival via competing CT exclusion, may also be

544 viewed as evidence of prophage selfishness and opportunism. We propose

545 that reported behaviours of intact prophages, whether as selfish genetic

546 elements pursuing their own survival and proliferation or as nearly

547 domesticated elements contributing to host adaptation and fitness, should not

548 be regarded as mutually exclusive. Instead, they should be recognized as

549 interconnected aspects of the nature of intact prophages, used to aid their

550 survival. This idea is supported by Quistad *et al*. [93], who reason that genetic

551 elements (e.g. intact prophages) with genes that promote host survival should

552 be called "mobile" rather than "selfish".

553

Conclusions

P. aeruginosa is one of the most common opportunistic human pathogens and can establish difficult-to-eradicate infections. Prophages are frequent components of this bacterium's genome and occasionally enhance its virulence. However, the contributing role of prophages in the evolution and fitness of the ubiquitous *P. aeruginosa* in its diverse niches has been explored less. Here we addressed the CF lung, an environment where *P. aeruginosa* communities are common, show increasing prevalence with patient age and can often persist despite antibiotic treatments. Among others, we found that (poly)lysogeny is widespread for CF *P. aeruginosa* isolates and that genomes rich in intact prophages are predicted less likely to contain active CRISPR-Cas systems. We also described 29 intact prophages with highly diverse genomes. The genomic diversity unlocked by the study is anticipated to aid towards a deeper understanding of the versatility and ecology of *Pseudomonas* prophages and of their interactions with their host. We observed that some of the studied intact prophages induced due to ciprofloxacin, an antibiotic often used for CF patient treatment. In addition to antibiotic-triggered induction, self-induction was observed. Such inducibility patterns could translate into a high availability of free prophages in the CF lung with various implications for the host CT competitiveness and its social interactions. Last but not least, we identified that 14 out of 15 of the inducible prophages displayed a high genomic conservation and long-term longitudinal frequency within the persistent CT. These two features, along with their facil inducibility, directly

577 point to their positive selection by the host and suggest that these prophages

578 enhance host fitness and persistence in the CF lung. In the cases of prophages

579 Ostro, Alize, Haboob, Riah, Solano, we bioinformatically predicted genes

580 encoding verified AMR and virulence factors, which offers additional support

581 to these prophages' essentiality for the host. Overall, we expect our findings

582 on the genomic diversity and persistence of certain prophages to assist in

583 developing novel antibacterial strategies and diagnostic solutions for *P.*

584 *aeruginosa* infections.

585

586 ## Data Availability Statement

587 Genomes of induced prophages and bacteria which are first presented in this

588 study have been uploaded to GenBank and their accession numbers can be

589 found in Supplementary Tables 4 and 6. The genomes and annotation files of

590 the 14 intact but uninduced prophages are available through Zenodo as

591 indicated in Supplementary Table 4.

592

593 ## Supplementary information

594 The supplementary files are available in the online version of this article.

595

Conflict of Interest Statement

None declared.

Ethics Declarations

The Scientific Ethics Committee at the Capital Region of Denmark (Region Hovedstaden) approved the use of featured CF isolates (registration number H-21078844) and related information regarding dates and genomic data. Permission to access the isolate biobank and patient isolate records was also given by the management of the Department of Clinical Microbiology at Rigshospitalet. All analyses were performed and is presented following pseudonymization. The authors declare no competing interests.

Acknowledgements

Contributions

621 Conceptualization, IK, SM; Methodology; all authors; Formal analysis, IK, JB,

622 AL, CL, RLM; Investigation, IK, JB, AL, CL; Validation, IK; Writing-original draft

623 preparation, IK; Review and editing, all authors; Visualization, IK, AL, RLM;

624 Project administration, IK, SM; Funding acquisition, IK, RL, HKJ.

625

References

627 1.   Argov T, Sapir SR, Pasechnek A, Azulay G, Stadnyuk O, Rabinovich L, et al.
628 Coordination of cohabiting phage elements supports bacteria–phage cooperation.
629 *Nature communications* 2019; **10**: 5288.

630 2.   Fortier L-C. The contribution of bacteriophages to the biology and virulence of
631 pathogenic clostridia. *Advances in Applied Microbiology*. 2017. Elsevier, pp 169–
632 200.

633 3.   Loh B, Chen J, Manohar P, Yu Y, Hua X, Leptihn S. A biological inventory of
634 prophages in A. baumannii genomes reveal distinct distributions in classes, length,
635 and genomic positions. *Frontiers in microbiology* 2020; **11**: 579802.

636 4.   de Sousa JA, Buffet A, Haudiquet M, Rocha EP, Rendueles O. Modular
637 prophage interactions driven by capsule serotype select for capsule loss under
638 phage predation. *The ISME Journal* 2020; **14**: 2980–2996.

639 5.   Ingmer H, Gerlach D, Wolz C. Temperate Phages of *Staphylococcus aureus*.
640 *Microbiol Spectr* 2019; **7**: 7.5.1.

641 6.   Lawrence D, Baldridge MT, Handley SA. Phages and human health: more than
642 idle hitchhikers. *Viruses* 2019; **11**: 587.

643   7.   Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al.
644   Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis
645   and non-cystic fibrosis individuals. *PloS one* 2009; **4**: e7370.

646   8.   Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Høiby N, et al.
647   Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary
648   perspective. *Nature Reviews Microbiology* 2012; **10**: 841–851.

649   9.   Rajan S, Saiman L. Pulmonary infections in patients with cystic fibrosis.
650   *Seminars in respiratory infections*. 2002. pp 47–56.

651   10.   Kung VL, Ozer EA, Hauser AR. The Accessory Genome of *Pseudomonas*
652   *aeruginosa*. *Microbiol Mol Biol Rev* 2010; **74**: 621–641.

653   11.   Nolan LM, Turnbull L, Katrib M, Osvath SR, Losa D, Lazenby JJ, et al.
654   Pseudomonas aeruginosa is capable of natural transformation in biofilms.
655   *Microbiology* 2020; **166**: 995.

656   12.   Johnson G, Banerjee S, Putonti C. Diversity of Pseudomonas aeruginosa
657   Temperate Phages. *mSphere* 2022; **7**: e01015-21.

658   13.   Shen K, Sayeed S, Antalis P, Gladitz J, Ahmed A, Dice B, et al. Extensive
659   Genomic Plasticity in *Pseudomonas aeruginosa* Revealed by Identification and
660   Distribution Studies of Novel Genes among Clinical Isolates. *Infect Immun* 2006; **74**:
661   5272–5283.

662   14.   Schroven K, Aertsen A, Lavigne R. Bacteriophages as drivers of bacterial
663   virulence and their potential for biotechnological exploitation. *FEMS microbiology*
664   *reviews* 2021; **45**: fuaa041.

665   15.   Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C,
666   Sanschagrin F, et al. Newly introduced genomic prophage islands are critical
667   determinants of in vivo competitiveness in the Liverpool Epidemic Strain of
668   Pseudomonas aeruginosa. *Genome research* 2009; **19**: 12–23.

669   16.   Davies EV, James CE, Kukavica-Ibrulj I, Levesque RC, Brockhurst MA,
670   Winstanley C. Temperate phages enhance pathogen fitness in chronic lung
671   infection. *The ISME journal* 2016; **10**: 2553–2555.

672   17.   Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and
673   adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nature*
674   *genetics* 2015; **47**: 57–64.

675   18.   Martin M. Cutadapt removes adapter sequences from high-throughput
676   sequencing reads. *EMBnet journal* 2011; **17**: 10–12.

677   19.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
678   SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell
679   Sequencing. *Journal of Computational Biology* 2012; **19**: 455–477.

680   20.   Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, et al. Prophage Hunter:
681   an integrative hunting tool for active prophages. *Nucleic Acids Research* 2019; **47**:
682   W74–W80.

683   21.   Knezevic P, Voet M, Lavigne R. Prevalence of Pf1-like (pro) phage genetic
684   elements among Pseudomonas aeruginosa isolates. *Virology* 2015; **483**: 64–71.

685   22.   Burgener EB, Sweere JM, Bach MS, Secor PR, Haddock N, Jennings LK, et al.
686   Filamentous bacteriophages are associated with chronic *Pseudomonas* lung
687   infections and antibiotic resistance in cystic fibrosis. *Sci Transl Med* 2019; **11**:
688   eaau9748.

689   23.   Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better,
690   faster version of the PHAST phage search tool. *Nucleic acids research* 2016; **44**:
691   W16–W21.

692   24.   Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST
693   Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 2008; **9**:
694   75.

695   25.   Besemer J, Borodovsky M. GeneMark: web software for gene finding in
696   prokaryotes, eukaryotes and viruses. *Nucleic acids research* 2005; **33**: W451–W454.

697   26.   Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes
698   and endosymbiont DNA with Glimmer. *Bioinformatics* 2007; **23**: 673–679.

699   27.   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
700   search tool. *Journal of molecular biology* 1990; **215**: 403–410.

701   28.   Gabler F, Nam S, Till S, Mirdita M, Steinegger M, Söding J, et al. Protein
702   Sequence Analysis Using the MPI Bioinformatics Toolkit. *CP in Bioinformatics* 2020;
703   **72**: e108.

704   29.   Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5:
705   genome-scale protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.

706   30.   Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic
707   Sequences. In: Kollmar M (ed). *Gene Prediction*. 2019. Springer New York, New York,
708   NY, pp 1–14.

709   31.   Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA
710   genes in nucleotide sequences. *Nucleic acids research* 2004; **32**: 11–16.

711   32.   Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al.
712   CARD 2023: expanded curation, support for machine learning, and resistome
713   prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic acids*
714   *research* 2023; **51**: D690–D699.

715   33.   Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-
716   base in 2022: A multi-species phenotype database for Pathogen–Host Interactions.
717   *Nucleic Acids Research* 2022; **50**: D837–D847.

718   34.   Gilchrist CL, Chooi Y-H. Clinker & clustermap. js: Automatic generation of
719   gene cluster comparison figures. *Bioinformatics* 2021; **37**: 2473–2475.

720   35.   Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server:
721   2018 update. *Nucleic acids research* 2018; **46**: W200–W204.

722   36.   Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore
723   sequencing and assembly of a human genome with ultra-long reads. *Nature*
724   *biotechnology* 2018; **36**: 338–345.

725   37.   Alvarez-Arevalo M, Sterndorff EB, Faurdal D, Jørgensen TS, Mourched A-S,
726   Vuksanovic O, et al. Extraction and Oxford Nanopore sequencing of genomic DNA
727   from filamentous Actinobacteria. *STAR protocols* 2023; **4**: 101955.

728   38.   Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for
729   FASTA/Q file manipulation. *PloS one* 2016; **11**: e0163962.

730   39.   Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads
731   using repeat graphs. *Nature biotechnology* 2019; **37**: 540–546.

732   40.   Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome
733   assembly from long uncorrected reads. *Genome research* 2017; **27**: 737–746.

734   41.   Huang Y-T, Liu P-Y, Shih P-W. Homopolish: a method for the removal of
735   systematic errors in nanopore sequencing by homologous polishing. *Genome Biol*
736   2021; **22**: 95.

737   42.   Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:
738   assessing genome assembly and annotation completeness with single-copy
739   orthologs. *Bioinformatics* 2015; **31**: 3210–3212.

740   43.   Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM:
741   assessing the quality of microbial genomes recovered from isolates, single cells,
742   and metagenomes. *Genome research* 2015; **25**: 1043–1055.

743   44.   Van der Auwera GA, O'Connor BD. Genomics in the cloud: using Docker,
744   GATK, and WDL in Terra. 2020. O'Reilly Media.

745   45.   Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of
746   Bacteriophages by Double Agar Overlay Plaque Assay. In: Clokie MRJ, Kropinski AM
747   (eds). *Bacteriophages*. 2009. Humana Press, Totowa, NJ, pp 69–76.

748   46.   Moineau S, Pandian S, Klaenhammer TR. Evolution of a Lytic Bacteriophage
749   via DNA Acquisition from the *Lactococcus lactis* Chromosome. *Appl Environ*
750   *Microbiol* 1994; **60**: 1832–1841.

751   47.   Bonilla N, Rojas MI, Cruz GNF, Hung S-H, Rohwer F, Barr JJ. Phage on tap–a
752   quick and efficient protocol for the preparation of bacteriophage laboratory stocks.
753   *PeerJ* 2016; **4**: e2261.

754   48.   Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and
755   quality control. *F1000Research* 2018; **7**.

756   49.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
757   sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

758   50.   Li H. Aligning sequence reads, clone sequences and assembly contigs with
759   BWA-MEM. 2013. arXiv.

760   51.   Okonechnikov K, Golosova O, Fursov M, Team U. Unipro UGENE: a unified
761   bioinformatics toolkit. *Bioinformatics* 2012; **28**: 1166–1167.

762   52.   Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid
763   core-genome alignment and visualization of thousands of intraspecific microbial
764   genomes. *Genome Biol* 2014; **15**: 524.

765   53.   Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-
766   analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.

767   54.   Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for
768   phylogenetic tree display and annotation. *Bioinformatics* 2007; **23**: 127–128.

769   55.   Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ.
770   CRISPRCasTyper: Automated Identification, Annotation, and Classification of
771   CRISPR-Cas Loci. *The CRISPR Journal* 2020; **3**: 462–469.

772   56.   Moraru C, Varsani A, Kropinski AM. VIRIDIC—A novel tool to calculate the
773   intergenomic similarities of prokaryote-infecting viruses. *Viruses* 2020; **12**: 1268.

774   57.   Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the
775   viral proteomic tree server. *Bioinformatics* 2017; **33**: 2379–2380.

776   58.   Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer.
777   *Bioinformatics* 2011; **27**: 1009–1010.

778   59.   Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia
779   S, et al. NGPhylogeny. fr: new generation phylogenetic services for non-specialists.
780   *Nucleic acids research* 2019; **47**: W260–W265.

781   60.   Nanda AM, Thormann K, Frunzke J. Impact of Spontaneous Prophage
782   Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. *J
783   Bacteriol* 2015; **197**: 410–419.

784   61.   Bertani G. STUDIES ON LYSOGENESIS I: The Mode of Phage Liberation by
785   Lysogenic Escherichia coli. *J Bacteriol* 1951; **62**: 293–300.

786   62.   Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like
787   particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 2015;
788   **16**: 7.

789   63.   Miller RV, Rubero VJ. Mucoid conversion by phages of Pseudomonas
790   aeruginosa strains from patients with cystic fibrosis. *J Clin Microbiol* 1984; **19**: 717–
791   719.

792   64.   Essoh C, Blouin Y, Loukou G, Cablanmian A, Lathro S, Kutter E, et al. The
793   susceptibility of Pseudomonas aeruginosa strains from cystic fibrosis patients to
794   bacteriophages. *PLoS One* 2013; **8**: e60575.

795   65.   Fothergill JL, Mowat E, Walshaw MJ, Ledson MJ, James CE, Winstanley C. Effect
796   of Antibiotic Treatment on Bacteriophage Production by a Cystic Fibrosis Epidemic
797   Strain of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 2011; **55**: 426–
798   428.

799   66.   James CE, Davies EV, Fothergill JL, Walshaw MJ, Beale CM, Brockhurst MA, et
800   al. Lytic activity by temperate phages of Pseudomonas aeruginosa in long-term
801   cystic fibrosis chronic lung infections. *The ISME journal* 2015; **9**: 1391–1398.

802   67.   Shah M, Taylor VL, Bona D, Tsao Y, Stanley SY, Pimentel-Elardo SM, et al. A
803   phage-encoded anti-activator inhibits quorum sensing in Pseudomonas aeruginosa.
804   *Molecular Cell* 2021; **81**: 571–583.

805   68.   Li G, Lu S, Shen M, Le S, Shen W, Tan Y, et al. Characterization and interstrain
806   transfer of prophage pp3 of Pseudomonas aeruginosa. *Plos one* 2017; **12**:
807   e0174429.

808   69.   Tariq MA, Everest FLC, Cowley LA, Wright R, Holt GS, Ingram H, et al.
809   Temperate Bacteriophages from Chronic Pseudomonas aeruginosa Lung Infections
810   Show Disease-Specific Changes in Host Range and Modulate Antimicrobial
811   Susceptibility. *mSystems* 2019; **4**: e00191-18.

812   70.   Plahe G. The effect of active prophage carriage on the virulence of
813   Pseudomonas aeruginosa. 2022. PhD Thesis, University of Salford (United Kingdom).

814   71.   Burns N, James CE, Harrison E. Polylysogeny magnifies competitiveness of a
815   bacterial pathogen *in vivo*. *Evolutionary Applications* 2015; **8**: 346–351.

816   72.   Bartell JA, Sommer LM, Marvig RL, Skov M, Pressler T, Molin S, et al. Omics-
817   based tracking of Pseudomonas aeruginosa persistence in "eradicated" cystic
818   fibrosis patients. *European Respiratory Journal* 2021; **57**.

819   73.   Wheatley RM, MacLean RC. CRISPR-Cas systems restrict horizontal gene
820   transfer in Pseudomonas aeruginosa. *The ISME Journal* 2021; **15**: 1420–1433.

821   74.   Nobrega FL, Walinga H, Dutilh BE, Brouns SJ. Prophages are associated with
822   extensive CRISPR–Cas auto-immunity. *Nucleic acids research* 2020; **48**: 12074–
823   12084.

824   75.   Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of
825   tRNAs in phages. *Genome research* 2007; **17**: 1486–1495.

826   76.   Hesse C, Schulz F, Bull CT, Shaffer BT, Yan Q, Shapiro N, et al. Genome-based
827   evolutionary history of *Pseudomonas* spp. *Environmental Microbiology* 2018; **20**:
828   2142–2159.

829   77.   van den Berg DF, van der Steen BA, Costa AR, Brouns SJ. Phage TRNAs evade
830   TRNA-targeting host defenses through anticodon loop mutations. *Elife* 2023; **12**:
831   e85183.

832   78.   Akhverdian VZ, Khrenova EA, Bogush VG, Gerasimova TV, Kirsanov NB. Wide
833   distribution of transposable phages in natural Pseudomonas aeruginosa populations.
834   *Genetika* 1984; **20**: 1612–1619.

835   79.   Tsao Y-F, Taylor VL, Kala S, Bondy-Denomy J, Khan AN, Bona D, et al. Phage
836   Morons Play an Important Role in Pseudomonas aeruginosa Phenotypes. *J Bacteriol*
837   2018; **200**.

838   80.   Jo J, Cortez KL, Cornell WC, Price-Whelan A, Dietrich LE. An orphan cbb 3-type
839   cytochrome oxidase subunit supports Pseudomonas aeruginosa biofilm growth and
840   virulence. *Elife* 2017; **6**: e30205.

841   81.   Liu Y, Liu Q, Qi L, Ding T, Wang Z, Fu J, et al. Temporal regulation of a
842   Salmonella Typhimurium virulence factor by the transcriptional regulator YdcR.
843   *Molecular & Cellular Proteomics* 2017; **16**: 1683–1693.

844   82.   Prada-Ramírez HA, Pérez-Mendoza D, Felipe A, Martínez-Granero F, Rivilla R,
845   Sanjuán J, et al. AMRZ regulates cellulose production in *P seudomonas syringae* pv.
846   tomato DC 3000. *Molecular Microbiology* 2016; **99**: 960–977.

847   83.   Waligora EA, Ramsey DM, Pryor EE, Lu H, Hollis T, Sloan GP, et al. AmrZ Beta-
848   Sheet Residues Are Essential for DNA Binding and Transcriptional Control of
849   *Pseudomonas aeruginosa* Virulence Genes. *J Bacteriol* 2010; **192**: 5390–5401.

850   84.   Pedersen SS, Høiby N, Espersen F, Koch C. Role of alginate in infection with
851   mucoid Pseudomonas aeruginosa in cystic fibrosis. *Thorax* 1992; **47**: 6–13.

852   85.   Burrows LL. *Pseudomonas aeruginosa* Twitching Motility: Type IV Pili in
853   Action. *Annu Rev Microbiol* 2012; **66**: 493–520.

854   86.   Lin J, Lau GW. DprA-Dependent Exit from the Competent State Regulates
855   Multifaceted Streptococcus pneumoniae Virulence. *Infect Immun* 2019; **87**: e00349-
856   19.

857   87.   Trasanidou D, Geros AS, Mohanraju P, Nieuwenweg AC, Nobrega FL, Staals
858   RH. Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs.
859   *FEMS microbiology letters* 2019; **366**: fnz098.

860   88.   Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T. The complete
861   nucleotide sequence of φCTX, a cytotoxin-converting phage of *Pseudomonas*
862   *aeruginosa* : implications for phage evolution and horizontal gene transfer via
863   bacteriophages. *Molecular Microbiology* 1999; **31**: 399–419.

864   89.   Khan A, Burmeister AR, Wahl LM. Evolution along the parasitism-mutualism
865   continuum determines the genetic repertoire of prophages. *PLoS Computational*
866   *Biology* 2020; **16**: e1008482.

867 90.   Harrison E, Brockhurst MA. Ecological and Evolutionary Benefits of Temperate
868 Phage: What Does or Doesn't Kill You Makes You Stronger. *BioEssays* 2017; **39**:
869 1700112.

870 91.   Bailey ZM, Wendling CC, Igler C. Prophage maintenance is determined by
871 environment-dependent selective sweeps rather than mutational availability.
872 *bioRxiv* 2023; 2023–03.

873 92.   Hampton TH, Thomas D, Van Der Gast C, O'Toole GA, Stanton BA. Mild Cystic
874 Fibrosis Lung Disease Is Associated with Bacterial Community Stability. *Microbiol*
875 *Spectr* 2021; **9**: e00029-21.

876 93.   Quistad SD, Doulcier G, Rainey PB. Experimental manipulation of selfish
877 genetic elements links genes to microbial community function. *Phil Trans R Soc B*
878 2020; **375**: 20190681.

879

880

Figure Legends:

881

882   Fig. 1: a) Overview of the double agar overlay outcomes. The supernatants of

883   induced or uninduced overnight cultures were plated against the lawns of

884   PAO1 and PaLo43. Visible lysis, *i.e.* lysis where individual plaques were

885   counted, is marked in green, clearing zones are judged as inefficient lysis and

886   marked in yellow, and lysis absence is marked in red. Prophage inductions are

887   mostly correctly implied or indicated via the noted clearing zones and plaques,

888   as confirmed by the quantitative sequencing results; b) Quantitative results of

889   induced prophages via mapping of lysate reads to gapless host genome

890   assemblies. Host strains are divided into two groups - those isolated from

891   patients with apparent monoclonal infections and those with polyclonal

892   infections. Different shades of blue indicate whether a given prophage was

893   only induced by an antibiotic or could also self-induce and a teal colour

894 denotes that a prophage was predicted as intact but not induced with the
895 methods tested. Lysogeny is the rule for the 12 clinical *P. aeruginosa* isolates
896 of the study.

897

898 Fig. 2: Bootstrap tree of predicted amino acid sequences of the concatenated
899 repressor/antirepressor built using NGPhylogeny.fr and the
900 "PhyML+SMS/OneClick" workflow. Branch tags are coloured according to
901 prophage. Prophage induction profiles are noted to the right of a tree branch
902 as cipro- (C), mitC- (M) or self-induced (S). The tree fails to cluster prophages
903 according to their induction patterns.

904

905 Fig. 3: a) Bootstrap tree based on SNPs among the genomes of the 12 studied
906 isolates and b) proteomic tree computed with tBLASTx in ViPTree shows the
907 clustering of study's prophages versus related literature prophages of *P.*
908 *aeruginosa* shown to exist as particles. Colours follow the scheme of Fig.2.
909 Intact and induced prophages are signified by a green star and likely intact
910 but uninduced prophages by a red circle. The clustering of this study's
911 prophages is not based on host phylogenetic relationships (a) and displays a
912 wide distribution across the tree, which indicates their high genomic diversity
913 (b).

914

915   Fig 4: Overview of the longitudinal isolates, whose genomes were scanned

916   here to identify longitudinally frequent prophage elements. A new isolate is

917   symbolised with "x" unless if it is one of the 12 studied isolates, in which case

918   a circle is used. Each horizontal line corresponds to all longitudinal isolates per

919   patient considered and the year of their isolation is found in the x axis. When

920   an isolate with a CT different to its predecessor's appears in the infection

921   history of the patient the CT number is again denoted on top of that isolate.

922   CT numbers in blue correspond to the persistent CT and when one of the

923   studied isolates belongs to the persistent CT it is also coloured in blue. At the

924   end of each patient history line, there is a plot displaying the number of

925   induced prophages (top bar) and the number of induced prophages that is also

926   found long-term in the genome of the persistent CT (bottom bar).

927

928   Fig. 5: BLASTn similarity genomic synteny maps designed with Easyfig. A

929   sequenced prophage from an early isolate is compared to its own genome

930   identified in a later longitudinal isolate of the persistent CT (See also

931   Supplementary Table 4). a) Prophage Alize induced from isolate F002 of a

932   monoclonally infected patient displays a high conservation of gene order and

933   high nucleotide similarity when compared to a prophage induced from a

934   longitudinal isolate eight years later; b) Prophage Meltemi is originally found

935   in isolate LRJ32 of a transient CT but, as for all other prophages found in LRJ32,

936   it is later retraced in an isolate of the persistent CT (DK67) from the same

937   polyclonally infected patient. Similarly to (a) the genome of Meltemi is virtually

938   unchanged nine years later when it is retraced in the persistent CT. The

939   remaining synteny maps are presented in Supplementary Figure 1.
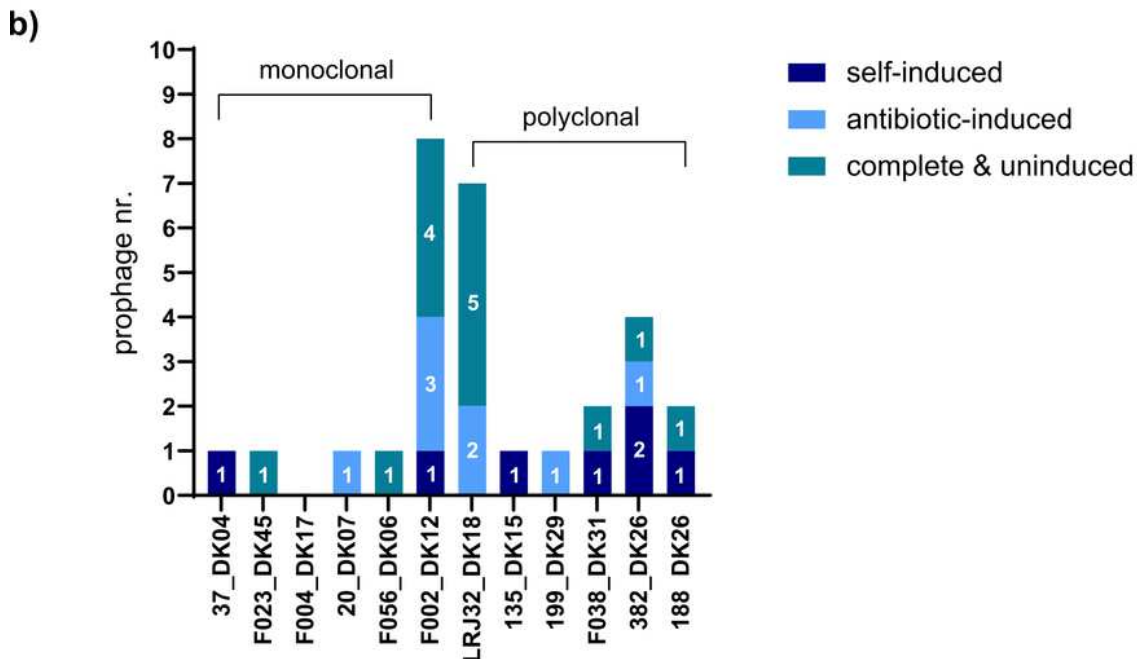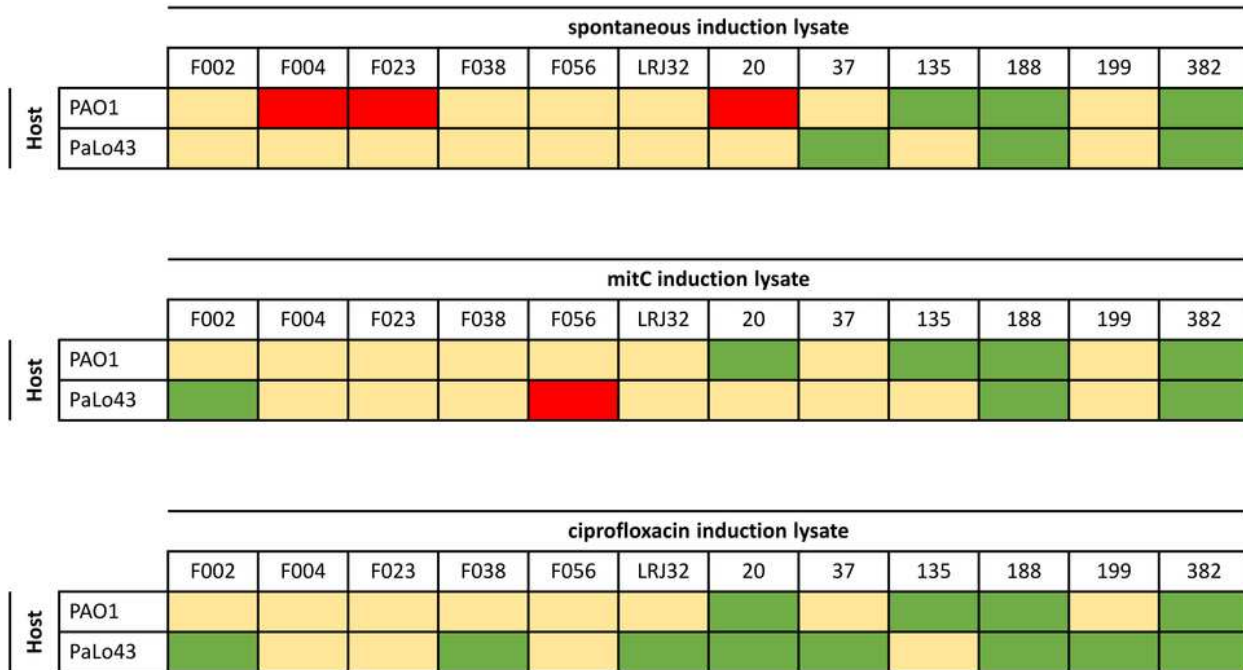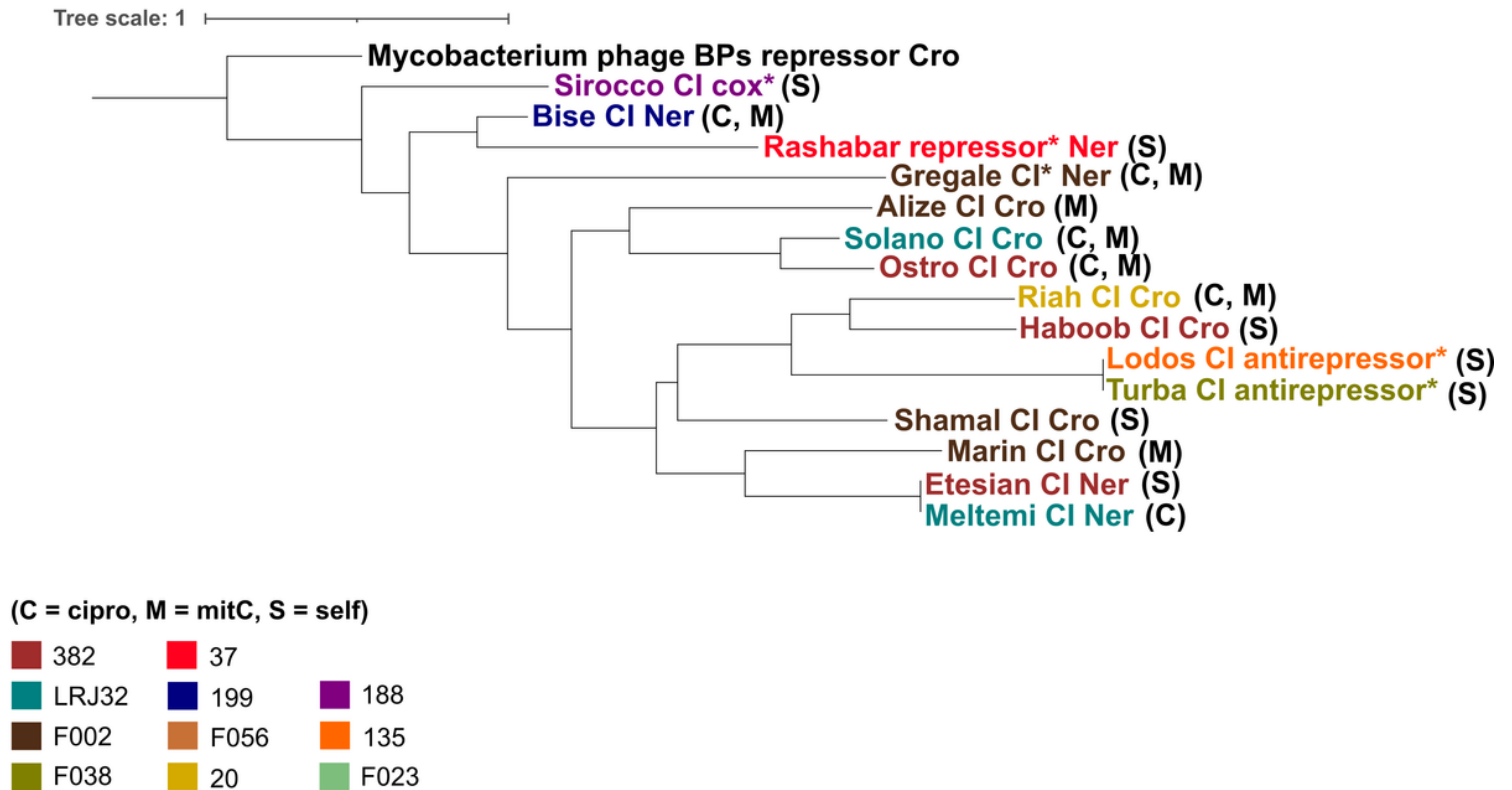
Figures

a)

b)

Figure 1

a) Overview of the double agar overlay outcomes. The supernatants of induced or uninduced overnight cultures were plated against the lawns of PAO1 and PaLo43. Visible lysis, *i.e.* lysis where individual plaques were counted, is marked in green, clearing zones are judged as inefficient lysis and marked in

yellow, and lysis absence is marked in red. Prophage inductions are mostly correctly implied or indicated via the noted clearing zones and plaques, as confirmed by the quantitative sequencing results; b) Quantitative results of induced prophages via mapping of lysate reads to gapless host genome assemblies. Host strains are divided into two groups - those isolated from patients with apparent monoclonal infections and those with polyclonal infections. Different shades of blue indicate whether a given prophage was only induced by an antibiotic or could also self-induce and a teal colour denotes that a prophage was predicted as intact but not induced with the methods tested. Lysogeny is the rule for the 12 clinical *P. aeruginosa* isolates of the study.

Repressor/antirepressor bootstrap tree

Figure 2

Bootstrap tree of predicted amino acid sequences of the concatenated repressor/antirepressor built using NGPhylogeny.fr and the "PhyML+SMS/OneClick" workflow. Branch tags are coloured according to prophage. Prophage induction profiles are noted to the right of a tree branch as cipro- (C), mitC- (M) or self-induced (S). The tree fails to cluster prophages according to their induction patterns.
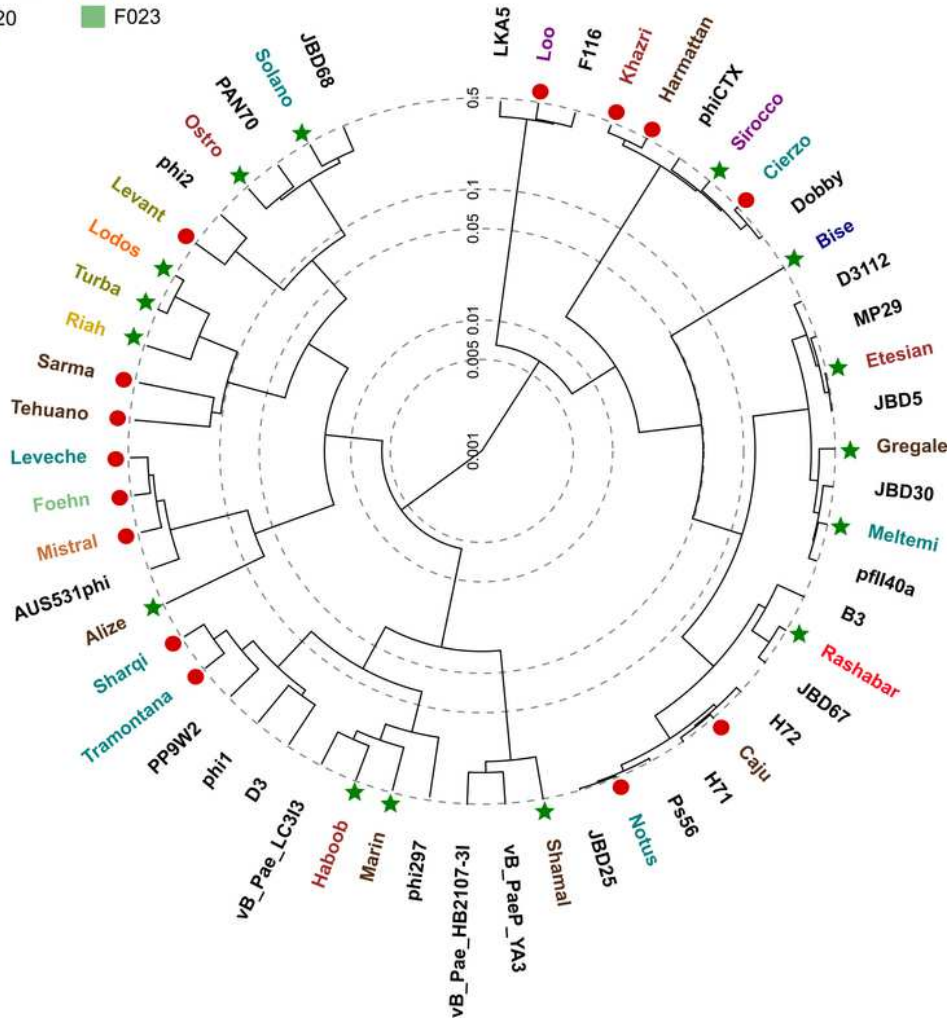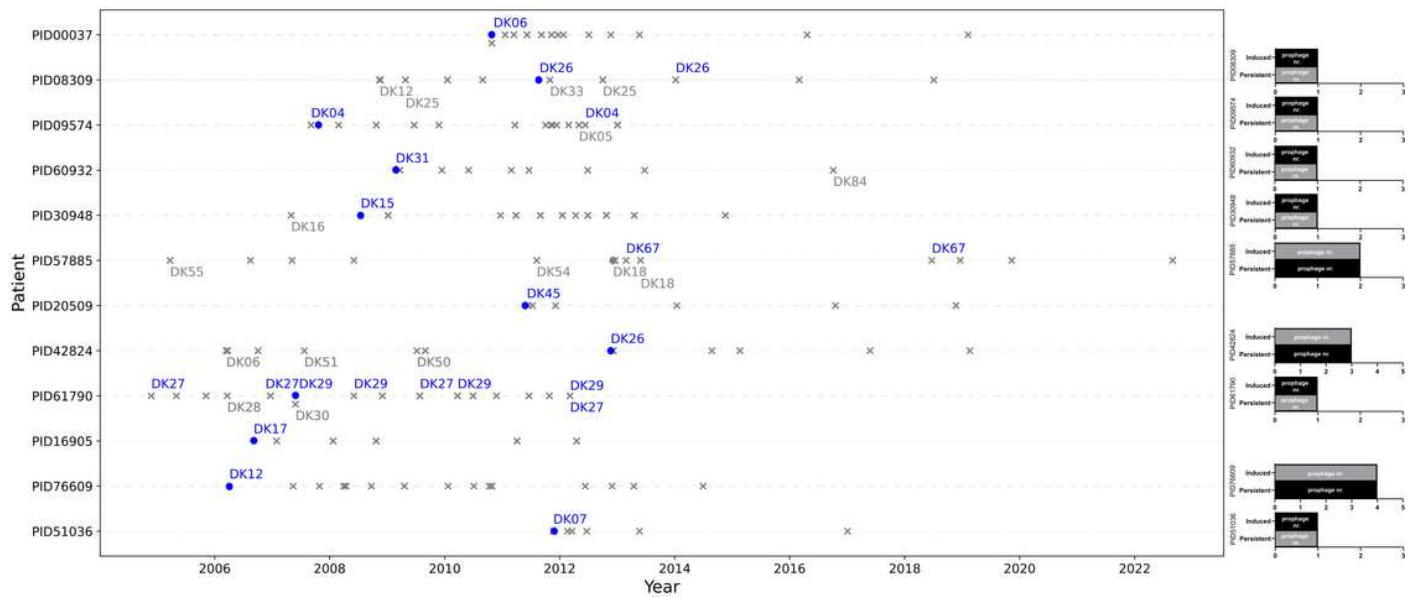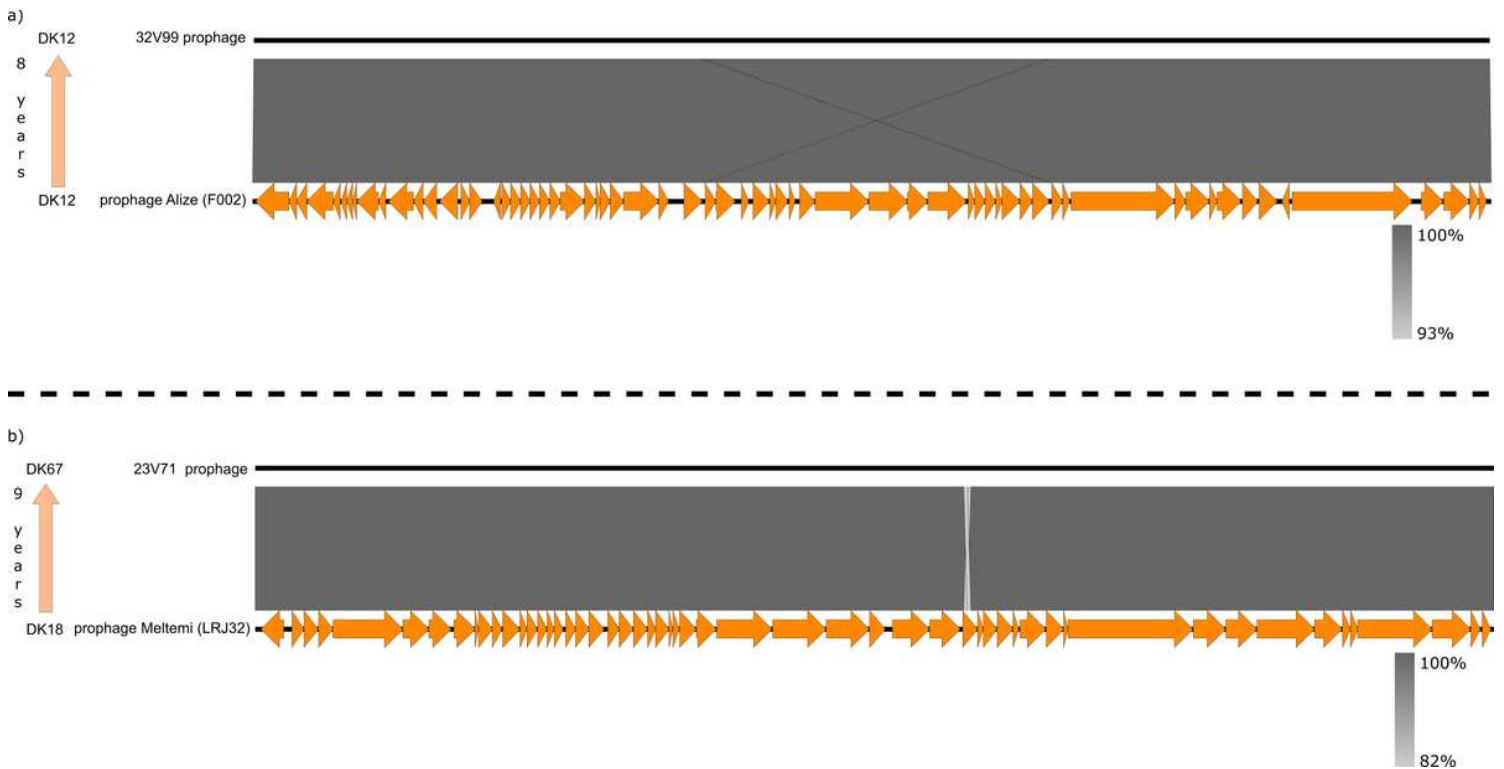
Figure 3
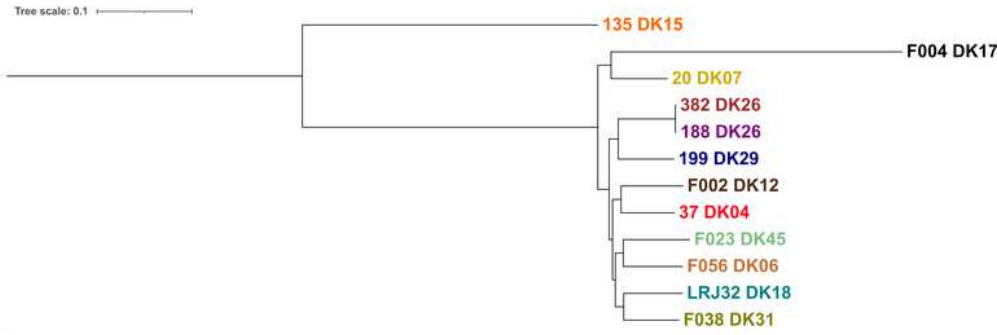
a) Bootstrap tree based on SNPs among the genomes of the 12 studied isolates and b) proteomic tree computed with tBLASTx in ViPTree shows the clustering of study's prophages versus related literature prophages of *P. aeruginosa* shown to exist as particles. Colours follow the scheme of Fig.2. Intact and induced prophages are signified by a green star and likely intact but uninduced prophages by a red circle.

The clustering of this study's prophages is not based on host phylogenetic relationships (a) and displays a wide distribution across the tree, which indicates their high genomic diversity (b).

Figure 4

Overview of the longitudinal isolates, whose genomes were scanned here to identify longitudinally frequent prophage elements. A new isolate is symbolised with "x" unless if it is one of the 12 studied isolates, in which case a circle is used. Each horizontal line corresponds to all longitudinal isolates per patient considered and the year of their isolation is found in the x axis. When an isolate with a CT different to its predecessor's appears in the infection history of the patient the CT number is again denoted on top of that isolate. CT numbers in blue correspond to the persistent CT and when one of the studied isolates belongs to the persistent CT it is also coloured in blue. At the end of each patient history line, there is a plot displaying the number of induced prophages (top bar) and the number of induced prophages that is also found long-term in the genome of the persistent CT (bottom bar).

Figure 5

BLASTn similarity genomic synteny maps designed with Easyfig. A sequenced prophage from an early isolate is compared to its own genome identified in a later longitudinal isolate of the persistent CT (See also Supplementary Table 4). a) Prophage Alize induced from isolate F002 of a monoclonally infected patient displays a high conservation of gene order and high nucleotide similarity when compared to a prophage induced from a longitudinal isolate eight years later; b) Prophage Meltemi is originally found in isolate LRJ32 of a transient CT but, as for all other prophages found in LRJ32, it is later retraced in an isolate of the persistent CT (DK67) from the same polyclonally infected patient. Similarly to (a) the genome of Meltemi is virtually unchanged nine years later when it is retraced in the persistent CT. The remaining synteny maps are presented in Supplementary Figure 1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.pdf
- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xls
- SupplementaryTable4.xlsx
- SupplementaryTable5.xlsx
- SupplementaryTable6.xlsx