Host adaptation mediated by intergenic evolution in a bacterial pathogen

S. M. Hossein Khademi[1] and Lars Jelsbak[1,a]

[1] Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Lyngby, Denmark

[a] Corresponding author: Lars Jelsbak, Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Lyngby, Denmark. Email: lj@bio.dtu.dk. Telephone: +45 45256129

Keywords: intergenic evolution, *cis*-regulatory elements, bacterial adaptation, gene expression, antibiotic resistance

Abstract

Bacterial pathogens evolve during the course of infection as they adapt to the selective pressures that confront them inside the host. Identification of adaptive mutations and their contributions to pathogen fitness remain a central challenge. Although mutations can either target intergenic or coding regions in the pathogen genome, studies of host adaptation have focused predominantly on molecular evolution within coding regions, whereas the role of intergenic mutations remains unclear. Here, we address this issue and investigate the extent to which intergenic mutations contribute to the evolutionary response of pathogens to host environments, and whether intergenic mutations have distinct roles in host adaptation. We characterize intergenic evolution in 44 lineages of a clinically important bacterial pathogen, *Pseudomonas aeruginosa*, as they adapt to their hosts. We identify 88 intergenic regions in which parallel evolution occurs. At the genetic level, we find that mutations in these regions under selection are located primarily within regulatory elements upstream of transcriptional start sites. At the functional level, we show that these mutations both create or destroy regulatory interactions in connection to transcriptional processes and are directly responsible for evolution of important pathogenic phenotypes including antibiotic sensitivity. Importantly, we find that intergenic mutations are more likely to be selected than coding region mutations and that intergenic mutations enable essential genes to become targets of evolution. In summary, our results highlight the evolutionary significance of intergenic mutations in creating host-adapted variants and that intergenic and coding regions have different qualitative and quantitative contributions to this process.

38

39 **Significance**

40 Pathogens adapt to their host during infection, but the contribution and function of
41 non-coding intergenic sequences to adaptation is poorly understood. Here, genome-
42 wide identification of adaptive mutations within intergenic regions demonstrates
43 that these sequences constitute an important part of the genetic basis for host
44 adaptation. We find that intergenic mutations are abundant relative to adaptive
45 mutations within coding sequences, and can contribute directly to evolution of
46 pathogen-relevant traits. Importantly, we find that intergenic mutations modify
47 expression of essential genes and thus make contributions that are functionally
48 distinct from coding mutations. These results improve our understanding of the
49 evolutionary processes *in vivo* and can potentially assist in refining predictions of
50 pathogen evolution, disease outcome and antibiotic resistance development.

51

52 **Introduction**

53 Bacterial pathogens evolve during infection as they adapt to the environment inside
54 the host (1). Since the bacterial phenotypes selected *in vivo* may have profound
55 impact on disease severity and progression (2, 3) as well as response to antibiotic
56 therapy (4), identification and analysis of the full range of beneficial genetic changes
57 that underlies host adaptation is of importance.

58 Although adaptive mutations may potentially change the sequences of either coding
59 regions or non-translated intergenic regions and thus affect protein function or
60 expression, respectively, studies of pathogen adaptation during infection of host
61 tissues have focused predominantly on molecular evolution within coding regions,
62 whereas the role of adaptive mutations in intergenic regions has received
63 comparably less attention. The shortage of systematic, genome-wide analyses of
64 intergenic evolution in bacterial pathogens is surprising, given the fact that these
65 regions are home to a large number of functional elements required for expression
66 of virulence and resistance determinants *in vivo* and that intergenic regions are
67 maintained by purifying selection in many bacterial species (5-7). Moreover, *cis*-
68 regulatory mutations are known to play an important role in phenotypic evolution
69 in eukaryotic organisms (8). Overall, it remains unclear to what extent intergenic
70 mutations contribute to the evolutionary response of pathogens to the host
71 environment and whether intergenic mutations have a qualitative distinct role in
72 host adaptation.

73 There are clear, albeit few, examples of intergenic regions that evolve under
74 selection within the host. For example, evolution of a novel regulatory interaction
75 between the virulence regulator SsrB and the promoter of the *sfrN* gene was shown

76 to result in enhanced within-host fitness in *Salmonella* Typhimurium (9). In
77 *Pseudomonas aeruginosa*, evolution of the intergenic region of the *phuR-phuSTUVW*
78 genes during host colonization was shown to increase expression of the Phu heme
79 uptake system and improve the ability of the pathogen to acquire iron from
80 hemoglobin (10). In *Mycobacterium tuberculosis*, evolution of ethambutol resistance
81 has been linked to the *empAB* promoter region acquiring mutations that enhance
82 expression of enzymes essential for the synthesis of cell wall arabinogalactan (11).
83 These and other examples point towards an evolutionary significant role
84 of mutations in intergenic regions in connection to bacterial pathogenesis and justify
85 a broader analysis of this type of mutations.
86 One reason for the paucity in genome-wide analysis of intergenic evolution is
87 probably related to the inherent difficulties in inferring function directly from the
88 sequence within intergenic regions and, consequently, to differentiate adaptive
89 mutations with functional effects from neutral mutations that have been fixed by
90 chance. Here, we harnessed the combination of parallel evolution and functional
91 genomics to identify intergenic regions under selection in the genome of the
92 opportunistic pathogen *P. aeruginosa* during the process of host adaptation in
93 multiple cystic fibrosis patients. Our study reveals that adaptive intergenic mutations
94 represent an egregiously underappreciated aspect of host adaptation in *P.*
95 *aeruginosa*, and that intergenic and coding region mutations contribute differently
96 to this process both qualitatively and quantitatively.
97

98 **Results**

99 *Parallel evolution in intergenic regions in P. aeruginosa.*

100 To investigate the contribution of intergenic mutations to bacterial adaptation to the
101 host environment, we considered data from seven studies (12-18) in which multiple
102 clonal *P. aeruginosa* isolates have been sampled and sequenced during the course of
103 infection in subjects with cystic fibrosis (CF). In CF infections, the host environments
104 in individual subjects represent parallel selective conditions by which evolution is
105 directed, and identification of parallel genetic evolution in bacteria from independent
106 infections is strongly suggestive of positive selection at these loci (16).

107 Here, we focused our analysis exclusively on intergenic regions in which mutations
108 were acquired during infection and included only intergenic regions also present in
109 the PAO1 reference genome. In total, we identified 3,489 mutations (2,024 SNPs and
110 1,465 indels) in the intergenic regions of the 44 different *P. aeruginosa* clone types
111 included in our dataset (Table S1). Since the majority of regulatory elements in the
112 bacterial genome range between 5-30 bp in length (19), we identified intergenic
113 regions under positive selection by only considering mutations found in parallel in
114 different clone types and distributed within a window of less than 30 bp (Materials
115 and Methods). Applying these criteria, we found 63 intergenic regions with parallel
116 genetic evolution (Figure 1).

117 Since certain *P. aeruginosa* clone types are transmissible and can form clinic-specific
118 outbreaks among patients (16), we also analyzed if distinct intergenic mutations had
119 accumulated in parallel among clonal isolates within each of the 44 clone types. We
120 identified 41 intergenic regions in which three or more distinct mutations (less than
121 30 bp apart) had accumulated in isolates of the same clone type (Figure 1).

122 Interestingly, 16 of these regions are also represented among the 63 regions
123 identified in our analysis of parallel mutations between clone types, providing further
124 support for the importance of these mutations in adaptation of *P. aeruginosa* to the
125 CF environment (Figure 1). In total, we identify 88 intergenic regions that evolved
126 under the pressure of natural selection within the hosts. The connection between
127 these 'pathoadaptive' regions and their flanking genes identify genetic systems with
128 importance for pathogen adaptation and thus provide insight into the selective forces
129 that operate on the pathogen.
130
131 *Intergenic mutations frequently target promoter sequences.*
132 We next analyzed the genomic distribution of the identified intergenic mutations.
133 Non-translated intergenic regions are distributed across the genome in three
134 possible orientations: 1) upstream of two genes, 2) downstream of two genes and 3)
135 upstream of one gene and downstream of one gene, where the latter may include
136 regions with no promoter and within an operon (Figure 2a). We found an over-
137 representation of mutations upstream of two genes among the pathoadaptive
138 regions (Fisher's exact test, $P = 0.010$, $n = 88$, Figure 2b). This bias towards selection
139 of intergenic mutations upstream of genes suggests that the majority of intergenic
140 mutations target potential *cis*-regulatory elements such as the core promoter,
141 transcription factor binding sites, ribo-regulators or translational elements, and
142 consequently influence protein expression levels by affecting transcriptional or post-
143 transcriptional processes.
144 To further explore this hypothesis, we analyzed the complete set of 88
145 pathoadaptive regions for the presence of known regulatory elements (Materials

146 and Methods) and mapped the overlap between these putative regulatory sites and
147 the identified adaptive mutations. While bacterial intergenic regions are home to a
148 wide range of regulatory elements many of which are not well characterized, we
149 nevertheless observed 28 regions (32%), in which the cluster of adaptive mutations
150 was positioned within one or more putative regulatory elements. The majority of
151 mutations within these 28 regions target the putative core promoter alone or in
152 combination with other elements (Figure 2c), suggesting that intergenic mutations
153 frequently target sequences important for transcriptional processes. In support of
154 this, we observed that intergenic mutations were more frequently located upstream
155 of known transcriptional start sites (TSS) (37 cases) than downstream (10 cases)
156 (Table S6).
157
158 *Pathoadaptive intergenic mutations change transcriptional activity of genes involved*
159 *in host interaction, metabolism and antibiotic susceptibility.*
160 To further explore the potential relationship between intergenic mutations and
161 transcription, we quantified the effects of a subset of intergenic mutations on
162 transcription of downstream genes. To this end, we constructed transcriptional
163 fusions of both wild type and mutant intergenic alleles with the luciferase reporter
164 (*luxCDABE*) genes and integrated single copies of the fusions at the neutral *attB* site
165 (10) in the chromosome of *P. aeruginosa* PAO1 (20). We measured the
166 transcriptional activity of 25 different intergenic regions in which pathoadaptive
167 mutations were located upstream of either one or two genes. This selection resulted
168 in a total of 32 transcriptional fusions, which represent 33% of all possible fusions
169 within the complete set of 88 pathoadaptive regions. In addition, for one of the

170 intergenic regions (*ampR//ampC*), we tested two alleles, each with different
171 mutations (Table S7 and Figure S1).
172 Measurements of *lux* expression during exponential growth in Luria-Bertani (LB)
173 medium and ABTGC minimal medium revealed significantly altered expressions in 16
174 of 34 tested fusions in at least one of the two conditions (Student *t* test, $P < 0.05$,
175 Figure 3). Altered expression was in most cases moderate (<3-fold change) and the
176 fold change ranged from -3.1 to 22.1 for the mutant allele compared to that of wild
177 type (Figure 3). Interestingly, 10 of these 16 fusions exhibited altered expressions
178 only in either LB or ABTGC minimal medium, but not in both conditions, which
179 suggests that many adaptive intergenic mutations alter transcriptional levels while
180 not interfering with conditional control mechanisms.
181 These results show that a substantial fraction of the intergenic mutations are
182 associated with functional (transcriptional) effects despite the fact that we recorded
183 these effects in the non-native PAO1 genetic background (*i.e.* with removal of
184 potential epistatic effects from the additional mutations found in the clinical isolate)
185 and in a narrow range of conditions, which most likely means that we are not
186 capturing the full spectrum of functional effects connected to the intergenic
187 mutations.
188 Several of the 16 fusions with altered expression relate to genes that encode
189 proteins with known functions in bacteria-host interactions, cellular metabolism and
190 antibiotic resistance. For example, *cerN* expresses a ceramidase involved in
191 utilization of host-produced sphingolipids (21), *exsC* expresses a protein involved in
192 positive regulation of the type III secretion system (22) and PA4837 is the first gene
193 in an operon (PA4837-34) involved in expression of a metalophore system essential

194 for survival in airway mucus secretions (23, 24). Other genes are known to play a role
195 in pyrimidine and aromatic amino acid metabolism (*pyrC* and *hmgA*, respectively).
196 Finally, two genes are linked to antibiotic resistance: *rluC* (25) and *ampR* (26). Seven
197 genes encode proteins of unknown functions and their role in relation to host
198 adaptation remains unclear.
199 Interestingly, expression changes were observed in both directions (seven mutant
200 alleles resulted in increased expression and nine mutant alleles resulted in decreased
201 expression) (Figure 3), suggesting that pathoadaptive intergenic mutations may
202 equally well either create or destroy regulatory interactions.
203
204 *Mutations upstream of ampR and ampC enhance resistance to several antibiotics*
205 Next, we explored the direct effects of intergenic mutations on the physiology of the
206 pathogen. As resistance towards antibiotics is a common phenotype that emerges
207 during CF infections, we selected the mutations found in the two alleles of the
208 *ampR//ampC* intergenic region for further study. Mutations in this intergenic region
209 resulted in enhanced expression of the global antibiotic resistance regulator AmpR,
210 but had no direct effect on expression of the AmpC β-lactamase (Figure 3). We
211 introduced these mutations in the genome of *P. aeruginosa* PAO1 through allelic
212 replacement (Materials and Methods). Since a SNP mutation (G7A) was present at
213 the start of *ampC* gene in one of the alleles, we also made an allelic replacement of
214 this mutation alone in the PAO1 genome to separate the effects caused by the
215 intergenic mutations (Figure S1). For each strain and their isogenic wild type, we
216 measured the Minimal Inhibitory Concentration (MIC) of various β-lactam antibiotics
217 such as imipenem, ceftazidime and ampicillin from carbapenem, cephalosporin and

218 penicillin classes of β-lactams, respectively. For both intergenic alleles, we observed
219 a small but significant increase in the MIC of imipenem and ampicillin (Student *t* test,
220 $P < 0.01$, Figure 4), but not ceftazidime. AmpR regulates β-lactam resistance both
221 through direct activation of AmpC expression as well as via an AmpC-independent
222 manner (26). Irrespective of the mechanism, our results show that acquisition of
223 intergenic mutations between *ampR* and *ampC* is directly linked to a host-relevant
224 phenotypic alteration (*i.e.* reduced β-lactam susceptibility).
225
226 *Intergenic evolution targets essential genes and contributes more to host adaptation*
227 *than intragenic evolution*
228 Finally, we compared the relative contribution of coding and intergenic mutations to
229 pathogen adaptation. We focused on the large fraction of isolates (n=474) included
230 in this study, in which 52 coding regions were found to be under positive selection
231 during host adaptation (17). In these isolates, we identified 35 pathoadaptive
232 intergenic regions (Materials and Methods, Table S8). Although coding region
233 mutations are numerically dominant over intergenic mutations, normalization to the
234 mutational targets available for intergenic and coding region mutations (89.8% of
235 the *P. aeruginosa* genome contains coding regions) reveals that intergenic regions
236 are 3.7 times more likely to be selected than cording regions and thus appear to play
237 a quantitatively more prominent role in host adaptation (Figure 5a).
238 We also analyzed qualitative differences between coding and intergenic mutations
239 by determining the presence of essential genes among the pathoadaptive coding and
240 intergenic regions. By cross-referencing the 35 pathoadaptive intergenic regions to
241 the list of 445 genes previously shown to be essential for survival of *P. aeruginosa*

242 PAO1 in CF sputum environment (27), we found that 7 of the 38 genes located
243 immediately downstream of the 35 pathoadaptive intergenic regions are essential
244 (Table S8). Two of these genes (*pyrC* and PA5492) showed altered expression as a
245 consequence of pathoadaptive mutations in their intergenic region, demonstrating
246 that such mutations can indeed modulate expression of essential genes (Figure 3).
247 Importantly, the association between pathoadaptive intergenic regions and essential
248 genes at a level of 18% represents a significant overrepresentation from the normal
249 prevalence of CF sputum essential genes in the *P. aeruginosa* PAO1 genome (Figure
250 5b, Fisher's exact test, $P = 0.029$). In contrast, there were no CF sputum essential
251 genes within 52 pathoadaptive coding regions demonstrating a significant
252 underrepresentation of these genes within adaptive coding regions (Figure 5b,
253 Fisher's exact test, $P = 0.033$).
254

255 **Discussion**

256 In this study, we present evidence that intergenic mutations constitute an important
257 part of the genetic basis for host adaptation in *P. aeruginosa*. Generally, the
258 contribution of intergenic regions to evolution of host-adapted variants has received
259 little attention. However, since the development of predictive models of pathogen
260 evolution and discovery of new therapeutic targets (28, 29) rely on understanding
261 the evolutionary response of pathogens to the host environment, identification of
262 the full range of adaptive mutations in both coding and non-coding regions is
263 important. Here, our genome-wide identification of intergenic regions under
264 selection within the host was made possible by combining analysis of parallel
265 evolution across a large number of infected individuals with functional genomics.
266 This approach may be useful for analysis of intergenic regions in connection to host
267 adaptation in other pathogens or niche adaptation in general.

268

269 Our identification of pathoadaptive intergenic regions provides insight into the
270 cellular functions targeted by intergenic mutations (Figure 1) and thus points to the
271 selective pressures that confront the pathogen within its CF host. For example,
272 adaptive mutations were found to alter expression of genes such as *cerN* (involved in
273 sphingolipid utilization) (21), *phuR-phuSTUVW* (involved in iron acquisition) (10) and
274 PA4837-34 (involved in zinc acquisition) (23), which strongly indicates that metabolic
275 adaptation for a better exploitation of available nutrients in the host is an important
276 evolutionary driver. Similarly, we observed that mechanisms of the development of
277 tolerance to antibiotics and other inhibitors in the host are also frequent targets of
278 intergenic molecular evolution. Similar functional categories have been found in

279 studies focusing on pathoadaptive mutations within coding regions (14, 16, 17),
280 suggesting that key selective pressures such as nutrient availability and antibiotic
281 stress can be mitigated both by intergenic and coding region evolution in *P.*
282 *aeruginosa*.
283
284 At the functional level, we show that intergenic evolution predominantly targets
285 transcriptional processes to alter the transcriptional activity of downstream genes.
286 However, we also found evidence of parallel evolution in two intergenic small RNAs,
287 four transcriptional terminators, and several cases of mutations located downstream
288 of transcriptional start sites (Table S5 and Table S6), which suggests that adaptive
289 mutations may also target elements that control protein expression at the post-
290 transcriptional level. Importantly, we have shown here and in a previous study that
291 intergenic mutations can be directly responsible for the evolution of important
292 pathogenic traits such as reduced sensitivity to antibiotics (Figure 4) and increased
293 iron uptake (10). Further studies, in particular of pathoadaptive regions upstream of
294 genes with unknown functions, will most likely uncover new mechanisms central to
295 CF host colonization and pathogenesis, and assist in identifying the full complement
296 of stressors present in the host, most of which are currently unknown.
297
298 Our study also reveals important qualitative differences between the intergenic and
299 coding region mutations. A generally accepted model is that intergenic mutations
300 would typically confer local and subtle regulatory effects primarily on the immediate
301 downstream genes, whereas mutations in coding regions - with their potential to
302 inactivate entire pathways – would be more likely to cause systemic changes of the

303 physiology of the cell (8, 30). One prediction from this model is that intergenic
304 mutations are associated with less antagonistic effects relative to coding region
305 mutations. While this prediction is difficult to test, our observation of enrichment of
306 essential genes for which intergenic evolution occurred is a clear illustration of this
307 point (Figure 5). The finding that intergenic mutations can bypass the deleterious
308 effects of coding region mutations, thus allowing essential genes to become targets
309 for evolutionary changes, reveals an important aspect of the role of intergenic
310 mutations as well as a key functional difference between intergenic and coding
311 region mutations.
312
313 From a quantitative perspective, we find that intergenic mutations are more likely to
314 be selected than coding region mutations during CF host adaptation in *P. aeruginosa*
315 (Figure 5). We hypothesize that the relative contribution of coding and intergenic
316 mutations is variable and depends on a set of identifiable factors of either
317 environmental nature (e.g. niche complexity) or intrinsic to the bacterial pathogen
318 (e.g. genome size, and the number of transcriptional regulatory systems and
319 essential genes encoded in the genome). Although the precise factors that influence
320 the relative contribution of the two types of mutations may be difficult to
321 disentangle, we speculate that in the case of *P. aeruginosa* CF infections, a major
322 contributing factor is the composition of the adaptive environment in the host. The
323 CF host niche is characterized by a complex combination of multiple stressors that
324 must be mitigated for successful bacterial colonization (31). In such environments,
325 mutations in intergenic regions that tune expression levels while maintaining
326 responsiveness to environmental and host-derived cues may result in less pleiotropic

327 effects than mutations that change protein structure or function (30). Further
328 studies of *P. aeruginosa* adaptation in other infections and host environments such
329 as chronic wounds and ulcerative keratitis (32) are required to identify factors that
330 may influence the relative contribution of intergenic and coding region evolution.

331

332 Our documentation of the evolutionary significance of intergenic mutations was
333 obtained in the particular genetic and ecological context of *P. aeruginosa* adaptation
334 to the cystic fibrosis airway niche. Nevertheless, our systematic study provides
335 insight into the contribution and functionality of intergenic versus intragenic
336 mutations, which is of broader relevance in connection to bacterial evolution in
337 natural environments. This is supported by recent observations indicating that other
338 contexts may also promote intergenic evolution. For example, our study resonate
339 well with results showing that adaptive intergenic mutations contribute to
340 innovation of novel metabolic functions in laboratory-evolving *E. coli* (33), evidence
341 of a signal of positive selection in *M. tuberculosis* intergenic regions (6) and the
342 suggestion that intergenic evolution may mitigate detrimental fitness effects
343 associated with acquisition of novel genetic material (34). We suggest that adaptive
344 mutations in intergenic regions represent an important but underappreciated aspect
345 of bacterial evolution not only in connection to host colonization but also niche
346 adaptation in other natural environments.

347

348

349 **Materials and Methods**

350 *Bacterial strains and growth conditions*

351 Luria-Bertani (LB) and ABT minimal medium supplemented with 1% glucose and 1%
352 casamino acids (ABTGC) (10) were routinely used for growth of *E. coli* and *P.*
353 *aeruginosa* strains. *E. coli* CC118 λpir was used for maintenance of recombinant
354 plasmids supplemented with 8 µg/ml of tetracycline. *P. aeruginosa* PAO1 strain was
355 used for phenotypic investigation of intergenic mutations. For final marker selection
356 of *P. aeruginosa,* 50 µg/ml of tetracycline was used.

357

358 *Identification of pathoadaptive intergenic regions*

359 The dataset used for analysis of intergenic evolution is described in SI Materials and
360 Methods. Presence of independent clonal lineages of *P. aeruginosa* was verified by
361 recording of MLST allele profiles of each isolate as described in SI Materials and
362 Methods. Pathoadaptive intergenic regions selected across 44 clones were defined
363 as regions containing multiple mutations from isolates of independent clones within
364 a narrow region of less than 30 bp. Similarly, intergenic regions containing multiple
365 distinct mutations from isolates of the same clone within a narrow 30 bp window
366 were also defined as pathoadaptive within that clone. Mutations within narrow
367 windows of pathoadaptive regions were more enriched than what is expected by
368 chance within the dataset. Detailed description is available in SI Materials and
369 Methods. Identification of putative intergenic elements within pathoadaptive
370 intergenic regions is described in SI Materials and Methods.

371

372 *Genetic techniques*

373 Genetic engineering of reporter fusion and allelic replacement plasmids are
374 described in SI Materials and Methods. Presence of mutated alleles within reporter
375 fusion and allelic replacement constructs was verified by Sanger sequencing.
376 Reporter fusion plasmids were electroporated into *P. aeruginosa* PAO1 as previously
377 described (35). Constructs with allelic replacements were introduced into *P.*
378 *aeruginosa* by triparental mating using the helper strain *E. coli* HB101/pRK600 (36).
379 The presence of mutated alleles in tetracycline-sensitive, sucrose-resistant colony
380 isolates was verified by PCR and Sanger sequencing.

381

382 *Phenotype assays*

383 Phenotypic expression of reporter fusion strains was investigated as described in SI
384 Materials and Methods. Continuous measurements of growth (OD_{600}) and
385 luminescence were recorded by Cytation 5 multimode reader (BioTek) at 200 rpm
386 shaking condition and 37 °C temperature. Lux expression normalized by cell density
387 for all strains was recorded and compared against reporter fusions containing wild
388 type alleles. Measurements were repeated three times for each reporter fusion
389 strain. Determination of MIC values for ceftazidime, imipenem and ampicillin is
390 described in SI Materials and Methods. MIC values were either measured by
391 standard microdilution method in Mueller-Hinton (MH) broth or using E-test on MH
392 agar plates. Measurements were repeated five times for each strain. Statistical
393 differences between means of replicates for reporter fusion and MIC resistance
394 assays were calculated by two-tailed Student *t* test.

395

Acknowledgements

We thank Lea M. Sommer and Anders Norman for technical advices in bioinformatics approaches, Esben V. Nisted for help in design of primers, Nicoline Uglebjerg and Caroline A. S. Lauridsen for assistance in identification of putative intergenic elements. We also thank Grith Hermansen, Pavelas Sazinas, Geoff Winsor, Ed Feil, Dominique Schneider and Thomas Hindré for valuable discussions. This work was supported by the Danish Council for Independent Research (6108-00300A) and the Villum Foundation (VKR023113).

Author contributions

S.M.H.K and L.J. conceived study and designed research. S.M.H.K. performed research. S.M.H.K and L.J. analyzed data and wrote the manuscript.

409 **References**

410 1. Didelot X, Walker AS, Peto TE, Crook DW, & Wilson DJ (2016) Within-host
411 evolution of bacterial pathogens. *Nat Rev Microbiol* 14(3):150-162.
412 2. Hoffman LR*, et al.* (2009) Pseudomonas aeruginosa lasR mutants are
413 associated with cystic fibrosis lung disease progression. *J Cyst Fibros* 8(1):66-
414 70.
415 3. Das S*, et al.* (2016) Natural mutations in a Staphylococcus aureus virulence
416 regulator attenuate cytotoxicity but permit bacteremia and abscess
417 formation. *Proc Natl Acad Sci U S A* 113(22):E3101-3110.
418 4. Honsa ES*, et al.* (2017) RelA Mutant Enterococcus faecium with
419 Multiantibiotic Tolerance Arising in an Immunocompromised Host. *MBio* 8(1).
420 5. Molina N & van Nimwegen E (2008) Universal patterns of purifying selection
421 at noncoding positions in bacteria. *Genome Res* 18(1):148-160.
422 6. Thorpe HA, Bayliss SC, Hurst LD, & Feil EJ (2017) Comparative Analyses of
423 Selection Operating on Nontranslated Intergenic Regions of Diverse Bacterial
424 Species. *Genetics* 206(1):363-376.
425 7. Kim D*, et al.* (2012) Comparative analysis of regulatory elements between
426 Escherichia coli and Klebsiella pneumoniae by genome-wide transcription
427 start site profiling. *PLoS Genet* 8(8):e1002867.
428 8. Wray GA (2007) The evolutionary significance of cis-regulatory mutations.
429 *Nat Rev Genet* 8(3):206-216.
430 9. Osborne SE*, et al.* (2009) Pathogenic adaptation of intracellular bacteria by
431 rewiring a cis-regulatory input function. *Proc Natl Acad Sci U S A*
432 106(10):3982-3987.
433 10. Marvig RL*, et al.* (2014) Within-host evolution of Pseudomonas aeruginosa
434 reveals adaptation toward iron acquisition from hemoglobin. *MBio*
435 5(3):e00966-00914.
436 11. Cui Z*, et al.* (2014) Mutations in the embC-embA intergenic region contribute
437 to Mycobacterium tuberculosis resistance to ethambutol. *Antimicrob Agents*
438 *Chemother* 58(11):6837-6843.
439 12. Chung JC*, et al.* (2012) Genomic variation among contemporary
440 Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis
441 patients. *J Bacteriol* 194(18):4857-4866.
442 13. Cramer N*, et al.* (2011) Microevolution of the major common Pseudomonas
443 aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol*
444 13(7):1690-1704.
445 14. Jeukens J*, et al.* (2014) Comparative genomics of isolates of a Pseudomonas
446 aeruginosa epidemic strain associated with chronic lung infections of cystic
447 fibrosis patients. *PLoS One* 9(2):e87611.
448 15. Marvig RL*, et al.* (2013) Draft Genome Sequences of Pseudomonas
449 aeruginosa B3 Strains Isolated from a Cystic Fibrosis Patient Undergoing
450 Antibiotic Chemotherapy. *Genome Announc* 1(5).
451 16. Marvig RL, Johansen HK, Molin S, & Jelsbak L (2013) Genome analysis of a
452 transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive
453 mutations and distinct evolutionary paths of hypermutators. *PLoS Genet*
454 9(9):e1003741.

455 17. Marvig RL, Sommer LM, Molin S, & Johansen HK (2015) Convergent evolution
456 and adaptation of Pseudomonas aeruginosa within patients with cystic
457 fibrosis. *Nat Genet* 47(1):57-64.
458 18. Smith EE*, et al.* (2006) Genetic adaptation by Pseudomonas aeruginosa to the
459 airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 103(22):8487-
460 8492.
461 19. Stewart AJ, Hannenhalli S, & Plotkin JB (2012) Why transcription factor
462 binding sites are ten nucleotides long. *Genetics* 192(3):973-985.
463 20. Stover CK*, et al.* (2000) Complete genome sequence of Pseudomonas
464 aeruginosa PAO1, an opportunistic pathogen. *Nature* 406(6799):959-964.
465 21. LaBauve AE & Wargo MJ (2014) Detection of host-derived sphingosine by
466 Pseudomonas aeruginosa is important for survival in the murine lung. *PLoS*
467 *Pathog* 10(1):e1003889.
468 22. Dasgupta N, Lykken GL, Wolfgang MC, & Yahr TL (2004) A novel anti-anti-
469 activator mechanism regulates expression of the Pseudomonas aeruginosa
470 type III secretion system. *Mol Microbiol* 53(1):297-308.
471 23. Mastropasqua MC*, et al.* (2017) Growth of Pseudomonas aeruginosa in zinc
472 poor environments is promoted by a nicotianamine-related metallophore.
473 *Mol Microbiol* 106(4):543-561.
474 24. Gi M*, et al.* (2015) A novel siderophore system is essential for the growth of
475 Pseudomonas aeruginosa in airway mucus. *Sci Rep* 5:14644.
476 25. Toh SM & Mankin AS (2008) An indigenous posttranscriptional modification
477 in the ribosomal peptidyl transferase center confers resistance to an array of
478 protein synthesis inhibitors. *J Mol Biol* 380(4):593-597.
479 26. Kumari H, Balasubramanian D, Zincke D, & Mathee K (2014) Role of
480 Pseudomonas aeruginosa AmpR on beta-lactam and non-beta-lactam
481 transient cross-resistance upon pre-exposure to subinhibitory concentrations
482 of antibiotics. *J Med Microbiol* 63(Pt 4):544-555.
483 27. Turner KH, Wessel AK, Palmer GC, Murray JL, & Whiteley M (2015) Essential
484 genome of Pseudomonas aeruginosa in cystic fibrosis sputum. *Proc Natl Acad*
485 *Sci U S A* 112(13):4110-4115.
486 28. Yen P & Papin JA (2017) History of antibiotic adaptation influences microbial
487 evolutionary dynamics during subsequent treatment. *PLoS Biol*
488 15(8):e2001586.
489 29. Smith PA & Romesberg FE (2007) Combating bacteria and drug resistance by
490 inhibiting mechanisms of persistence and adaptation. *Nat Chem Biol*
491 3(9):549-556.
492 30. Coombes BK (2013) Regulatory evolution at the host-pathogen interface. *Can*
493 *J Microbiol* 59(6):365-367.
494 31. Folkesson A*, et al.* (2012) Adaptation of Pseudomonas aeruginosa to the
495 cystic fibrosis airway: an evolutionary perspective. *Nat Rev Microbiol*
496 10(12):841-851.
497 32. Winstanley C*, et al.* (2005) Genotypic and phenotypic characteristics of
498 Pseudomonas aeruginosa isolates associated with ulcerative keratitis. *J Med*
499 *Microbiol* 54(Pt 6):519-526.

500 33. Blank D, Wolf L, Ackermann M, & Silander OK (2014) The predictability of
501 molecular evolution during functional innovation. *Proc Natl Acad Sci U S A*
502 111(8):3044-3049.
503 34. McNally A*, et al.* (2016) Combined Analysis of Variation in Core, Accessory
504 and Regulatory Genome Regions Provides a Super-Resolution View into the
505 Evolution of Bacterial Populations. *PLoS Genet* 12(9):e1006280.
506 35. Choi KH & Schweizer HP (2006) mini-Tn7 insertion in bacteria with single
507 attTn7 sites: example Pseudomonas aeruginosa. *Nat Protoc* 1(1):153-161.
508 36. Kessler B, de Lorenzo V, & Timmis KN (1992) A general system to integrate
509 lacZ fusions into the chromosomes of gram-negative eubacteria: regulation of
510 the Pm promoter of the TOL plasmid studied with all controlling elements in
511 monocopy. *Mol Gen Genet* 233(1-2):293-301.
512

513 **Figure Legends**

514 **Figure 1:** Pathoadaptive intergenic regions. Regions targeted by mutations involved
515 in host adaptation through parallel evolution across or within clone types. The black
516 squares in the matrix show the intergenic region with parallel mutations in isolates
517 of the respective clone type. The red squares in the matrix show the intergenic
518 region with parallel mutations within isolates of a clone type alone. Squares with
519 striped red color indicate regions in which mutations has been selected both within
520 isolates of that clone type and across other clone types.

521

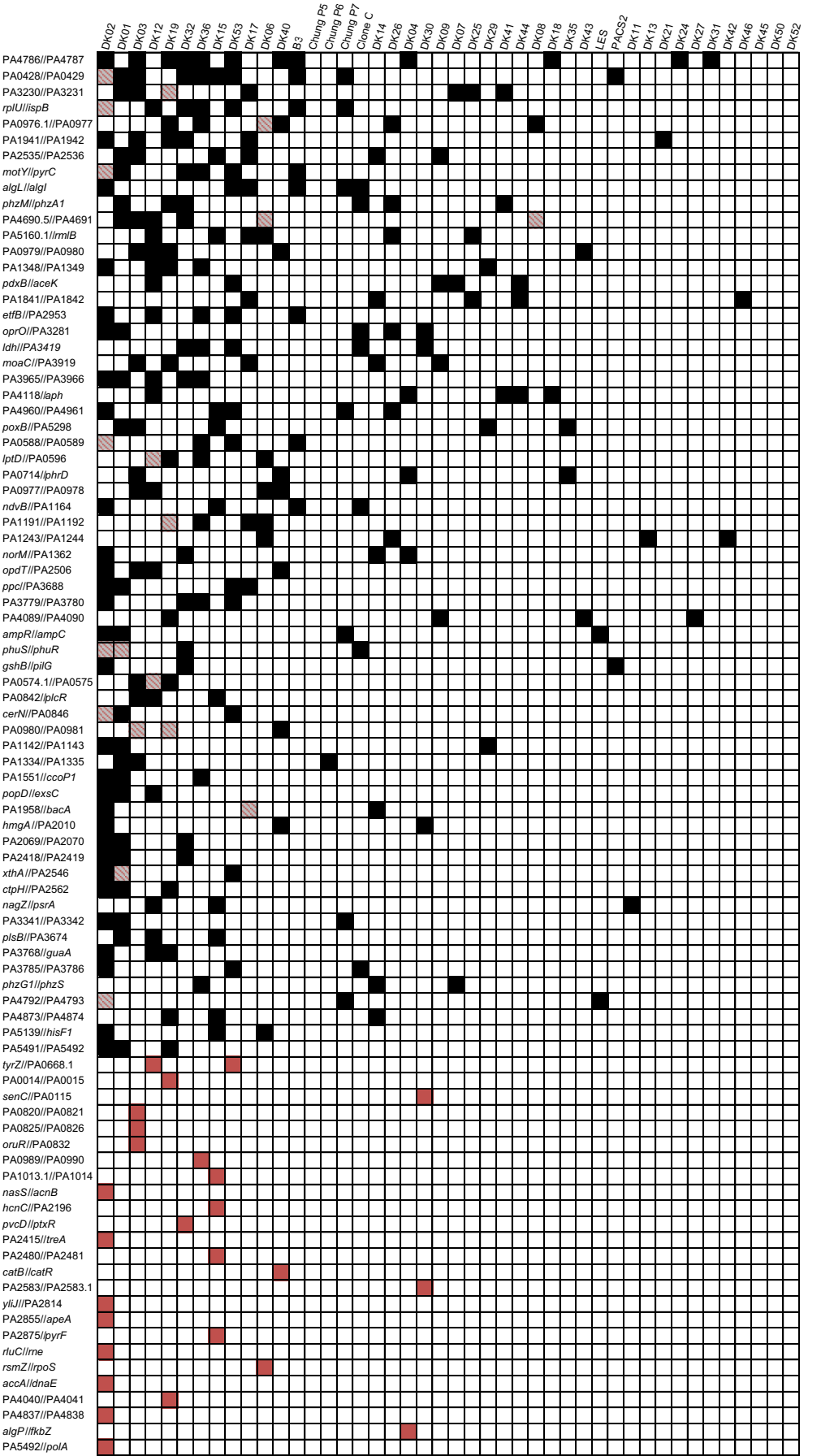
522 **Figure 2:** Orientation and regulatory elements in intergenic regions. a) Overview of
523 the three different orientations of intergenic regions and the possible location of
524 potential elements within each type. b) Distribution of different orientations of
525 intergenic regions (I-III) within PAO1 genome and the pathoadaptive intergenic
526 regions. Two-tailed Fisher's exact test is performed to analyze over-representation
527 or under-representation of certain orientations within pathoadaptive intergenic
528 regions (n = 88). c) Pie chart demonstrating the distribution of putative intergenic
529 elements targeted by pathoadaptive intergenic mutations among regions where the
530 mutation cluster was within any known element (n = 28).

531

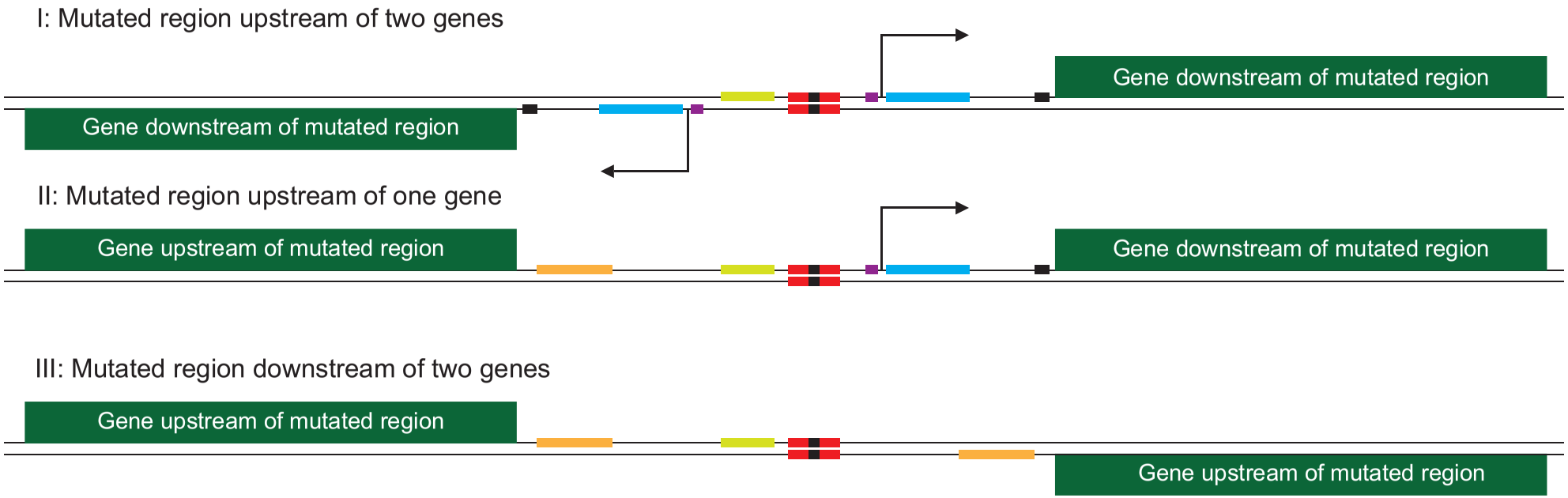
532 **Figure 3:** Intergenic mutations with functional effects on transcription. Expression of
533 *lux* from transcriptional fusions with mutated and wild type alleles were measured at
534 OD_{600} = 0.15 and normalized by cell density. Transcriptional fusions were examined
535 in Luria-Bertani (LB) and ABTGC minimal media. Mean luminescence was calculated
536 for three biological replicates and the relative fold change of mutant versus wild type

537 allele calculated. Statistical analysis of the difference between two means was
538 performed by a two-tailed Student *t* test and the asterisk denotes $P < 0.05$.
539
540 **Figure 4:** Mutations in the intergenic region between *ampC* and *ampR* cause an
541 increased tolerance towards imipenem and ampicillin. The values for Minimal
542 Inhibitory Concentration (MIC) and the constructed mutations in each strain of PAO1
543 are shown. Mutation G-98A upstream *ampC* derives from isolate DK2-CF173-1995.
544 Three mutations G-38A, C-66T and G-78A upstream of *ampC* originate from isolate
545 DK1-P43-M2-2002. A SNP mutation at the start of *ampC* (G7A) in DK1-P43-M2-2002
546 was also constructed in laboratory strain PAO1 to isolate the effect of this mutation
547 and the effect of intergenic mutations from DK1-P43-M2-2002. Error bars indicate
548 standard deviation from three different biological replicates. Double asterisk indicate
549 significant difference between mean MIC of the strains (Two-tailed Student *t* test, P
550 < 0.01).
551
552 **Figure 5:** Contribution and function of coding and intergenic mutations. 52 coding
553 regions are under positive selection in a selected subset of our dataset comprising of
554 474 long-term CF adapted isolates of *P. aeruginosa* (17). In the same isolates, we
555 identify 35 intergenic regions under positive selection for adaptive mutations (Table
556 S8). a) To calculate the relative presence of adaptive mutations within coding or
557 intergenic regions, we divided the number of pathoadaptive features by mutational
558 targets available within each category (89.8% coding and 10.2% intergenic). Final
559 relative frequency for selection of each pathoadaptive feature is shown in the
560 barplot. b) 445 genes are essential for survival of *P. aeruginosa* in CF sputum

561 environment (27). The percentage of these essential genes within 52 pathoadaptive
562 genes and 38 genes downstream of 35 pathoadaptive intergenic regions is
563 demonstrated. Asterisk denotes $P < 0.05$ from two-tailed Fisher's exact test. Red
564 dashed line indicates the percentage of CF essential genes in PAO1.
565


DK02
DK01
DK03
DK12
DK19
DK32
DK36
DK15
DK53
DK17
DK06
DK40
B3
Chung P5
Chung P6
Chung P7
Clone C
DK14
DK26
DK04
DK30
DK09
DK07
DK25
DK29
DK41
DK44
DK08
DK18
DK35
DK43
LES
PACS2
DK11
DK13
DK21
DK24
DK27
DK31
DK42
DK46
DK45
DK50
DK52
PA4786//PA4787
PA0428//PA0429
PA3230//PA3231
rplU//rplB
PA0976.1//PA0977
PA1941//PA1942
PA2535//PA2536
motY//pyrC
algL//algI
phzM//phzA1
PA4690.5//PA4691
PA5160.1//rmlB
PA0979//PA0980
PA1348//PA1349
pdxB//aceK
PA1841//PA1842
etfB//PA2953
oprO//PA3281
ldh//PA3419
moaC//PA3919
PA3965//PA3966
PA4118//aph
PA4960//PA4961
poxB//PA5298
PA0588//PA0589
lptD//PA0596
PA0714//phrD
PA0977//PA0978
ndvB//PA1164
PA1191//PA1192
PA1243//PA1244
norM//PA1362
opdT//PA2506
ppcl//PA3688
PA3779//PA3780
PA4089//PA4090
ampR//ampC
phuS//phuR
gshB//pilG
PA0574.1//PA0575
PA0842//plcR
cerN//PA0846
PA0980//PA0981
PA1142//PA1143
PA1334//PA1335
PA1551//ccoP1
popD//exsC
PA1958//bacA
hmgA//PA2010
PA2069//PA2070
PA2418//PA2419
xthA//PA2546
ctpH//PA2562
nagZ//psrA
PA3341//PA3342
plsB//PA3674
PA3768//guaA
PA3785//PA3786
phzG1//phzS
PA4792//PA4793
PA4873//PA4874
PA5139//hisF1
PA5491//PA5492
tyrZ//PA0668.1
PA0014//PA0015
senC//PA0115
PA0820//PA0821
PA0825//PA0826
oruR//PA0832
PA0989//PA0990
PA1013.1//PA1014
nasS//acnB
hcnC//PA2196
pvcD//ptxR
PA2415//treA
PA2480//PA2481
catB//catR
PA2583//PA2583.1
yliJ//PA2814
PA2855//apeA
PA2875//pyrF
rluC//rne
rsmZ//rpoS
accA//dnaE
PA4040//PA4041
PA4837//PA4838
algP//fkbZ
PA5492//polA

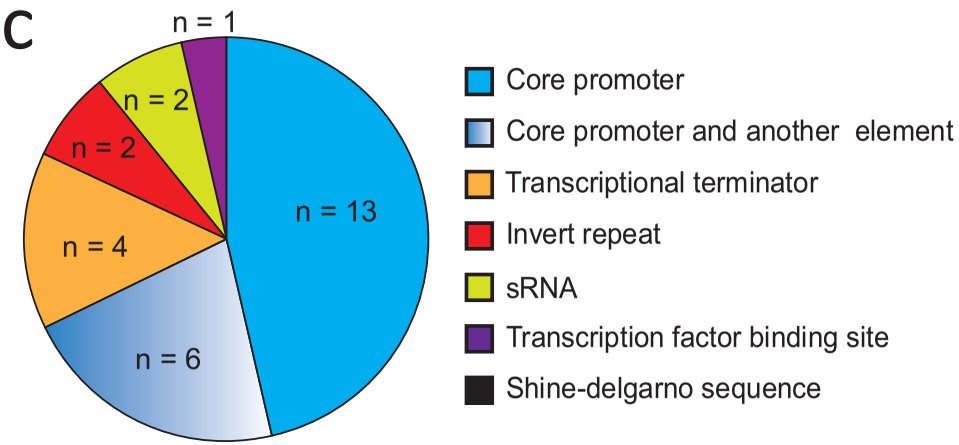
a

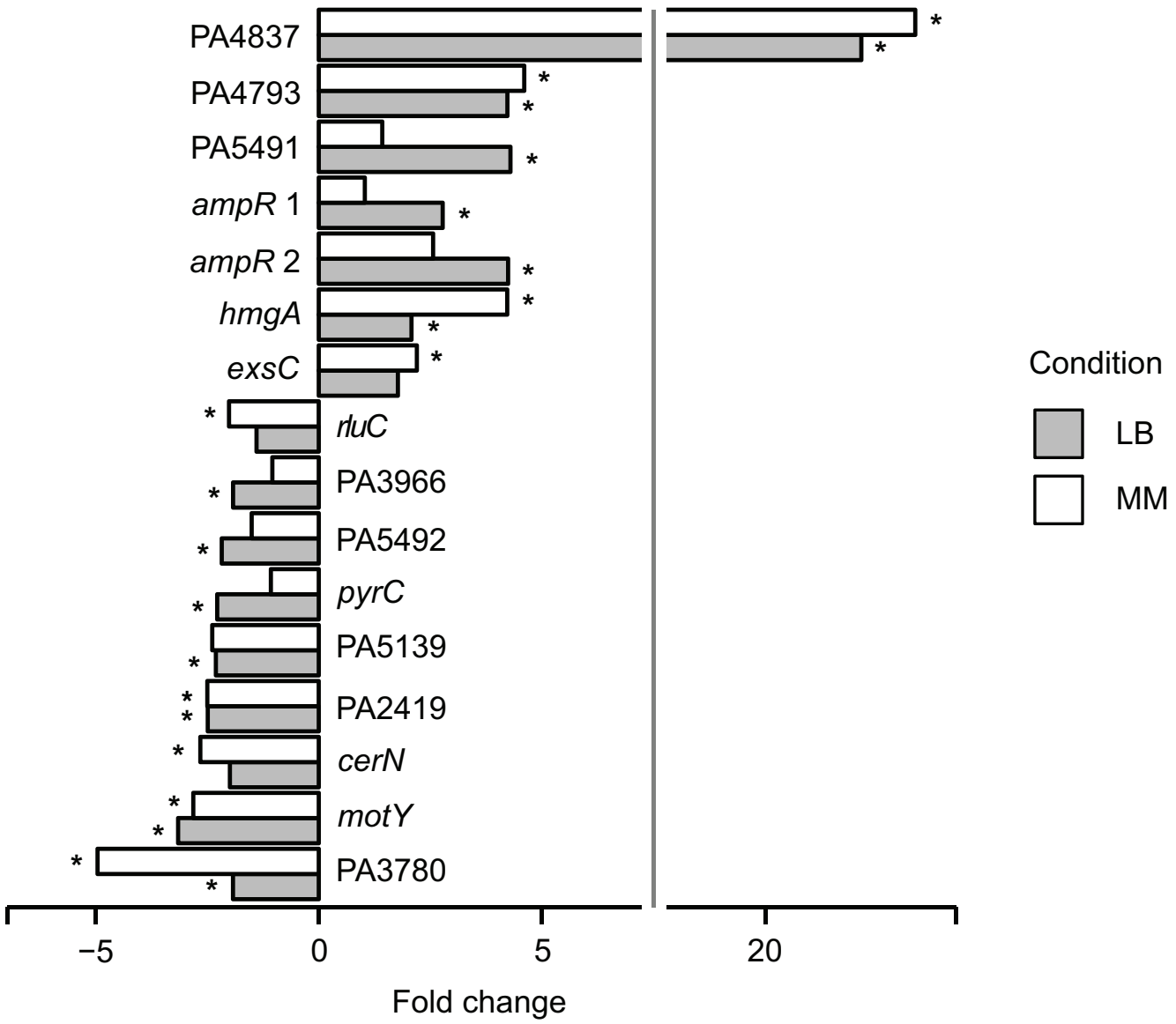


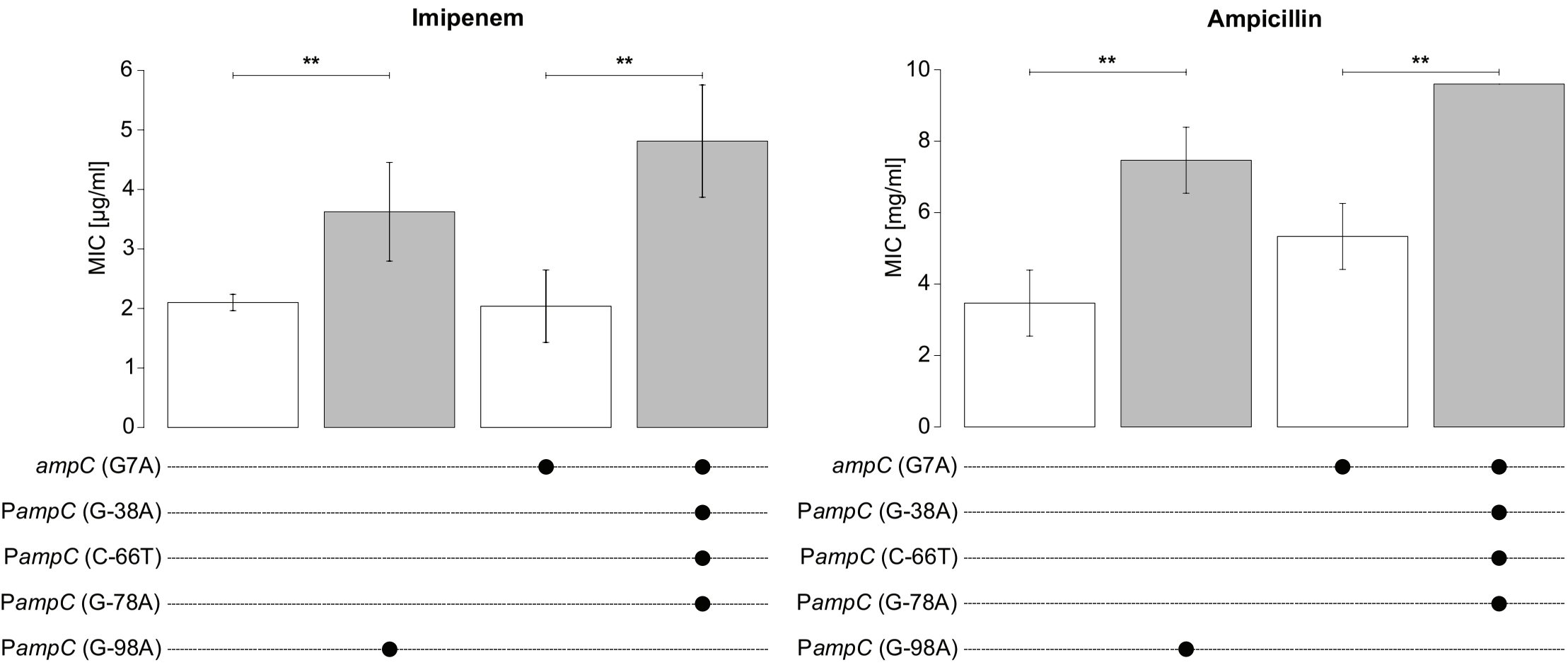
b

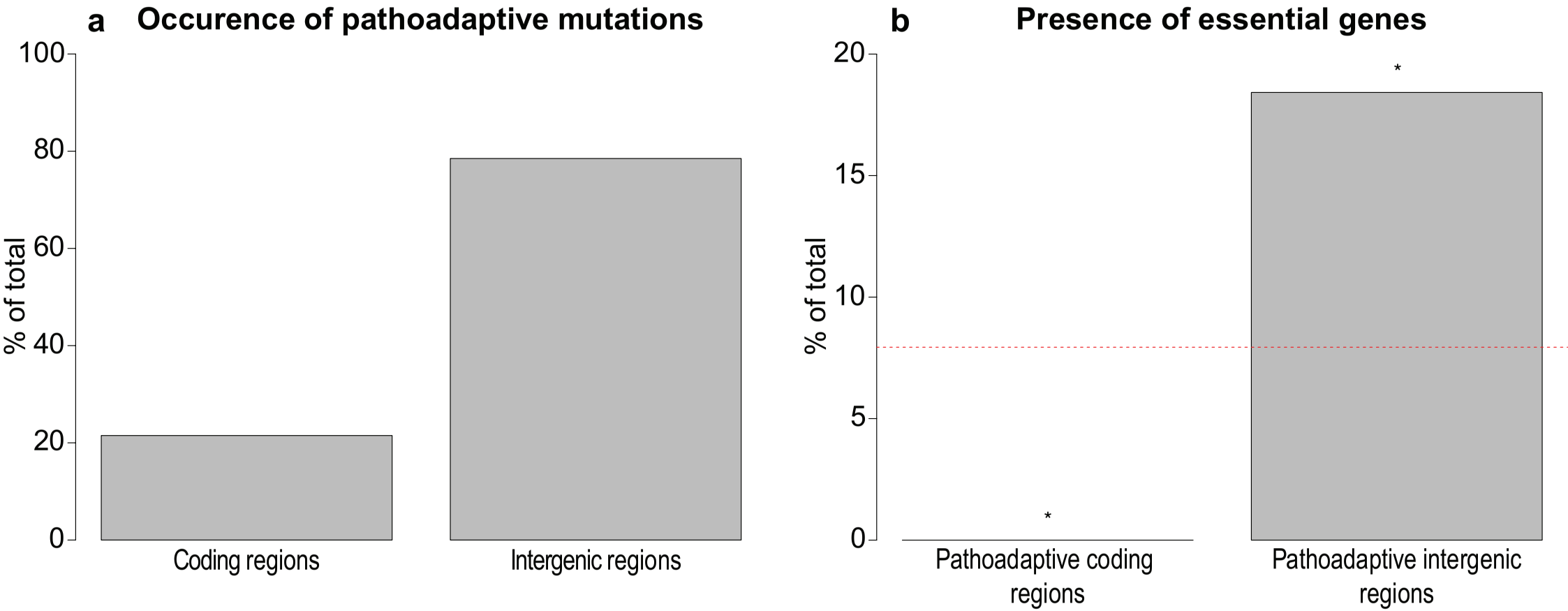
	Present	% of total	PAO1 genome	% of total	Enrichment	*P-value*
I	25	28	812	17	1,64	0,010
II	15	17	755	16	1,06	0,771
III	48	55	3115	67	0,82	0,022

c




PA4837
PA4793
PA5491
ampR 1
ampR 2
hmgA
exsC
rluC
PA3966
PA5492
pyrC
PA5139
PA2419
cerN
motY
PA3780
−5
0
5
20
Fold change
Condition
LB
MM


Imipenem
MIC [µg/ml]
0
1
2
3
4
5
6
**
**
ampC (G7A)
PampC (G-38A)
PampC (C-66T)
PampC (G-78A)
PampC (G-98A)
Ampicillin
MIC [mg/ml]
0
2
4
6
8
10
**
**
ampC (G7A)
PampC (G-38A)
PampC (C-66T)
PampC (G-78A)
PampC (G-98A)


a
Occurence of pathoadaptive mutations
% of total
100
80
60
40
20
0
Coding regions
Intergenic regions
b
Presence of essential genes
% of total
20
15
10
5
0
*
*
Pathoadaptive coding regions
Pathoadaptive intergenic regions