1   # When is a bacterial "virulence factor" really virulent?

2

3

4   Elisa T. Granato,[a#] Freya Harrison,[b] Rolf Kümmerli,[a] Adin Ross-Gillespie[a]*

5

6   Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland[a],

7   Centre for Biomolecular Sciences, University of Nottingham, Nottingham, United

8   Kingdom[b]

9

10   #Address correspondence to Elisa T. Granato, egranato@botinst.uzh.ch.

11

12   *Present address: Adin Ross-Gillespie, SIB Swiss Institute of Bioinformatics,

13   Lausanne, Switzerland.

ABSTRACT

Bacterial traits that contribute to disease are termed 'virulence factors' and there is much interest in therapeutic approaches that disrupt such traits. However, ecological theory predicts disease severity to be multifactorial and context dependent, which might complicate our efforts to identify the most generally important virulence factors. Here, we use meta-analysis to quantify disease outcomes associated with one well-studied virulence factor – pyoverdine, an iron-scavenging compound secreted by the opportunistic pathogen *Pseudomonas aeruginosa*. Consistent with ecological theory, we found that the effect of pyoverdine, albeit frequently contributing to disease, varied considerably across infection models. In many cases its effect was relatively minor, suggesting that pyoverdine is rarely essential for infections. Our work demonstrates the utility of meta-analysis as a tool to quantify variation and overall effects of purported virulence factors across different infection models. This standardised approach will help us to evaluate promising targets for anti-virulence approaches.

INTRODUCTION

30

31   Understanding which bacterial characteristics contribute most to disease is a major

32   area of research in microbiology and infection biology (1–3). Bacterial characteristics

33   that reduce host health and/or survival are considered 'virulence factors'. Such

34   factors include structural features like flagella and pili that facilitate attachment to

35   host cells (4, 5), as well as secreted products like toxins and enzymes that degrade

36   host tissue (6, 7), or siderophores that scavenge iron from the host (8). Research on

37   virulence factors has not only increased our fundamental understanding of the

38   mechanisms underlying virulence, but has also identified potential novel targets for

39   antibacterial therapy. There is indeed much current interest in developing 'anti-

40   virulence' drugs to disrupt virulence factor production – the idea being that by simply

41   disarming pathogens rather than killing them outright, we could ostensibly elicit

42   weaker selection for drug resistance (9–11).

43

44   Although our understanding of different types of virulence factors and their

45   interactions is continuously deepening, it is still unclear just how generalizable this

46   assembled knowledge is. It is often assumed, for reasons of parsimony, that a given

47   structure or secreted molecule central to the virulence of a particular bacterial strain

48   in a specific host context will similarly enhance virulence in another bacterial strain,

49   or in a different host (12). Yet, ecological theory predicts that the effects of a given

50   trait will frequently vary in response to the environment (78). In the context of

51   infections, this may be particularly true for opportunistic pathogens, which face very

52   heterogeneous environments: they can live in environmental reservoirs (e.g. soil,

53   household surfaces), as commensals of healthy hosts, or, when circumstances allow,

54   as pathogens, causing serious infections in a range of different hosts and host

3

55   tissues (13–15). Opportunistic pathogens underlie many hospital-acquired infections,

56   especially in immune-compromised patients (16–18), and the treatment of such

57   infections is often challenging because, as generalists, such pathogens are pre-

58   selected to be tenacious and highly adaptable. Thus, in designing new anti-virulence

59   drugs against opportunistic pathogens, we need to know not only whether the

60   targeted trait is indeed associated with pathogenicity, but also the generality of this

61   association across different pathogen strain backgrounds, host species, and infection

62   types.

63

64   Here we show how a meta-analysis approach can be used to quantify variation and

65   overall effects of virulence factors across host environments. As a test case, we

66   focus on pyoverdine, a siderophore secreted by the opportunistic pathogen

67   *Pseudomonas aeruginosa* to scavenge iron from the host environment (19). Table 1

68   provides an overview of the workflow of our meta-analysis, where we combined the

69   outcomes of 76 individual virulence experiments from 23 studies (12, 20–41, see also

70   Tables S1 and S2 in the supplemental material). Using a weighted meta-analysis

71   approach, we were able to investigate the evidence for pyoverdine's contribution to

72   virulence across eight host species, including vertebrates, invertebrates and plants,

73   five tissue infection models and various *P. aeruginosa* genotypes. We chose

74   pyoverdine production as the model trait for our analysis because: (i) it has been

75   extensively studied across a range of *Pseudomonas* strains (42); (ii) its virulence

76   effects have been examined in a large number of host species; (iii) *P. aeruginosa* is

77   one of the most troublesome opportunistic human pathogens, responsible for many

78   multi-drug resistant nosocomial infections (43, 44); and (iv) multiple anti-virulence

79   drugs have been proposed to target pyoverdine production and uptake (22, 23, 45).

4

80 The applied question here, then, is whether targeting pyoverdine could generally and

81 effectively curb pathogenicity.

82

83 **RESULTS**

84 **Literature search and study characteristics.** We searched the literature for

85 papers featuring infections of whole live host organisms with *P. aeruginosa* strains

86 known to vary in pyoverdine phenotype. Following a set of inclusion/exclusion rules

87 (see materials and methods for details), we were able to include data from a total of

88 76 experiments from 23 original papers in our meta-analysis (Table 1; see also Fig.

89 S1 and Tables S1 and S2 in the supplemental material). These experiments featured

90 a range of host organisms, including mammals (mice and rabbits, n = 32), the

91 nematode *Caenorhabditis elegans* (n = 32), insects (fruit fly, silk worm and wax

92 worm, n = 8) and plants (wheat and alfalfa, n = 4). Experiments further differed in the

93 way infections were established and in the organs targeted. The most common

94 infection types were gut (n = 34), systemic (n = 16), respiratory (n = 8) and skin

95 infections (n = 6), but we also included some other types of infections (n = 12). Each

96 experiment compared infections with a control *P. aeruginosa* strain (which produced

97 wildtype levels of pyoverdine) to infections with a mutant strain defective for

98 pyoverdine production. The most common control strains used were PAO1 (n = 53)

99 and PA14 (n = 19), which are both well-characterized clinical isolates. However,

100 some experiments used less well-characterized wildtype strains, such as FRD1 (n =

101 2) and PAO6049 (n = 2). Twenty-six experiments used mutant strains with clean

102 deletions or transposon Tn5 insertions in genes encoding the pyoverdine

103 biosynthesis pathway. In these cases, pleiotropic effects are expected to be relatively

104 low – i.e. presumably only pyoverdine production was affected. The other 50

105 experiments used mutants where pleiotropic effects were likely or even certain. For

106 example, some mutant strains carried mutations in *pvdS*, which encodes the main

107 regulator of pyoverdine production that also regulates the production of toxins and

108 proteases (46, 47). Others carried mutations in *pvdQ*, encoding an enzyme known to

109 degrade quorum-sensing molecules in addition to its role in pyoverdine biosynthesis

110 (27).

111

112 **Relationship between effect sizes and moderator variables.** We combined data

113 from the set of experiments described above in a meta-analysis to determine the

114 extent to which pyoverdine's effect on virulence varied across four moderator

115 variables: (i) host taxa, (ii) tissue types, (iii) pathogen wildtype background, and (iv)

116 pyoverdine-mutation type. To obtain a comparable measure of virulence across

117 experiments, we extracted in each instance the number of cases where a given

118 infection type did or did not have a virulent outcome (i.e. dead vs. alive, or with vs.

119 without symptoms) for both the mutant (m) and the wildtype (w) strain for each

120 experiment (see materials and methods for details). We then took as our effect size

121 the log-odds-ratio, i.e. $\ln((m_{virulent} / m_{non\text{-}virulent}) / (w_{virulent} / w_{non\text{-}virulent}))$ (see Table S2

122 in the supplemental material), a commonly-used measure especially suitable for

123 binary response variables like survival (48).

124 Consistent with the theoretical prediction that host-pathogen interactions and host

125 ecology are important modulators of virulence, we found considerable variation in

126 the effect sizes across experiments and subgroups of all moderators (Fig. 1).

127 Pyoverdine-deficient mutants showed substantially reduced virulence in invertebrate

128 and mammalian hosts, whereas there was little evidence for such an effect in plants

129 (Fig. 1A). Overall, evidence for pyoverdine being an important virulence factor was

130   weak for taxa with a low number of experiments (i.e. for plants, and the insect models

131   *Drosophila melanogaster* and *Galleria mellonella*). We found that pyoverdine-

132   deficient mutants exhibited reduced virulence in all organs and tissues tested, with

133   the exception of plants (Fig. 1B). Comparing the effect sizes across wildtype strain

134   backgrounds, we see that pyoverdine deficiency reduced virulence in experiments

135   featuring the well-characterized PA14 and PAO1 strains (Fig. 1C) whereas the

136   reduction was less pronounced in experiments with less well-characterized wildtype

137   strains. This could be due to sampling error (only a few experiments used these

138   strains) or it may be that these strains really behave differently from PA14 and PAO1.

139   Finally, we observed that the nature of the pyoverdine-deficiency mutation matters

140   (Fig. 1D). Infections with strains carrying well-defined mutations known to exclusively

141   (or at least primarily) affect pyoverdine production showed a relatively consistent

142   reduction in virulence. Conversely, where mutants were poorly-defined, or carried

143   mutations likely to affect other traits beyond pyoverdine, here the virulence pattern

144   was much more variable, with both reduced and increased virulence relative to

145   wildtype infections (Fig. 1D). We posit that at least some of the differences in

146   observed virulence between these mutants and their wildtype counterparts was likely

147   due to pleiotropic differences in phenotypes unrelated to pyoverdine.

148

149   **Assessing the relative importance of moderator variables.** Fig. 1 highlights

150   that we are dealing with an extremely heterogeneous dataset (a random meta-

151   analyses of the full dataset without moderators yielded heterogeneity measures I^2

152   = 98.1% and H = 7.28). Much of the variation we observe is probably due to other

153   factors beyond those explored in Fig 1. The issue is that (a) we do not know what

154   all these additional factors might be, and (b) the probably patchy distribution of

7

155 experiments across the levels and ranges of these other factors would leave us

156 with limited power to test for their effects. Accordingly, we decided to focus our

157 attention on quantifying the impact of the four previously described moderators by

158 using a more homogenous core dataset (n = 50), where rare and poorly

159 characterized subgroups were removed. Specifically, we excluded experiments

160 involving plants and / or undefined wildtype strains (n = 6), experiments reporting

161 tissue damage as a measure of virulence (n = 12), and experiments where the hosts

162 were likely not colonized by bacteria but died from exposure to bacterial toxins (n =

163 8). This leaves us with a core dataset comprising only those experiments where

164 animal host models were infected with strains from well-defined PA14 or PA01

165 wildtype background, and survival vs. death was used as a virulence endpoint.

166

167 Using this restricted dataset, we performed a series of meta-regression models to

168 test for significant differences between subgroups of our moderator factors, and we

169 also estimated the share of total variance in effect sizes that is explained by each

170 moderator variable (Fig. 2). These models revealed that infection type is the

171 variable that explains the largest share of total variance (25.6%). For instance, in

172 systemic infection models the pyoverdine-defective mutants showed strongly

173 reduced virulence compared to the wild-type, whereas this difference was less

174 pronounced in gut infections. Host taxon explained only 7.9% of the total variance

175 in effect sizes, and there was no apparent difference in the mean effect size

176 among invertebrate vs. mammalian host models. Finally, the wildtype strain

177 background and the likelihood of pleiotropy in the mutant strain both explained less

178 than 1% of the overall effect size variation, and accordingly, there were no

179 apparent differences between subgroups (Fig. 2). Note that even with the inclusion

180   of these moderator factors in the model, substantial heterogeneity remained in our

181   restricted data set (I^2 = 96.8%, H = 5.60).

182

183   **Publication bias.** In any field, there is a risk that studies with negative or

184   unanticipated results may be less likely to get published (e.g. in our case,

185   pyoverdine-deficient mutants showing no change or increased levels of virulence)

186   (49). Especially when negative or unanticipated results are obtained from

187   experiments featuring low sample sizes (and thus high uncertainty), the scientists

188   responsible may be less inclined to trust their results, and consequently opt not to

189   publish them. This pattern could result in a publication bias, and an overestimation of

190   the effect size. To test whether such a publication bias exists in our dataset, we

191   plotted the effect size of each experiment against its (inverted) standard error (Fig.

192   3). If there is no publication bias, we would expect to see an inverted funnel, with

193   effect sizes more or less evenly distributed around the mean effect size, irrespective

194   of the uncertainty associated with each estimate (i.e. position on the y-axis). Instead,

195   we observed a bias in our dataset, with many lower-certainty experiments that show

196   strongly negative effect sizes (i.e. supporting the hypothesis that pyoverdine is

197   important for virulence; Fig. 3) but a concomitant paucity of lower-certainty

198   experiments that show weakly negative, zero or positive effect sizes (i.e. not

199   supporting the hypothesis).

DISCUSSION

What we can conclude from this meta-analysis. Our meta-analysis reveals that

pyoverdine-deficient strains of the opportunistic pathogen *P. aeruginosa* typically

showed reduced virulence across a wide range of host species and bacterial

genotypes. This confirms that iron limitation is a unifying characteristic of the host

environment, making siderophores an important factor for pathogen establishment

and growth within the host (50, 51). Conversely, we also saw that the extent to which

pyoverdine deficiency reduced virulence varied considerably, and was quite modest

in many instances. Pyoverdine-deficient mutant strains were typically more benign,

owing to a reduced capacity for *in vivo* growth and/or a reduced capacity for inflicting

damage on their host. Nonetheless, these mutants were typically still able to

establish a successful infection, and, in many cases, could still kill their host (21, 22,

25, 40). These results support ecological theory predicting that the effect of a certain

phenotype (i.e. producing pyoverdine in our case) should vary in response to the

environment (i.e. the host and infection context). Our findings have consequences for

any therapeutic approaches targeting this particular virulence factor as they reveal a

possible trade-off: such treatments could have wide applicability, but their (clinical)

impact would likely vary across infection contexts, and be limited to attenuating rather

than curing the infection. This would mean that for *P. aeruginosa* infections, at least,

therapies targeting siderophore production could be helpful but should probably still

be accompanied by other therapeutic measures (52).

Our work demonstrates how meta-analyses can be used to quantitatively synthesize

data from different experiments carried out at different times by different researchers

using different designs. Such an analytical approach goes beyond a classical review,

10

225   where patterns are typically summarized in a qualitative manner. For instance, a

226   recent study proposed that three different virulence factors (pyocyanin, protease,

227   swarming) of *P. aeruginosa* are host-specific in their effects (12). Here we use a

228   meta-analytic approach to quantitatively derive estimates of the overall virulence

229   potential of a given bacterial trait and investigate variables that affect infection

230   outcomes. We assert that such quantitative comparisons are essential to identify

231   those virulence factors that hold greatest promise as targets for effective broad-

232   spectrum anti-virulence therapies.

233

234   Our finding that effect sizes vary considerably across our assembled experiments

235   provides a different perspective compared to that which one would obtain from a

236   cursory reading of the literature. For instance, the first study investigating pyoverdine

237   in the context of an experimental infection model (38) reported that pyoverdine is

238   essential for virulence. Although this experiment and its message have been widely

239   cited (including by ourselves), it may no longer be the strongest representative of the

240   accumulated body of research on this topic. As we see in Figure 1, the effect size it

241   reports is associated with a high uncertainty due to a comparatively low sample size.

242   Moreover, the observed effect cannot unambiguously be attributed to pyoverdine

243   because an undefined UV-mutagenized mutant was used. We highlight this example

244   not to criticise it, but rather because it serves to demonstrate why drawing inferences

245   from (appropriately weighted) aggregations of all available evidence is preferable to

246   focusing solely on the results of a single study.

247

248   **What we could conclude with additional data.** Our meta-analytic approach not

249   only provides information on the overall importance of pyoverdine for *P. aeruginosa*

11

250   virulence, but it also allows us to identify specific gaps in our knowledge and potential

251   biases in the published literature. First, most experiments in our dataset employed

252   acute infection models, even though *P. aeruginosa* is well known for its persistent,

253   hard-to-treat chronic infections. This raises the question to what extent insights on

254   the roles of virulence factors important in acute infections can be transferred to

255   chronic infections. In the case of pyoverdine, we know that in chronically-infected

256   cystic fibrosis airways, pyoverdine production is often selected against (54–56).

257   Although the selective pressure driving this evolutionary loss is still under debate

258   (current explanations include pyoverdine disuse, competitive strain interactions

259   and/or a switch to alternative iron-uptake systems (56–58)), this example illustrates

260   that the role of pyoverdine might differ in acute versus chronic infections.

261

262   Second, our comparative work shows that experiments were predominantly carried

263   out with the well-characterized strains PAO1 and PA14. While these strains were

264   initially isolated from clinical settings, they have subsequently undergone evolution in

265   the laboratory environment (59–61), and might now substantially differ from the

266   clinical strains actually causing acute infections in hospitals. Therefore, while we

267   found no overall differences between the lab strains used in our data set, we argue

268   that it would still be useful to carry out additional studies on a range of clinical

269   isolates to be able to make firm conclusions on the general role of pyoverdine as a

270   virulence factor.

271

272   Finally, our data analysis revealed that low-certainty studies showing no or small

273   effects of pyoverdine on virulence were under-represented in our data set, which

274   points towards a systematic publication bias. It remains to be seen whether such

12

275   biases are common with regard to research on virulence factors, and whether they

276   result in a general overestimation of the effect these factors have on host survival or

277   tissue damage. With regard to pyoverdine, further studies are clearly needed to

278   obtain a more accurate estimate of the true effect size.

279

280   **Guidelines for future studies.** While our study demonstrates the strength of

281   quantitative comparative approaches, it is important to realize that extracting effect

282   sizes is one of the biggest challenges in any meta-analysis. This challenge was

283   particularly evident for the experiments we found, which profoundly varied in the way

284   data was collected and reported. As a consequence, we had to exclude many studies

285   because they used measures of virulence that were only reported by a minority of

286   studies, or because their reporting of results was unclear (for a selected list of

287   examples, see Table S3 in the supplemental material). To amend this issue for future

288   studies, we would like to first highlight the problems we encountered and then

289   provide general guidelines of how data reporting could be improved and

290   standardized. One main problem we experienced was incomplete data reporting (i.e.

291   mean treatment values, absolute values and/or sample size was not reported), which

292   prevents the calculation of effect sizes and uncertainty measures. Another important

293   issue was that different studies measured virulence using very different metrics.

294   Some measured virulence at the tissue level (i.e. the extent of damage inflicted),

295   while others focused on the whole host organism. Others focused on the dynamics of

296   the bacteria themselves, taking this as a proxy for the eventual damage to the host.

297   There were both quantitative measures (e.g. extent of damage), and qualitative

298   measures (e.g. assignments to arbitrary categories of virulence). Survival data was

299   sometimes presented as a timecourse, sometimes as an endpoint; sometimes as raw

13

300   counts, sometimes as proportions. In most cases, the time scales over which survival

301   was assessed were fairly arbitrary. Compiling such diverse measures of virulence is

302   not simply time consuming, but it also generates extra sources of heterogeneity in

303   the dataset, which might interfere with the basic assumptions of meta-analytical

304   models (62, 63).

305

306   How can these problems be prevented in future studies? We propose the following.

307   (a) Whenever possible, time-to-event data (e.g. death, organ failure, etc.) should be

308   recorded in a form that preserves both the outcome and the times to event per

309   subject. (b) The number of replicates used (hosts) and a measure of variance among

310   replicates must be provided to be able to calculate a confidence estimate for the

311   experiment. (c) If data are scaled in some way (e.g. relative to a reference strain), the

312   absolute values should still be reported, because these are crucial for the calculation

313   of effect sizes. Finally, (d) studies leading to unexpected or negative results (e.g. no

314   difference in virulence between a wildtype and a mutant) should still be published, as

315   they are needed to estimate a true and unbiased effect size. In summary, all findings,

316   irrespective of their magnitude or polarity, should be presented "as raw as possible"

317   (e.g. in supplementary files or deposited in online data archives). This will make

318   comparisons across studies much easier and will provide a useful resource for future

319   meta-analytic studies.

320

321   **Conclusions.** Currently, bacterial traits are subject to a binary categorisation

322   whereby some are labelled as virulence factors while others are not. We demonstrate

323   that traits' effects on virulence are anything but binary. Rather, they strongly depend

324   on the infection context. Our study affirms meta-analysis as a powerful tool to

14

325   quantitatively estimate the overall effect of a specific virulence factor and to compare

326   its general importance in infections across different bacterial strains, hosts, and host

327   organs. Such quantitative comparisons provide us with a more complete picture on

328   the relative importance of specific virulence factors. Such knowledge is especially

329   valuable for opportunistic pathogens, which have a wide range of virulence factors at

330   their disposal, and infect a broad range of host organisms (13–15). Meta-analytical

331   comparisons could thus inform us on which traits would be best suited as targets for

332   anti-virulence therapies. Ideal traits would be those with high effect sizes and general

333   importance across pathogen and host organisms.

MATERIALS AND METHODS

Literature search. We conducted an extensive literature search, using a combination of two online databases: Web of Science and Google Scholar. The following terms were used to search abstracts and full texts: "aeruginosa" in combination with "pyoverdin" or "pyoverdine" and in combination with "virulence" or "infection" or "pathogen" or "disease" or "mortality" or "lethality". This search was first performed on May 19th 2014 and it was repeated periodically until Sep 21st 2015 in order to include more recent publications. In addition, the reference lists of all shortlisted studies were scanned for relevant publications. We further contacted the corresponding authors of several publications to ask for unpublished datasets.

Inclusion criteria. The database search yielded a total of 442 studies, and we identified 10 additional records through other sources. These 452 studies were then scanned for relevant content according to the following set of inclusion criteria. Studies were considered potentially eligible for inclusion if they contained original research, were written in English and provided data that compared the virulence of a wildtype pyoverdine-producing *P. aeruginosa* strain with that of a mutant strain demonstrating impaired pyoverdine production. We defined virulence as a decrease in host fitness, measured as an increase in mortality or tissue damage when infected with bacteria. We defined "wildtype strains" as strains that were originally clinical isolates, have been widely used in laboratories as virulent reference strains and have not been genetically modified. Strains with impaired pyoverdine production included strains that were completely deficient in pyoverdine production and strains that were only partially deficient, i.e. that produced less than the wildtype strain under identical experimental conditions. We considered both genetically engineered knock-out

16

359   strains and clinical isolates with reduced pyoverdine production. There were 31

360   original publications containing 115 experiments that satisfied these criteria and were

361   thus considered appropriate for in-depth examination.

362

363   We screened these 115 experiments using a second set of rules to identify those

364   experiments that contain comparable quantitative data, which is essential for a meta-

365   analysis. Inclusion criteria were: (i) virulence was measured directly (and not inferred

366   indirectly via genetic analysis); (ii) virulence was measured *in vivo* (and not *in vitro* via

367   virulence factor production); (iii) virulence was measured quantitatively as direct

368   damage to the host caused by bacterial infections, and not by indirect or qualitative

369   measures such as bacterial growth performance in the host, threshold infective dose

370   required to kill a host, the damage associated with virulence factor administration, or

371   resistance to macrophage-like predation (53, 64, 65); and (iv) absolute virulence data

372   were presented (and not only data scaled relative to the wildtype without information

373   on the absolute risk of mortality, since effect sizes cannot be calculated from such

374   data). This second set of rules was fulfilled by 23 original publications containing 76

375   individual experiments (see Tables S1 and S2 in the supplemental material). For an

376   overview of the whole selection process, see Fig. S1 in the supplemental material.

377

378   **Data extraction and effect size calculations.** From all of these 76 experiments, we

379   extracted information on: (i) the host organism; (ii) the type of infection; (iii) the

380   observation period of infected hosts; (iv) the identity of the control (wildtype) strain;

381   (v) the identity of the pyoverdine-deficient strain; (vi) the mutated gene in the

382   pyoverdine-defective strain; (vii) the mutation type (e.g. insertion/deletion); (viii) the

383   sample size used for the wildtype and mutant experiments; and (ix) the relevant

384   virulence measure (host survival or tissue damage) for wildtype and mutant strains

385   (see Table S1 in the supplemental material). This information was used to

386   categorize the experiments and identify potentially important moderator variables

387   (see below).

388

389   Next, we extracted quantitative data from these experiments so we could calculate

390   effect sizes for the virulence associated with pyoverdine production. For mortality

391   assays, we extracted raw counts of how many individuals died and how many

392   survived following infection with a dose of *P. aeruginosa* wildtype or, alternatively, a

393   mutant strain known to be deficient for pyoverdine production. For experiments on

394   tissue damage, we extracted information on the number of individuals with and

395   without the symptoms related to tissue damage (e.g. a lesion in an organ). In cases

396   with zero counts (i.e. either all or none of the individuals in a particular treatment

397   group died or experienced tissue damage), we converted counts to 0.5 to avoid

398   having zero denominators in the subsequent calculations of the (log-odds ratio) effect

399   sizes (66). In cases where data from multiple time-points or survival curves were

400   available, we concentrated on the time point with the largest difference between the

401   wildtype and the mutant infection.

402

403   Using this count data, we calculated the effect size for each experiment as the log-

404   odds-ratio = ln $((m_{virulent} / m_{non-virulent}) / (w_{virulent} / w_{non-virulent}))$, where $m_{virulent}$ and

405    $w_{virulent}$ are the number of individuals that died or experienced tissue damage when

406   infected by the mutant and the wildtype strain, respectively, and $m_{non-virulent}$ and

407    $w_{non-virulent}$ are the number of individuals that survived or remained unharmed by the

408   infection. Information on the sample size was used to calculate the 95% confidence

18

409   interval for each effect size and for weighting effect sizes relative to one another (see

410   details below). Where experiments reported a range of sample sizes, we used the

411   arithmetic mean. Some studies reported only a minimum sample size. In those

412   cases, we used this number. For experiments using *C. elegans*, infections were often

413   carried out on replicate petri dishes in a large number of individuals. In these cases,

414   we used the total number of individual worms used in each treatment group as

415   sample size, and not the number of replica plates.

416

417   For some experiments conducted in mammals, data on *in vivo* growth of a wildtype

418   and a pyoverdine deficient strain were available in addition to virulence measures (see

419   Table S4). We also calculated effect sizes (standardized mean differences) for this set

420   of studies (Fig. S4). This limited dataset shows a similar pattern to the main dataset

421   shown in Fig. 1, but was not included in the main analysis because we were primarily

422   interested in quantifying virulence effects (i.e. host damage and/or mortality) and not

423   pathogen growth.

424

425   **Moderator variables.** We considered four moderator variables (host taxon, infection

426   type, wildtype strain background, likelihood of pleiotropy associated with pyoverdine

427   deficiency) that could potentially explain variation in virulence. In cases where

428   information was missing for a specific moderator variable, we contacted the authors

429   to obtain additional information. For each moderator, we defined the following

430   relevant subgroups.

431   *Host organism* – We first split experiments into broad taxonomic units (mammals,

432   invertebrates, plants), and then classified hosts by genus.

433   *Infection type* – We classified experiments according to the organ or body region

434   targeted by the infection. Major categories include infections of the host organisms'

435   respiratory tract, digestive system, skin (including burn wounds), and infections that

436   generated a non-localized infection of the body cavity (systemic infection).

437   Experiments that did not fit in any of these categories, such as infections of whole

438   seedlings, were classified as 'other infection types'.

439   *Wildtype strain background* – Four different *P. aeruginosa* wildtypes (PAO1, PA14,

440   FRD1 and PAO6049) were used for infection experiments. Although it is well

441   established that even standard strains such as PAO1 can substantially differ between

442   labs, there was not enough information available to take such strain-level variation into

443   account.

444   *Likelihood of pleiotropy* – The focal phenotype investigated in this meta-analysis is the

445   production of pyoverdine, the main siderophore of *P. aeruginosa*. Mutants exhibiting

446   reduced or no pyoverdine production can be generated either by deleting a specific

447   pyoverdine-synthesis gene, or through untargeted mutagenesis (e.g. UV light)(67).

448   The latter mutants are likely to have mutations in other genes unrelated to pyoverdine

449   synthesis. These mutations are typically unknown but could also affect virulence. In

450   principle, even single gene deletions can have pleiotropic effects on the phenotype,

451   via disruption of interactions with other genes. Depending on the locus in question,

452   certain genetic modifications are more likely to induce pleiotropy than others. To

453   account for these complications, we inferred on a case-by-case basis whether the

454   mutation used was likely to only induce a change in (or loss of) pyoverdine production

455   (i.e. pleiotropy less likely) or was likely to induce a change in other phenotypes as well

456   (i.e. pleiotropy more likely). In the biosynthesis of pyoverdine, multiple enzymes are

457   involved in non-ribosomal peptide synthesis (19)**.** Two gene clusters, the *pvc* operon

458 and the *pvd* locus, encode proteins involved in the synthesis of the chromophore and

459 peptide moieties, respectively (19, 68). In most of these genes, a mutation or deletion

460 leads to a complete loss of pyoverdine production, and most likely does not affect any

461 other trait. Accordingly, we assigned mutants carrying mutations in these genes to the

462 category "pleiotropy less likely". An exception is *pvdQ*, a gene coding for a periplasmic

463 hydrolase, which is required for pyoverdine production, but is also involved in the

464 degradation of N-acyl-homoserine lactone quorum-sensing molecules (27). Strains

465 with deletions in this gene were therefore assigned to the category "pleiotropy more

466 likely". Other strains falling into this category included: (i) mutants where the key

467 regulator of pyoverdine synthesis, PvdS, was deleted, leading to deficiencies in toxin

468 and protease production, in addition to a complete loss of pyoverdine production (69);

469 (ii) strains that carry a deletion in a central metabolic gene and only coincidentally show

470 no (or strongly reduced) pyoverdine production; (iii) double mutants that carry deletions

471 in both the pyoverdine and the pyochelin synthesis pathway (pyochelin is the

472 secondary siderophore of *P. aeruginosa*) (70); and (iv) pyoverdine mutants created via

473 non-targeted (e.g. UV) mutagenesis.

474

475 **Core dataset.** To quantify the impact of the moderator variables, we removed

476 experiments belonging to rare or poorly characterized subgroups to generate a more

477 homogenous core dataset. We excluded experiments involving plants and / or

478 undefined wildtype strains (n = 6), experiments reporting tissue damage as a measure

479 of virulence (n = 12), and experiments where the hosts were likely not colonized by

480 bacteria but died from exposure to bacterial toxins (n = 8). This resulted in a core

481 dataset comprising 50 experiments that was used for subsequent analyses on the

482 influence of moderator variables.

483

484   **Statistical analysis.** Analyses were performed in R version 3.2.3 (71), using

485   functions from packages 'meta' (72) and 'metafor' (73). We used the 'metabin'

486   function to transform the count data into the (log-) odds ratio described above. We

487   then weighted these values by the inverse of their respective squared standard

488   errors, and pooled them to obtain a single distribution of effect sizes. We reasoned

489   that the variability of the effect sizes in our dataset probably reflects more than simple

490   sampling error around a single true mean. Rather, we assume that our effect sizes

491   represent a random sample from a larger distribution comprising all possible true

492   effect size estimates. As such, we inferred that a random effects meta-analysis would

493   be more appropriate for our dataset than a fixed effects model (for further discussion,

494   see (62, 63). In a random effects meta-analysis, we partition the total heterogeneity

495   observed in our dataset (described by the statistic Q) into two constituent parts –

496   within-experiment variation (ε) and between-experiment variation (ζ). The latter

497   component, scaled appropriately to account for the weightings intrinsic to meta-

498   analysis, is quantified as the τ^2 statistic. There are several different algorithms one

499   can use to effect this partitioning of variance. We chose a restricted maximum

500   likelihood (REML) approach. The use of a random model, rather than a simpler fixed

501   model, affects the weights accorded to each constituent effect size, which in turn

502   changes our estimates for pooled means and their associated errors. We further

503   slightly broadened confidence intervals and weakened test statistics using Knapp

504   and Hartung's algorithm (74) – a widely-used and conservative adjustment designed

505   to account for the inherent uncertainty associated with the partitioning of

506   heterogeneity we perform in the course of fitting a random effects model.

507

508   We assessed the degree of residual heterogeneity in our dataset using statistics I^2

509   and H. I^2 estimates the approximate proportion of total variability across experiments

510   that is attributable to unexplained heterogeneity, as opposed to simple sampling error

511   (chance). H reports 'excess' heterogeneity as a fold difference compared to the

512   baseline amount of variability we would have expected if the sample were

513   homogenous (75).

514

515   Both metrics described above indicated considerable residual heterogeneity in our

516   dataset, so we suspected that, beyond the random- and sampling error, some

517   measurable characteristics of the experiments in our dataset could be contributing, in

518   predictable ways, to the observed heterogeneity of our assembled effect sizes. We

519   investigated four potential moderators, namely the host taxon, the type of infection,

520   the wildtype background of the infecting strains, and the type of mutant involved (i.e.

521   whether more or less pleiotropy was expected). In a first approach, we split the

522   dataset into subgroups representing different levels of these moderators, and

523   estimated pooled means within these different subgroups. For this, we again used

524   random-effect meta-analysis (as above), but we set the level of between-experiment

525   heterogeneity (τ^2) to be common across all subgroups.

526

527   In a second approach, we fitted a series of meta-regression models that extended

528   our basic model to additionally consider the contributions of multiple moderator

529   factors. Our models were able to estimate moderators' additive effects only, because

530   the distribution of data across different combinations of factor levels was too patchy

531   to permit a proper investigation of moderators' interactive effects. Moderators'

532   alterations of the expected (i.e. baseline) effect size could be quantified as

23

533   coefficients, which could, when standardized as *t*-statistics, be tested for significant

534   differences from zero. In addition, we could test whether, collectively, the inclusion of

535   moderators in our meta-analysis model significantly reduced the residual

536   heterogeneity relative to a situation with no moderators.

537

538   To estimate what share of the residual heterogeneity in our dataset could be

539   individually attributable to each of the respective moderators, we performed a series

540   of likelihood ratio tests comparing, in each case, a full model including all four

541   moderators, against a reduced model that excluded one of the moderators. Variance

542   component estimation in these models used maximum likelihood instead of REML

543   because nested REML models cannot be compared in this way. From each pairwise

544   comparison, we obtained a pseudo-R^2 value, which reflects the difference in τ^2

545   (between-experiment heterogeneity) between the two models, scaled by the τ^2 of the

546   simpler model.

547

548   To test for putative publication bias in our dataset, we compared effect sizes against

549   their respective standard errors, the idea being that if there is no bias, there should

550   be no link between the magnitude of the result from a given experiment, and the

551   'noisiness' or uncertainty of that particular result. If there is bias, we could find an

552   overrepresentation of noisier experiments reporting higher magnitude results. Using

553   the 'metabias' function of the R package 'meta', we performed both (weighted) linear

554   regressions and rank correlations to test for this pattern (76, 77).

AUTHOR CONTRIBUTIONS

E.G., F. H., R.K., and A.R.G. conceived the study; E.G. conducted the literature search and compiled the data set; A.R.G. conducted statistical analysis; E.G., F.H., R.K., and A.R.G. interpreted the data and wrote the paper.

REFERENCES

1. **Jimenez PN**, **Koch G**, **Thompson J a.**, **Xavier KB**, **Cool RH**, **Quax WJ**. 2012. The Multiple Signaling Systems Regulating Virulence in Pseudomonas aeruginosa. Microbiol Mol Biol Rev **76**:46–65.

2. **Lebeaux D**, **Ghigo JM**, **Beloin C**. 2014. Biofilm-related infections: bridging the gap between clinical management and fundamental aspects of recalcitrance toward antibiotics. Microbiol Mol Biol Rev **78**:510–543.

3. **Rahme LG**, **Stevens EJ**, **Wolfort SF**, **Shao J**, **Tompkins RG**, **Ausubel FM**. 1995. Common virulence factors for bacterial pathogenicity in plants and animals. Science (80-) **268**:1899–1902.

4. **Josenhans C**, **Suerbaum S**. 2002. The role of motility as a virulence factor in bacteria. Int J Med Microbiol **291**:605–614.

5. **Kazmierczak BI**, **Schniederberend M**, **Jain R**. 2015. Cross-regulation of Pseudomonas motility systems: The intimate relationship between flagella, pili and virulence. Curr Opin Microbiol **28**:78–82.

6. **Vasil M**, **Chamberlain C**, **Grant C**. 1986. Molecular studies of Pseudomonas exotoxin A gene. Infect Immun **52**:538–548.

7. **Lebrun I**, **Marques-Porto R**, **Pereira a S**, **Pereira A**, **Perpetuo E a**. 2009. Bacterial toxins: an overview on bacterial proteases and their action as virulence factors. Mini Rev Med Chem **9**:820–828.

8. **Miethke M**, **Marahiel MA**. 2007. Siderophore-based iron acquisition and pathogen control. Microbiol Mol Biol Rev **71**:413–51.

9. **Clatworthy AE**, **Pierson E**, **Hung DT**. 2007. Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol **3**:541–548.

589   10. **Rasko DA**, **Sperandio V**. 2010. Anti-virulence strategies to combat bacteria-
590   mediated disease. Nat Rev Drug Discov **9**:117–128.

591   11. **Pepper JW**. 2008. Defeating pathogen drug resistance: Guidance from
592   evolutionary theory. Evolution (N Y) **62**:3185–3191.

593   12. **Dubern JF**, **Cigana C**, **De Simone M**, **Lazenby J**, **Juhas M**, **Schwager S**,
594   **Bianconi I**, **Döring G**, **Eberl L**, **Williams P**, **Bragonzi A**, **Cámara M**. 2015.
595   Integrated whole-genome screening for Pseudomonas aeruginosa virulence
596   genes using multiple disease models reveals that pathogenicity is host specific.
597   Environ Microbiol **17**:4379–4393.

598   13. **Kurz CL**, **Chauvet S**, **Andres E**, **Aurouze M**, **Vallet I**, **Michel GPF**, **Uh M**,
599   **Celli J**, **Filloux A**, **De Bentzmann S**, **Steinmetz I**, **Hoffmann JA**, **Finlay BB**,
600   **Gorvel JP**, **Ferrandon D**, **Ewbank JJ**. 2003. Virulence factors of the human
601   opportunistic pathogen Serratia marcescens identified by in vivo screening.
602   EMBO J **22**:1451–1460.

603   14. **Staniszewska M**, **Bondaryk M**, **Piłat J**, **Siennicka K**, **Magda U**,
604   **Kurzątkowski W**. 2012. Virulence Factors of Candida Albicans. Trends
605   Microbiol **66**:629–633.

606   15. **He J**, **Baldini RL**, **Déziel E**, **Saucier M**, **Zhang Q**, **Liberati NT**, **Lee D**, **Urbach**
607   **J**, **Goodman HM**, **Rahme LG**. 2004. The broad host range pathogen
608   Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands
609   harboring plant and animal virulence genes. Proc Natl Acad Sci U S A
610   **101**:2530–5.

611   16. **Obritsch MD**, **Fish DN**, **MacLaren R**, **Jung R**. 2005. Nosocomial infections
612   due to multidrug-resistant Pseudomonas aeruginosa: epidemiology and
613   treatment options. Pharmacotherapy **25**:1353–1364.

614   17.   **Länger F**, **Kreipe HH**. 2011. Immunodeficiency and immunocompromised

615   patients. Pathologe **32**:411–417.

616   18.   **Gaynes R**, **Edwards JR**. 2005. Overview of nosocomial infections caused by

617   gram-negative bacilli. Clin Infect Dis **41**:848–54.

618   19.   **Visca P**, **Imperi F**, **Lamont IL**. 2007. Pyoverdine siderophores: from

619   biogenesis to biosignificance. Trends Microbiol **15**:22–30.

620   20.   **Lin J**, **Cheng J**, **Chen K**, **Guo C**, **Zhang W**, **Yang X**, **Ding W**, **Ma L**, **Wang Y**,

621   **Shen X**. 2015. The icmF3 locus is involved in multiple adaptation- and

622   virulence-related characteristics in Pseudomonas aeruginosa PAO1. Front Cell

623   Infect Microbiol **5**:1–15.

624   21.   **Lopez-Medina E**, **Fan D**, **Coughlin L a.**, **Ho EX**, **Lamont IL**, **Reimmann C**,

625   **Hooper L V.**, **Koh AY**. 2015. Candida albicans Inhibits Pseudomonas

626   aeruginosa Virulence through Suppression of Pyochelin and Pyoverdine

627   Biosynthesis. PLOS Pathog **11**:e1005129.

628   22.   **Ross-Gillespie A**, **Weigert M**, **Brown SP**, **Kümmerli R**. 2014. Gallium-

629   mediated siderophore quenching as an evolutionarily robust antibacterial

630   treatment. Evol Med Public Heal 18–29.

631   23.   **Imperi F**, **Massai F**, **Facchini M**, **Frangipani E**, **Visaggio D**, **Leoni L**. 2013.

632   Repurposing the antimycotic drug flucytosine for suppression of Pseudomonas

633   aeruginosa pathogenicity. Proc Natl Acad Sci **110**:7458–63.

634   24.   **Oliver KE**. 2011. Characterization of the Role of dadA in Pseudomonas

635   aeruginosa Virulence Factor Production. PhD Thesis.

636   25.   **Romanowski K**, **Zaborin A**, **Fernandez H**, **Poroyko V**, **Valuckaite V**, **Gerdes**

637   **S**, **Liu DC**, **Zaborina OY**, **Alverdy JC**. 2011. Prevention of siderophore-

638   mediated gut-derived sepsis due to P. aeruginosa can be achieved without iron

28

639 provision by maintaining local phosphate abundance: role of pH. BMC

640 Microbiol **11**:212.

641 26. **Carter MEK**, **Fothergill JL**, **Walshaw MJ**, **Rajakumar K**, **Kadioglu A**,

642 **Winstanley C**. 2010. A subtype of a Pseudomonas aeruginosa cystic fibrosis

643 epidemic strain exhibits enhanced virulence in a murine model of acute

644 respiratory infection. J Infect Dis **202**:935–942.

645 27. **Jimenez PN**, **Koch G**, **Papaioannou E**, **Wahjudi M**, **Krzeslak J**, **Coenye T**,

646 **Cool RH**, **Quax WJ**. 2010. Role of PvdQ in Pseudomonas aeruginosa

647 virulence under iron-limiting conditions. Microbiology **156**:49–59.

648 28. **Zaborin A**, **Romanowski K**, **Gerdes S**, **Holbrook C**, **Lepine F**, **Long J**,

649 **Poroyko V**, **Diggle SP**, **Wilke A**, **Righetti K**, **Morozova I**, **Babrowski T**, **Liu**

650 **DC**, **Zaborina O**, **Alverdy JC**. 2009. Red death in Caenorhabditis elegans

651 caused by Pseudomonas aeruginosa PAO1. Proc Natl Acad Sci U S A

652 **106**:6327–32.

653 29. **Papaioannou E**, **Wahjudi M**, **Nadal-Jimenez P**, **Koch G**, **Setroikromo R**,

654 **Quax WJ**. 2009. Quorum-quenching acylase reduces the virulence of

655 Pseudomonas aeruginosa in a Caenorhabditis elegans infection model.

656 Antimicrob Agents Chemother **53**:4891–4897.

657 30. **Attila C**, **Ueda A**, **Cirillo SLG**, **Cirillo JD**, **Chen W**, **Wood TK**. 2008.

658 Pseudomonas aeruginosa PAO1 virulence factors and poplar tree response in

659 the rhizosphere. Microb Biotechnol **1**:17–29.

660 31. **Harrison F**, **Browning LE**, **Vos M**, **Buckling A**. 2006. Cooperation and

661 virulence in acute Pseudomonas aeruginosa infections. BMC Biol **4**:21.

662 32. **Salunkhe P**, **Smart CHM**, **Morgan AW**, **Panagea S**, **Walshaw MJ**, **Hart A**,

663 **Geffers R**, **Tu B**, **Winstanley C**. 2005. A Cystic Fibrosis Epidemic Strain of

664   Pseudomonas aeruginosa Displays Enhanced Virulence and Antimicrobial

665   Resistance. J Bacteriol **187**:4908–4920.

666   33.   **Ochsner UA**, **Snyder A**, **Vasil AI**, **Vasil ML**. 2002. Effects of the twin-arginine

667   translocase on secretion of virulence factors, stress response, and

668   pathogenesis. Proc Natl Acad Sci U S A **99**:8312–8317.

669   34.   **Silo-Suh L**, **Suh S-J**, **Sokol P a**, **Ohman DE**. 2002. A simple alfalfa seedling

670   infection model for Pseudomonas aeruginosa strains associated with cystic

671   fibrosis shows AlgT (sigma-22) and RhlR contribute to pathogenesis. Proc Natl

672   Acad Sci U S A **99**:15699–15704.

673   35.   **Gallagher L a**, **Manoil C**. 2001. Pseudomonas aeruginosa PAO1 Kills

674   Caenorhabditis elegans by Cyanide Poisoning. J Bacteriol **183**:6207 – 6214.

675   36.   **Takase H**, **Nitanai H**, **Hoshino K**. 2000. Impact of Siderophore Production on

676   Pseudomonas aeruginosa Infections in Immunosuppressed Mice. Infect Immun

677   **68**:1834–9.

678   37.   **Xiong YQ**, **Vasil ML**, **Johnson Z**, **Ochsner U a**, **Bayer a S**. 2000. The

679   oxygen- and iron-dependent sigma factor pvdS of Pseudomonas aeruginosa is

680   an important virulence factor in experimental infective endocarditis. J Infect Dis

681   **181**:1020–6.

682   38.   **Meyer JM**, **Neely A**, **Stintzi A**, **Georges C**, **Holder I a**, **Meyer J**, **Neely A**,

683   **Stintzi A**, **Georges C**. 1996. Pyoverdin is essential for virulence of

684   Pseudomonas aeruginosa. Infect Immun **64**:518–523.

685   39.   **Okuda J**, **Okamoto M**, **Hayashi N**, **Sawada S**, **Minagawa S**, **Gotoh N**. 2012.

686   Complementation of the exoS gene in the pvdE pyoverdine synthesis gene-

687   deficient mutant of Pseudomonas aeruginosa results in recovery of the pvdE

688   gene-mediated penetration through the intestinal epithelial cell barrier but not

30

689   the pvdE-mediated virulence . J Infect Chemother **18**:332–340.

690   40.   **Kirienko N V.**, **Kirienko DR**, **Larkins-Ford J**, **Wählby C**, **Ruvkun G**, **Ausubel**

691   **FM**. 2013. Pseudomonas aeruginosa disrupts Caenorhabditis elegans iron

692   homeostasis, causing a hypoxic response and death. Cell Host Microbe

693   **13**:406–416.

694   41.   **Feinbaum RL**, **Urbach JM**, **Liberati NT**, **Djonovic S**, **Adonizio A**, **Carvunis**

695   **A-R**, **Ausubel FM**. 2012. Genome-wide identification of Pseudomonas

696   aeruginosa virulence-related genes using a Caenorhabditis elegans infection

697   model. PLoS Pathog **8**:e1002813.

698   42.   **Meyer JM**, **Stintzi A**, **De Vos D**, **Cornelis P**, **Tappe R**, **Taraz K**, **Budzikiewicz**

699   **H**. 1997. Use of siderophores to type pseudomonads: The three Pseudomonas

700   aeruginosa pyoverdine systems. Microbiology **143**:35–43.

701   43.   **Hauser A**, **Rello J**. 2003. Severe infections caused by Pseudomonas

702   aeruginosa. Springer, New York.

703   44.   **Hirsch EB**, **Tam VH**. 2011. Impact of multidrug-resistant Pseudomonas

704   aeruginosa infection on patient outcomes. Expert Rev Pharmacoecon

705   Outcomes Res **10**:1–18.

706   45.   **Kaneko Y**, **Thoendel M**, **Olakanmi O**, **Britigan BE**, **Singh PK**. 2007. The

707   transition metal gallium disrupts Pseudomonas aeruginosa iron metabolism

708   and has antimicrobial and antibiofilm activity. J Clin Invest **117**:877–888.

709   46.   **Wilderman PJ**, **Vasil AI**, **Johnson Z**, **Wilson MJ**, **Cunliffe HE**, **Lamont IL**,

710   **Vasil ML**. 2001. Characterization of an Endoprotease (PrpL) Encoded by a

711   PvdS-Regulated Gene in Pseudomonas aeruginosa. Infect Immun **69**:5385–

712   5394.

713   47.   **Ochsner UA**, **Johnson Z**, **Lamont IL**, **Cunliffe HE**, **Vasil ML**. 1996. Exotoxin

714   A production in Pseudomonas aeruginosa requires the iron-regulated pvdS

715   gene encoding an alternative sigma factor. Mol Microbiol **21**:1019–1028.

716 48. **Szumilas M**. 2010. Odds ratios. J Can Acad Child Adolesc Psychiatry **19**:227–

717   229.

718 49. **Dwan K**, **Altman DG**, **Arnaiz JA**, **Bloom J**, **Chan AW**, **Cronin E**, **Decullier E**,

719   **Easterbrook PJ**, **Von Elm E**, **Gamble C**, **Ghersi D**, **Ioannidis JPA**, **Simes J**,

720   **Williamson PR**. 2008. Systematic review of the empirical evidence of study

721   publication bias and outcome reporting bias. PLoS One **3**.

722 50. **Parrow NL**, **Fleming RE**, **Minnick MF**. 2013. Sequestration and scavenging of

723   iron in infection. Infect Immun **81**:3503–3514.

724 51. **Becker KW**, **Skaar EP**. 2014. Metal limitation and toxicity at the interface

725   between host and pathogen. FEMS Microbiol Rev.

726 52. **Banin E**, **Lozinski A**, **Brady KM**, **Berenshtein E**, **Butterfield PW**, **Moshe M**,

727   **Chevion M**, **Greenberg EP**, **Banin E**. 2008. The potential of desferrioxamine-

728   gallium as an anti-Pseudomonas therapeutic agent. Proc Natl Acad Sci U S A

729   **105**:16761–16766.

730 53. **Kirienko N V.**, **Ausubel FM**, **Ruvkun G**. 2015. Mitophagy confers resistance

731   to siderophore-mediated killing by Pseudomonas aeruginosa. Proc Natl Acad

732   Sci **112**:1821–1826.

733 54. **Wiehlmann L**, **Wagner G**, **Cramer N**, **Siebert B**, **Gudowius P**, **Morales G**,

734   **Köhler T**, **van Delden C**, **Weinel C**, **Slickers P**, **Tümmler B**. 2007. Population

735   structure of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A **104**:8101–6.

736 55. **Jiricny N**, **Molin S**, **Foster K**, **Diggle SP**, **Scanlan PD**, **Ghoul M**, **Johansen**

737   **HK**, **Santorelli LA**, **Popat R**, **West SA**, **Griffin AS**. 2014. Loss of social

738   behaviours in populations of Pseudomonas aeruginosa infecting lungs of

739   patients with cystic fibrosis. PLoS One **9**.

740 56. **Andersen SB**, **Marvig RL**, **Molin S**, **Krogh Johansen H**, **Griffin AS**. 2015.

741   Long-term social dynamics drive loss of function in pathogenic bacteria. Proc

742   Natl Acad Sci **112**:10756–10761.

743 57. **Kümmerli R**. 2015. Cheat invasion causes bacterial trait loss in lung infections:

744   Fig. 1. Proc Natl Acad Sci **112**:10577–10578.

745 58. **Marvig R**, **Damkiær S**, **Khademi S**. 2014. Within-Host Evolution of

746   Pseudomonas aeruginosa Reveals Adaptation toward Iron Acquisition from

747   Hemoglobin. MBio **5**.

748 59. **Bragonzi A**, **Paroni M**, **Nonis A**, **Cramer N**, **Montanari S**, **Rejman J**, **Di Serio**

749   **C**, **Döring G**, **Tümmler B**. 2009. Pseudomonas aeruginosa microevolution

750   during cystic fibrosis lung infection establishes clones with adapted virulence.

751   Am J Respir Crit Care Med **180**:138–45.

752 60. **Frydenlund Michelsen C**, **Hossein Khademi SM**, **Krogh Johansen H**,

753   **Ingmer H**, **Dorrestein PC**, **Jelsbak L**. 2015. Evolution of metabolic divergence

754   in Pseudomonas aeruginosa during long-term infection facilitates a proto-

755   cooperative interspecies interaction. ISME J 1–14.

756 61. **Klockgether J**, **Cramer N**, **Wiehlmann L**, **Davenport CF**, **Tümmler B**. 2011.

757   Pseudomonas aeruginosa genomic structure and diversity. Front Microbiol

758   **2**:1–18.

759 62. **Borenstein M**, **Hedges L V.**, **Higgins JPT**, **Rothstein HR**. 2010. A basic

760   introduction to fixed-effect and random-effects models for meta-analysis. Res

761   Synth Methods **1**:97–111.

762 63. **Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins HRR**. 2009.

763   Introduction to Meta-Analysis. John Wiley, West Sussex, UK.

764   64.   **Ryan RP**, **Lucey J**, **O'Donovan K**, **McCarthy Y**, **Yang L**, **Tolker-Nielsen T**,

765   **Dow JM**. 2009. HD-GYP domain proteins regulate biofilm formation and

766   virulence in Pseudomonas aeruginosa. Environ Microbiol **11**:1126–36.

767   65.   **Jeukens J**, **Boyle B**, **Kukavica-Ibrulj I**, **Ouellet MM**, **Aaron SD**, **Charette SJ**,

768   **Fothergill JL**, **Tucker NP**, **Winstanley C**, **Levesque RC**. 2014. Comparative

769   genomics of isolates of a Pseudomonas aeruginosa epidemic strain associated

770   with chronic lung infections of cystic fibrosis patients. PLoS One **9**:1–15.

771   66.   **Cox DR**. 1970. The continuity correction. Biometrika **57**:217–219.

772   67.   **Hohnadel D**, **Haas D**, **Meyer J-M**. 1986. Mapping of mutations affecting

773   pyoverdine production in Pseudomonas aeruginosa. FEMS Microbiol Lett

774   **36**:195–199.

775   68.   **Stintzi A**, **Johnson Z**, **Stonehouse M**, **Ochsner U**, **Meyer JM**, **Vasil ML**,

776   **Poole K**. 1999. The pvc gene cluster of Pseudomonas aeruginosa: Role in

777   synthesis of the pyoverdine chromophore and regulation by PtxR and PvdS. J

778   Bacteriol **181**:4118–4124.

779   69.   **Beare P a**, **For RJ**, **Martin LW**, **Lamont IL**. 2003. Siderophore-mediated cell

780   signalling in Pseudomonas aeruginosa: divergent pathways regulate virulence

781   factor production and siderophore receptor synthesis. Mol Microbiol **47**:195–

782   207.

783   70.   **Braud A**, **Hannauer M**, **Mislin GL a**, **Schalk IJ**. 2009. The Pseudomonas

784   aeruginosa pyochelin-iron uptake pathway and its metal specificity. J Bacteriol

785   **191**:3517–25.

786   71.   **R Development Core Team**. 2008. R: A language and environment for

787   statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

788   72.   **Schwarzer G**. 2015. meta: General Package for Meta-Analysis. R package

34

789 version 4.3-0.

790 73. **Viechtbauer W**. 2010. Conducting meta-analyses in R with the metafor

791 package. J Stat Softw **36**:1–48.

792 74. **Knapp G**, **Hartung J**. 2003. Improved tests for a random effects meta-

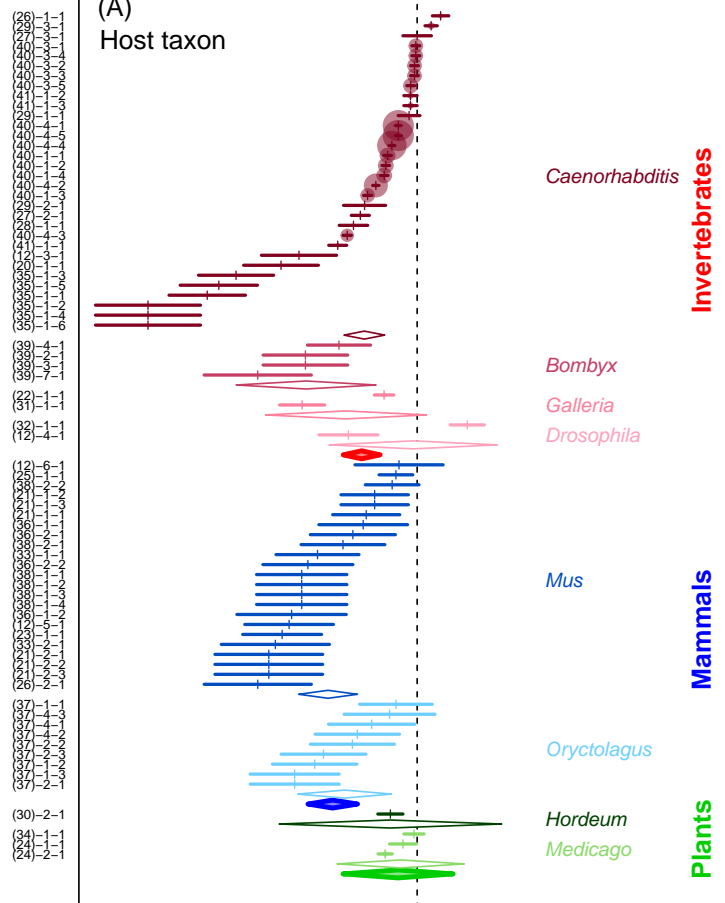793 regression with a single covariate. Stat Med **22**:2693–2710.

794 75. **Higgins JPT**, **Thompson SG**. 2002. Quantifying heterogeneity in a meta-

795 analysis. Stat Med **21**:1539–1558.

796 76. **Egger M**, **Davey Smith G**, **Schneider M**, **Minder C**. 1997. Bias in meta-

797 analysis detected by a simple, graphical test. BMJ **315**:629–34.

798 77. **Begg CB**, **Mazumdar M**. 1994. Operating characteristics of a rank correlation

799 test for publication bias. Biometrics **50**:1088–101.

800 78. **Lambrecht L**, **Fellous S, Koella J**. 2006. Coevolutionary interactions between

801 host and parasite genotypes. Trends in Parasitology **22**:12–16.

Study

(A) Host taxon

Caenorhabditis — **Invertebrates**

Bombyx
Galleria
Drosophila

Mus — **Mammals**

Oryctolagus

Hordeum — **Plants**
Medicago

Study

(B) Infection model

gut — **Invertebrates**

systemic

other

gut — **Mammals**

respiratory

skin

systemic

other — **Plants**

(C) Strain background

PA14

PAO1

PAO6049
FRD1

(D) Pleiotropy expected given mutation type

less

more

Log−odds−ratio comparing mortality from infections with pyoverdine−deficient mutant strain vs. infections with wildtype bacteria

relative virulence of mutant vs. wildtype: Less   Same   More   Less   Same   More

FIG. 1 Forest plots depicting the variation in effect size across experiments on pyoverdine as a virulence factor in *P. aeruginosa*. All panels display the same effect sizes originating from the 76 experiments involved in the meta-analysis, but grouped differently according to four moderator variables, which are: (A) host taxon; (B) infection type; (C) wildtype strain background; and (D) the likelihood of pleiotropy in the pyoverdine-deficient strain. Effect sizes are given as log-odds-ratio ± 95% confidence interval. Negative and positive effect sizes indicate lower and higher virulence of the pyoverdine-deficient mutant relative to the wildtype, respectively. Diamonds represent the mean effect sizes (obtained from meta-regression analysis) for each subgroup of a specific moderator variable. IDs of the individual experiments are listed on the Y-axis (for details, see Table S1 in the supplemental material). The numbers in brackets on the Y-axis correspond to the citation number of the corresponding publication.

FIG. 2 Test for differences between subgroups of moderator variables with regard to the effect sizes for pyoverdine as a virulence factor in *P. aeruginosa*. Our baseline condition for all comparisons is the following: gut infections in invertebrate hosts, using the *P. aeruginosa* wildtype strain PA14 vs a pyoverdine-deficient PA14 mutant with a low expected level of pleiotropy. The effect size for this baseline scenario is set to zero. All other scenarios had more extreme (negative) effect sizes, and are therefore scaled relative to this baseline condition. Comparisons reveal that virulence in pyoverdine-deficient strains was significantly more reduced in systemic compared to gut infections, and that most effect size variation is explained by the infection type. There were no significant effect size differences between any of the other subgroups. Bars show the difference in log odds-ratio (± 95% confidence interval) between the baseline and any of the alternate conditions. Values given in brackets indicate percentage of effect size heterogeneity explained by a specific moderator.
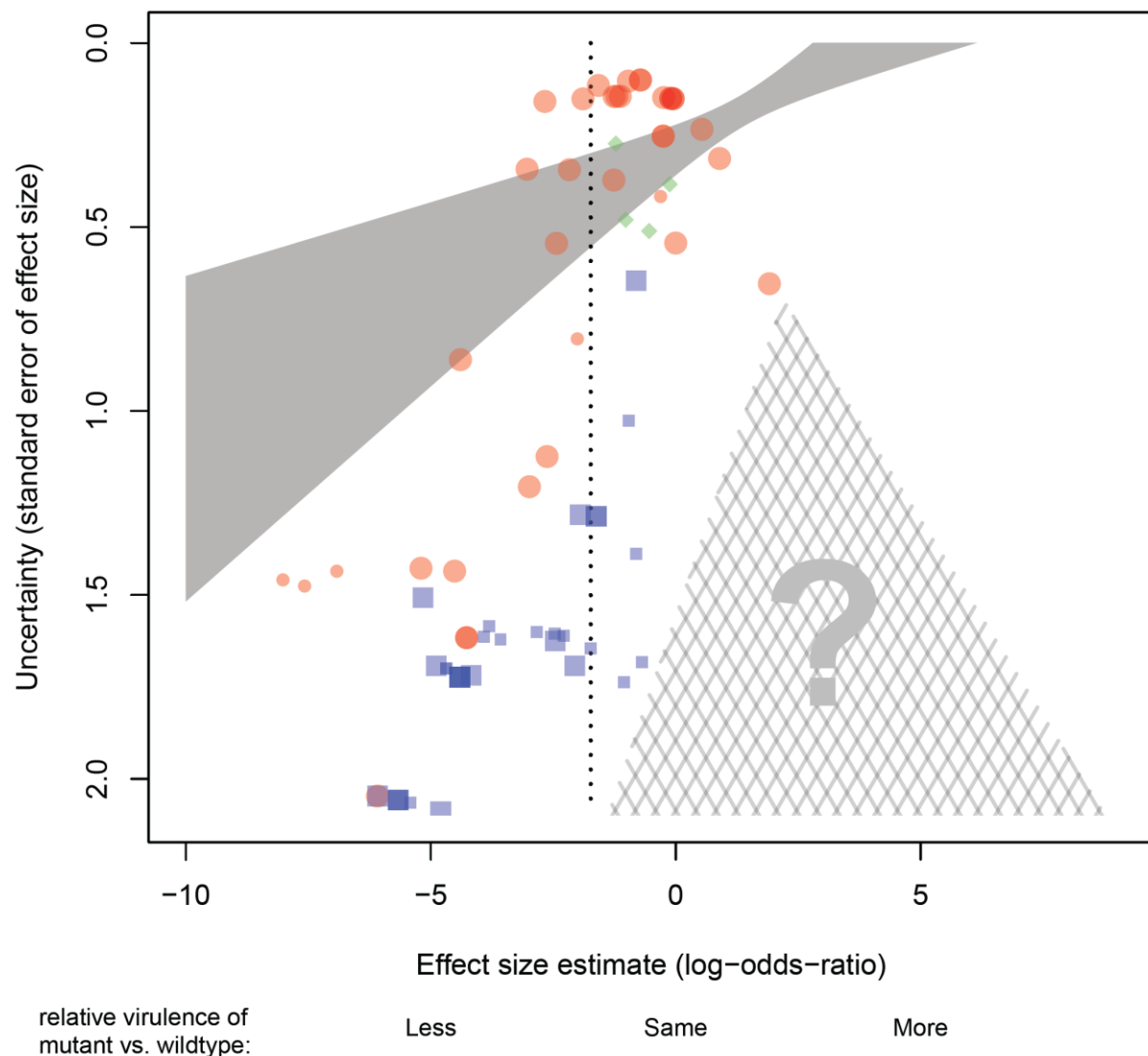
FIG. 3 Association between effect sizes and their standard errors across 76 experiments examining the role of pyoverdine production for virulence in _P. aeruginosa._ In the absence of bias, we should see an inverted funnel-shaped cloud of points, more or less symmetrically distributed around the mean effect size (vertical dotted line). Instead, we see an over-representation of low-certainty experiments associated with strong (negative) effect sizes. This suggests a significant publication bias: experiments with low-certainty and weak or contrary effects presumably do exist, but are under-represented here (note the absence of data points in the cross-hatched triangle). Effect sizes are given as log-odds-ratio. Each symbol represents a single experiment. Symbol colours and shapes stand for different host organisms (red circles = invertebrates; blue squares = mammals; green diamonds = plants). Large symbols denote the experiments included in the core dataset. The solid shaded area represents the 95% confidence interval for the weighted linear regression using the complete dataset. Note that due to the stronger weights accorded to high certainty experiments (i.e. the points towards the top of the plot), many of the lower-weighted (higher-uncertainty) points towards the bottom of the plot lie quite far from the regression line and also outside the confidence interval.

TABLE 1 Meta-analysis workflow for this study

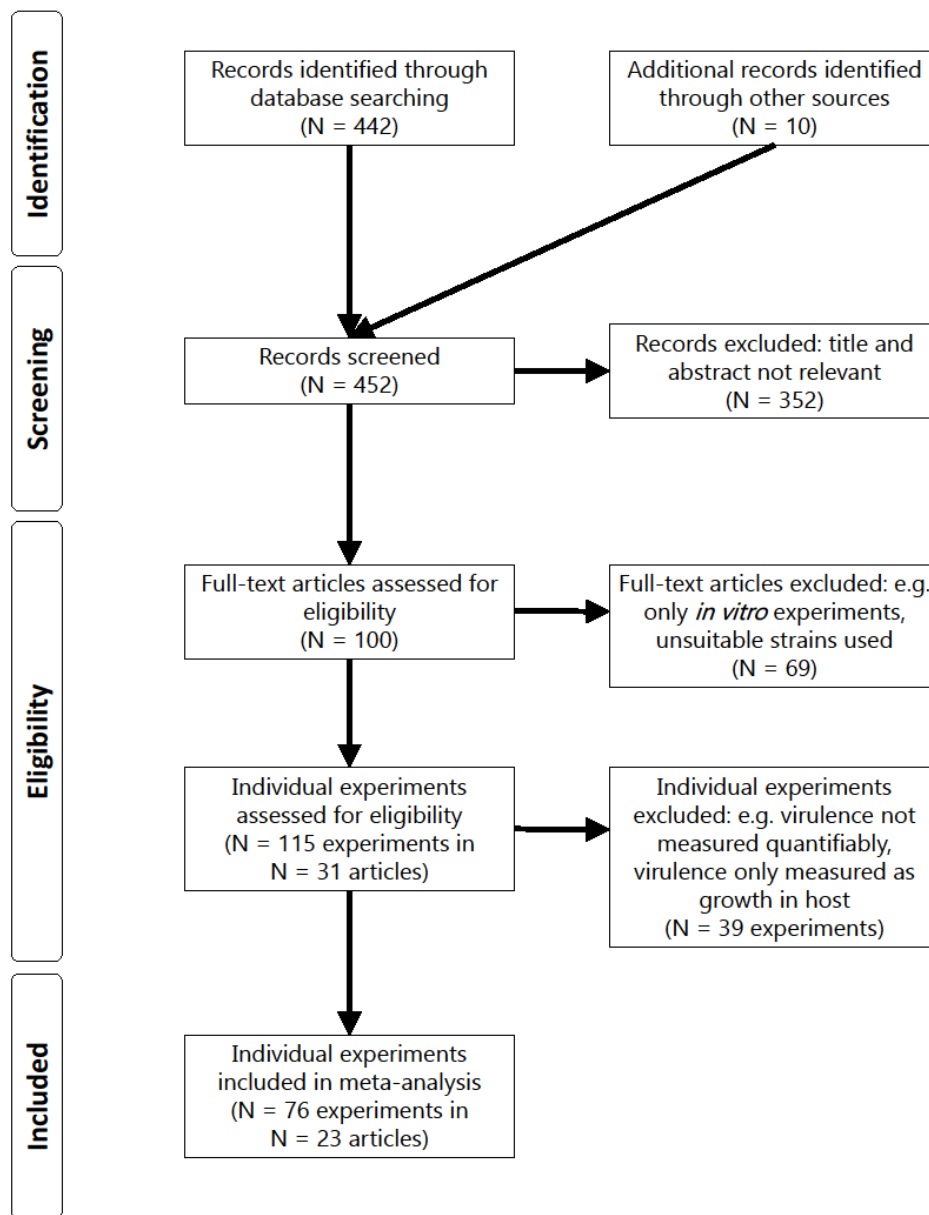| In general	This study	Details
1. Formulate hypothesis and predictions	HYPOTHESIS: Pyoverdine is an important virulence factor for *Pseudomonas aeruginosa* PREDICTION: Pyoverdine-defective mutants cause less virulence than wildtype strains	
2. Systematically search for relevant studies	We searched (see details in main text) for any reports of experiments featuring monoclonal infections of whole live host organisms with *P. aeruginosa* strains known to vary in pyoverdine phenotype, where virulence was quantified in terms of host mortality.	Tables S1+S2
3. Extract and standardize effect sizes and their standard errors	For each case reporting host survival, we calculated the (log) ratio of mortality odds from pyoverdine-mutant infections vs. wildtype infections – i.e. the (log) odds-ratio.	
4a. Check heterogeneity across studies	Our assembled effect sizes were more heterogeneous than expected from chance – even when we allowed that some of this variation could be due to random noise.	Figure 1
4b. Consider putative moderator variables (optional)	We tested for evidence of distinct sub-groups in our dataset, within which the effect sizes might be more homogeneous. We identified four putative moderators and codified each study for the following: (i) host taxon; (ii) infection type; (iii) strain background; and (iv) level of pleiotropy expected, given the particular mutation(s) involved.	
4c. Check for publication bias (optional)	We found that smaller / lower-powered studies were more likely to report large effect sizes in support of the hypothesis, whereas larger / higher-powered studies tended to report smaller effect sizes.	Figure 2 Figure 3
5. Derive mean effect size(s); quantify influence of moderator variables (if applicable)	Despite the steps taken (see above), our dataset still showed substantial heterogeneity. Estimates of mean effect sizes (in / across subgroups) and moderator coefficients should therefore be viewed as best approximations.	

FIG. S1 Flow diagram (PRISMA format) of the screening and selection process for studies investigating the association between pyoverdine production and virulence in *P. aeruginosa*.
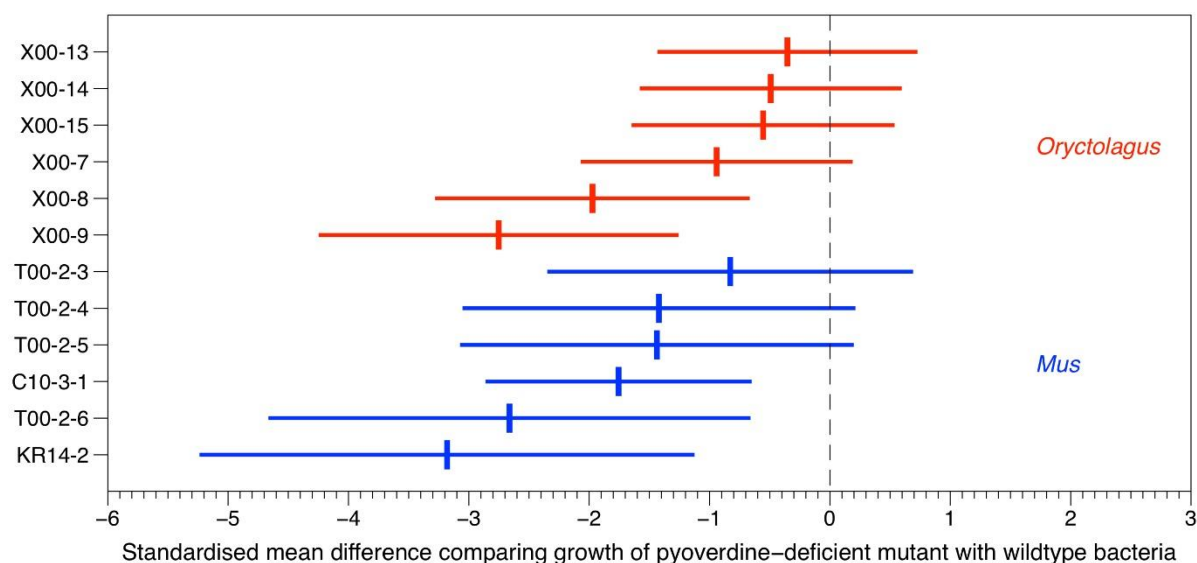
FIG. S2 Forest plot depicting the variation in effect size across experiments on the effect of pyoverdine on the growth of *P. aeruginosa* in mammalian hosts. Effect sizes are given as standardized mean difference ± 95% confidence interval and are grouped by host genus. Negative and positive effect sizes indicate lower and higher *in vivo* growth of the pyoverdine-deficient mutant relative to the wildtype, respectively. IDs of the individual experiments are listed on the Y-axis (for details, see Table S4 in the supplemental material).

SUPPLEMENTARY FILE LEGEND

TABLES S1-S4 Full dataset collected for meta-analysis, including calculated effect sizes and a list of excluded experiments.