Evolution and regulation of microbial secondary metabolism

Guillem Santamaria[1,3+], Chen Liao[1+], Zhe Wang[2], Kyu Rhee[2], Francisco Pinto[3], Jinyuan Yan[1]*, Joao B. Xavier[1]*

[1]Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[2]Department of Medicine, Weill Cornell Medical College, New York, NY

[3]BioISI – Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisboa, Lisboa, Portugal

[+]these authors contributed equally

[*]Correspondence: yanj2@mskcc.org and xavierj@mskcc.org

Abstract

Microbes have disproportionate impacts on the macroscopic world. This is in part due to their ability to grow to large groups and cooperatively secrete massive amounts of secondary metabolites that impact their environment. Yet, the conditions enabling secondary metabolism without compromising primary needs remain unclear. Here we investigated the biosynthesis of rhamnolipids, a secondary metabolite that *Pseudomonas aeruginosa* makes to decrease the surface tension of surrounding liquid. Using a combination of genomics, metabolomics, transcriptomics, and mathematical modeling we show that biosynthesis of rhamnolipids from glycerol varies inconsistently across the phylogenetic tree; instead, non-producer lineages are also those worse at reducing the oxidative stress of primary glycerol metabolism. The link to oxidative stress explains the inconsistent distribution across the *P. aeruginosa* tree, adding a new

23 layer to the regulation of rhamnolipids—a microbial secondary metabolite important for fitness
24 in natural and clinical settings.

25 Significance

26 The bacterium *Pseudomonas aeruginosa* is a major source of hospital-acquired infections. This
27 pathogen's knack for virulence relies on its ability to multiply and secrete massive amounts of
28 substances that overwhelm microbial competitors and weaken host defenses. It remains unclear
29 how the bacteria conciliate their need to grow and multiply—a need at the individual-level—
30 with their ability to secrete products—a need of the population. Here we combined genomics,
31 metabolomics and mathematical modeling to study the biosynthesis of rhamnolipids, a surfactant
32 that *P. aeruginosa* makes to reduce the surface tension of surrounding liquids. Our study reveals
33 a new link between oxidative stress and rhamnolipid synthesis, which helps explain how this
34 important bacterial product has evolved and how it persists in many lineages of pathogens.

35 Introduction

36 Microorganisms touch practically every human activity, in harmful and beneficial ways:
37 They cause infections (1) and help us digest food (2), make bioterrorism weapons (3) and
38 synthesize life-saving drugs (4), contribute to climate change (5) and clean our wastewater (6).
39 The outsize impact of microbes on the macroscopic world that surrounds them comes from their
40 ability to multiply to populations of billions that cooperate to transform chemicals in vast
41 amounts. Many of those chemicals are not essential to microbial growth, development, and
42 division: they are products of secondary metabolism. But despite the importance, it remains
43 unclear why microbes invest in secondary metabolism (8, 9).

44 Secondary pathways divert precursors, cofactors, and energy from primary metabolism
45 (10), a potential fitness cost to the microorganism that should be disfavored by natural selection.
46 Secondary metabolites secreted extracellularly provide functions to the population such as cell-
47 cell signaling (11), fighting other microbial populations (4), invading a host (12) and scavenging
48 nutrients (13); these functions could bring a fitness advantage to the population (14). Still, a
49 fitness benefit to the population is insufficient to explain the evolution and maintenance of a trait
50 costly to the individual (15). Social evolution theory could provide a quantitative framework for
51 the evolution of microbial secondary metabolism (9): If C is the fitness cost to the cell making
52 the secondary metabolite and B is the average benefit it brings to the population, then secondary
53 metabolism is favored when $rB>C$, where r is the relatedness among different genotypes across
54 the population (16). The relatedness r is often low in microbial populations due to strain mixing
55 and frequent mutations (17, 18), and this means that the cost-to-benefit ratio (C/B) must also be
56 low.
57 Social evolution theory, then, explains why natural selection favors secondary pathways
58 that minimize their burden on the primary metabolism (8). But social evolution does not tell
59 how—the molecular mechanism—microbes achieved a low cost. Many secondary products
60 require multiple enzymes for their biosynthesis (19), and the burden of those pathways on
61 primary metabolism is likely hard to minimize. Microorganisms evolved sophisticated regulatory
62 networks that can sense and respond to—often various—environmental stimuli (20–23). The
63 sophisticated regulation of secondary metabolic pathways explains conditional phenotypes in
64 laboratory culture, such as when secondary metabolites depend on the type and quantity of the
65 nutrients provided, especially the carbon source (24), and when production occurs in specific
66 growth phases, typically the stationary phase (25). The molecular mechanism regulating

67 secondary metabolism provides a level of understanding of the same biological phenomenon
68 complementary to evolutionary theory (26).

69 Here we investigate both the evolution and the regulation of rhamnolipid synthesis in
70 *Pseudomonas aeruginosa*. Rhamnolipids are secondary metabolites that decrease the surface
71 tension of the surrounding liquid (27). A *P. aeruginosa* cell can produce 20% or more of its dry
72 weight in rhamnolipids (28), which is potentially a huge cost to the individual cell (29). But the
73 rhamnolipids secreted collectively by many bacteria enable spectacular social behaviors like
74 swarming (30), climbing over walls (31), killing microbial competitors (32) and breaching the
75 epithelial barriers of a host (33).

76 The first step in rhamnolipid biosynthesis is catalyzed by the RhlA enzyme, which takes
77 a metabolic intermediate from fatty acid synthesis, β-hydroxyacyl-ACP, and produces 3-(3-
78 hydroxyalkanoyloxy) alkanoic acids (HAAs) (34). The RhlB and RhlC enzymes each add one
79 unit of rhamnose to the HAAs to produce mono-rhamnolipids and di-rhamnolipids (35). The
80 expression of the *rhlAB* operon is regulated by a sophisticated network that includes at least three
81 quorum sensing signals (36, 37) and is conditional to nutrients in the media, requiring a high
82 carbon-to-nitrogen or carbon-to-iron ratio (29, 38, 39). This regulation produces a strategy called
83 metabolic prudence that delays the expression of rhamnolipid biosynthetic genes to times when
84 the population is large enough, but the individual cell also has sufficient carbon source (29).
85 Compared to non-regulated constitutive production, the regulation of rhamnolipid biosynthesis
86 genes make it robust against cheating by lowering the *C/B* ratio even when the relatedness *r* is
87 relatively low (40).

88 Our understanding of rhamnolipids biosynthesis comes mostly from laboratory strains,
89 and this leaves clinically relevant questions unanswered. The three genes encoding the secondary

90 pathway—*rhlA*, *rhlB* and *rhlC*—are conserved across the *Pseudomonas* and the *Burkholderia*
91 genera (42). However, the ability to make rhamnolipids can vary widely, even among closely
92 related strains of *P. aeruginosa* (41). What explains this phenotypic variation?

93 Here we investigate the phenotypic diversity of set of *P. aeruginosa* strains isolated from
94 patients with cancer. We use glycerol as carbon source, a relevant nutrient for *P. aeruginosa*
95 infection (43, 44) that puts a substantial strain on *Pseudomonas* primary metabolism (45). We
96 show that the lineages that lost their ability to make rhamnolipids from glycerol grow not less but
97 slower on glycerol, compared to lineages capable of rhamnolipid production. Then we use
98 metabolomics and mathematical modeling to link the inability to reduce oxidative stress
99 produced by primary metabolism in glycerol with the lack of rhamnolipid synthesis. This link
100 adds a new layer to the regulation of rhamnolipid synthesis and explains the inconsistent
101 variation of the surfactant production phenotype across the phylogenetic tree.

102 Results

103 **Surfactant activity lost in strains with impaired primary metabolism.** We investigated the
104 phenotypic variability among clinical isolates in a diverse panel of 31 *P. aeruginosa* strains,
105 including 28 isolates from infected patients (46) and the three type strains PAO1, PA14 and PA7,
106 all of which have their genomes sequenced (**Fig. 1A**). We grew each strain in the same synthetic
107 media using glycerol as the sole carbon source and ammonium sulfate as the sole nitrogen
108 source, producing a carbon-to-nitrogen molar ratio of 7.0, which enabled rhamnolipid synthesis
109 in strain PA14 (38). We classified each strain as non- (and very low, n=8), mild- (n=6) and
110 strong-producers (n=17) based on drop collapse assay, which assess the surfactant activity of
111 each strain's secretions. The strong producers showed similar surface tension to our laboratory

112 strain PA14, non-producers showed surfactant activity indistinguishable to cell-free media, and
113 mild-producers show activities in between. There was no obvious relation between the body site
114 infected in the patient and the surfactant activity phenotype. There was also no coherent
115 distribution with the isolates' phylogeny, which we built using the sequence variation in core
116 genes (**Fig. 1B**; also supported by a poor correlation in Moran's I test, P=0.14). Ancestral
117 reconstruction the surfactant activity trait supports that the ability to synthesize surfactants from
118 glycerol was present in the common ancestor of all strains (**Supplementary Fig. 1**). The genes
119 *rhlA*, *rhlB* and *rhlC*, the genes encoding the rhamnolipid biosynthesis pathway, were present in
120 all 31 isolates, and their sequence was 100% conserved. This indicates that the absence of
121 surfactant activity was not due to a loss of the rhamnolipid biosynthetic pathway. Instead, we
122 hypothesized that the phenotypic variation could be caused by some lineages losing their ability
123 by rewiring their regulatory network to overcome specific selective pressures faced during their
124 evolution.

125 We investigated whether variation in the accessory genome (the set of genes missing in at
126 least one strain) could explain the phenotype differences. Among the 8,290 accessory genes, 354
127 are only absent in those non-producers but present in all mild- and strong-producers
128 (**Supplementary Table 1**). A principal component analysis of the presence-absence profiles
129 showed that our isolates are not grouped by their surfactant phenotype (**Supplementary Fig. 2**),
130 consistent with the conclusion drawn from the core genome.
131 To search for a general explanation, we conducted analyses that showed producers have
132 faster primary metabolisms in the same media of drop collapse assay. This finding was
133 unexpected because, if anything, secondary metabolism should divert resources from primary
134 metabolism (10): strains that produce surfactants would grow slower, not faster. We started by

135 noting intriguing differences in the growth curves of the 31 strains: although there were no
136 notable differences in the population densities reached, the cultures varied widely in their lag-
137 phase and the exponential phase (**Fig. 1C**). The shape of the growth curve, as a whole, was not
138 predictive of surfactant activity: hierarchical clustering using the entire time series failed to
139 separate producers and non-producers (**Supplementary Fig. 3**). But local differences revealed
140 that it's how fast a strain grows, not how much it grows, that distinguishes producers from non-
141 producers. Non-negative matrix factorization (47) decomposed each growth curve as a weighted
142 sum of three basis functions (i.e., features) and showed that—although the growth curves of the
143 rhamnolipid producers (orange lines; including mild- and strong-producers) and non-producers
144 (blue lines) largely overlapped (**Fig. 1C**)—producers had higher weights for base 1 which agrees
145 with a faster growth rate in the exponential phase (**Fig. 1D**). This was also confirmed by dividing
146 each growth curve into three phases (**Fig. 1E** and **Supplementary Fig. 4**) and defining 7
147 quantitative features to characterize each growth phase (**Fig. 1F** and **Supplementary Table 2**).
148 Random Forest classification revealed that the top two features associated with rhamnolipid
149 production were the maximum and the average specific growth rates in phase I, the exponential
150 growth phase (**Fig. 1G**). Features of phase III (the decay phase) such as the phase starting OD
151 which represents the final biomass produced by different strains in phase II, were among the
152 least significant.

153 These analyses suggested that the biosynthesis of rhamnolipids, which peaks when
154 growth slows down from exponential (37–39), is associated with the speed of the exponential
155 phases and hence an efficient primary metabolism on glycerol. The result supports the idea that
156 the non-producer lineages had particular life histories that selected for specific metabolic

157 adaptations, and that those adaptations had the collateral damage of impacting their ability to
158 grow on glycerol as a sole carbon source.

159 **Metabolomics shows differences in intracellular metabolomes associated with loss of**
160 **surfactant biosynthesis.** We then investigated metabolomic differences that distinguish
161 producer and non-producer lineages. We extracted the intracellular metabolome of all our strains
162 except for two non-producers (M55212 and F23197), which grew too slowly during the
163 transition between phase I and phase II when rhamnolipid production begins (**Materials and**
164 **Methods**). LC-MS (Liquid chromatography-Mass spectrometry) revealed 67 metabolites whose
165 identities are known, and their abundances varied significantly across the tested strains
166 (**Supplementary Fig. 5A-B**). Hierarchical clustering and principal component analysis
167 (**Supplementary Fig. 6A-B**) showed a reasonable separation between producers and non-
168 producers. Fitting the data using the Orthogonal Projections to Latent Structures-Discriminant
169 Analysis (OPLS-DA) (48) confirmed a significant correlation between metabolite amount and
170 the surfactant phenotype (**Fig. 3A**, $R^2 = 0.82$, $Q^2 = 0.66$, *p*-value = 5e-4). A Mann-Whitney U
171 test revealed 15 such metabolites statistically significantly different in producers vs non-
172 producers, which we then used to identify pathways perturbed in non-producers using the
173 FELLA algorithm for pathway enrichment analysis (49) (**Supplementary Fig. 7**,
174 **Supplementary Table 3**). The most perturbed pathways were the TCA cycle and amino acid
175 metabolism (**Fig. 3B**). Three out of six metabolites in the TCA cycle—fumarate, succinate and
176 citrate—were significantly changed: fumarate and malate had positive loadings, while those of
177 citrate, succinate, and to a lesser extent cis-aconitate and alpha-ketoglutarate were negative (**Fig.**
178 **3B**). Pyruvate remained relatively constant across all the strains (**Supplementary Fig. 5A**),

179 implying that the differential responses in the TCA cycle were independent from the changes in
180 its upstream central carbon metabolism.

181 Besides the TCA cycle metabolites, most of the annotated compounds in the metabolism
182 of branched chain amino acids (leucine/isoleucine, valine) and sulfur-containing amino acids
183 (cysteine/methionine) were more abundant in non-producers than producers (**Supplementary**
184 **Fig. 7**). A notable exception was formylmethionine (fMet), which was lower in non-producers.
185 We went back to previous metabolomics data (50) and found that the Δ*rhlA* mutant of *P.*
186 *aeruginosa* PA14 also had lower levels of fMet compared with its wild-type (**Supplementary**
187 **Fig. 8**) . Since Δ*rhlA* mutant grows just as fast in glycerol as the wild-type (51) the link between
188 lower fMet and lack of rhamnolipid production should not be simply due to a growth defect.

189 **Theory formalizes the link between oxidative stress, slower growth and loss of rhamnolipid**
190 **synthesis.** Several theoretical arguments support that non-producers might have slower primary
191 metabolisms because they suffer from higher oxidative stress than the producers. First, the major
192 differences include TCA intermediates (**Fig. 3B**). The TCA cycle harbors five enzymes with Fe-
193 S clusters, aconitase A, aconitase B, succinate dehydrogenase, fumarase A, fumarase B (52),
194 which are among the most vulnerable to oxidative stress. If non-producers are less successful
195 with removing the oxidative stress from growth in glycerol, then the stress accumulated could
196 reduce flux through the TCA cycle, which would slow down growth. The accumulation of
197 succinate and depletion of fumarate in non-producers could be due to a reduced activity of
198 succinate dehydrogenase (SDH) under oxidative stress: SDH is a membrane-bound
199 dehydrogenase linked to the respiratory chain—a major site of ROS production in the cell—and
200 also a member of the TCA cycle that catalyzes the oxidation of succinate into fumarate (53).

201 Since SDH contains [2Fe-2S], [3Fe-4S], and [4Fe-4S] clusters (54), ROS that damages Fe-S
202 clusters could decrease SDH activity *in vivo*.
203 Second, we saw that non-producers also included intermediate metabolites in amino acid
204 biosynthetic pathways which are also substrates of Fe-S containing enzymes (**Fig. 3B**). For
205 example, the glutamate synthetase has a large chain (encoded by *gltB*) and a small chain
206 (encoded by *gltD*), and both subunits contain Fe-S clusters that catalyze the production of
207 glutamate from oxoglutarate and glutamine (55). Non-producers had slightly higher oxoglutarate
208 and glutamine as well as lower glutamate than producers.
209 Finally, we used a whole-genome reconstruction to model the intracellular metabolic
210 fluxes during growth in the glycerol medium (56, 57) which supported our observations.
211 Rhamnolipids are produced when carbon is in excess (38, 39). In the simulations, rhamnolipids
212 secretion occurs when C:N flux ratio > 6.3 (**Supplementary Fig. 9**). We then set C:N to 10.0,
213 which exceeds the minimum threshold, and we simulated various levels of redox stress level by
214 changing the flux levels of three redox molecules—NADH (reduced nicotinamide adenine
215 dinucleotide), NADPH (reduced nicotinamide adenine dinucleotide phosphate) and GSH
216 (reduced glutathione)—responsible for the bulk of cellular electron transfer and the main sources
217 of reactive oxygen species (ROS) (58). Although redox stress is ultimately caused by high ROS
218 levels, not fluxes, and ROS was not an explicit variable in our model, we assumed that ROS can
219 perturb intracellular metabolic fluxes, particularly the fluxes of redox molecules, and we then
220 asked how cell growth and rhamnolipid secretion responded to these ROS-mediated indirect
221 perturbations.
222 For all three redox molecules, we found that the maximum growth rate was maintained at
223 an intermediate flux range (redox homeostasis; **Fig. 4A-C**, upper panels, white area). The growth

224 rate decreased gradually as the flux of redox molecules changed in either direction away from
225 this homeostatic interval (gray shading). The reason for the gradual decrease in growth rate is
226 that deficiency or excess of these redox molecules which participate in the redox control of a
227 great variety of biological processes disturbs the primary metabolism for growth. Importantly,
228 before the growth starts slowing down due to an excess in redox species there was an abrupt
229 shutdown in the overflow on carbon into many central carbon metabolites and—importantly— in
230 the secretion of HAAs, mono- and di-rhamnolipids (**Fig. 4A-C**, lower panels). The shutdown of
231 overflow reactions under excess of redox molecules mimics the metabolic state in the
232 rhamnolipid non-producers who lose metabolic plasticity to overflow metabolites without
233 compromising growth. When we forced these secretion fluxes to zero—to simulate a Δ*rhlA*
234 mutant—the simulation did not predict a growth benefit. This result agrees with data showing
235 that the *ΔrhlA* mutant of *P. aeruginosa* strain PA14 has no growth benefit in glycerol media
236 compared with wild-type (51). This also suggests that regaining rhamnolipid secretion should not
237 restore fast growth in glycerol to non-producers, which agrees with the idea that loss of
238 rhamnolipid synthesis is a consequence—not a cause—of the slower growth in non-producers.
239 Notably, most of the overflowed metabolites are carbon-rich molecules, including HAA, mono-
240 and di-rhamnolipid (59). Collectively, these theoretical arguments support a link between
241 rhamnolipids secretion and oxidative stress, where fast growth and rhamnolipid secretion are
242 both impacted by the accumulation of ROS produced by primary metabolism.

243 **Experiments confirm that non-producers are worse at reducing oxidative stress.** Our model
244 suggested that non-producers suffer from higher oxidative stress. *Pseudomonas* have a variety of
245 antioxidant enzymes including catalases, glutathione reductase, NADH peroxidase and cysteine-
246 based peroxidases (60), which were only secreted when the growth rate maintained its maximum

247 value under excess carbon, a finding also consistent with experiments in PA14 and PAO1 (29,
248 38, 39). Isogenic mutants lacking some of those mechanisms are impaired in rhamnolipid
249 synthesis (61). If the slower growth in non-producers is due to impairments in some of those
250 mechanisms this could affect surfactant biosynthesis. To test this hypothesis we compared the
251 dynamics of hydrogen peroxide (H_2O_2) (**Fig. 3D,E**). H_2O_2 is a representative ROS that can
252 diffuse freely between cells and the environment; we can access its dynamics by measuring the
253 fluorescence by the Amplex assay, where the difference between cell cultures and baseline
254 reveals each strain's net ability to reduce H_2O_2 (62). All strains except for M1608 quenched the
255 signal over the first 18 hours (**Fig. 3E**). As expected, the non-producers (blue lines) removed less
256 H_2O_2 than mild rhamnolipid producers (green lines) and strong producers (red lines). To factor
257 out the possibility that the non-producers performed worse due to lower cell density, we
258 calculated the H_2O_2 removal rate per cell and observed a similar relative trend of H_2O_2
259 degradation between non-producers and producers (**Fig. 3F**): a linear mixed-effect model
260 quantified the difference that mild and strong producers have in per-cell H_2O_2 removal rates,
261 which significantly higher than in non-producers ($P < 0.001$; **Fig. 3G**). This supports that
262 rhamnolipid non-producers are suffering from primary metabolism generated oxidative species,
263 which can be caused by high production or slow degradation. Interestingly, the worst strain in
264 this assay, M1608, lacks *katE* (**Supplementary Table 1**), a catalase that degrades H_2O_2 (63).
265 Another non-producer, F5677, lacks the redox sensor *soxR (64)*.
266
267 **RNAseq reveals oxidative stress response in non-producers.** To understand the gene
268 expression changes in rhamnolipid non-producers compared to the producers, we selected 11
269 representative strains (6 strong producers, 3 mild producers and 2 non-producers), in proportion

270 to the number in each category. We then assess their transcriptomics in Phase II when
271 rhamnolipids biosynthesis starts in the producers (**Materials and Methods**). A PCA of the gene
272 expression profiles could not cluster the strains according to their surfactant phenotypes
273 (**Supplementary Fig. 10**). We then turned to RLQ analysis (**Materials and Methods**) to find
274 both genes and pathways associated with rhamnolipid production. At the gene function level, the
275 top 5 pathways enriched in each group do not overlap: The strong producers were enriched in
276 acarbose and validamycin biosynthesis (pau00525), phenazine biosynthesis (pau00405),
277 polyketide sugar unit biosynthesis (pau00523), starch and sucrose metabolism (pau00500), and
278 pentose and glucuronate interconversions (pau00040) (**Fig. 4A**, **Supplementary Table 4**). By
279 contrast, biosynthesis of siderophore group nonribosomal peptides (pau01053), ascorbate and
280 aldarate metabolism (pau00053), monobactam biosynthesis (pau00261), D-glutamine and D-
281 glutamate metabolism (pau00471), and sulfur metabolism (pau00920) were the top 5 pathways
282 enriched in the non-producers (**Fig. 4A**, **Supplementary Table 4**).
283 As expected, the genes in the biosynthesis of rhamnose (*rmlA, rmlB, rmlC, rmlD*),
284 rhamnolipid biosynthesis (*rhlA, rhlB, rhlC, rhlR, rhlI*), quorum sensing (*lasA, lasB, lasI,*
285 *pqsABCDEH*) were significantly lower in the non-producers (**Fig. 4B, Supplementary Fig. 11**).
286 Interestingly, *in vitro* data shows that the expression of genes from the above list, including *lasI,*
287 *lasR, rhlI, rhlR, pqsA, pqsR*, in *P. aeruginosa* PAO1, can be significantly reduced by quorum
288 sensing quencher and ROS molecule H_2O_2 (65), supporting that the loss of the phenotype may be
289 linked to oxidative stress. More importantly, the *zwf* gene encoding glucose 6-phosphate
290 dehydrogenase (G6PDH) was expressed higher in non-producers (**Supplementary Fig. 11**).
291 G6PDH is a major source of NADPH and contributes to antioxidant defenses by reducing
292 oxidized glutathione, and its higher expression supports our hypothesis that non-producers

293 experience higher oxidative stress. Other genes including the peroxide-sensing transcriptional
294 regulator (*oxyR*), catalases (*katA, katB*), alkyl hydroperoxide reductases (*ahpC, ahpF*),
295 glutaredoxin (*grx*), glutathione reductase (*gor*), and thioredoxins (*trxA, trxB1, trxB2*) were not
296 highly expressed in the producers (**Supplementary Fig. 11**), suggesting that the antioxidant
297 defense may be regulated at the metabolic level.
298
299 **Exogenous oxidative stress induces rhamnolipid biosynthesis in a strain that can deal with**
300 **stress produced by primary metabolism**. So far, we have shown that surfactant producers
301 reduce oxidative species better, and we provided a model to explain the link between a faster
302 exponential growth and the ability to make rhamnolipids, indicating that rhamnolipid production
303 may be co-evolving with primary metabolism through the regulation of oxidative stress. Another
304 proof of this co-evolving mechanism is that rhamnolipid production evolves to be part of a wide
305 oxidative stress response. This possibility was suggested by experiments showing that the
306 induction of rhamnolipids synthesis in an engineered strain of *P. putida* reduced the release of
307 CO_2, which was produced using O_2 as electron acceptor, suggesting that rhamnolipids could
308 function as a secondary electron sink (66). To investigate this possibility, we tracked gene
309 expression in one of the rhamnolipid producers—the strain PA14—using a reporter fusion
310 (P_{rhlAB}-GFP) to track the expression of *rhlAB* after adding different levels of H_2O_2 (**Fig. 5**,
311 **Supplementary Fig. 12**). We added H_2O_2 when the cells started entering the exponential phase,
312 instead of in the middle of the exponential phase, to avoid strong perturbation of the measuring
313 signal. Indeed, H_2O_2 delayed cell growth (**Figure 5A, C**) but no difference was observed for the
314 exponential growth rate, indicating that this strain can sustain its primary metabolism despite the
315 oxidative stress. However, the *rhlAB* promoter activity increased as H_2O_2 was added compared

316 to a strain carrying the constitutive promoter ($P_{A1/04/03}$) (**Fig 5F**, orange). This indicated that
317 adding an exogenous redox stress to a strain that can deal with it—since it did not slow down the
318 speed of cell growth in phase I and phase II—can give feedback to increase *rhlAB* expression in
319 the PA14 strain. We next compared this with the strain *ΔrhlA* (**Supplementary Fig. 13**). We
320 observed a similar increase of P_{rhlAB}-GFP expression with increasing H_2O_2 in *ΔrhlA* strain
321 (**Figure 5E,F**, blue). This suggests an attempt to increase rhamnolipid biosynthesis, which this
322 strain is incapable of because it lacks the *rhlA* gene. However, we saw no decrease in the growth
323 rate, indicating that the lack of rhamnolipid synthesis had no detrimental impact on primary
324 metabolism. Interestingly, however, the baseline of the promoter activity increased significantly
325 (1.14x, $p<0.01$) compared to the wild type. This higher expression is not due to the genetic
326 background, because the promoter activity between PA14 and *ΔrhlA* was not significantly
327 different in the absence of H_2O_2 ($p=0.3$). Taken together, these results indicate that rhamnolipid
328 biosynthesis can be induced by additional redox stress, and that the signal for the induction is
329 stronger when rhamnolipid synthesis fails to start, while primary metabolism is unaffected in
330 these conditions.

331 Discussion

332 Our results suggest an expanded model for the regulation of the rhamnolipid secondary
333 pathway to explain the variation of surfactant activity observed in *P. aeruginosa* isolates (**Fig. 6**).
334 A microbe can produce secondary metabolites such as rhamnolipids only after it meets its
335 primary requirements of energy production, biomass synthesis and reduction of oxidative stress
336 (**Fig. 6A**) and has excess carbon source. The lineages that retained their ability to synthesize
337 rhamnolipids were also those capable of reducing the oxidative stress created by primary
338 metabolism in glycerol, and they could overflow any excess carbon to secondary rhamnolipid

339 metabolism (**Fig. 6B**). The loss of rhamnolipids in non-producer lineages was distributed
340 inconsistently across the phylogenetic tree, possibly due to the life histories of each lineage.
341 Consistent with the model, those non-producers had impaired abilities to reduce oxidative stress,
342 a slower exponential growth indicating a deficient primary metabolism, perturbed levels of TCA
343 cycle metabolites consistent with oxidative damage and accumulated intermediates of amino acid
344 synthesis—also consistent with a slower primary metabolism. Those burdens prevent the cell
345 from producing rhamnolipids as the resources can be prioritized for primary metabolism, as we
346 observed that the OD at the end of exponential growth is not a good distinguisher for producers
347 and non-producers (**Fig. 6C**).

348 What were the particular selective pressures that led to the non-producer lineages? The
349 answer to this question remains less clear. Non-producers lacked genes (**Supplementary Table**
350 **1**) that suggested each non-producer lineage had a particular path which led to their phenotype.
351 Importantly, though, all three genes of the secondary pathway (*rhlA*, *rhlB*, *rhlC*) remained intact,
352 and even non-producer lineages might be able to make surfactants in conditions different from
353 those we analyzed here. More likely, the impaired ability to reduce oxidative stress—which we
354 linked to the inability to make rhamnolipids from glycerol—is a maladaptive consequence of
355 specific adaptations experienced by non-producer lineages. Since the strains were isolated from
356 hospitalized patients—a host-associated environment where we may expect stresses such as ROS
357 imposed by the immune system in their fight against pathogens (67) and by antibiotics treatment
358 (65)—it is unlikely that the selection was for a worse response to oxidative stress. The producer
359 lineages could quench oxidation faster than non-producers. Our mathematical model explains
360 this by noting that accumulation of oxidative species burdens primary metabolism and shuts
361 down carbon overflow to rhamnolipids.

362 Experiments in PAO1 and clinical isolates showed previously that *P. aeruginosa*
363 decreases expression of the key quorum-sensing genes (*lasI*, *lasR*, *rhlI*, *rhlR*, *pqsA* and *pqsR*) in
364 response to oxidative stress caused by exposure to H_2O_2 (65). Considering that quorum sensing
365 regulates hundreds of genes (68), shutting down rhamnolipid secretion could be part of a broader
366 response to reduce oxidative stress, and saves resources for primary metabolism. But for a strain
367 capable of reducing the stress produced by glycerol metabolism, adding an extrinsic oxidative
368 stress induced expression of the rhamnolipid synthesis operon *rhlAB*. Lacking the Δ*rhlA* gene
369 increased the baseline expression, but not the rate of *rhlAB* increase with H_2O_2, suggesting an
370 intricate regulation. The lack of *rhlA* did not impact growth compared to the wild type even in
371 the highest H_2O_2 concentrations. This is also consistent with our flux-balance analysis of
372 *Pseudomonas* metabolism, where shutting down rhamnolipid secretion did not impact the growth
373 rate because the excess carbon could be alternatively secreted in other forms such as acetate.
374 We detected nuanced consequences of deleting rhamnolipid synthesis in the strain PA14,
375 one of the lineages capable of dealing with glycerol stress. The Δ*rhlA* isogenic mutant of PA14
376 had lower levels of fMet (50) which, interestingly, also happened in the non-producer lineages
377 analyzed here. fMet plays a role in translation initiation, and in the quality control mechanisms
378 that degrade misfolded proteins (69). *Bacillus subtilis* lacking Formyl-Methionine Transferase—
379 the enzyme attaching a formyl group to methionine loaded on $tRNA^{fmet}$—are more sensitive to
380 H_2O_2 and defective in swarming (70). This link between fMet, oxidative stress and the loss of
381 rhamnolipid synthesis is intriguing and warrants further research. Also the Δ*rhlA* mutant had
382 higher levels of gamma-Glutamylcysteine (**Supplementary Fig. 8**), the immediate precursor of
383 glutathione. Since glutathione is a well-known antioxidant that protects cells from oxidative
384 damage (71), the result could indicate a mild oxidative stress. The Δ*rhlA* mutant expressed *rhlAB*

385 at a higher level than the wild-type, suggesting a futile attempt to synthesize rhamnolipids. The
386 fatty acid biosynthesis pathway that provides precursors for rhamnolipid production regenerates
387 NAD(P)+ from NAD(P). Lack of rhamnolipid synthesis could perturb this pathway causing
388 minor increases in oxidative stress.

389 The new model augments our model of rhamnolipid secondary metabolism (**Fig. 6**), helps
390 interpret the inconsistent distribution of rhamnolipid synthesis across the *P. aeruginosa* tree, and
391 improves our understanding the sophisticated regulation that microbes have to balance the
392 individual costs and population benefits of secondary metabolism. There are many products of
393 secondary metabolism that contribute to the outsize impact that microbes have on the
394 macroscopic world. Products of one microbial species can impact other species—cooperatively
395 or competitively—potentially amplifying microbial functions in diverse communities such as the
396 human microbiome (59). We've only sampled the space of microbial secondary metabolism (7),
397 but a general guiding principle emerges: natural selection will favor sophisticated regulation that
398 conciliates the individual-level costs and the population-level benefits of secondary metabolism.

399

400 ## Materials and Methods

401 **Media.** Glycerol synthetic media were prepared with 800 ml of Milipore water, 200 ml of 5X
402 minimal salts buffer, 1ml of 1 M magnesium sulphate, 0.1 ml of calcium sulphate with 0.5 gN/l
403 nitrogen (ammonium sulphate), iron at 5 μM (iron III sulphate) and 3.0 gC/l glycerol, respectively..
404 5X stock minimal salts buffer was prepared with 12 g of Na_2HPO4 (anhydrous), 15 g of KH_2PO4
405 (anhydrous) and 2.5 g of NaCl into 1 l water.

406 **Growth curve assays.** The clinical isolates were inoculated in 3 mL of lysogen broth (LB) and
407 incubated at 37ºC overnight with shaking at 250 rpm. 500 µL of cell culture was pelleted and
408 washed 3 times with PBS. 0.0025 OD_{600} units were inoculated into 3.0 gC/l glycerol minimal
409 medium in BD Falcon (BD Biosciences, San Jose, CA) 96 well flat-bottom plates, with 150 µL of
410 suspension per well. The plate was incubated during 48 hours at 37ºC in a Tecan Infinite M1000
411 or Tecan Infinite M1000 Pro plate reader (Männedorf, Switzerland), with an orbital shaking of 4
412 mm of amplitude. OD_{600} was measured in 10 minutes intervals.

413 **Rhamnolipid production quantification.** The production of rhamnolipids was assessed by drop-
414 collapse assay (72, 73). We placed 50 µL of the culture's supernatant on a polystyrene surface (the
415 lid of a 96 well plate). The presence of rhamnolipids decreases the surface tension of the liquid,
416 making the drop collapse. We considered a strain rhamnolipid non-producer if the drop does not
417 spread, a mild-producer if spreads, but not enough to pass over the lid's circle that corresponds to
418 the wells, and producer if spreads over it.

419 **Pangenome analysis** The pangenomes of our clinical *P. aeruginosa* isolates were identified using
420 the roary tool (74) using the output of genome annotation from prokka (75). The core genes need
421 to have at least 95% identity using blastp and to be present in at least 99% of all genomes, and
422 then were aligned with the MAFFT method (76).

423 **Phylogenetic analysis.** The phylogenetic tree of clinical *P. aeruginosa* strains was constructed
424 from core genomes as previously described (46). Moran's I test was carried out using the ape
425 package in R (77). The ancestral state of swarming and rhamnolipid production was reconstructed
426 using corHMM package (78).

427 **Growth curve analysis.** We first normalized the growth curves in different experiments (plates)
428 by matching their corresponding controls (PA14). The plate-specific coefficients were obtained by
429 solving a linear regression model 'log(OD of PA14) ~ Plate + Time' (the Wilkinson notation)
430 using Matlab function fitlm, where Plate is plate ids and Time is time points (both are categorical
431 variables). Then the growth curves in each plate were normalized by multiplying the adjustment
432 coefficient of the plate. These normalized growth curves were then smoothed by Savitzky-Golay
433 filter (sgolayfilt function in Matlab). The growth phases (phase I, II, III) were determined by the
434 following rules: (phase I): first-order derivative >0 and second-order derivative >0; (phase II) first-
435 order derivative >0 and second-order derivative <0; (phase III) first-order derivative <0. The mean
436 and maximum specific growth rate for each growth phase were estimated from the best-fit model
437 (highest R2) among three models of bacterial growth curves: (1) exponential model $y = y_0 +$
438 $y_1 e^{\mu t}$; (2) Zwietering-Logistic model $y = y_0 + (A - y_0)\left(1 + e^{\frac{4\mu}{A}(\lambda - t)+2}\right)^{-1}$; (3) Zwietering-
439 Gompertz model $y = y_0 + (A - y_0)e^{-e^{\frac{\mu e}{A}(\lambda - t)+1}}$, where $y_0, y_1, \mu, \lambda, A$ are the parameters to fit.
440 Both non-negative factorization and random forest regression were performed in python, by
441 sklearn.decomposition.NMF and sklearn.ensemble.RandomForestClassifier respectively.

442 **Metabolite extraction.** All *P. aeruginosa* strains were grown until the end of exponential phase
443 of growth in glycerol minimal medium. Bacteria was then loaded into 0.25 µm nylon membranes
444 (Millipore) using vacuum, transferred to pre-warmed hard agar plates with the same medium
445 composition and incubated at 37ºC during 2.5 h. The filters were then passed to 35 mm polystyrene
446 dishes (Falcon) with 1 mL of 2:2:1 methanol:acetonitrile:H_2O quenching buffer and incubated
447 there during 15 minutes on dry ice. Cells were removed by scraping and the lysate containing

448 quenching buffer was transferred to 1.5 mL tubes and centrifuged at 16000 rpm for 10 minutes at
449 4ºC. Supernatant transferred to fresh tubes and stored at -80ºC.

450 **Metabolomic data preprocessing.** The extracts were profiled using liquid-chromatography
451 coupled to mass spectrometry (LC-MS), identifying a total of 92 compounds (Supplementary Fig.
452 5). Some compounds contained missing values. These missing values in metabolite abundance can
453 be (1) truly missing; (2) present in a sample but its level is below detection limit; (3) present in a
454 sample at a level above the detection limit but missing due to failure of algorithms in data
455 processing. Here we assume that a metabolite with missing values in all three replicates is truly
456 missing in the sample and removed from our analysis (Supplementary Fig. 5). However, if the
457 missing values were only found in one or two replicates, the missing values were imputed by the
458 average of the non-missing values. After that imputation all compounds with missing values were
459 removed (Supplementary Fig. 5).

460 The peak areas were normalized using Cross-Contribution Compensating Multiple Standard
461 Normalization (CCMN) (79) with NormalizeMets R package (80). This method relies on the use
462 of multiple internal standards. Since LC-MS lacks such internal standards, we used instead a set
463 of metabolites assumed to be constant across all the strains. They were selected with a Kuskal-
464 Wallis test, adjusting the *p*-value with Benjamini-Hochberg method. The ones with a *p*-value
465 above 0.05 were considered constant (pyruvate, methylglyoxal, (S)-2-Acetolactate, Tyramine, D-
466 Glucose, (S)-Lactate, N-acetyl-L-glutamate 5-semialdehyde, 4-Aminobutyraldehyde and
467 Glycine), therefore after the normalization step they were removed (indicated in red,
468 **Supplementary Fig. 5A**). The processed area peaks for all metabolites are included in
469 **Supplementary Table 5**.

470 **Unsupervised analysis of metabolomic data.** The hierarchical clustering of the normalized
471 metabolomic data was performed using gplots R package (81), with Euclidean distance and Ward's
472 aggregation method. The clustering was performed with all metabolites; however, 16 metabolites
473 (indicated in black, **Supplementary Figure 4**) with unknown compound identity were removed from
474 **Supplementary Figure 5** and downstream analyses. These unknown compounds are not artifacts
475 arising from LC-MS instruments since their peaks were detected in all strains. Fumarate and
476 guanosine, which were only putative initially, were bioinformatically confirmed as all other
477 compounds with the same molecular weight are produced by enzymatic reactions missing in our
478 clinical isolates. The Principal Component Analysis was done with R's base method, using also
479 the normalized data and previously removing the 16 metabolites with unknown identity. The plot
480 was obtained using ggplot2 R package (82).

481 **Metabolic pathway enrichment.** The differential metabolites between rhamnolipid producers and
482 non-producers was determined by a Mann-Whitney test, with *p*-values adjusted with Benjamini-
483 Hochberg method and a significance level of 0.05. These compounds were fed to FELLA
484 algorithm (49, 83). FELLA retrieves a graph describing the relationships among pathway, module,
485 enzyme, reaction and compounds of *Pseudomonas aeruginosa* strain UCBPP-PA14 from the
486 KEGG database. The graph was then used as an input network of differential compounds for its
487 diffusion algorithms (84). The output of FELLA consists of all subnetworks of the entries predicted
488 to have a high connectivity with the differential compounds. We filtered the subnetwork entries to
489 keep only metabolic pathways. The entries shown in the supplementary table 3 are the ones with
490 a significant probability of receiving part of the simulated flux.

491 **OPLS-DA model.** OPLS-DA model of metabolomics data was built using ropls R package (85),
492 fixing the number of orthogonal components to 3. R^2 and Q^2, key parameters for assessing the

493 validity of the model, were assessed with 7-fold cross validation. The significance of the model
494 was determined by permutation test (n = 2000). The *p*-value corresponds to the proportion of Q^2_{perm}
495 above Q^2. With a *p*-value below 0.05 we considered the model significant. The loadings of the
496 predictive component of the model were extracted to determine how each metabolite contributes
497 to the separation according to the phenotype.

498 **Genome-scale modeling.** Custom Python codes were developed with the COBRApy package (86)
499 to carry out all metabolic flux modeling and simulations in the paper. Since iJN1411 model was
500 developed for *Pseudomonas putida*, we removed genes and associated reactions that are missing
501 in all our strains but present in the iJN1411 model. The futile cycles involving NADH, NADPH,
502 and GSH were also removed. The modified iJN1411 model was further expanded by adding
503 rhamnolipid biosynthesis pathway involving 9 new metabolites and 12 new reactions.

504 The boundary fluxes of the model were set to mimic the composition of the glycerol minimum
505 medium. For C:N=10, the lower bounds of glycerol and ammonium fluxes were set to -10 and -3
506 respectively. For C:N=3, the lower bounds were set to -3 and -3 respectively. The flux unit is
507 mmol/gDW/h throughout the paper. To constrain the total producing flux of NADH (the same for
508 NADPH and GSH) at a certain value C, we first defined a binary variable X_k for each NADH-
509 involving reaction k to indicate whether NADH is produced by this reaction. Given the
510 stoichiometric coefficient of NADH in this reaction (s_k) and its flux value (f_k), the mathematical
511 constraints for X_k was set by $X_k = 1$ for $s_k f_k > 0$ and $X_k = 0$ otherwise. Therefore, the constraint
512 that equalizes the total NADH producing flux and a constant C is simply $\sum_k s_k X_k f_k = C$. However,
513 both X_k and f_k are variables and such quadratic constraint has not yet been supported by
514 COBRApy. We overcame this difficulty by defining $A_k = X_k f_k$ and linearized the product with

515 the following two inequalities: $X_k l_b \leq A_k \leq X_k u_b$ and $f_k - (1 - X_k)u_b \leq A_k \leq f_k - (1 -$
516 $X_k)l_b$, where l_b and u_b are the lower and upper bounds of f_k. The two constraints ensure that $A_k =$
517 0 when $X_k = 0$ and $A_k = f_k$ when $X_k = 1$. The minimum/maximum flux values of byproduct
518 secretion were simulated by flux variability analysis at maximum growth rate.

519 **Detection of hydrogen peroxide (H_2O_2)**. The H_2O_2 level in the extracellular medium was
520 quantified with Amplex® Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen, Carlsbad,
521 USA, Catalog no. A22188). 500 µL of cell suspension were spinned down after overnight growth
522 in LB medium, washed twice in PBS and normalized to OD 1. Each reaction was done in a final
523 volume of 100 µL, with a final concentration of 50 µM of Amplex® Red reagent, 0.1 U/mL of
524 HRP (Horseradish Peroxidase) and 0.2 cell OD, in glycerol synthetic medium, in BD Falcon (BD
525 Biosciences, San Jose, CA) 96 well flat-bottom plates. The first column of the plate corresponded
526 to the reaction without cells (the volume was substituted by PBS), and the last column instead of
527 cells contained H_2O_2 (final concentration 10 µM). OD_{600} and fluorescence 530/590 nm was
528 measured in 10 minutes intervals (48 h 37ºC).

529 **RNA-seq and data process.** The *P. aeruginosa* strains were inoculated into LB and grew with
530 agitation for 16h. The overnight culture was diluted in glycerol synthetic medium the next day and
531 grew to late exponential phase to harvest RNA. Cells were harvested by centrifuge and then
532 immediately resuspended in RNAprotect (QIAGEN). The RNA was extracted using the RNeasy
533 Mini Kit from QIAGEN. The standard RNA sequencing was done by Genewiz (Genewiz, Inc;
534 South Plainfield, NJ) using pair-end indices of 150bp read length on an Illumina HiSeq platform.
535 The raw fastq files first went through quality check by fastQC (87), and aligned with STAR (88).
536 The final output was generated using DEseq2 (89) package in R. The read counts for each gene is

537 included in **Supplementary Table 6**. The SRA accession number for the RNAseq data is listed in
538 **Supplementary Table 7**.

539 **RLQ analysis.** To test for the functional associations between genes and rhamnolipid production,
540 we adopted the RLQ analysis (ade4 package in R) that was developed for identifying the
541 associations between traits and environmental variables. Here we applied it to RNAseq analysis to
542 understand the main co-structures between gene functions and phenotypes mediated gene
543 expression. The theory of RLQ analysis is described in detail elsewhere (90). Briefly, it builds on
544 the simultaneous ordination of three tables: a R table (categorical) describing rhamnolipid
545 production (strong-, mild- and non-producers) for 11 selected isolates (each replicate was treated
546 as an independent observation), a L table (continuous) describing the RNAseq counts of 1,474
547 core genes shared among the 11 isolates, and a Q table (binary) describing 120 KEGG pathways
548 for the 1,474 genes. We first applied corresponding analysis (CA) to table L and Q and principal
549 component analysis (PCA) to table R. In the CA of R and Q, each strain and gene were weighted
550 by their total RNA expression levels respectively. The three separate ordinations are then used as
551 inputs to the rlq function. The output of the function includes coefficients for the gene functions
552 and phenotypes (loadings) as well as the scores of strains (out of phenotypes) and scores of genes
553 (out of gene functions). RLQ analysis maximizes the squared cross-covariances between the two
554 sets of scores weighted by gene expressions.

555 **Hydrogen peroxide (H_2O_2) degradation.** H_2O_2 was removed from environment by cells which
556 degrade H_2O_2 intracellularly. The cumulative H_2O_2 removal curve was calculated by subtracting
557 the values of emission of the wells containing each strain from the values of emission of the wells
558 without cells or H_2O_2. Then they were normalized to the relative scale by dividing their values by
559 the corresponding values (at the same time points) of the control strain (PA14) in the same

experiment (plate). These normalized curves were then smoothed by the Savitzky-Golay filter (sgolayfilt function in Matlab). The H_2O_2 removal rate per cell was calculated by dividing the first-order derivative of the cumulative removal curves by OD values. To estimate the effect of rhamnolipid production on per-cell H_2O_2 removal rate, we developed a linear mixed-effect model that extends the simple linear regression model by considering both fixed and random effects. The model was specified by the Wilkinson notation 'RemovalRate ~ Time + RHL + (1|CurveID)', where RemovalRate is per-cell H_2O_2 removal rate, Time is time point (categorical variable), RHL is categorical rhamnolipid production (non-producer, mild-producer, and strong-producer), and CurveID is the unique ID for each strain and replicate. The expression '(1|CurveID)' indicates that CurveID was modeled as random effects. The model was solved by fitlme in Matlab.

Promoter activity and growth rate comparison. *P. aeruginosa* strains were inoculated into LB and overnight cultures were harvested and washed with PBS three times. Cells were diluted to OD_{600}=0.1 in synthetic media and monitored using TECAN plate reader for both OD_{600} and GFP. Upon cells entering the exponential phase, H_2O_2 was added to the cells, which continued to be monitored over 48h. After identifying growth phase II, the promoter activity was calculated as: $P = \frac{dGFP}{dt} \cdot \frac{1}{OD}$. The ratio of promoter activity in each experiment was extracted as $R_p = \frac{median(PrhlAB)}{median(PA1/04/03)}$. Similarly, , the growth rate was calculated as $G = \frac{dOD}{dt} \cdot \frac{1}{OD}$, and the ratio of strains with different promoters was calculated as $R_G = \frac{median(G_{PrhlAB})}{median(G_{\frac{PA1}{04}/03})}$. The effect of H_2O_2 was estimated by fitlme in Matlab as: $\log(R) \sim a_0 + \beta \cdot [H_2O_2] + \epsilon$, where R is the measured ratio of either promoter activity or growth rate, and ϵ is denoted as the random effect of experiments.

Data availability

All experimental and simulation data that support the conclusions of this study are available from the main text and Supplementary Information. The source codes for generating the figures of this study are available from:
https://github.com/guisantagui/code_PA_paper;
https://github.com/liaochen1988/Source_code_for_Pseudomonas_Metabolomics_Paper;
https://github.com/Jinyuan1998/PA_metabolomics_rhamnolipids_SourceCode.

Acknowledgements

This work was supported by NIH grants U01 AI124275 and R01 AI137269-01 to J.B.X. F.R.P. and G.S. were partially supported by UIDB/04046/2020 and UIDP/04046/2020 Centre grants from FCT, Portugal (to BioISI) and by the LungCARD project (Grant Agreement nº: 734790, H2020-MSCA-RISE-2016). G.S. is recipient of a fellowship from BioSys PhD programme PD65-2012 (Ref SFRH/BD/142899/2018) from FCT (Portugal). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. D. M. Morens, G. K. Folkers, A. S. Fauci, The challenge of emerging and re-emerging infectious diseases. *Nature* **430**, 242–249 (2004).
2. I. Rowland, *et al.*, Gut microbiota functions: metabolism of nutrients and other food components. *Eur. J. Nutr.* **57**, 1–24 (2018).
3. D. A. Henderson, The looming threat of bioterrorism. *Science* **283**, 1279–1282 (1999).
4. A. L. Demain, A. Fang, The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **69**, 1–39 (2000).

603 5. R. D. Bardgett, C. Freeman, N. J. Ostle, Microbial contributions to climate change through
604 carbon cycle feedbacks. *ISME J.* **2**, 805–814 (2008).

605 6. B. Kartal, J. G. Kuenen, M. C. M. van Loosdrecht, Engineering. Sewage treatment with
606 anammox. *Science* **328**, 702–703 (2010).

607 7. A. L. Demain, Importance of microbial natural products and the need to revitalize their
608 discovery. *J. Ind. Microbiol. Biotechnol.* **41**, 185–201 (2014).

609 8. T. Cavalier-Smith, Origins of secondary metabolism. *Ciba Found. Symp.* **171**, 64–80;
610 discussion 80 (1992).

611 9. D. B. Kell, A. S. Kaprelyants, A. Grafen, Pheromones, social behaviour and the functions
612 of secondary metabolism in bacteria. *Trends Ecol. Evol.* **10**, 126–129 (1995).

613 10. Q. Yan, *et al.*, Secondary Metabolism and Interspecific Competition Affect Accumulation
614 of Spontaneous Mutants in the GacS-GacA Regulatory System in *Pseudomonas protegens*.
615 *MBio* **9** (2018).

616 11. N. Dufour, R. P. Rao, Secondary metabolites and other small molecules as intercellular
617 pathogenic signals. *FEMS Microbiol. Lett.* **314**, 10–17 (2011).

618 12. T. I. Nicas, B. H. Iglewski, The contribution of exoproducts to virulence of *Pseudomonas*
619 *aeruginosa*. *Can. J. Microbiol.* **31**, 387–392 (1985).

620 13. M. Albelda-Berenguer, M. Monachon, E. Joseph, Siderophores: From natural roles to
621 potential applications. *Adv. Appl. Microbiol.* **106**, 193–225 (2019).

622 14. R. A. Maplestone, M. J. Stone, D. H. Williams, The evolutionary role of secondary
623 metabolites--a review. *Gene* **115**, 151–157 (1992).

624 15. K. R. Foster, The sociobiology of molecular systems. *Nat. Rev. Genet.* **12**, 193–203
625 (2011).

626 16. S. A. West, A. S. Griffin, A. Gardner, S. P. Diggle, Social evolution theory for
627 microorganisms. *Nat. Rev. Microbiol.* **4**, 597–607 (2006).

628 17. S. B. Andersen, R. L. Marvig, S. Molin, H. Krogh Johansen, A. S. Griffin, Long-term
629 social dynamics drive loss of function in pathogenic bacteria. *Proc Natl Acad Sci USA* **112**,
630 10756–10761 (2015).

631 18. S. B. Andersen, *et al.*, Privatisation rescues function following loss of cooperation. *eLife* **7**
632 (2018).

633 19. A. Osbourn, Secondary metabolic gene clusters: evolutionary toolkits for chemical
634 innovation. *Trends Genet.* **26**, 449–457 (2010).

635 20. A. A. Brakhage, Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **11**, 21–
636 32 (2013).

637 21. M. J. Bibb, Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.*
638 **8**, 208–215 (2005).

639 22. Ö. Bayram, *et al.*, Changes of global gene expression and secondary metabolite
640 accumulation during light-dependent Aspergillus nidulans development. *Fungal Genet.*
641 *Biol.* **87**, 30–53 (2016).

642 23. A. L. Lind, *et al.*, Examining the evolution of the regulatory circuit controlling secondary
643 metabolism and development in the fungal genus Aspergillus. *PLoS Genet.* **11**, e1005096
644 (2015).

645 24. B. Ruiz, *et al.*, Production of microbial secondary metabolites: regulation by the carbon

646 source. *Crit. Rev. Microbiol.* **36**, 146–167 (2010).

647 25. V. S. Malik, Microbial secondary metabolism. *Trends Biochem. Sci.* **5**, 68–72 (1980).

648 26. P. Bateson, K. N. Laland, Tinbergen's four questions: an appreciation and an update.
649 *Trends Ecol. Evol.* **28**, 712–718 (2013).

650 27. A. M. Abdel-Mawgoud, F. Lépine, E. Déziel, Rhamnolipids: diversity of structures,
651 microbial origins and roles. *Appl. Microbiol. Biotechnol.* **86**, 1323–1336 (2010).

652 28. L. Guerra-Santos, O. Käppeli, A. Fiechter, *Pseudomonas aeruginosa* biosurfactant
653 production in continuous culture with glucose as carbon source. *Appl. Environ. Microbiol.*
654 **48**, 301–305 (1984).

655 29. J. B. Xavier, W. Kim, K. R. Foster, A molecular mechanism that stabilizes cooperative
656 secretions in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **79**, 166–179 (2011).

657 30. J. Tremblay, A.-P. Richardson, F. Lépine, E. Déziel, Self-produced extracellular stimuli
658 modulate the *Pseudomonas aeruginosa* swarming motility behaviour. *Environ. Microbiol.*
659 **9**, 2622–2630 (2007).

660 31. J. Q. Yang, Evidence for biosurfactant-induced flow in corners and bacterial spreading in
661 unsaturated porous media. *PNAS* (2021).

662 32. A. V. Sotirova, D. I. Spasova, D. N. Galabova, E. Karpenko, A. Shulga, Rhamnolipid-
663 biosurfactant permeabilizing effects on gram-positive and gram-negative bacterial strains.
664 *Curr. Microbiol.* **56**, 639–644 (2008).

665 33. L. Zulianello, *et al.*, Rhamnolipids are virulence factors that promote early infiltration of
666 primary human airway epithelia by Pseudomonas aeruginosa. *Infect. Immun.* **74**, 3134–
667 3147 (2006).

668 34. K. Zhu, C. O. Rock, RhlA converts beta-hydroxyacyl-acyl carrier protein intermediates in
669 fatty acid synthesis to the beta-hydroxydecanoyl-beta-hydroxydecanoate component of
670 rhamnolipids in *Pseudomonas aeruginosa. J. Bacteriol.* **190**, 3147–3154 (2008).

671 35. H. Chong, Q. Li, Microbial production of rhamnolipids: opportunities, challenges and
672 strategies. *Microb. Cell Fact.* **16**, 137 (2017).

673 36. S. Mukherjee, *et al.*, The PqsE and RhlR proteins are an autoinducer synthase-receptor pair
674 that control virulence and biofilm development in *Pseudomonas aeruginosa. Proc Natl*
675 *Acad Sci USA* **115** (2018).

676 37. A. Latifi, M. Foglino, K. Tanaka, P. Williams, A. Lazdunski, A hierarchical quorum-
677 sensing cascade in *Pseudomonas aeruginosa* links the transcriptional activators LasR and
678 RhIR (VsmR) to expression of the stationary-phase sigma factor RpoS. *Mol. Microbiol.*
679 **21**, 1137–1146 (1996).

680 38. K. E. Boyle, H. Monaco, D. van Ditmarsch, M. Deforet, J. B. Xavier, Integration of
681 metabolic and quorum sensing signals governing the decision to cooperate in a bacterial
682 social trait. *PLoS Comput. Biol.* **11**, e1004279 (2015).

683 39. B. Mellbye, M. Schuster, Physiological framework for the regulation of quorum sensing-
684 dependent public goods in *Pseudomonas aeruginosa. J. Bacteriol.* **196**, 1155–1164 (2014).

685 40. L. de Vargas Roditi, K. E. Boyle, J. B. Xavier, Multilevel selection analysis of a microbial
686 social trait. *Mol. Syst. Biol.* **9**, 684 (2013).

687 41. M. M. Müller, B. Hörmann, M. Kugel, C. Syldatk, R. Hausmann, Evaluation of
688 rhamnolipid production capacity of Pseudomonas aeruginosa PAO1 in comparison to the
689 rhamnolipid over-producer strains DSM 7108 and DSM 2874. *Appl. Microbiol.*

690 *Biotechnol.* **89**, 585–592 (2011).

691 42. A. Germer, *et al.*, Exploiting the Natural Diversity of RhlA Acyltransferases for the
692 Synthesis of the Rhamnolipid Precursor 3-(3-Hydroxyalkanoyloxy)Alkanoic Acid. *Appl.*
693 *Environ. Microbiol.* **86** (2020).

694 43. J. Scoffield, L. Silo-Suh, Glycerol metabolism promotes biofilm formation by
695 *Pseudomonas aeruginosa*. *Can. J. Microbiol.* **62**, 704–710 (2016).

696 44. A. M. Abdel-Mawgoud, F. Lépine, E. Déziel, A stereospecific pathway diverts β-oxidation
697 intermediates to the biosynthesis of rhamnolipid biosurfactants. *Chem. Biol.* **21**, 156–164
698 (2014).

699 45. I. Poblete-Castro, C. Wittmann, P. I. Nikel, Biochemistry, genetics and biotechnology of
700 glycerol utilization in Pseudomonas species. *Microb. Biotechnol.* **13**, 32–53 (2020).

701 46. J. Yan, *et al.*, Bow-tie signaling in c-di-GMP: Machine learning in a simple biochemical
702 network. *PLoS Comput. Biol.* **13**, e1005677 (2017).

703 47. D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization.
704 *Nature* **401**, 788–791 (1999).

705 48. O. Cloarec, *et al.*, Statistical total correlation spectroscopy: an exploratory approach for
706 latent biomarker identification from metabolic 1H NMR data sets. *Anal. Chem.* **77**, 1282–
707 1289 (2005).

708 49. S. Picart-Armada, F. Fernández-Albert, M. Vinaixa, O. Yanes, A. Perera-Lluna, FELLA:
709 an R package to enrich metabolomics data. *BMC Bioinformatics* **19**, 538 (2018).

710 50. K. E. Boyle, *et al.*, Metabolism and the evolution of social behavior. *Mol. Biol. Evol.* **34**,
711 2367–2379 (2017).

712 51. D. van Ditmarsch, J. B. Xavier, High-resolution time series of *Pseudomonas aeruginosa*
713 gene expression and rhamnolipid secretion through growth curve synchronization. *BMC*
714 *Microbiol.* **11**, 140 (2011).

715 52. B. Py, F. Barras, Building Fe-S proteins: bacterial strategies. *Nat. Rev. Microbiol.* **8**, 436–
716 446 (2010).

717 53. L. Hederstedt, L. Rutberg, Succinate dehydrogenase--a comparative review. *Microbiol.*
718 *Rev.* **45**, 542–555 (1981).

719 54. C. Ayala-Castro, A. Saini, F. W. Outten, Fe-S cluster assembly pathways in bacteria.
720 *Microbiol. Mol. Biol. Rev.* **72**, 110–25, table of contents (2008).

721 55. B. Curti, M. A. Vanoni, E. Verzotti, G. Zanetti, Glutamate synthase: a complex iron-
722 sulphur flavoprotein. *Biochem. Soc. Trans.* **24**, 95–99 (1996).

723 56. J. Nogales, S. Gudmundsson, E. Duque, J. L. Ramos, B. O. Palsson, Expanding The
724 Computable Reactome In *Pseudomonas putida* Reveals Metabolic Cycles Providing
725 Robustness. *BioRxiv* (2017) https:/doi.org/10.1101/139121.

726 57. J. Nogales, *et al.*, High-quality genome-scale metabolic modelling of *Pseudomonas putida*
727 highlights its broad metabolic capabilities. *Environ. Microbiol.* **22**, 255–269 (2020).

728 58. W. Xiao, J. Loscalzo, Metabolic responses to reductive stress. *Antioxid. Redox Signal.* **32**,
729 1330–1347 (2020).

730 59. J. B. Xavier, Social interaction in synthetic and natural microbial communities. *Mol. Syst.*
731 *Biol.* **7**, 483 (2011).

732 60. S. Mishra, J. Imlay, Why do bacteria use so many enzymes to scavenge hydrogen
733 peroxide? *Arch. Biochem. Biophys.* **525**, 145–160 (2012).

734 61. T. Vinckx, Q. Wei, S. Matthijs, P. Cornelis, The *Pseudomonas aeruginosa* oxidative stress
735 regulator OxyR influences production of pyocyanin and rhamnolipids: protective role of
736 pyocyanin. *Microbiology (Reading, Engl)* **156**, 678–686 (2010).

737 62. B. Zhao, F. A. Summers, R. P. Mason, Photooxidation of Amplex Red to resorufin:
738 implications of exposing the Amplex Red assay to light. *Free Radic. Biol. Med.* **53**, 1080–
739 1087 (2012).

740 63. M. R. Mulvey, P. A. Sorby, B. L. Triggs-Raine, P. C. Loewen, Cloning and physical
741 characterization of *katE* and *katF* required for catalase HPII expression in *Escherichia coli*.
742 *Gene* **73**, 337–345 (1988).

743 64. E. Hidalgo, V. Leautaud, B. Demple, The redox-regulated SoxR protein acts from a single
744 DNA site as a repressor and an allosteric activator. *EMBO J.* **17**, 2629–2636 (1998).

745 65. F. A. Mohamed, G. H. Shaker, M. M. Askoura, Oxidative Stress Influences *Pseudomonas*
746 *aeruginosa* Susceptibility to Antibiotics and Reduces Its Pathogenesis in Host. *Curr.*
747 *Microbiol.* **77**, 479–490 (2020).

748 66. T. Tiso, *et al.*, Creating metabolic demand as an engineering strategy in *Pseudomonas*
749 *putida* - Rhamnolipid synthesis as an example. *Metab. Eng. Commun.* **3**, 234–244 (2016).

750 67. K. Zhu, S. Chen, T. A. Sysoeva, L. You, Universal antibiotic tolerance arising from
751 antibiotic-triggered accumulation of pyocyanin in *Pseudomonas aeruginosa. PLoS Biol.*
752 **17**, e3000573 (2019).

753 68. M. Whiteley, K. M. Lee, E. P. Greenberg, Identification of genes controlled by quorum
754 sensing in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **96**, 13904–13909 (1999).

755 69. K. I. Piatkov, T. T. M. Vu, C.-S. Hwang, A. Varshavsky, Formyl-methionine as a

756 degradation signal at the N-termini of bacterial proteins. *Microb. Cell* **2**, 376–393 (2015).

757 70. Y. Cai, P. Chandrangsu, A. Gaballa, J. D. Helmann, Lack of formylated methionyl-tRNA
758 has pleiotropic effects on *Bacillus subtilis. Microbiology (Reading, Engl)* **163**, 185–196
759 (2017).

760 71. B. Ezraty, A. Gennaris, F. Barras, J.-F. Collet, Oxidative stress, protein damage and repair
761 in bacteria. *Nat. Rev. Microbiol.* **15**, 385–396 (2017).

762 72. D. K. Jain, D. L. Collins-Thompson, H. Lee, J. T. Trevors, A drop-collapsing test for
763 screening surfactant-producing microorganisms. *J. Microbiol. Methods* **13**, 271–279
764 (1991).

765 73. C.-Y. Chen, S. C. Baker, R. C. Darton, The application of a high throughput analysis
766 method for the screening of potential biosurfactants from natural sources. *J. Microbiol.*
767 *Methods* **70**, 503–510 (2007).

768 74. A. J. Page, *et al.*, Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*
769 **31**, 3691–3693 (2015).

770 75. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069
771 (2014).

772 76. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple
773 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066
774 (2002).

775 77. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary
776 analyses in R. *Bioinformatics* **35**, 526–528 (2019).

777 78. J. M. Beaulieu, B. C. O'Meara, M. J. Donoghue, Identifying hidden rate changes in the

evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst. Biol.* **62**, 725–737 (2013).

79. H. Redestig, *et al.*, Compensation for systematic cross-contribution improves normalization of mass spectrometry based metabolomics data. *Anal. Chem.* **81**, 7974–7980 (2009).

80. A. M. De Livera, G. Olshansky, J. A. Simpson, D. J. Creek, NormalizeMets: assessing, selecting and implementing statistical methods for normalizing metabolomics data. *Metabolomics* **14**, 54 (2018).

81. G. Warnes, *et al.*, gplots: Various R Programming Tools for Plotting Data. *undefined* (2015).

82. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

83. S. Picart-Armada, *et al.*, Null diffusion-based enrichment for metabolomics data. *PLoS ONE* **12**, e0189012 (2017).

84. F. Vandin, E. Upfal, B. J. Raphael, Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* **18**, 507–522 (2011).

85. E. A. Thévenot, A. Roux, Y. Xu, E. Ezan, C. Junot, Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).

86. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).

800 87. S. Andrews, *FastQC: A Quality Control Tool for High Throughput Sequence Data*
801 (QUBES, 2010) (May 28, 2019).

802 88. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
803 (2013).

804 89. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
805 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

806 90. J. Thioulouse, *et al.*, *Multivariate Analysis of Ecological Data with ade4* (Springer New
807 York, 2018) https:/doi.org/10.1007/978-1-4939-8850-1.

808

809

Figures

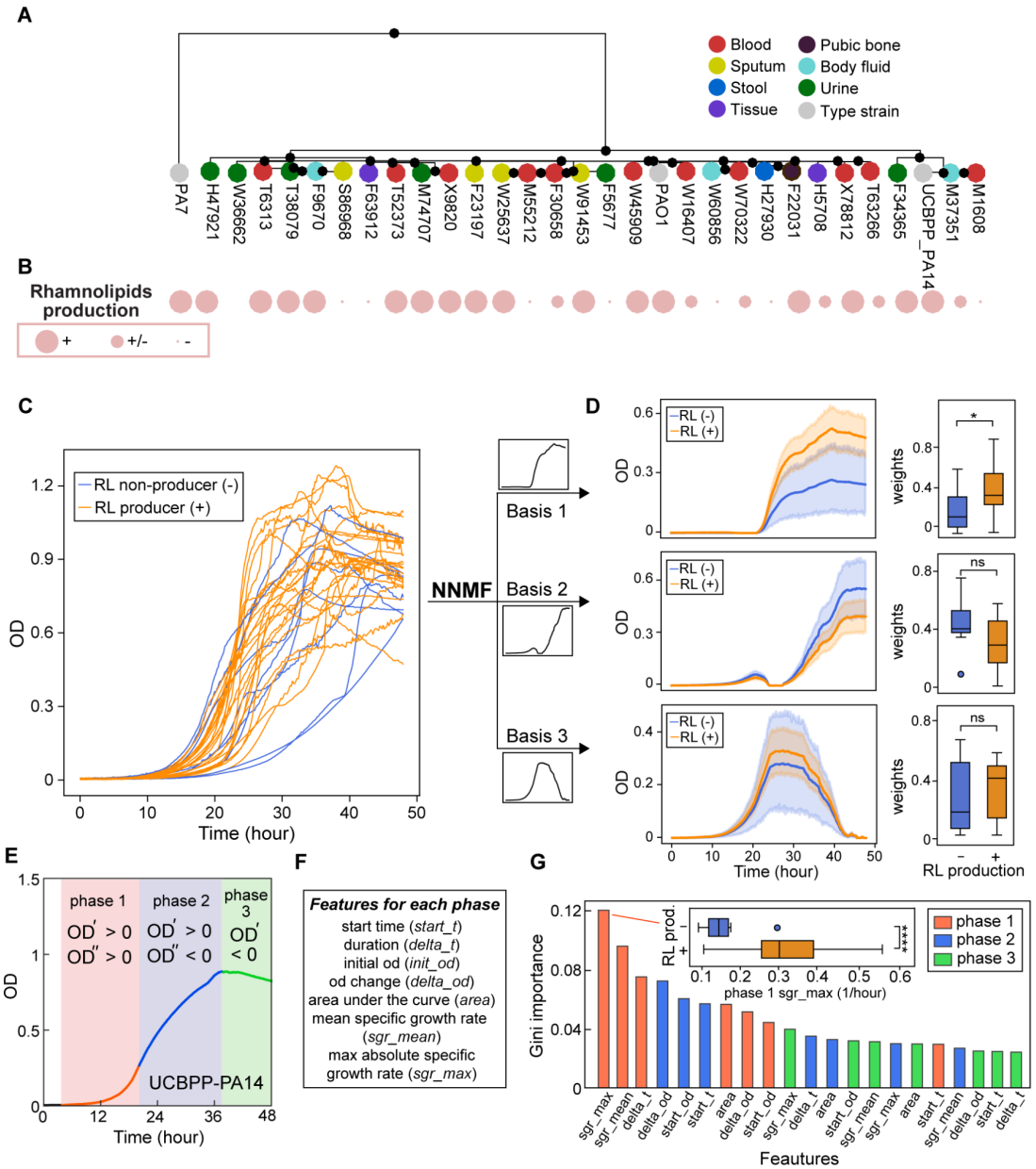


Figure 1. Diversity of rhamnolipid production across the *P. aeruginosa* phylogeny of core genomes and growth curve features distinguish rhamnolipid producers from non-producers. **A.** Phylogeny of clinical isolates obtained from patients with cancer at MSKCC (21, 75) together with reference type strains PAO1, PA14 and PA7. The tissue where each isolate was originally isolated is labeled by circle colors. **B.** Rhamnolipid production phenotypes. The ability of producing rhamnolipids of these strains is qualitatively

817 indicated by circle sizes (right column). **C,D.** Unsupervised feature selection using non-negative matrix
818 factorization (NNMF), which decomposes growth curves of all *Pseudomonas* isolates into three additive
819 basis functions (features) such that each growth curve can be approximately represented by the weighted
820 sum of these functions. **C**. Growth curves from both rhamnolipid producers (orange) and non-producers
821 (blue). **D**. Decomposed components (basis function multiplied by weights; left panels) and weights (right
822 panels) from NNMF grouped by rhamnolipid (RL) production. The shaded areas represent 95% bootstrap
823 confidence intervals of the mean. **E-G**. Supervised feature selection using Random Forest classifier. **E,F**.
824 Feature extraction method. Each growth curve (excluding the initial lag phase) was divided into three
825 phases (**E**) and each phase was described by 7 quantitative features (**F**). **G.** Ranking of feature importance
826 in classifying rhamnolipid producers. Inset: boxplot of maximum specific growth rate of phase I grouped
827 by rhamnolipid production. Welch's t-test was used in (**D**) and (**G**) for significance testing. ****, p-value
828 ≤ 0.0001; *, p-value ≤ 0.05; ns, p-value > 0.05.

829

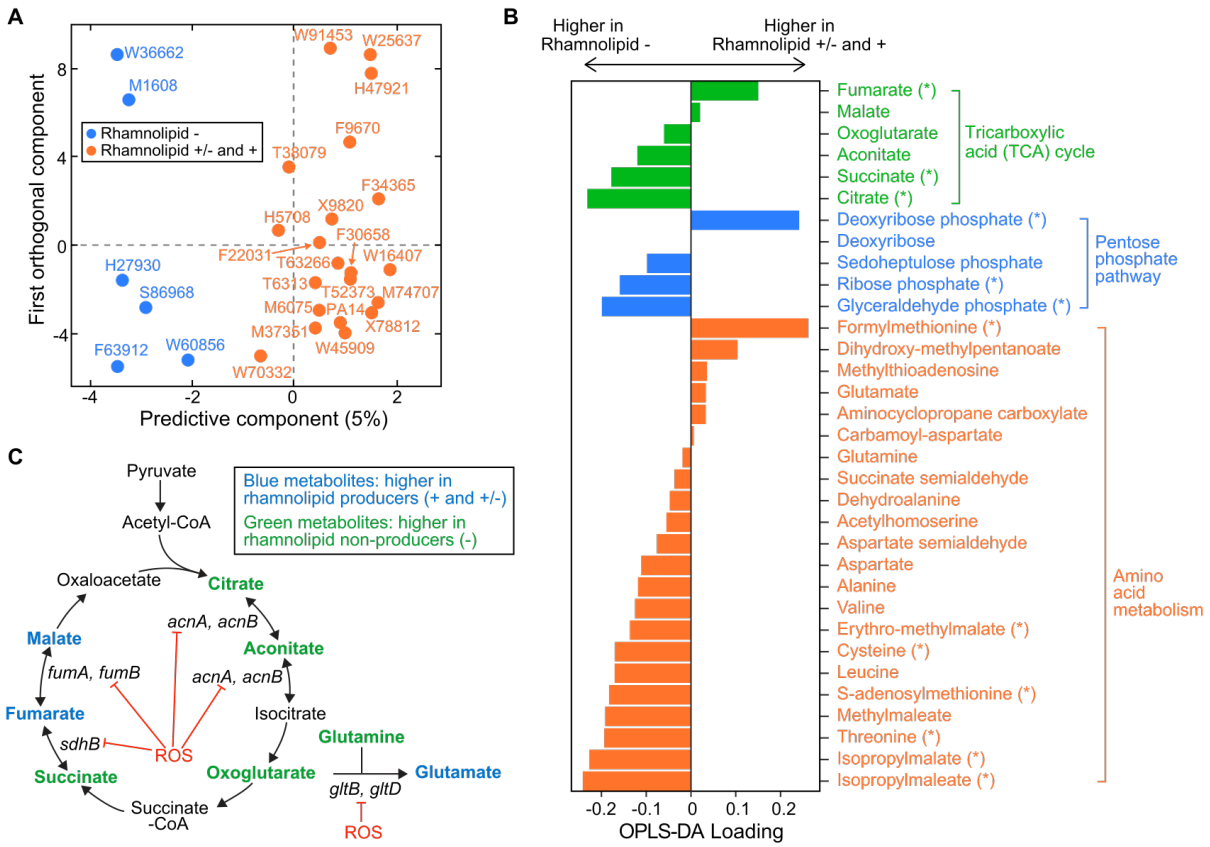
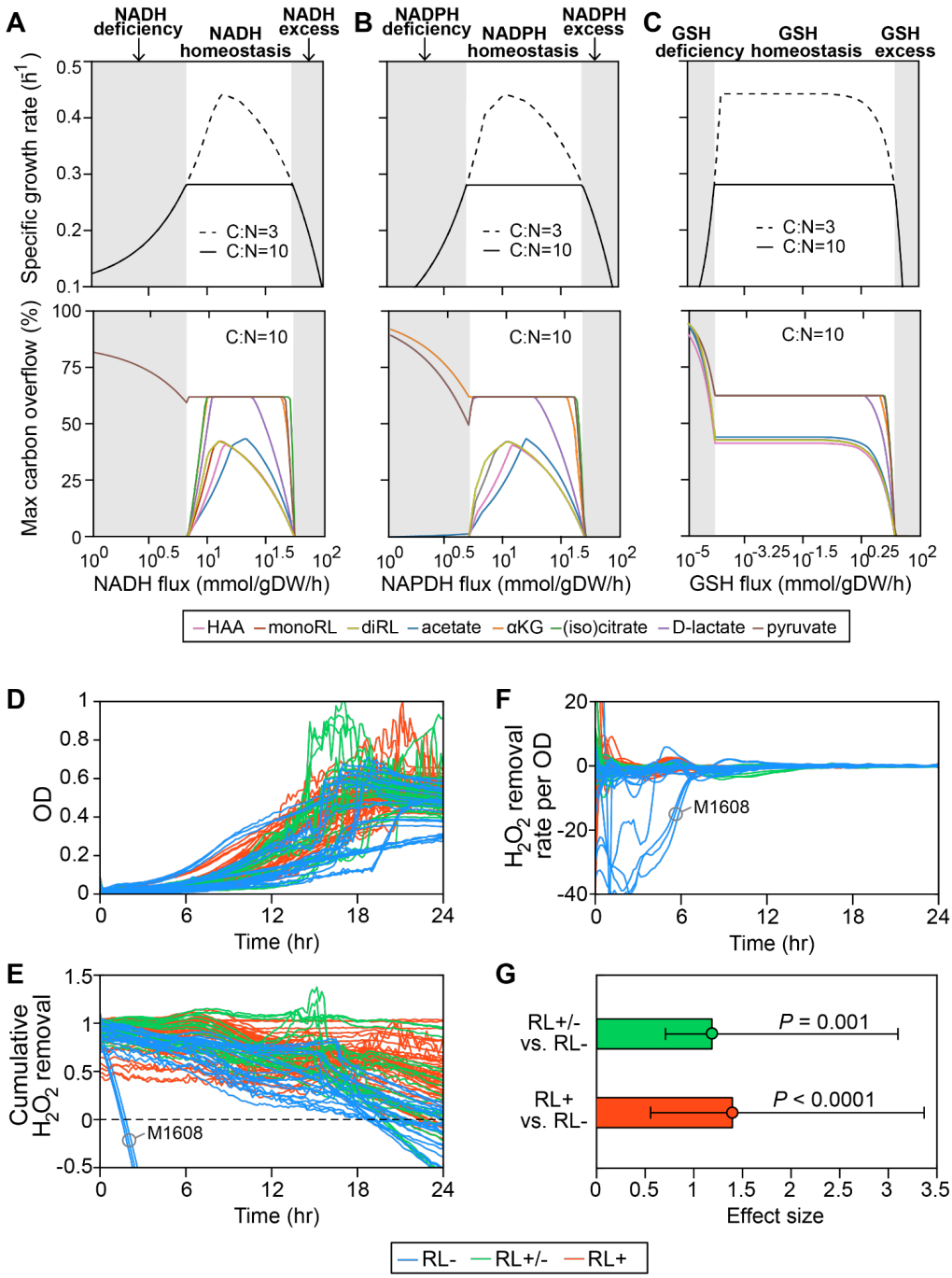


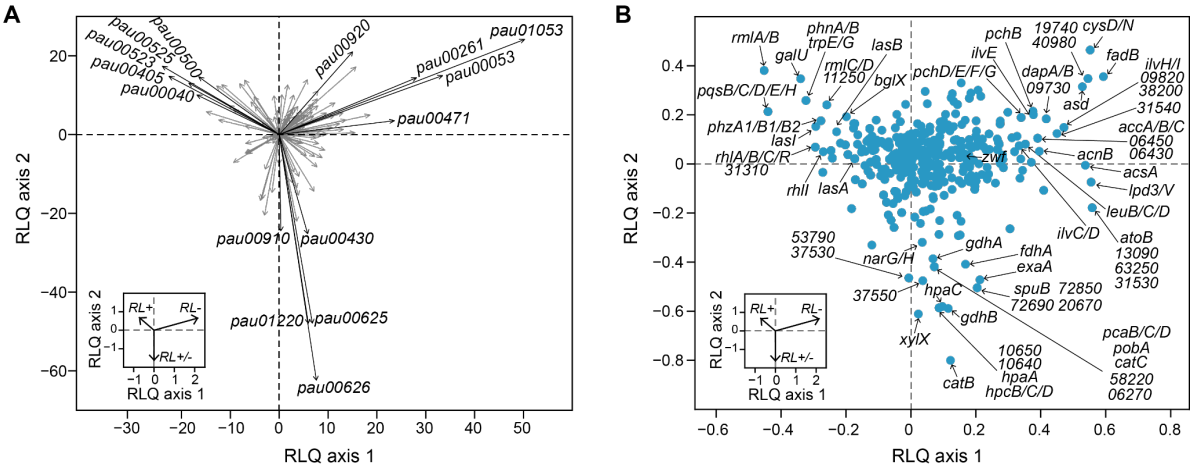
Figure 2. Comparative metabolomics between rhamnolipid producers (include both mild and strong production levels) and non-producers. A. The scores plot of the OPLS-DA (orthogonal partial least squares discriminant analysis) model. The first principle component separated the rhamnolipid non-producers from producers. **B.** OPLS-DA loading values for the first principle component of a selected number of metabolites. Metabolites with asterisk (*) are significantly different between producers and non-producers (adjusted p-value of a Mann-Whitney test < 0.05). **C**. Differential abundance of metabolites involved in reactions catalyzed by some Fe-S-containing enzymes whose activities are inhibited by reactive oxygen species (ROS). Abbreviations: *acnA* (aconitate hydratase A), *acnB* (aconitate hydratase B), *sdhB* (succinate dehydrogenase subunit), *fumA* (fumarase A), *fumB* (fumarase B), *gltB* (glutamate synthase subunit), *gltD* (glutamate synthase subunit).



842

843 **Figure 3** Dependence of rhamnolipid secretion on intracellular redox status. **A-C**. Model simulation. The
844 redox stress levels are perturbed by altering fluxes of (**A**) NADH (reduced nicotinamide adenine
845 dinucleotide), (**B**) NADPH (reduced nicotinamide adenine dinucleotide phosphate) and (**C**) GSH (reduced
846 glutathione). Upper panels are predicted maximum growth rates and lower panels are predicted maximum

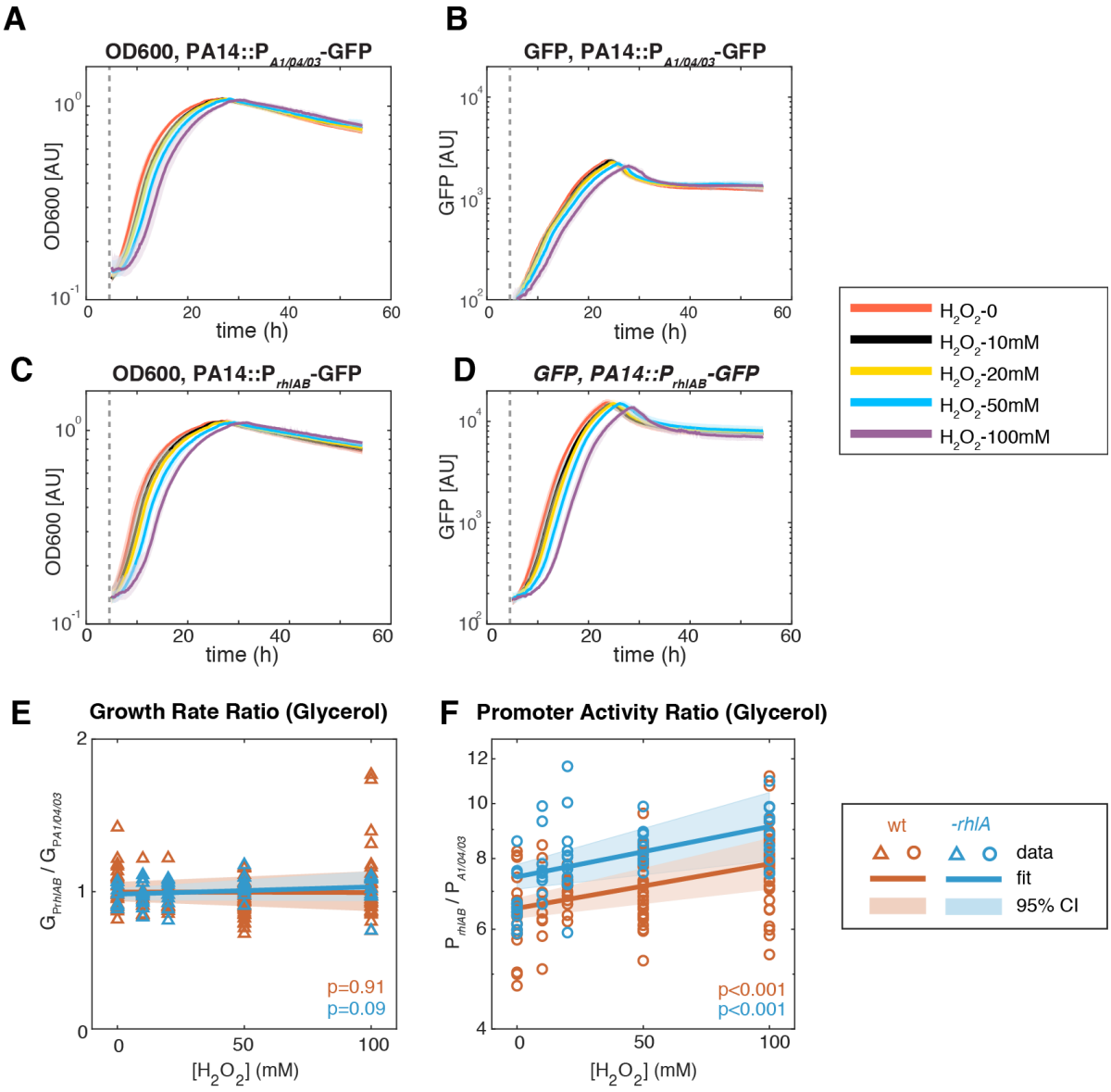
847 byproduct secretion fluxes. C:N indicates the carbon-to-nitrogen influx ratio provided by culture medium.
848 C:N=3 and C:N=10 represent carbon- and nitrogen-limiting conditions respectively. Abbreviations: HAA:
849 3-(3-hydroxyalkanoyloxy) alkanoate; monoRL: monorhamnolipid; diRL: dirhamnolipid; aKG: alpha-
850 ketoglutarate. **D-G**. Experimental validation via comparison of the ability of removing culture medium
851 hydrogen peroxide (H_2O_2) among strong rhamnolipid producers (+), weak producers (+/-), and non-
852 producers (-). **D.** Population density. **E.** The total amount of hydrogen peroxide removed from the
853 environment. Negative values indicate net cellular production of hydrogen peroxide released to the
854 environment. **F.** The specific hydrogen peroxide removal rate. In both **E** and **F**, each trajectory of H_2O_2
855 fluorescence intensity was normalized to the averaged trajectory of the wild-type UCBPP-PA14 strain. **G.**
856 Effect size of rhamnolipid production as a predictor of H_2O_2 removal rate per OD in a mixed-effect
857 analysis.



858

859 **Figure 4.** Functional pathways (**A**) and individual gene expressions (**B**) associated with rhamnolipid
860 production. RLQ axes 1 and 2 are the first two axes of RLQ analysis. The insets in both A and B represent
861 the rhamnolipid synthesis phenotypes in the principal RLQ axes. Genes and pathways aligning with the
862 direction of each phenotypic category are associated with the phenotype. In panel A, the top 5 pathways
863 ("pau" are KEGG pathway IDs) associated with each phenotype are highlighted in black. The five-digit
864 number in panel B represents the PA14 locus ID. RL+: strong rhamnolipid producer; RL+/-: mild-
865 rhamnolipid producer; RL-: rhamnolipid non-producer. Description of the KEGG pathways: pau01053
866 (Biosynthesis of siderophore group nonribosomal peptides), pau00053 (Ascorbate and aldarate
867 metabolism), pau00261 (Monobactam biosynthesis), pau00471 (D-Glutamine and D-glutamate
868 metabolism), pau00920 (Sulfur metabolism), pau00626 (Naphthalene degradation), pau01220
869 (Degradation of aromatic compounds), pau00625 (Chloroalkane and chloroalkene degradation), pau00430
870 (Taurine and hypotaurine metabolism), pau00910 (Nitrogen metabolism), pau00525 (Acarbose and
871 validamycin biosynthesis), pau00405 (Phenazine biosynthesis), pau00523 (Polyketide sugar unit
872 biosynthesis), pau00500 (Starch and sucrose metabolism), pau00040 (Pentose and glucuronate
873 interconversions).

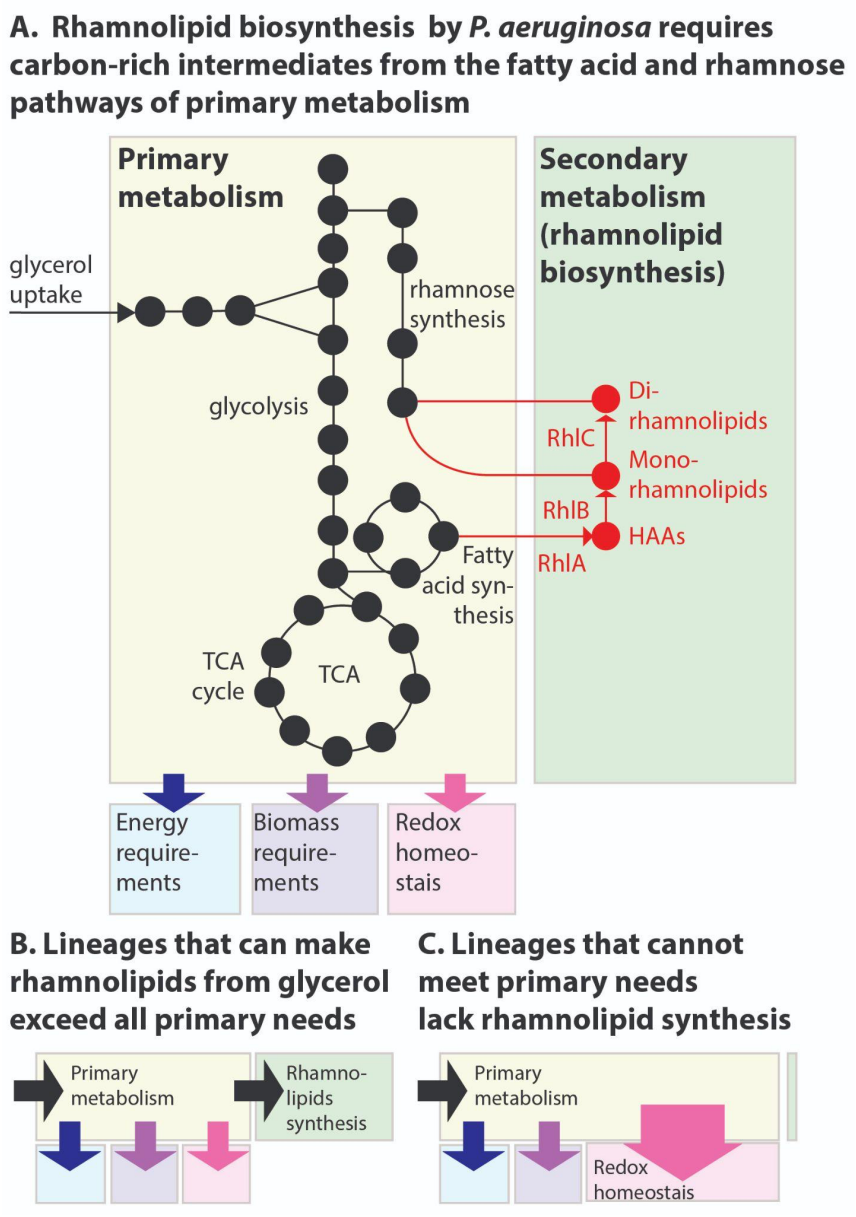
874



875

876 **Figure 5. Comparing the effect of different oxidative stress on the *rhlAB* gene expression and cell**
877 **growth in synthetic media of different carbon sources**. (A-D) Demonstration of the cell growth (A, C)
878 and GFP expression (B, D) of PA14::$P_{A1/04/03}$-GFP (constitutive promoter) and PA14::P_{rhlAB}-GFP when
879 different levels of H_2O_2 was added to the media at the beginning of exponential growth (gray dash line)
880 using one experiment. The shaded area represents the range of all replicates from the same experiment. (E,
881 F) The ratio of growth rate (E) and promoter activity (F) of PA14::P_{rhlAB}-GFP to PA14::$P_{A1/04/03}$-GFP in
882 phase II under different levels of H_2O_2 when glycerol was used as the carbon source of the synthetic media.

The p-value represents how significant the coefficient (slope) is using fitglme. Each dot represents one replicate.



885

886 **Figure 6**. The allocation of metabolic resources made from glycerol in primary vs secondary metabolism.

887 **A**. Growth and rhamnolipid biosynthesis on glycerol as the sole carbon source places a strong burden on

888 *P. aeruginosa*: The bacterium has to make all the molecules needed for energy, biomass and redox

889 homeostasis, and rhamnolipid biosynthesis—a secondary metabolic pathway—competes for those

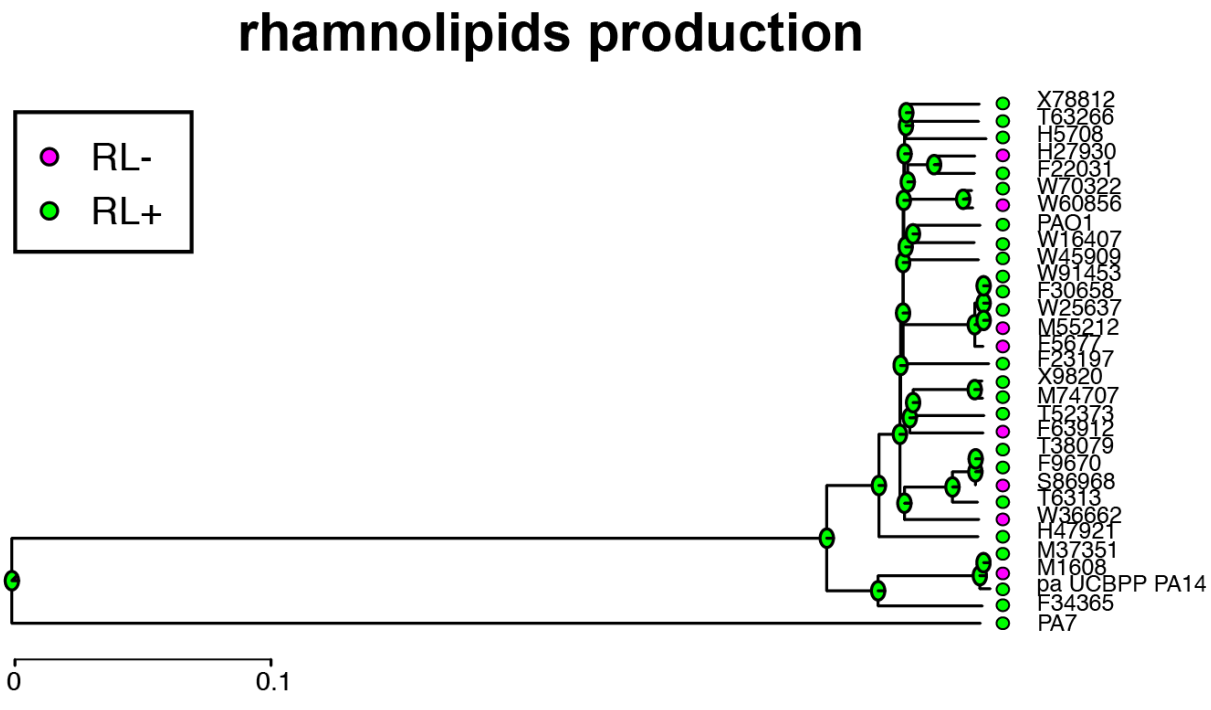
890 resources. **B**. *P. aeruginosa* lineages that retain the ability to make rhamnolipids from glycerol meet their

891 primary needs and use excess resources for secondary metabolism. **C**. *P. aeruginosa* lineages that lost the

892 ability to make rhamnolipids from glycerol were also worse at reducing the oxidative stress produced

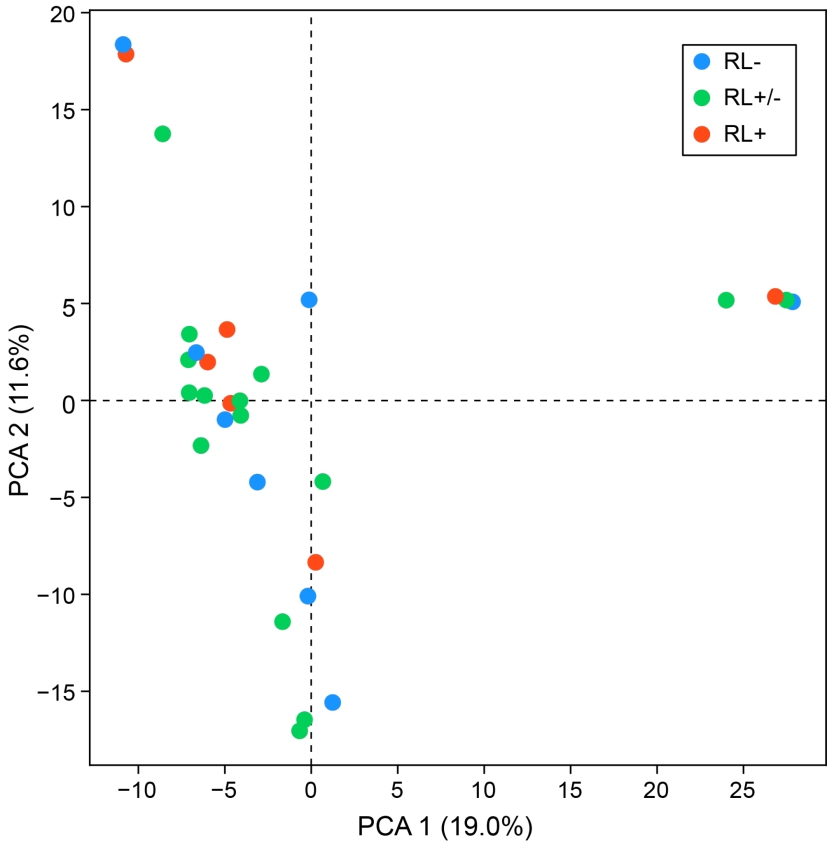
893 from growth in glycerol; the needs imposed on primary metabolism, such as maintaining redox
894 homeostasis, may leave insufficient resources for secondary metabolism, explaining the loss of
895 rhamnolipid synthesis.

896 **Supplementary Figures**

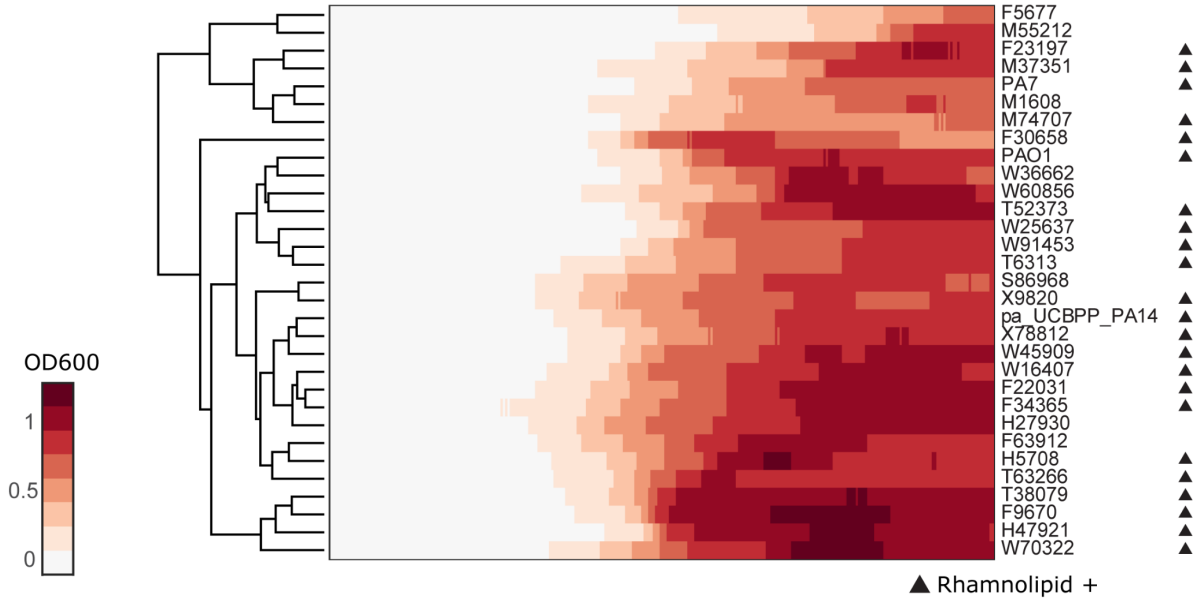


897

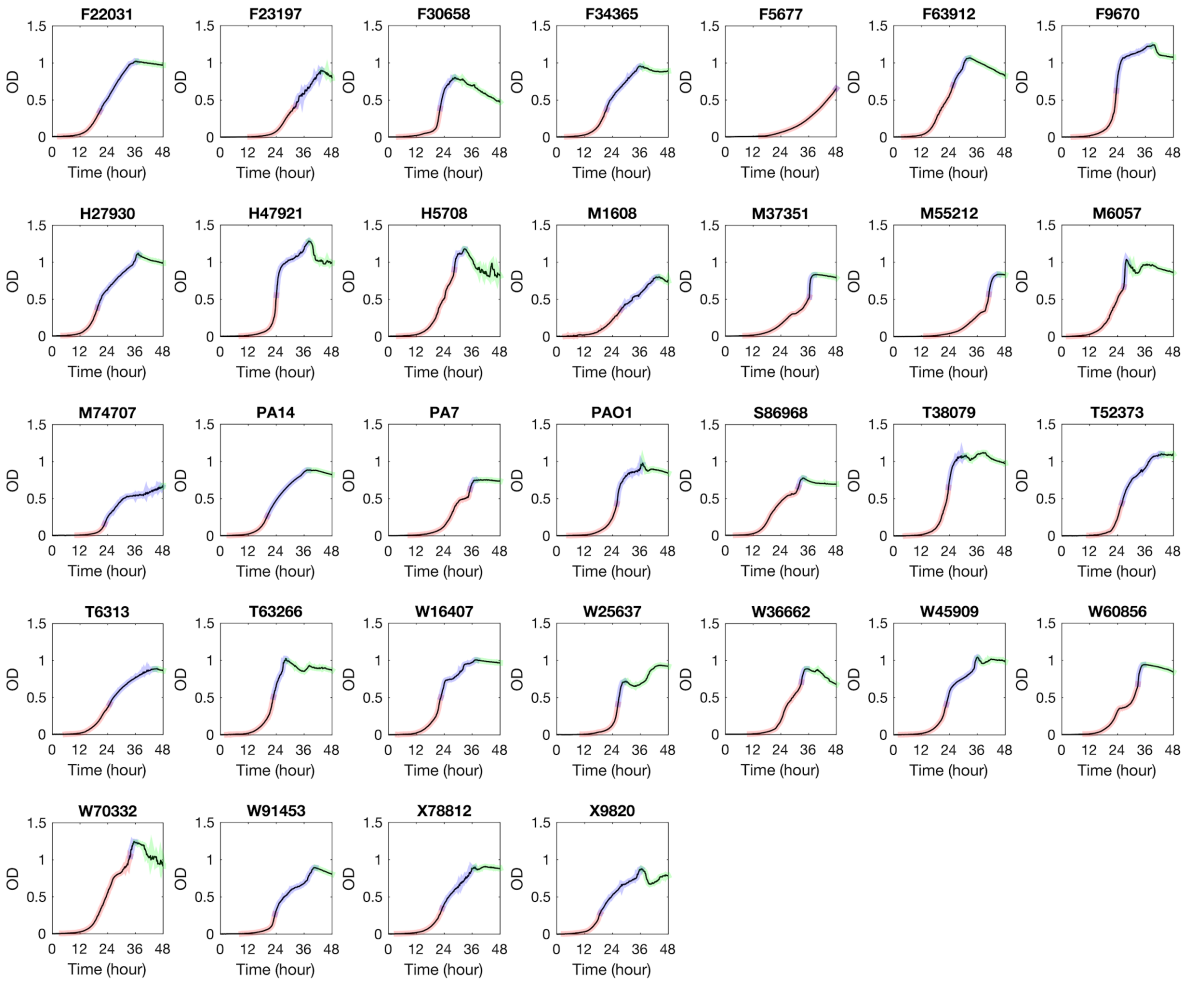
898 **Supplementary Figure 1.** Phylogenetic ancestor state reconstruction of rhamnolipid production. Pie charts
899 at the ancestor nodes of branches represent the relative likelihood proportion of each possible phenotypic
900 state. We rooted the tree with PA7, a *P. aeruginosa* isolate that is often used as an outlier to root
901 phylogenetic trees. RL, rhamnolipid production.



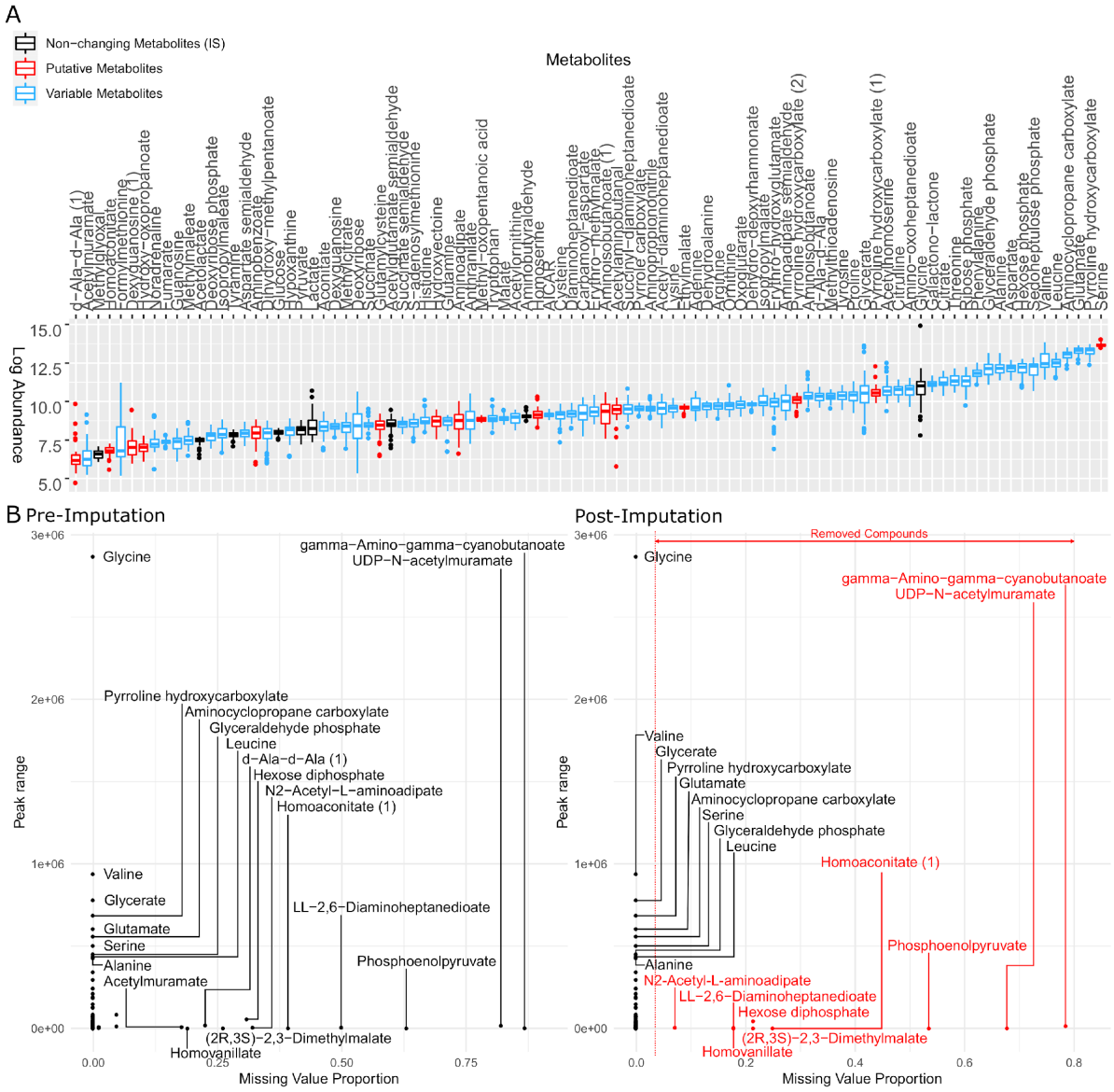
Supplementary Figure 2. Principal component analysis (PCA) of the presence/absence of the accessory genomes.



Supplementary Figure 3. Clustergram of growth curves of *P. aeruginosa* clinical isolates and three type strains PA14, PAO1 and PA7 in glycerol minimal medium. Euclidean distance was used as the measure of similarity.



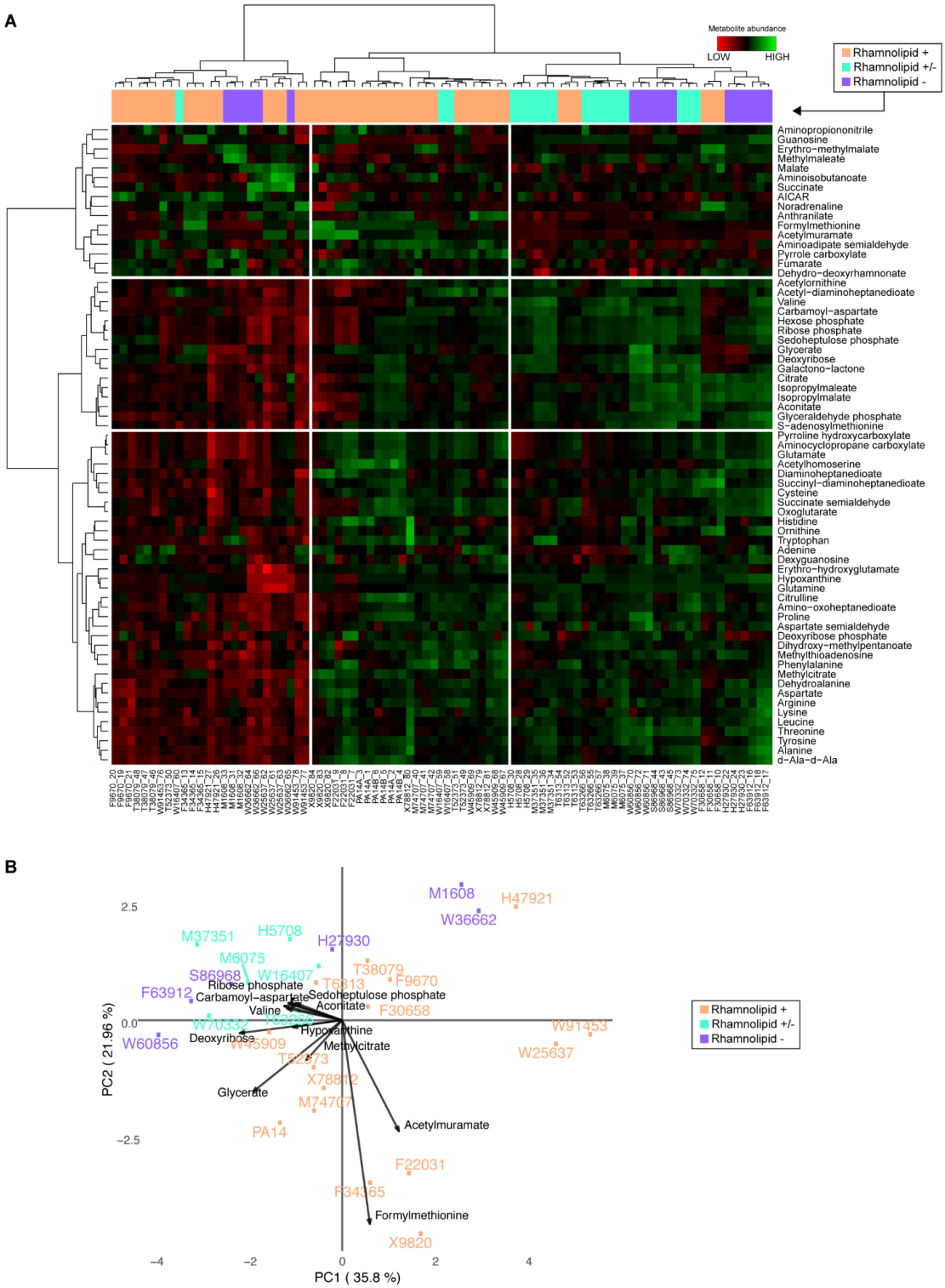
Supplementary Figure 4. Growth curve of *P. aeruginosa* strains in glycerol minimal medium. Phase I, II, and III are colored by red, blue and green respectively.



Supplementary Figure 5. A. Logarithm of raw peak area of all metabolites identified by LC-MS (Liquid chromatography-Mass spectrometry). The metabolites in black were used as internal standard in the data normalization step but removed from the dataset afterwards. The metabolites in red are putative metabolites that were initially included in clustering metabolomics but later removed from the heatmap (**Supplementary Figure 6**). The metabolites in blue are variable among clinical isolates and therefore used for later on analysis. **B.** Metabolite peak area before (left panel) and after (right panel) imputation. The missing value proportion of a metabolite in the x-axis represents the frequency of missing values of the

922 metabolite across all replicates of all strains in our study. The peak range of a metabolite in the y-axis is
923 defined as the maximum peak area minus the minimum peak area among all non-missing values of the
924 same metabolite across all our strains.

925



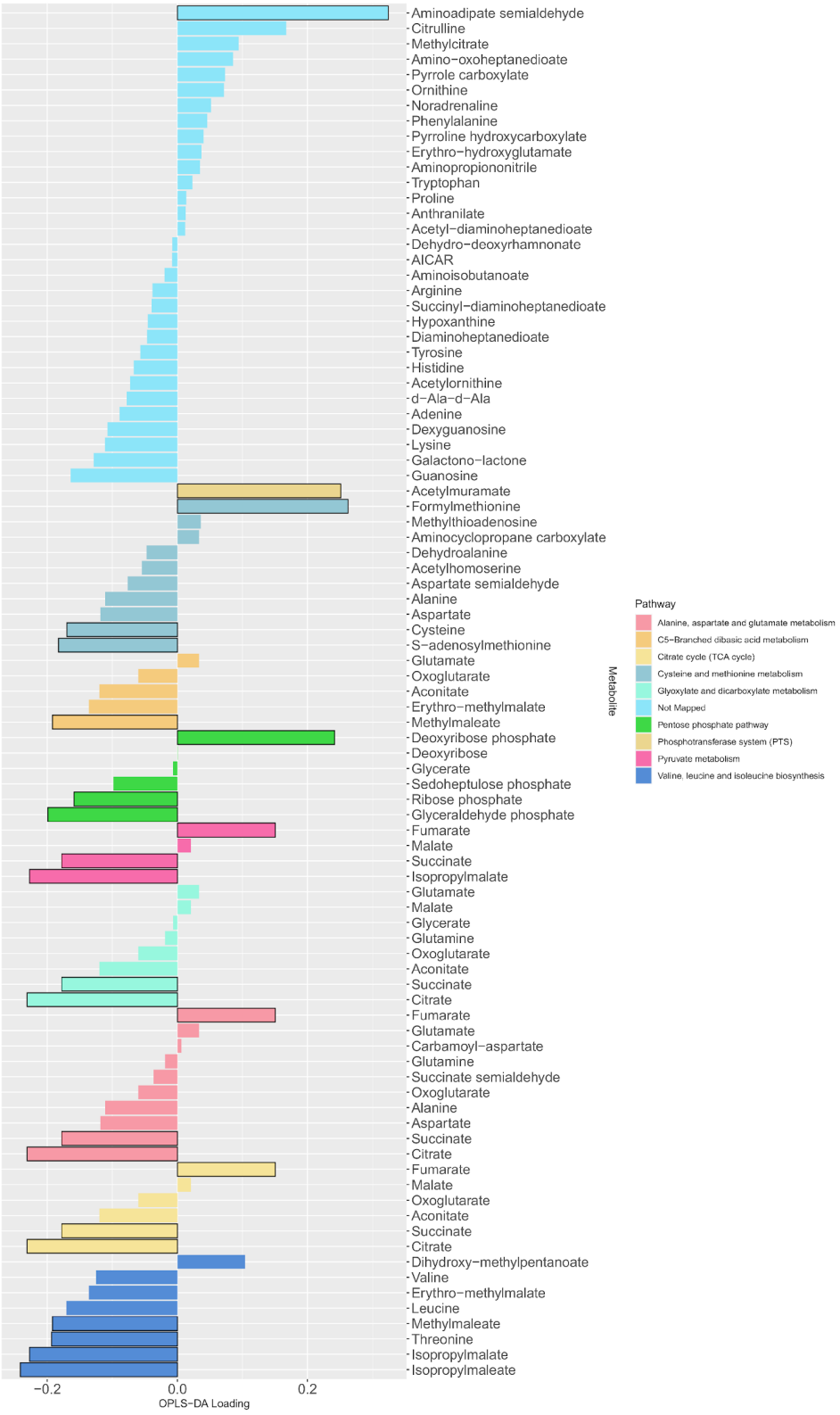
926

927 **Supplementary Figure 6. Variance and clustering analysis of metabolomics across *Pseudomonas***

928 ***aeruginosa* strains. A.** Hierarchical clustering of the metabolic profiles using Euclidean distance and Ward

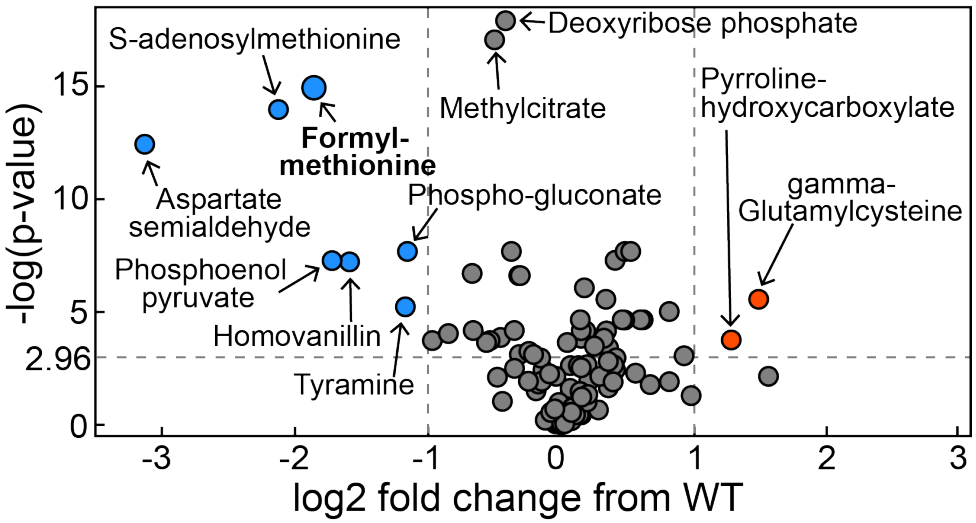
929 aggregation method. Each row represents one metabolite and each column represents a specific sample
930 (both strain and sample number are indicated). **B.** PCA plot of metabolites from LC-MS. The rhamnolipid
931 strong producer (+) occupies different territory but the separation among the three groups can not be
932 explained using PC1 or PC2. Strains from the purple (-) and the green (+/-) groups overlap.

933


Aminoadipate semialdehyde
Citrulline
Methylcitrate
Amino−oxoheptanedioate
Pyrrole carboxylate
Ornithine
Noradrenaline
Phenylalanine
Pyrroline hydroxycarboxylate
Erythro−hydroxyglutamate
Aminopropiononitrile
Tryptophan
Proline
Anthranilate
Acetyl−diaminoheptanedioate
Dehydro−deoxyrhamnonate
AICAR
Aminoisobutanoate
Arginine
Succinyl−diaminoheptanedioate
Hypoxanthine
Diaminoheptanedioate
Tyrosine
Histidine
Acetylornithine
d−Ala−d−Ala
Adenine
Dexyguanosine
Lysine
Galactono−lactone
Guanosine
Acetylmuramate
Formylmethionine
Methylthioadenosine
Aminocyclopropane carboxylate
Dehydroalanine
Acetylhomoserine
Aspartate semialdehyde
Alanine
Aspartate
Cysteine
S−adenosylmethionine
Glutamate
Oxoglutarate
Aconitate
Erythro−methylmalate
Methylmaleate
Deoxyribose phosphate
Deoxyribose
Glycerate
Sedoheptulose phosphate
Ribose phosphate
Glyceraldehyde phosphate
Fumarate
Malate
Succinate
Isopropylmalate
Glutamate
Malate
Glycerate
Glutamine
Oxoglutarate
Aconitate
Succinate
Citrate
Fumarate
Glutamate
Carbamoyl−aspartate
Glutamine
Succinate semialdehyde
Oxoglutarate
Alanine
Aspartate
Succinate
Citrate
Fumarate
Malate
Oxoglutarate
Aconitate
Succinate
Citrate
Dihydroxy−methylpentanoate
Valine
Erythro−methylmalate
Leucine
Methylmaleate
Threonine
Isopropylmalate
Isopropylmaleate
−0.2
0.0
0.2
OPLS−DA Loading
Metabolite
Pathway
Alanine, aspartate and glutamate metabolism
C5−Branched dibasic acid metabolism
Citrate cycle (TCA cycle)
Cysteine and methionine metabolism
Glyoxylate and dicarboxylate metabolism
Not Mapped
Pentose phosphate pathway
Phosphotransferase system (PTS)
Pyruvate metabolism
Valine, leucine and isoleucine biosynthesis

934

935 **Supplementary Figure 7.** The loading values of all predictive metabolites of the OPLS-DA model. The
936 differential metabolites between producers and non-producers were determined by a Mann Whitney test
937 (adjusted *p*-values with Benjamini-Hochberg method) with a level of significance of 0.05 (bars with black
938 outline) and used as input for a metabolic pathway enrichment with FELLA algorithm. The colors indicate
939 the mapped metabolite pathway for each metabolite.

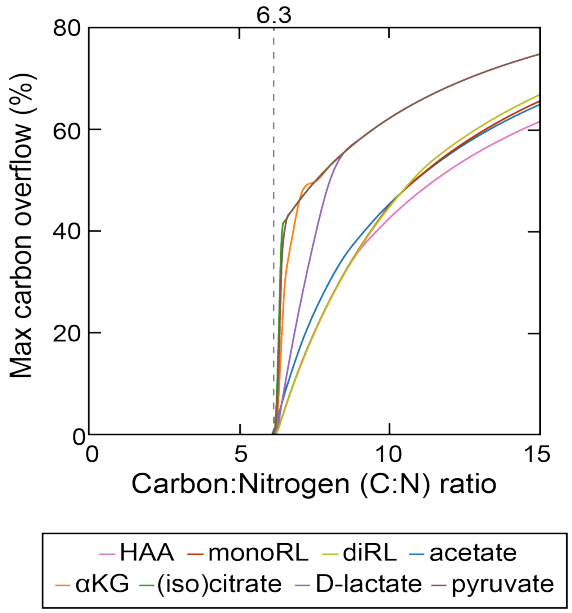


940

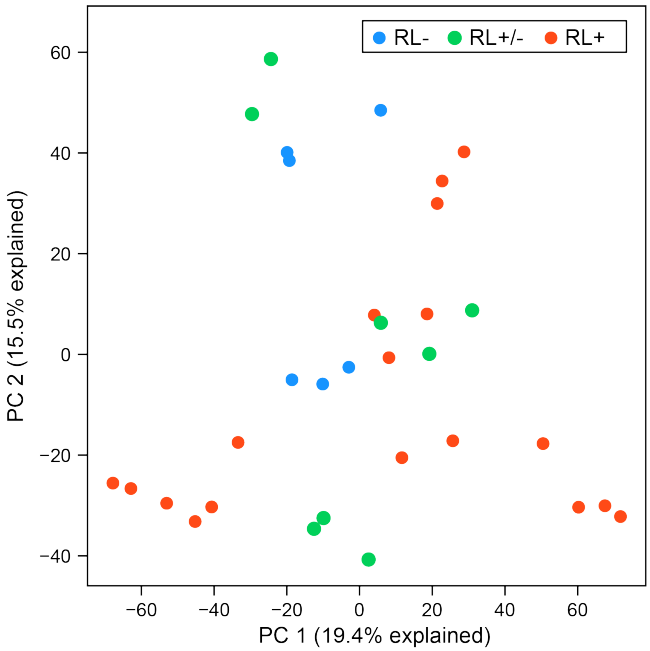
941 **Supplementary Figure 8.** Volcano plot of metabolomics data between wild-type *P. aeruginosa* UCBPP-

942 PA14 strain and its *rhlA* mutant grown in glycerol minimal medium (replotted with permission from (20)).

943



Supplementary Figure 9. Theoretical estimation of threshold carbon (glycerol):nitrogen (ammonium) ratio above which carbon is in excess in the sense that carbon release through rhamnolipids and central carbon metabolites does not compromise biomass production. Abbreviations: HAA: 3-(3-hydroxyalkanoyloxy)alkanoate; monoRL: monorhamnolipid; diRL: dirhamnolipid; aKG: alpha-ketoglutarate.



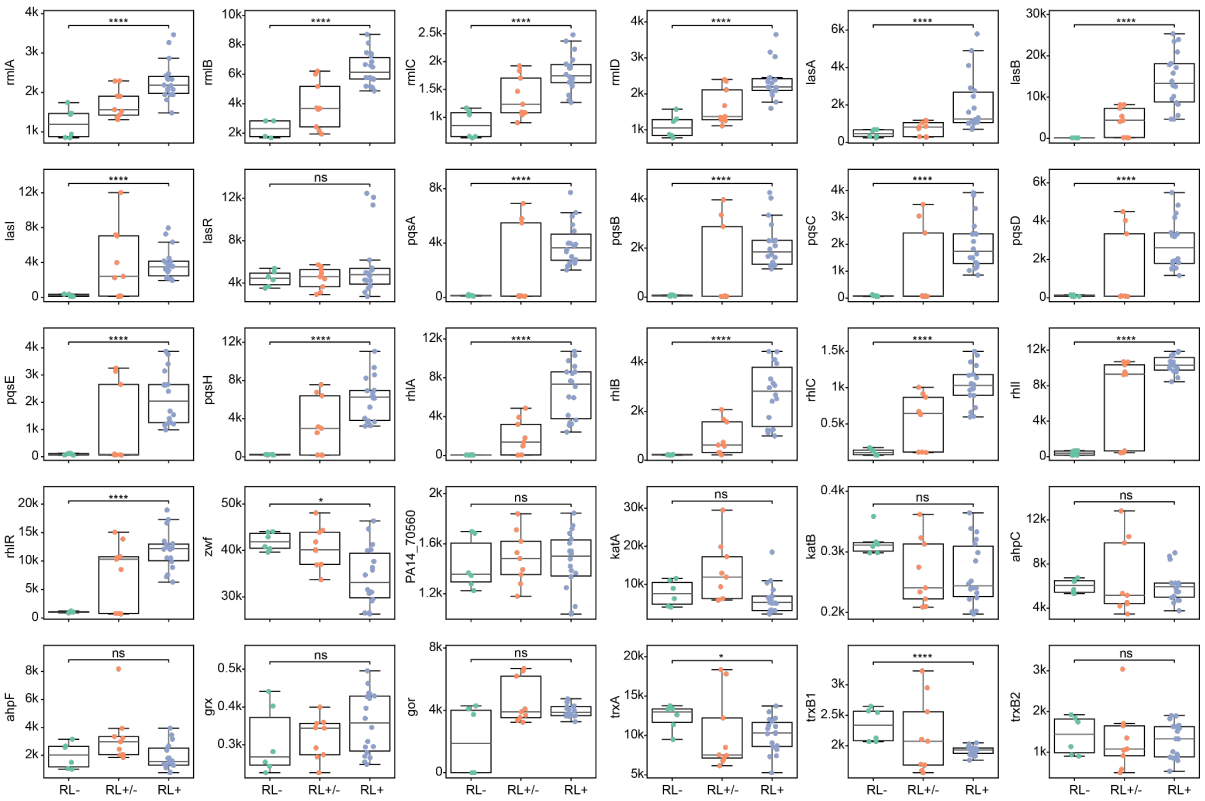
950

951 **Supplementary Figure 10.** PCA plot of RNA expression colored by rhamnolipid production phenotypes.

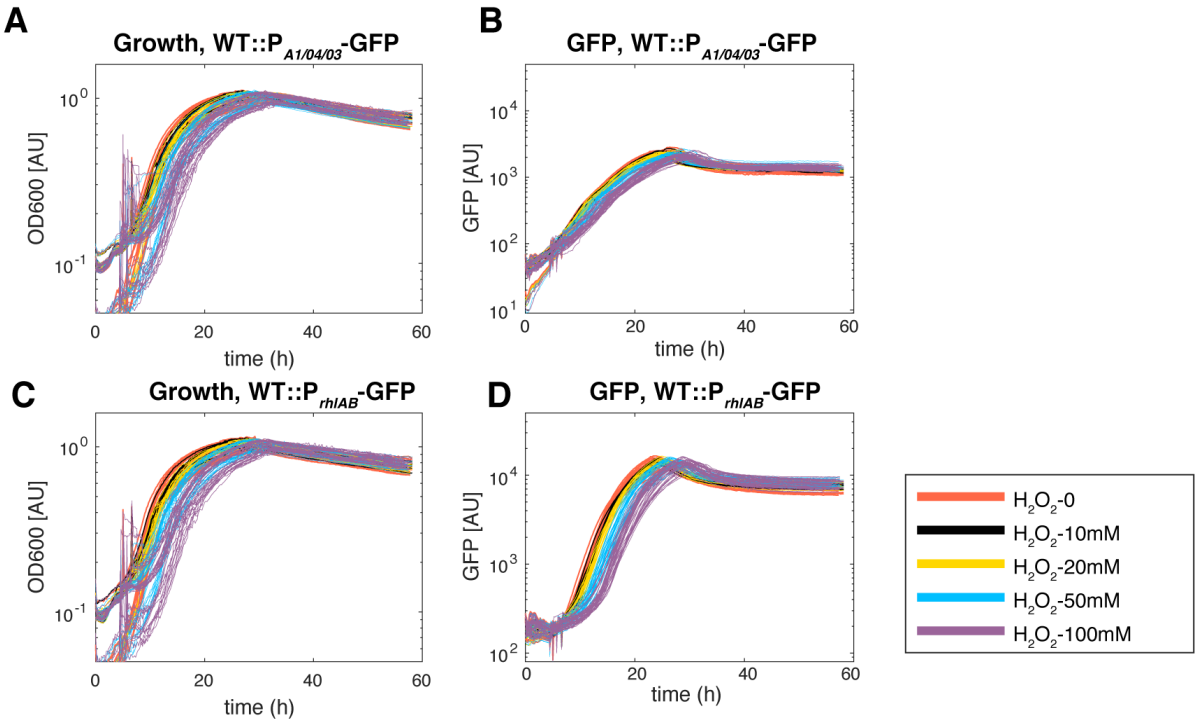
952 Each dot represents a replicate of an RNA-seq experiment. PC: principal component; RL-: rhamnolipid

953 non-producer; RL+/-: mild rhamnolipid producer; RL+: strong rhamnolipid producer.

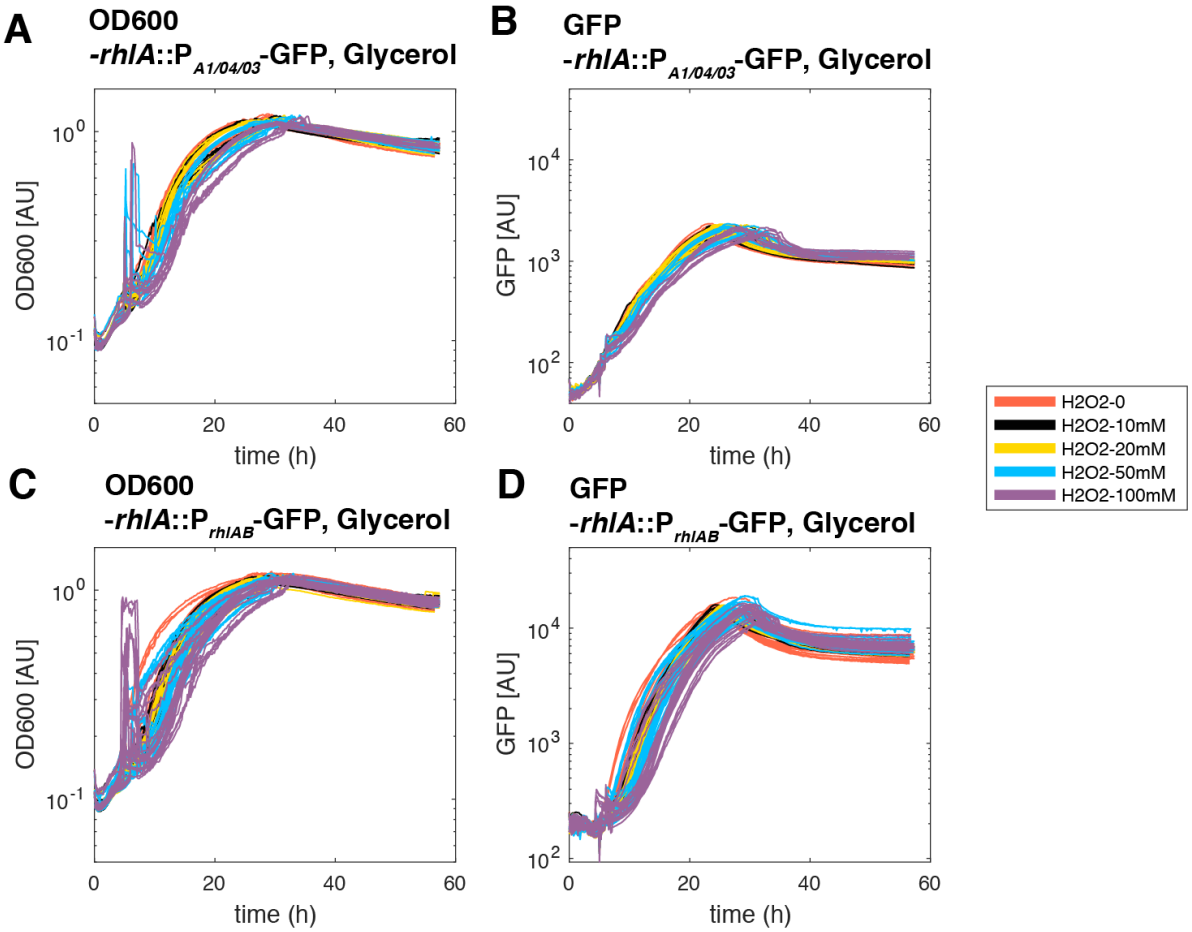
954



Supplementary Figure 11. Comparisons of expression for selected genes across strains with different abilities of rhamnolipid biosynthesis. Each dot represents a replicate of an RNA-seq experiment. RL-: rhamnolipid non-producer; RL+/-: mild rhamnolipid producer; RL+: strong rhamnolipid producer. Mann-Whitney U test with FDR method for multitest correction. ****, P<0.0001; *, P<0.05. ns, non-significant.



Supplementary Figure 12. Cell growth and *rhlAB* expression under extra oxidative stress in glycerol minimal media with wild type strain. Each line represents the data from a single well. The disturbance of the lines were caused by added H_2O_2 to the media.



Supplementary Figure 13. Cell growth and *rhlAB* expression under extra oxidative stress in glycerol minimal media with *ΔrhlA* strain. Each line represents the data from a single well. The disturbance of the lines were caused by added H_2O_2 to the media.

973 **Supplementary Data**

974 **Supplementary Table 1. Presence and absence of genes across the genomes of our clinical**
975 **isolates**

976 **Supplementary Table 2. Quantitative values for the 7 local features of phase I, II and III**
977 **(phase start time point, phase duration, phase initial OD, OD change, area under the curve,**
978 **mean specific growth rate, maximum specific growth rate) of growth curves.**

979 **Supplementary Table 3. Different pathways that are significantly different in rhamnolipid**
980 **producers and non-producers identified by FELLA .**

981 **Supplementary Table 4. Correlations between gene expressions or pathways with**
982 **rhamnolipid production in RLQ analysis.** For each rhamnolipid production category (strong-,
983 mild- and non-producers), its correlation value with a single gene or a functional pathway was
984 computed as the dot products between the arrow of phenotypic category and the arrow of the gene
985 or pathway in RLQ axes.

986 **Supplementary Table 5. Normalized peak area of metabolomics**

987 **Supplementary Table 6. Read counts of *P. aeruginosa* RNAseq aligned to the PA14 genome**

988 **Supplementary Table 7. Metadata of RNAseq archived in NCBI SRA database**

989

990

991