

A Systems-Level Approach for Investigating *Pseudomonas aeruginosa* Biofilm Formation

Zhaobin Xu¹, Xin Fang², Thomas K. Wood³, Zuyi Jacky Huang^{1,4,5*}

1 Department of Chemical Engineering, Villanova University, Villanova, Pennsylvania, United States of America, **2** The Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, Maryland, United States of America, **3** Departments of Chemical Engineering and Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, United States of America, **4** The Center for Nonlinear Dynamics & Control, Villanova University, Villanova, Pennsylvania, United States of America, **5** Villanova Center for the Advancement of Sustainability in Engineering (VCASE), Villanova University, Villanova, Pennsylvania, United States of America

Abstract

Prevention of the initiation of biofilm formation is the most important step for combating biofilm-associated pathogens, as the ability of pathogens to resist antibiotics is enhanced 10 to 1000 times once biofilms are formed. Genes essential to bacterial growth in the planktonic state are potential targets to treat biofilm-associated pathogens. However, the biofilm formation capability of strains with mutations in these essential genes must be evaluated, since the pathogen might form a biofilm before it is eliminated. In order to address this issue, this work proposes a systems-level approach to quantifying the biofilm formation capability of mutants to determine target genes that are essential for bacterial metabolism in the planktonic state but do not induce biofilm formation in their mutants. The changes of fluxes through the reactions associated with the genes positively related to biofilm formation are used as soft sensors in the flux balance analysis to quantify the trend of biofilm formation upon the mutation of an essential gene. The essential genes whose mutants are predicted not to induce biofilm formation are regarded as gene targets. The proposed approach was applied to identify target genes to treat *Pseudomonas aeruginosa* infections. It is interesting to find that most essential gene mutants exhibit high potential to induce the biofilm formation while most non-essential gene mutants do not. Critically, we identified four essential genes, *lysC*, *cysH*, *adk*, and *galU*, that constitute gene targets to treat *P. aeruginosa*. They have been suggested by existing experimental data as potential drug targets for their crucial role in the survival or virulence of *P. aeruginosa*. It is also interesting to find that *P. aeruginosa* tends to survive the essential-gene mutation treatment by mainly enhancing fluxes through 8 metabolic reactions that regulate acetate metabolism, arginine metabolism, and glutamate metabolism.

Citation: Xu Z, Fang X, Wood TK, Huang ZJ (2013) A Systems-Level Approach for Investigating *Pseudomonas aeruginosa* Biofilm Formation. PLoS ONE 8(2): e57050. doi:10.1371/journal.pone.0057050

Editor: Tarek Msadek, Institut Pasteur, France

Received: July 16, 2012; **Accepted:** January 16, 2013; **Published:** February 22, 2013

Copyright: © 2013 Xu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Z.H. was supported by Villanova University (grant SRF-RSG 217030-7252; VCASE seed grant 420492). T.W. was supported by the National Institutes of Health (R01 GM089999). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zuyi.huang@villanova.edu

Introduction

Biofilms have been frequently associated with human diseases such as osteomyelitis [1], chronic wound infections [2,3] and cystic fibrosis [4], as they facilitate the survival of pathogens in hostile environments. It is reported that nearly 65% of all nosocomial infections in the USA are associated with biofilms [5]. When exposed to stress, such as that imposed by antibiotic treatments or limited nutrients, pathogens adhere to each other to form biofilms for the purpose of survival [6]. The development of biofilms generally comprises the following four stages, i.e., the initial attachment of planktonic pathogens to a surface, the accumulation of biofilms through the production of extracellular polysaccharide substance (EPS) that interconnects and transiently immobilizes biofilm cells, the maturation of biofilm architecture, and the dispersal of single cells from the biofilm [7]. The first few stages play a key role in treating the biofilm-associated pathogens, as the ability of pathogens to resist antibiotics is significantly enhanced once they form biofilms [5,7]. Therefore, significant effort in the biofilm research community has been devoted to the investigation of bacterial metabolism and signaling which are involved in the transition from planktonic growth to biofilm growth [8,9].

Elucidating the mechanisms of biofilm formation is far from trivial: hundreds of highly interacted molecules such as metabolites, metabolic enzymes, and signaling proteins are involved in regulating this process. Most current research is focused on the experimental investigation of the impact of individual molecules, such as regulators, on biofilm formation [10,11]. This is insufficient for characterizing the biofilm formation process, as a systems-level characterization of interactions between molecules involved in biofilm formation is required to fully understand biofilm formation mechanisms and thus manipulate the metabolism of microorganisms in biofilms. Genome-scale metabolic modeling has been commonly used for systemically studying microorganism metabolism, as evidenced by its wide application in identifying genes that are essential for the growth of *Escherichia coli* [12], *Staphylococcus aureus* [13], *Mycobacterium tuberculosis* [14], *Acinetobacter baumannii* [15], and *Pseudomonas aeruginosa* [16]. A recent study by Thiele and her coworkers [17] shows the first systems biology approach to identifying candidate drug targets for treating *P. aeruginosa* in biofilms. This approach mainly applies single/double gene inhibition simulations to determine the growth of *P. aeruginosa* in specific microenvironments that imitate microbial communities associated with biofilm formation. How-

ever, certain issues have not been addressed in this approach, including the quantification of the trend for mutants to form biofilms, and the identification of metabolic reactions that facilitate biofilm formation in mutants. This forms the motivation of this work, that is, to consider the trend of biofilm formation for mutants in the identification of drug-target genes.

Here, we propose an approach to identifying drug targets against a biofilm-forming pathogen, by identifying genes that satisfy two requirements: 1) these genes are crucial for the growth of the pathogen in the planktonic state, that is, the mutation of any of these genes can eliminate the planktonic pathogen; and 2) the inhibition of the function of these genes does not induce biofilm formation.

In particular, we first use the essential planktonic-growth genes presented in Oberhardt et al., 2008 [16] as the initial set of genes satisfying the first requirement, and further, from the initial set, identify genes also satisfying the second requirement. We perform the search by pinpointing a set of biofilm associated reactions from the genes that are reported to be positively related to *P. aeruginosa* biofilm formation in Msken et al., 2010 [18], and using the flux changes through these reactions as soft-sensors to quantify the impact of the mutation of each gene from the initial set on biofilm formation. The rationale for the use of biofilm-associated reactions is that a large enhancement of fluxes (i.e., activity levels) through these reactions reflects a large potential for a mutant to form biofilms. Specifically, to obtain the flux changes through biofilm-associated reactions, we simulate the metabolism of a gene mutant via flux balance analysis (FBA) [19], and sample the biofilm-associated reaction fluxes via the artificially-center-hit-and-run method (ACHR), an efficient sampling approach for a linearly constrained space [20]. In addition, metabolic reactions whose fluxes significantly increase in most mutants can be determined and regarded as reactions that facilitate biofilm formation. *P. aeruginosa* is chosen as the reference pathogen in this work, because *P. aeruginosa* is one of the leading causes of nosocomial infections in hospitalized patients and *P. aeruginosa* is resistant to a wide array of antibiotics by forming biofilms during chronic infections [21,22].

Results

An illustrative example

The key innovation of the proposed approach is to take the biofilm formation of a single mutant into account when identifying gene targets to treat biofilm-associated pathogens. While the detail of the proposed approach is given in the Material and Methods section, an example is shown in Figure 1 to illustrate the steps for the identification of drug targets that impair the growth of *P. aeruginosa* without inducing biofilm formation. The genes positively associated with biofilm formation (obtained from Msken et al., 2010 [18] for *P. aeruginosa* in LB medium) are first overlaid with the *P. aeruginosa* metabolic network to identify reactions associated with biofilm formation (Figure 1, Step 1). The metabolic model presented in Oberhardt et al., 2008 [16], which contains 1056 genes and 883 metabolic reactions, is used in this work, as it is the most comprehensive metabolic model for *P. aeruginosa* and it has been used to correctly predict catabolism of various substrates such as amino acids and common sugars via flux balance analysis. Further, we quantitatively evaluated the impact of the mutation of each essential planktonic-growth gene (obtained from Oberhardt et al., 2008 [16] for *P. aeruginosa* in LB medium, including the PA1756 gene) on biofilm associated reactions, by calculating relative activity changes of these reactions via Steps 2 and 3. The obtained relative activity change profile quantitatively suggests the capability of a mutant to form biofilms, that is, large values for the

changes imply a potential induction of the mutant's biofilm formation. In Step 4, the similarity in the shape and magnitude of relative activity profiles is used to categorize the essential planktonic-growth genes [16] into different clusters. The clusters of essential planktonic-growth genes whose mutants have low potential to induce biofilm formation are identified as drug-target genes (such as PA1756), because the mutation of these genes can eliminate planktonic pathogens. Metabolic reactions whose activity levels significantly increase in most single mutants are identified in Step 5. These reactions illustrate how *P. aeruginosa* adjusts its metabolism to form biofilms upon the mutation of genes essential for planktonic growth

Identification of genes and reactions positively associated with *P. aeruginosa* biofilm formation

A set of biofilm-associated metabolic reactions are obtained via Step 1 in Figure 1. The experimental data of genetic determinants of *P. aeruginosa* biofilm have been presented in Msken et al., 2010 [18]. 394 genes are reported to be positively associated with biofilm formation, and 64 of them are involved in the metabolic network. None of these 64 genes are essential for bacterial growth in the planktonic state. Flux balance analysis showed that inhibition of 37 of these 64 genes doesn't affect bacterial growth rate. These genes are thus not directly involved in the bacterial biomass synthesis. Fluxes of reactions associated with these 37 genes are the ideal virtual sensors that monitor the trend of *P. aeruginosa* to form biofilms, as these genes are only associated with biofilm formation. The 39 reactions associated with these 37 genes are considered as biofilm-associated reactions (see Table 1).

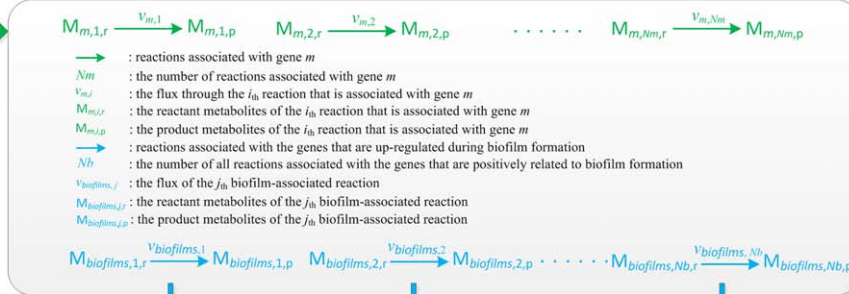
It can be seen from Table 1 that the biofilm-associated reactions are mainly involved in the consumption of nitrite, the regulation of acetate, tricarboxylic acid (TCA) cycle, the generation of carbon dioxide and ammonia, iron metabolism, the regulation of hydrogen peroxide, arginine metabolism, the production of glutamate, pyrimidine metabolism, and oxidative phosphorylation. While these reactions were identified by overlaying the 37 genes that are reported to be associated with biofilm formation in Msken et al., 2010 [18] onto the metabolic network of *P. aeruginosa*, their important role in biofilm formation is further confirmed by the following independent evidence:

- It has been shown that accumulation of nitrite is required to inhibit biofilm formation in staphylococcal strains [23]. The consumption of nitrite thus facilitates biofilm formation.
- Acetate is found to be engaged in the switch from a physiology program that permits a rapid growth in the presence of abundant nutrients to a program related to biofilm formation that can enhance the survival of *E. coli* in the absence of nutrients [24].
- Certain genes in TCA cycle are up-regulated in the staphylococcal biofilm [25].
- The concentration of carbon dioxide can be used as an indicator of biofilm formation [26], and carbon dioxide and ammonia influence the pH value in the cytosol, which in turn affect biofilm formation [18].
- The important role of iron in the *E. coli* biofilm formation has been reported by several previous studies [27,28,29].
- Hydrogen peroxide has been recognized to trigger the initiation of biofilm in *Streptococcus gordonii* via the release of the extracellular DNA [30].
- The arginine fermentation regulated by the arginine deaminase is the major metabolic process for the generation of ATP in *P. aeruginosa* in the anaerobic condition [31].

Essential planktonic-growth gene m
(e.g., PA1756) reported by Oberhardt
et al., 2008

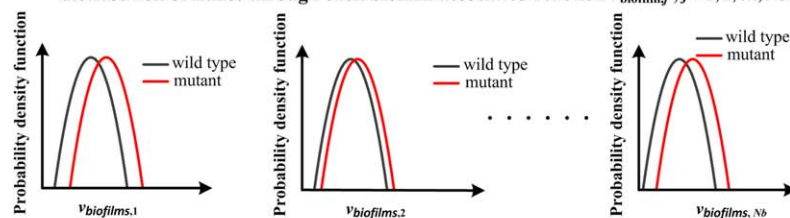
Genes positively associated with *P.aeruginosa* biofilm formation
reported by Mücken et al., 2010

Genome scale metabolic network of *P. aeruginosa*

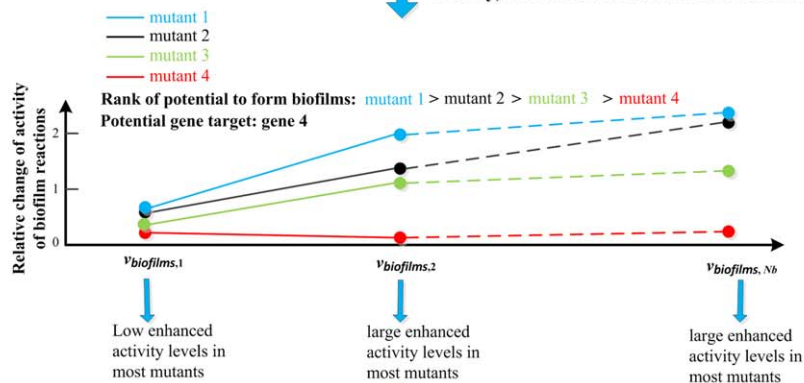


Step 1: Determine reactions positively associated with biofilm formation

Step 2: Upon the mutation of each essential gene (e.g. gene PA1756), quantify the distribution of fluxes through each biofilm-associated reaction $v_{biofilms,j}$, $j = 1, 2, \dots, Nb$



Step 3: Quantify biofilm formation capability of mutants (e.g., mutant 4 shown below is for gene PA1756) by determining the change of flux (e.g. activity) across all biofilm-associated reactions



Step 4: Cluster essential genes according the biofilm formation capability of their mutants

Step 5: Identify reactions facilitating biofilm formation

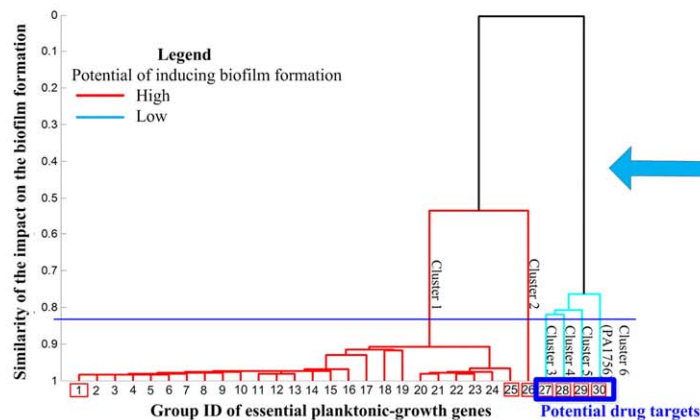


Figure 1. Schematic description of quantifying the biofilm formation capability of single mutants of essential genes for *P. aeruginosa*. The genes positively associated with *P. aeruginosa* biofilm formation are used to determine biofilm-associated reactions in Step 1. The essential gene *m* (e.g., PA1756 gene) is used as a reference gene in Steps 2 through 3 to illustrate the proposed approach for quantifying the potential of a single mutant to form biofilms. Specifically, reactions associated with gene *m* are partially mutated, and the flux distributions of all biofilm-associated reactions for both wide-type and mutant strains are quantified in Step 2. The relative change of activity through all biofilm-associated reactions for the gene *m* mutant is quantified in Step 3. A large enhancement in the activity of biofilm-associated reactions implies a large potential for the single mutant to form biofilms. The profiles for four mutants are given as examples in Step 3. Mutant 4 exhibits the lowest potential to form biofilms. The relative activity change profiles for all single mutants are used to categorize essential genes into different clusters in Step 4. Mutant 4 is assigned to a different cluster from that for the other three mutants, as its relative activity profile is not similar to those for the other mutants in both the shape and the magnitude. The essential planktonic-growth genes whose mutants might not induce biofilm formations are identified from the cluster results and regarded as potential target genes. For example, gene PA1756, which corresponds to mutant 4 in Step 3, is one potential gene target due to the low enhanced activities through those biofilm-associated reactions in its mutant. The biofilm-associated reactions whose activity levels are apparently enhanced in most mutants are identified in Step 5. These reactions indicate the underlying mechanisms for *P. aeruginosa* biofilm formation.

doi:10.1371/journal.pone.0057050.g001

- Glutamate synthesis is reported to be positively correlated with biofilm formation of *Mycobacterium bovis* [32].
- Pyrimidine metabolism is reported to participate in the modulation of the *E. coli* biofilm formation [33]. It has been experimentally verified that pyrimidine uracil influences quorum sensing and biofilm formation in *P. aeruginosa* [34].
- Oxidative phosphorylation is coupled with the ATP production for pathogens in biofilms. It has been demonstrated that enzymes that participate in the oxidative phosphorylation are up-regulated in *Bordetella pertussis* biofilms [35].

Investigation of biofilm formation capability for single-mutants

Given the above biofilm-associated reactions, we obtained a relative activity change profile of these reactions for each single mutant of the 136 essential planktonic-growth genes obtained from Oberhardt et al., 2008 [16] (via Steps 2 and 3), and, subsequently, based upon the similarity in relative activity change profiles for different single mutants, went through Step 4 to hierarchically categorize the essential planktonic-growth genes into six different clusters (as shown in Figure 2). Clusters 1 and 2 are located in one branch, while the other four clusters are associated with the other branch. The hierarchical clustering algorithm allows us to further separate genes in each cluster into different groups according to the similarity in their relative activity change profiles. The genes associated with each group shown in Figure 2 are listed in Table 2.

The profiles of relative activity change for the mutants of the representative groups for different clusters are shown in Figure 3. Figure 3 (A) and (B) show the profiles for the mutants of the PA0945 (from Group 1, Cluster 1) and PA5038 genes (from Group 25, Cluster 1), respectively. The profiles of these two mutants are used to represent the whole spectrum of Cluster 1, because, although they are from the two groups (Groups 1 and 25) with the lowest similarity in the cluster, their profiles of relative activity change in Figure 3 (A) and (B) are similar, suggesting that the single mutations of genes in Cluster 1 have similar impacts on biofilm-associated reactions and thus potentially on biofilm formation. Figure 3 (C) through (G) show the relative activity change profiles of representative groups from Clusters 2 through 6, respectively. Since there is only one group (actually one gene per group) associated with each of these clusters, these groups are regarded as the representatives of the corresponding clusters (as shown in Figure 2).

Single mutants of genes from different clusters have different potentials to induce biofilm formation. As shown in Figure 3 (A) and (B), the activity levels of certain biofilm-associated reactions have significant increases upon the mutation of a gene from Cluster 1, suggesting that the gene mutants from Cluster 1 might

form biofilms and in turn prevent the elimination of *P. aeruginosa*. The comparison of Figure 3 (C) to Figure 3 (A) and (B) shows that mutants from Cluster 2 might still form biofilms but to a lesser extent. In particular, the mutant from Cluster 2 (PA4031) and those from Cluster 1 (PA0945 and PA5038) have similar relative activity changes in the biofilm-associated reactions Rxn#1 through Rxn#32, but the mutant from Cluster 2 has much smaller changes in Rxn#33 through Rxn#39 than those from Cluster 1. Interestingly, as shown in Table 2, there are totally 132 essential genes associated with Cluster 1 and Cluster 2, while only the remaining four are from Clusters 3 through 6, suggesting that the mutation of most essential planktonic-growth genes might induce the formation of biofilms. Therefore, not every essential planktonic-growth gene is a good target to treat biofilm-associated pathogens.

The mutation of any gene from Clusters 3 through 6 might not induce biofilm formation, as approximately 96% of the biofilm-associated reactions for these mutants are of small relative activity changes (e.g. most less than 0.5). In addition, the relative activity changes of certain biofilm-associated reactions are reversed and have negative values. Even though the amplitudes of these negative relative fold changes are minor (e.g., most less than 0.5), the metabolic activities are not distributed in the direction for promoting biofilm formation for the mutants from Clusters 3 through 6. Therefore, the four genes in these clusters, i.e., PA0904 (*bysC*), PA1756 (*cysH*), PA3686 (*adk*), and PA2023 (*galU*), are regarded potential gene targets to treat *P. aeruginosa*, because they are essential to planktonic *P. aeruginosa* and their mutants might not induce biofilm formation.

Validation of the predicted results with existing experimental data

Experimental data were collected from the literature to validate the aforementioned prediction results, which shows that the mutants from Clusters 1 and 2 might induce biofilm formation while the mutants from Clusters 3 through 6 might not. The essential genes contained in Clusters 1 and 2 are mainly involved in vitamin and cofactor synthesis, amino acid catabolism, cell wall synthesis, central metabolism, the membrane transport system, nucleotide synthesis, nucleotide salvage, and lipid synthesis. The involvement of these genes in the aforementioned biological systems is listed in the Table S1. The mutants of some genes in Cluster 1 have been reported to form biofilms. For example, the inhibitor of the enzyme encoded by the *glmU* gene (Group 4) is able to enhance the formation of biofilms of *P. aeruginosa* PAO1 [36]. Another example is given that the *pgsA* null mutant (Group 4) of *E. coli* activates the Rcs signal transduction pathway that is crucial for *E. coli* biofilm formation [37]. In addition, the down-regulation of the following six essential genes, i.e., *pyrH* (Group 6),

Table 1. Biofilm-associated genes and reactions that are identified via the overlay of the genes reported by Msken, et al., 2010 [18] to be positively associated with biofilm formation onto the metabolic network presented by Oberhardt, et al., 2008 [16].

Reactions	Genes	Biological subsystems
Rxn#1: dGTP + H ₂ O → Deoxyguanosine + Inorganic triphosphate	PA1124 (<i>dgt</i>)	pyrimidine metabolism
Rxn#2: GTP + H ₂ O → Guanosine + Inorganic triphosphate	PA1124 (<i>dgt</i>)	pyrimidine metabolism
Rxn#3: H ⁺ + Malonate → Acetate + CO ₂	PA0208 (<i>mdcA</i>)	carbon dioxide production
Rxn#4: Coenzyme A + 3-Oxoadipyl-CoA → Acetyl-CoA + Succinyl-CoA	PA0228 (<i>pcaF</i>)	acetate metabolism
Rxn#5: 4-Aminobutanoate + 2-Oxoglutarate → L-Glutamate + Succinic semialdehyde	PA0266 (<i>gabT</i>)	glutamate production
Rxn#6: 2 S-Adenosyl-L-methionine + Uroporphyrinogen III → 2 S-Adenosyl-L-homocysteine + H ⁺ + Dihydrosirohydrochlorin	PA0510 (<i>nirE</i>)	coenzyme B12 synthesis
Rxn#7: Ferrocyclochrome c + 2 H ⁺ + Nitrite → Ferricycyclochrome c + H ₂ O + Nitric oxide	PA0511 (<i>nirJ</i>)	nitrite consumption
Rxn#8: Fumarate + H ₂ O ↔ L-Malate	PA0854 (<i>fumC2</i>)	TCA cycle
Rxn#9: cobalt ₂ [e] → cobalt ₂ [c]	PA0913 (<i>mgtE</i>)	cobalt transport
Rxn#10: mg ₂ [e] → mg ₂ [c]	PA0913 (<i>mgtE</i>)	magnesium transport
Rxn#11: 1-Aminopropan-2-ol + Adenosyl-cobyrinic acid ↔ Adenosyl cobinamide + H ₂ O	PA1275 (<i>cobD</i>)	coenzyme B12 synthesis
Rxn#12: L-Threonine → 2-Oxobutanoate + Ammonium	PA1326 (<i>ilvA2</i>)	production of ammonia
Rxn#13: 4-Maleylacetoacetate → 4-Fumarylacetoacetate	PA2007 (<i>maiA</i>)	acetate metabolism
Rxn#14: Benzoate + 2 H ⁺ + Nicotinamide adenine dinucleotide - reduced + O ₂ → Cis-1, 2-dihydroxycyclohexa-3,5-diene-1-carboxylate + Nicotinamide adenine dinucleotide	PA2518 (<i>xylX</i>)	benzoate degradation
Rxn#15: Isocitrate + Nicotinamide adenine dinucleotide phosphate ↔ 2-Oxoglutarate + CO ₂ + Nicotinamide adenine dinucleotide phosphate - reduced	PA2623 (<i>icd</i>)	TCA and carbon dioxide production
Rxn#16: 3 H ⁺ + Nicotinamide adenine dinucleotide - reduced + Ubiquinone-8 → Nicotinamide adenine dinucleotide + Ubiquinol-8 + 2 H ⁺	PA2642 (<i>nuoG</i>)	oxidative phosphorylation
Rxn#17: ATP + H ₂ O + Phosphonate[e] → ADP + H ⁺ + Phosphate + Phosphonate[c]	PA3383 (<i>phnD</i>)	phosphate transport
Rxn#18: S-Adenosyl-L-methionine + Butyryl-[acyl-carrier protein] → 5-Methylthioadenosine + Acyl carrier protein + H ⁺ + N-butyryl-L-homoserine lactone	PA3476 (<i>rhII</i>)	homoserine synthesis
Rxn#19: 2 Ferrocyclochrome c + 4 H ⁺ + 0.5 O ₂ → 2 Ferricycyclochrome c + H ₂ O + 2 H ⁺	PA4133 (<i>ccoN</i>)	iron metabolism
Rxn#20: ATP + H ₂ O + Fe-enterobactin → ADP + Fe-enterobactin + H ⁺ + Phosphate	PA4160 (<i>fepD</i>)	iron metabolism
Rxn#21: Chorismate → Isochorismate	PA4231 (<i>pchA</i>)	oxidative phosphorylation
Rxn#22: 2 Hydrogen peroxide → 2 H ₂ O + O ₂	PA4236 (<i>katA</i>)	hydrogen peroxide consumption
Rxn#23: Alpha-D-Ribose 5-phosphate + Uracil ↔ H ₂ O + Pseudouridine 5'-phosphate	PA4544 (<i>rluD</i>)	pyrimidine metabolism
Rxn#24: 1.5 O ₂ + Protoporphyrinogen IX → 3 H ₂ O + Protoporphyrin	PA4664 (<i>hemK</i>)	coenzyme B12 synthesis
Rxn#25: Alpha-Oxo-benzeneacetic acid ↔ Benzaldehyde + CO ₂	PA4901 (<i>mdlC</i>)	carbon dioxide production
Rxn#26: Reduced glutathione + Methylglyoxal → (R)-S-Lactoylglutathione	PA5111 (<i>gloA3</i>)	glutathione metabolism
Rxn#27: N(omega)-(L-Arginino)succinate ↔ L-Arginine + Fumarate	PA5263 (<i>argH</i>)	arginine metabolism
Rxn#28: ATP + H ₂ O + Phosphate[e] → ADP + H ⁺ + 2 Phosphate[c]	PA5368 (<i>pstC</i>)	phosphate transport
Rxn#29: Citrate[e] + H ⁺ [e] ↔ Citrate[c] + H ⁺ [c]	PA5476 (<i>citA</i>)	TCA cycle
Rxn#30: 5,6-dihydrouracil + H ₂ O ↔ N-Carbamoyl-beta-alanine + H ⁺	PA0441 (<i>dht</i>)	pyrimidine metabolism
Rxn#31: Nicotinamide adenine dinucleotide + O-Phospho-4-hydroxy-L-threonine → 2-Amino-3-oxo-4-phosphonooxybutyrate + H ⁺ + Nicotinamide adenine dinucleotide - reduced	PA0593 (<i>pdxA</i>)	pyridoxine metabolism
Rxn#32: 2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate + 4-Methyl-5-(2-phosphoethyl)-thiazole + H ⁺ → Diphosphate + Thiamin monophosphate	PA3976 (<i>thiE</i>)	thiamin metabolism
Rxn#33: ATP + Coenzyme A + Succinate ↔ ADP + Phosphate + Succinyl-CoA	PA1588 (<i>sucC</i>)	TCA cycle
Rxn#34: L-Aspartate + ATP + L-Citrulline → AMP + N(omega)-(L-Arginino)succinate + H ⁺ + Diphosphate	PA3525 (<i>argG</i>)	arginine metabolism
Rxn#35: Acetate + ATP + Coenzyme A → Acetyl-CoA + AMP + Diphosphate	PA4733 (<i>acsB</i>)	acetate metabolism
Rxn#36: 2 ATP + L-Glutamine + H ₂ O + Bicarbonate → 2 ADP + Carbamoyl phosphate + L-Glutamate + 2 H ⁺ + Phosphate	PA4758 (<i>carA</i>)	arginine metabolism
Rxn#37: 2 H ⁺ + H ₂ O + Urea → CO ₂ + 2 Ammonium	PA4867 (<i>ureB</i>)	arginine metabolism, ammonia production, carbon dioxide production

Table 1. Cont.

Reactions	Genes	Biological subsystems
Rxn#38: $\text{H}_2\text{O} + \text{Urocanate} \rightarrow 4\text{-Imidazolone-5-propanoate}$	PA5100 (<i>hutU</i>)	glutamate metabolism
Rxn#39: $\text{ATP} + \text{Oxaloacetate} \rightarrow \text{ADP} + \text{CO}_2 + \text{Phosphoenolpyruvate}$	PA5192 (<i>pckA</i>)	carbon dioxide production

doi:10.1371/journal.pone.0057050.t001

tklA (Group 22), *tpiA* (Group 19), *rpiA* (Group 24), *dapD* (Group 4), and *mliA* (Group 4), is reported as the driving force for *Streptococcus* biofilm formation [38].

As seen from Figure 3, the activity levels of all the biofilm-associated reactions are small for the genes from Clusters 3 through 6, i.e., PA0904 (*lysC*), PA1756 (*cysH*), PA2023(*galU*), and PA3686 (*adk*). Hence, they are regarded as ideal gene targets to treat *P. aeruginosa* infections because the inactivation of any of them can kill the planktonic pathogen cells but cannot promote the switch from planktonic growth to biofilm formation. The following experimental evidence further confirms these four genes are good targets to treat biofilm-associated *P. aeruginosa*:

- The enzyme encoded by *lysC* is aspartate kinase (AK) that catalyzes the phosphorylation of aspartic acid, the first step in the biosynthesis of the aspartic amino-acid family, lysine,

threonine, and methionine. Aspartate kinase can also influence the synthesis of diaminopimelic acid, which is an intermediate in the lysine biosynthesis and meanwhile a key compound participating in cell-wall synthesis in most bacteria [39]. Furthermore, it has been found that the mutation of *lysC* is unable to form biofilms in *Vibrio cholerae* [40].

- *cysH* encodes adenosine 5'-phosphosulfate reductase that catalyzes the first committed step of reductive sulfate assimilation in *P. aeruginosa*. This enzyme is required for the survival and virulence of *P. aeruginosa* in biofilms [41].
- Gene *galU* encodes glucose-1-phosphate uridylyltransferase, which mediates the reversible reaction of glucose-1-phosphate (G1P) and UTP into UDP-glucose and pyrophosphate. The function of glucose-1-phosphate uridylyltransferase in micro-organisms is to synthesize capsular polysaccharide that is

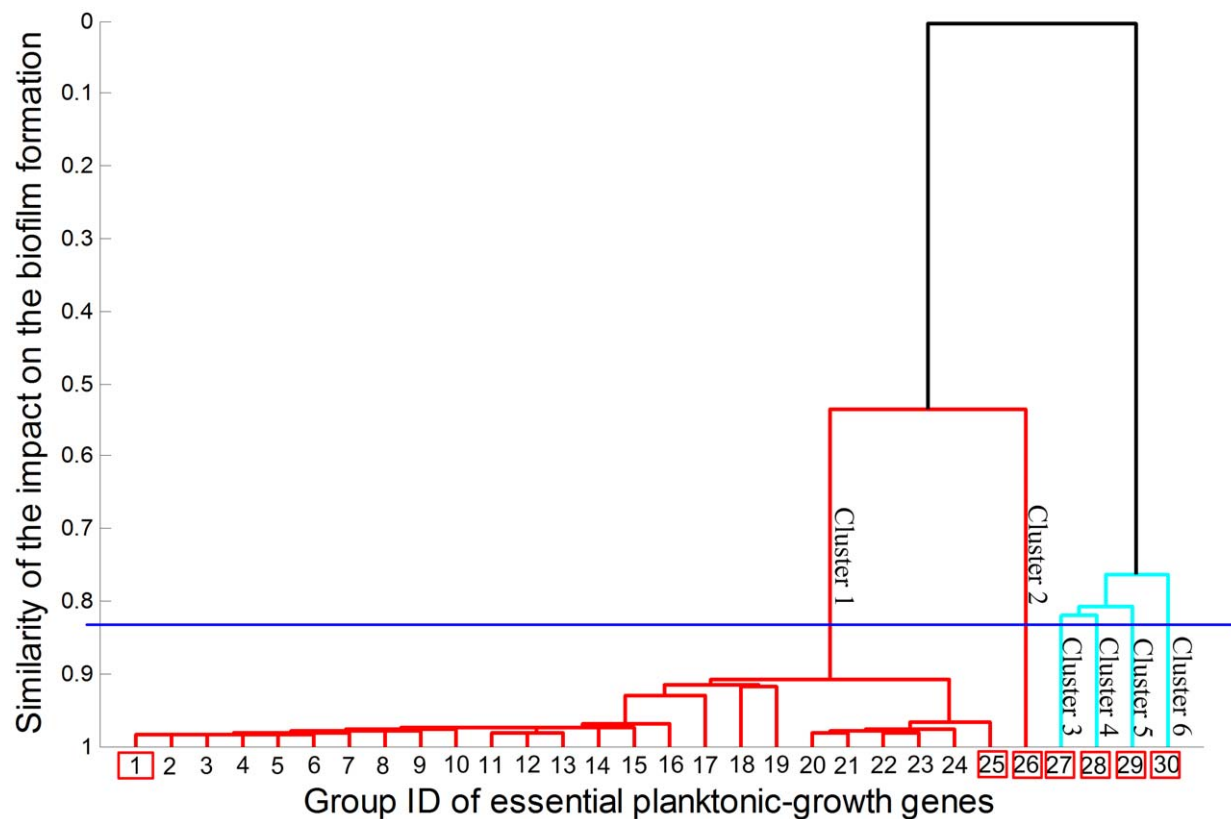


Figure 2. Clustering result for essential planktonic-growth genes based upon the ability of their mutants to form biofilms. These essential genes are separated into 30 groups by the hierarchical clustering program. The genes in each group are listed in Table 2. They are categorized into six clusters by selecting a threshold marked by the blue color line. The groups marked by red rectangles are regarded as the representatives of the cluster of genes. Two groups with the lowest similarity in the same cluster are selected as representatives if more than one group of genes are involved in that cluster.

doi:10.1371/journal.pone.0057050.g002

Table 2. Genes associated with each group shown in Figure 2.

Group ID	Cluster	# of genes in each group	Essential planktonic-growth genes
1	1	6	PA0945(<i>purM</i>) , PA4693(<i>pssA</i>), PA4855(<i>purD</i>), PA4957(<i>psd</i>), PA5164(<i>rmlC</i>), PA5549(<i>glmS</i>)
2	1	1	PA5161(<i>rmlB</i>)
3	1	1	PA3639(<i>accA</i>)
4	1	99	PA0005(<i>lptA</i>), PA0006(<i>yaeD</i>), PA0280(<i>cysA</i>), PA0281(<i>cysW</i>), PA0282(<i>cysT</i>), PA0283(<i>sbp</i>), PA0342(<i>thyA</i>), PA0350(<i>folA</i>), PA0363(<i>coaD</i>), PA0430(<i>metF</i>), PA0582(<i>folB</i>), PA0724(<i>coaA</i>), PA0761(<i>nadB</i>), PA1004(<i>nadA</i>), PA1013(<i>purC</i>), PA1162(<i>dapE</i>), PA1376(<i>aceK</i>), PA1393(<i>cysC</i>), PA1493(<i>cysP</i>), PA1758(<i>pabB</i>), PA1796(<i>folD</i>), PA1806(<i>fabI</i>), PA1959(<i>bacA</i>), PA2165(<i>glgA</i>), PA2584(<i>pgsA</i>), PA2629(<i>purB</i>), PA2962(<i>tmk</i>), PA2964(<i>pabC</i>), PA2967(<i>fabG</i>), PA2977(<i>murB</i>), PA2979(<i>kdsB</i>), PA2981(<i>lpxK</i>), PA3088(<i>yfjB</i>), PA3108(<i>purF</i>), PA3111(<i>folC</i>), PA3112(<i>accD</i>), PA3163(<i>cmk</i>), PA3337(<i>rfaD</i>), PA3636(<i>kdsA</i>), PA3637(<i>pyrG</i>), PA3643(<i>lpxB</i>), PA3644(<i>lpxA</i>), PA3651(<i>cdsA</i>), PA3666(<i>dapD</i>), PA3673(<i>plsB</i>), PA4006(<i>nadD</i>), PA4050(<i>pgpA</i>), PA4053(<i>ribE</i>), PA4054(<i>ribB</i>), PA4055(<i>ribC</i>), PA4056(<i>ribD</i>), PA4201(<i>ddlA</i>), PA4397(<i>panE</i>), PA4406(<i>lpxC</i>), PA4410(<i>ddlB</i>), PA4411(<i>murC</i>), PA4412(<i>murG</i>), PA4414(<i>murD</i>), PA4415(<i>mraY</i>), PA4416(<i>murF</i>), PA4417(<i>murE</i>), PA4425(<i>yraO</i>), PA4442(<i>cysN</i>), PA4443(<i>cysD</i>), PA4450(<i>murA</i>), PA4457(<i>KdsD</i>), PA4524(<i>nadC</i>), PA4529(<i>coaE</i>), PA4561(<i>ribF</i>), PA4655(<i>hemH</i>), PA4666(<i>hemA</i>), PA4670(<i>prs</i>), PA4729(<i>panB</i>), PA4730(<i>panC</i>), PA4749(<i>glmM</i>), PA4750(<i>folP</i>), PA4759(<i>dapB</i>), PA4847(<i>accB</i>), PA4848(<i>accC</i>), PA4854(<i>purH</i>), PA4920(<i>nadE</i>), PA4938(<i>purA</i>), PA4988(<i>waaA</i>), PA4996(<i>rfaE</i>), PA5009(<i>waaP</i>), PA5010(<i>waaG</i>), PA5011(<i>waaC</i>), PA5012(<i>waaF</i>), PA5034(<i>hemE</i>), PA5162(<i>rmlD</i>), PA5163(<i>rmlA</i>), PA5175(<i>cysQ</i>), PA5243(<i>hemB</i>), PA5259(<i>hemD</i>), PA5260(<i>hemC</i>), PA5278(<i>dapF</i>), PA5320(<i>coaC</i>), PA5336(<i>gmk</i>), PA5552(<i>glmU</i>)
5	1	2	PA3646(<i>lpxD</i>), PA3807(<i>ndk</i>)
6	1	1	PA3654(<i>pyrH</i>)
7	1	1	PA2968(<i>fabD</i>)
8	1	1	PA0654(<i>speD</i>)
9	1	1	PA1614(<i>gpsA</i>)
10	1	1	PA1609(<i>fabB</i>)
11	1	1	PA3763(<i>purl</i>)
12	1	1	PA4662(<i>murI</i>)
13	1	2	PA4458(<i>yrbI</i>), PA5008(<i>waaX</i>),
14	1	1	PA3659(<i>dapC</i>)
15	1	1	PA1687(<i>speE</i>)
16	1	1	PA3296(<i>phoA</i>)
17	1	1	PA0546(<i>metK</i>)
18	1	1	PA2053(<i>cynT</i>)
19	1	1	PA4748(<i>tpiA</i>)
20	1	2	PA0025(<i>aroE</i>), PA5039(<i>aroK</i>)
21	1	1	PA1681(<i>aroC</i>)
22	1	1	PA0548(<i>tktA</i>)
23	1	1	PA3164(<i>pseudogene</i>)
24	1	1	PA0330(<i>rpiA</i>)
25	1	1	PA5038(<i>aroB</i>)
26	2	1	PA4031(<i>ppa</i>)
27	3	1	PA0904(<i>lysC</i>)
28	4	1	PA2023(<i>galU</i>)
29	5	1	PA3686(<i>adk</i>)
30	6	1	PA1756(<i>cysH</i>)

The representative genes are underlined and marked in bold.

doi:10.1371/journal.pone.0057050.t002

essential for the survival and virulence of *P. aeruginosa* in biofilm [42].

- The enzyme encoded by *adk* is adenylate kinase, a ubiquitous enzyme that contributes to the homeostasis of adenine nucleotides in eukaryotic and prokaryotic cells. It is interesting to find that adenylate kinase functions as a *P. aeruginosa* virulence factor [43]. In particular, adenylate kinase is secreted by *P. aeruginosa*, and it in turn causes the death of human macrophage cell.

Investigation of metabolic reactions that facilitate biofilm formation

An interesting finding by inspecting Figure 3 (A) through (C) is that certain biofilm-associated reactions have significantly increased activity levels (e.g., more than three-fold) upon the single mutations of all the three representative genes from Cluster 1 and Cluster 2, i.e., PA0945 (*purM*), PA5038 (*aroB*), and PA4031 (*ppa*). This implies that, although these metabolic reactions are all associated with biofilm formation, certain parts of them are more

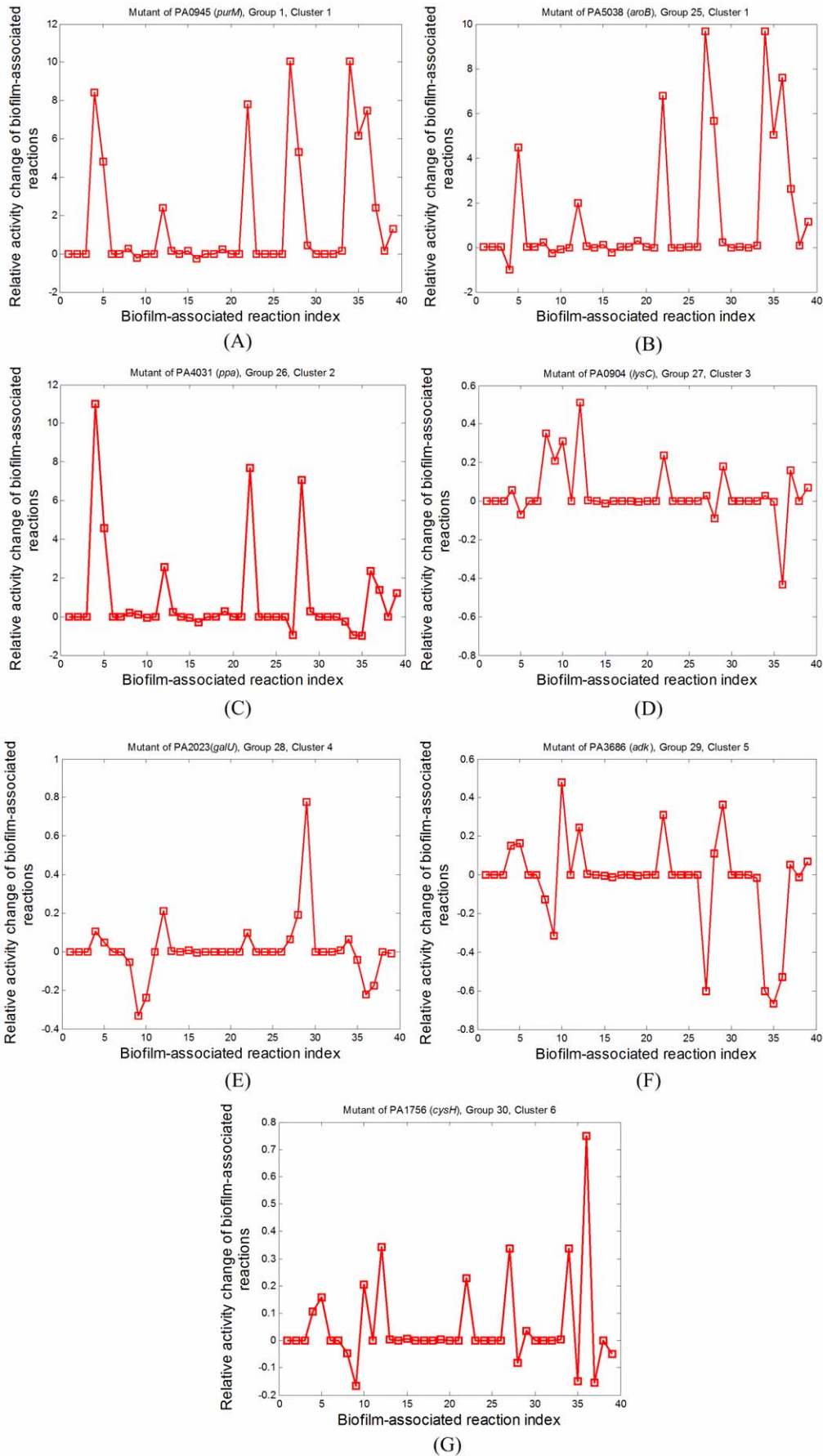


Figure 3. The relative activity change of biofilm-associated reactions for representative mutants. (A) PA0945 from Group 1 in Cluster 1, (B) PA5038 from Group 25 in Cluster 1, (C) PA4031 from Group 26 in Cluster 2, (D) PA0904 from Group 27 in Cluster 3, (E) PA2023 from Group 28 in Cluster 4, (F) PA3686 from Group 29 in Cluster 5, and (G) PA1756 from Group 30 in Cluster 6. The relative activity change of a biofilm-associated reaction is quantified by the relative flux change of this reaction upon the mutation of a single essential gene. The biofilm formation capability of a mutant is indicated by the relative activity change across all biofilm-associated reactions. If certain biofilm-associated reactions exhibit significantly enhanced activity levels, the mutant has large potential to form biofilms. Approximately 10 of the biofilm-associated reactions are of significantly increased activity levels upon the mutation of the genes from Cluster 1, and the activity levels of 6 biofilm-associated reactions surge in the mutants of the genes from Cluster 2. The mutants from Clusters 1 and 2 thus have high potential to form biofilms. Compared to the large relative activity increases of biofilm-associated reactions for mutants associated with Clusters 1 and 2, the activity level increases are minor for the mutants of the genes from Clusters 3 through 6. These mutants thus have low potential to form biofilms.
doi:10.1371/journal.pone.0057050.g003

likely to be utilized by *P. aeruginosa* upon the stress such as that imposed by the mutation of essential planktonic growth genes.

Therefore, we analyzed the relative activity changes of each biofilm-associated reaction across all single mutants, and categorized the 39 biofilm-associated reactions into two types: 1) the reactions with minor increase of activity levels for most essential mutants and 2) those with significant activity increase. Figure 4 (A) and (B) show the typical relative activity changes of two biofilm-associated reactions: a minor-increase reaction (Reaction Rxn# 1) and a “significant-increase” reaction (Reaction Rxn# 4). The relative activity changes of other biofilm-associated reactions are similar to either the one for Rxn# 1 or that for Rxn# 4 (data not shown). Reactions associated with each of the two types of reactions are listed in Table 3.

The “significant-increase” type of reactions have significant increases in activity levels for the mutants of most essential genes, and might reveal the underlying mechanisms for biofilm formation when *P. aeruginosa* is treated by the antimicrobial agents that attack the essential genes. It can be seen from Table 3 that these reactions are mainly involved in acetate metabolism (mainly via Rxn# 4 and 35), arginine metabolism (mainly via Rxn#27 and 34), glutamate metabolism (mainly via Rxn# 5 and 36), the regulation of hydrogen peroxide (mainly via Rxn# 22), and the phosphate transport (mainly via Reaction Rxn #28). These findings are supported by existing experimental data listed as follows.

- The microarray data presented in Prüss et al, 2010 [44] show that acetate metabolism acts as a metabolic sensor for adjusting the biofilm structure of *E. coli* K12 to the change of the surrounding environment.
- It is reported by Beenken et al., 2004 [45], that the arginine plays an important role in generation of ammonia that can neutralize acids generated by bacterial glycolysis in *Staphylococcus aureus* biofilms.
- The isomerization process of glutamate is one of the crucial steps influencing the transport and accumulation of extracellular substances for biofilm formation of *Bacillus amyloquelificans* [46].
- Hydrogen peroxide is involved in the lysine oxidase activity that causes cell death within micro-colonies during biofilm formation of both *Marinomonas mediterranea* and *Pseudoalteromonas tunicata* [47].
- It is reported by Monds et al., 2007 [48], that the extracellular phosphate in the environmental conditions is involved in the regulation of *Pseudomonas fluorescens* Pf0-1 biofilm formation. The phosphate transport thus influences the biofilm development by coordinating the extra and intracellular phosphate concentrations.

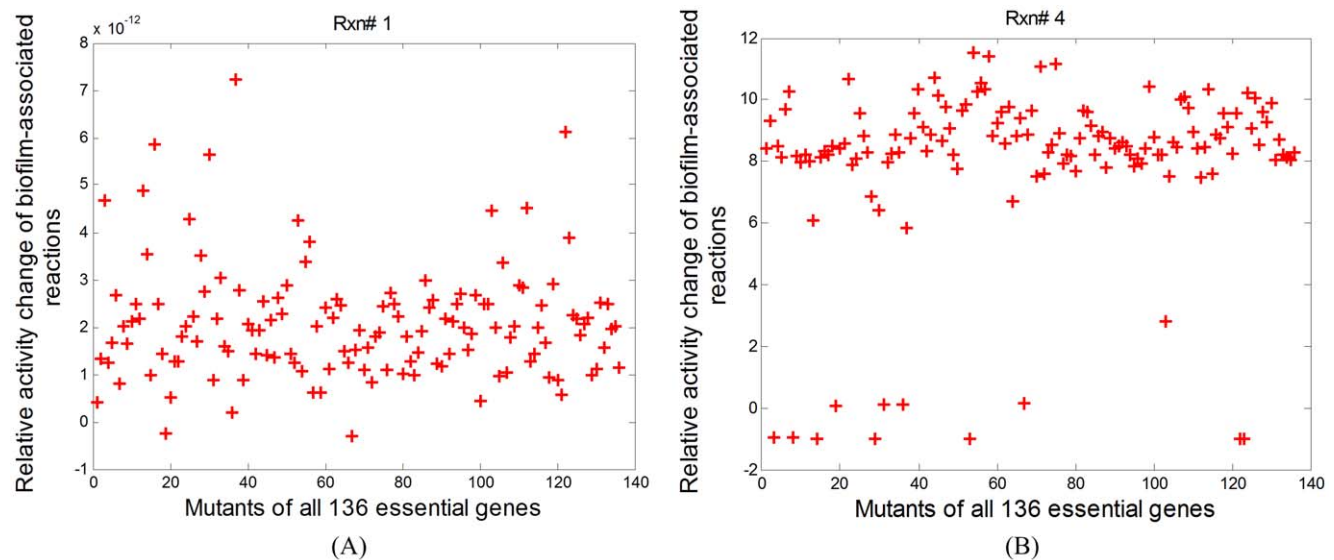


Figure 4. Relative fold change in activity levels of two biofilm-associated reactions for all single mutants. (A) Reaction Rxn# 1, and (B) Reaction Rxn# 4. The activity level of Rxn#1 is of little change for most essential mutants, while the fluxes are significantly re-distributed through Rxn# 4 for most mutants. In other words, Rxn# 4 is of much higher activity levels upon the mutation of most essential genes. This implies that Rxn# 4 stands for the mechanism utilized by mutants to form biofilms. Based upon the relative activity change profiles, the biofilm-associated reactions are categorized into two types, one with minor flux changes upon the mutation of most essential genes (represented by Rxn# 1), and the other with large flux changes upon the mutation of most essential genes (represented by Rxn# 4).
doi:10.1371/journal.pone.0057050.g004

Table 3. Categorization of the biofilm-associated reactions based upon their relative activity changes in the mutants of 136 essential planktonic-growth genes.

Category of biofilm-associated reactions	Biofilm-associated reactions included in each cluster
"Minor-increase": reactions with minor activity changes for most 136 single mutants	Rxn#1~3, 6~21, 23~26, 29~33, 37~39 in Table 1
"Significant-increase": reactions with large increase (e.g., more than three-fold relative change) in their activity levels for most single mutants	Rxn#4: Coenzyme A + 3-Oxoadipyl-CoA → Acetyl-CoA + Succinyl-CoA
	Rxn#5: 4-Aminobutanoate + 2-Oxoglutarate → L-Glutamate + Succinic semialdehyde
	Rxn#22: 2 Hydrogen peroxide → 2 H ₂ O + O ₂
	Rxn#27: N(omega)-(L-Arginino)succinate ↔ L-Arginine + Fumarate
	Rxn#28: ATP + H ₂ O + Phosphate[e] → ADP + H ⁺ + 2 Phosphate[c]
	Rxn#34: L-Aspartate + ATP + L-Citrulline → AMP + N(omega)-(L-Arginino)succinate + H ⁺ + Diphosphate
	Rxn#35: Acetate + ATP + Coenzyme A → Acetyl-CoA + AMP + Diphosphate
	Rxn#36: 2 ATP + L-Glutamine + H ₂ O + Bicarbonate → 2 ADP + Carbamoyl phosphate + L-Glutamate + 2 H ⁺ + Phosphate

doi:10.1371/journal.pone.0057050.t003

While all the aforementioned experiments focus on the investigation of the impact of individual metabolic modules on biofilm formation, this work represents the first mathematical modeling approach for the systemic identification of the underlying metabolic mechanisms that facilitate biofilm formation of single-mutants.

Discussion

The formation of biofilms facilitates the survival of disease-causing pathogens in hostile environmental conditions. Therefore, preventing the pathogen's transition from the planktonic state to the biofilm growth mode is one of the most important steps to treat biofilm-associated pathogens. Since the metabolism of pathogens is determined by the interaction of hundreds of metabolic reactions, genes, and enzymes, systems biology approaches can facilitate gene target identification for preventing the planktonic to biofilms transition. In this work, a systems-level analysis approach was presented to answer the following question that still remains unanswered, that is, how to mathematically quantify the capability of a pathogen to form biofilms upon the mutation of a specific gene. The fluxes through the reactions associated with an essential planktonic-growth gene are limited to 10% of their nominal values in flux balance analysis to mimic the mutation of the gene in this work. Although the approach used here is a partial shutdown instead of a complete gene-knockout mutation, it better reflects the response of pathogens to the treatment of antimicrobial agents, which generally cannot immediately eliminate the biological functions of the target genes. The mutant might form biofilm before antimicrobial agents completely eliminate the bacterial metabolism. In addition, the pathogen might be treated by a sub-inhibition dose, which can be mimicked by setting the allowable fluxes to 10% of their nominal values. The terminology "mutant" or "mutation" is still used in this work, although an *in silico* partial shutdown mutation is performed to identify gene targets from those essential planktonic-growth genes of *P. aeruginosa*.

Based upon the results in this work, it is interesting to hypothesize that the mutations of essential genes are more likely to induce *P. aeruginosa* biofilm than those of non-essential ones. For essential genes, the mutations of 132 out of the 136 essential planktonic-growth genes were predicted to induce biofilm

formation, and the down-regulation of 8 of them, i.e., *pyrH*, *tklA*, *tpiA*, *rpiA*, *dapD*, and *rmlA* [38], *glmU* [36], and *pgsA* [37], has been experimentally proven to be positively correlated with biofilm formation. In contrast, we find that only two out of 920 non-essential mutants might induce biofilm formation, when applying the proposed approach to evaluate the biofilm formation capability of mutants of the nonessential genes in the *P. aeruginosa* model reported by Oberhardt et al., 2008 [16]. For the 920 non-essential genes, we have followed the experiment approaches presented in Ueda et al., 2009 [34], to carry out comprehensive screening experiment for altered biofilm mutants from the PA14 non-redundant transposon mutant library (Liberati et al., 2006, [49]). The library contains 835 of the 920 non-essential genes, and of these 835 mutants, 823 have been verified not to induce biofilm formation (data can be provided upon request). While we predicted a lower number of mutants that form biofilms (i.e., two by prediction versus 12 by experiment), this is probably because we use a conservative approach in mimicking partial shutdown mutation, in which a reaction that is catalyzed by multiple enzymes is not inhibited for single mutants, but on the whole the results were verified for the vast majority of the bacteria with mutations in non-essential genes. The potential reason for explaining the hypothesized strong biofilm formation capability of most essential mutants is that the mutation of an essential planktonic-growth gene might cause limited nutrient uptakes, which strongly enhance biofilm formation [6], and make the pathogen feel the stress, which is reported as one of the major driving forces for biofilm formation [50]. On the contrary, when a non-essential gene is inhibited, the pathogen does not feel very much stress, because non-essential genes are not crucial for the biomass synthesis. This hypothesis as well as the findings of the biofilm formation phenotypes of single mutants forms the foundation for the further experimental investigation.

This work was mainly focused on investigating the biofilm formation capability of single mutants, as the single gene inhibition is easier to implement than the multiple gene inhibition. While it is possible to apply the proposed approach to multiple-mutants, the ACHR sampling approach needs to be upgraded to improve the computational efficiency. It takes approximately 15 minutes to obtain 20,000 samples of fluxes of the biofilm-associated reactions upon the mutation of a single gene for a desktop computer with

Intel Core i5 2.5 GHz CPU and 8 GB RAM. Since 920 nonessential genes have been reported in Oberhardt et al., 2008 [16], it is very time-consuming to study the biofilm formation ability of all possible multiple-mutants. Nevertheless, the study of the impact of the multiple-gene mutation on biofilm formation is an interesting topic for the future research.

To the best knowledge of the authors, this work represents the first systems-level approach to incorporate the quantification of biofilm formation capability in evaluation of genes as the targets to treat biofilm-associated pathogens. Four essential planktonic-growth genes, i.e., PA0904 (*lysC*), PA1756 (*cysH*), PA3686 (*adk*), and PA2023(*galU*), were identified as the potential gene targets to treat *P. aeruginosa*, as the mutation of any of these genes can eliminate the pathogen without inducing biofilm formation. This finding is implied by existing experimental data. Based upon the relative activity change of biofilm-associated reactions over all the single mutants, it is interesting to find that the fluxes of approximately 8 biofilm-associated reactions significantly increase for most single essential mutants. They are mainly associated with acetate metabolism, arginine metabolism, and glutamate metabolism. All these findings can be used to generate hypotheses for experiment design. In addition, the proposed approach can be applied to identify gene targets for treating any other biofilm-associated pathogen if the genes positively associated with biofilm formation have been identified and the metabolic model has been developed for the pathogen.

Materials and Methods

In this section, the systems-level approach whose outline is shown in the illustrative example (Figure 1) is described. This approach is used to quantify the biofilm formation capability of single essential mutants, cluster essential genes according to the biofilm formation capacity of their mutants, and systematically identify the metabolic reactions whose activity levels significantly increase for most single essential mutants. The detail of the proposed approach is shown step by step as follows.

- Step 1: Genes positively associated with biofilm formation of the target pathogen are identified and are overlaid onto the metabolic network of the pathogen to determine the reactions positively associated with biofilm formation. The fluxes of these biofilm-associated reactions are used as soft sensors to monitor the ability of a mutant to form biofilm. In particular, if the fluxes (i.e., activity levels) of certain biofilm-associated reactions significantly increase upon the mutation of an essential planktonic-growth gene, the pathogen might form biofilms before it is eliminated by the antimicrobial agents which attack that essential gene. In other words, the essential gene is not a good target to treat the pathogen.
- Step 2: An integrated flux balance analysis and ACHR sampling approach is applied to determine the relative change of the flux distribution of individual biofilm-associated reactions upon the mutation of an essential gene. FBA is a standard tool to determine the reaction fluxes through the entire metabolic network under the assumption that the microorganism utilizes available nutrients to get the maximum growth rate [19,51]. FBA can be regarded as a linear programming problem given by Equation (1).

$$\max_{\mathbf{v}} v_{biomass}$$

Subject to :

$$\sum \mathbf{S} \cdot \mathbf{v} = 0$$

$$lb_i \leq v_i \leq ub_i$$

where $v_{biomass}$ is the biomass growth rate, $\mathbf{S} \in \mathbb{R}^{n_m \times n_r}$ is the stoichiometric matrix, n_m and n_r are the number of metabolites and metabolic reactions respectively, $\mathbf{v}$ is a vector containing fluxes for all reactions in the metabolic network, lb_i and ub_i are the lower and upper bounds of flux v_i . FBA is performed to predict the maximum microorganism growth rate under a specific nutrient environment specified by the lower and upper bounds of fluxes. Equation (1) can be solved by the program package GNU Linear Programming Kit (GLPK) [52].

Since the number of metabolic reactions (i.e., n_r) is generally larger than the number of metabolites (i.e., n_m) in the reaction network, the stoichiometric matrix $\mathbf{S}$ is underdetermined. The optimal solution for Equation (1) is typically not unique [53]. In this work, the solution space corresponding to the maximum growth rate is sampled using ACHR sampling. In particular, the ACHR sampling routine provided by the Cobra toolbox [19,51] is used to get N flux samples for each metabolic reaction from the solution space, where N is a large number (e.g., 20,000) that can make the sampling process converge. For each metabolic reaction, its N sampled flux values are analyzed to quantify the mean value (represented by μ_{v_i}), and the probability density function of fluxes (represented by f_{v_i}) in the solution space. i is the index of a metabolic reaction.

The aforementioned approach is used to sample fluxes of biofilm-associated reactions for the wild-type strain and the single essential mutants. The corresponding mean value, and probability density function quantified from the sampled fluxes for each biofilm-associated reaction are represented by $\mu_{v_n}^{wild}$ and $f_{v_n}^{wild}$ for the wild-type strain, and $\mu_{v_n}^{gene\ m}$ and $f_{v_n}^{gene\ m}$ for gene m mutant. n is the index of a biofilm-associated reaction. To mimic the mutation of an essential gene m in flux balance analysis, the upper bounds of fluxes of reactions associated with the essential gene m are set to 10% of the mean value of their N sampled fluxes obtained for the wild type strain. If a reaction is reversible, the lower bound of its flux is set to -10% of the mean value of its N sampled fluxes obtained for the wild type strain. The pathogen might form biofilms before the gene target is totally inhibited. A 90% flux reduction can mimic the growth pattern of the pathogen before antimicrobial agents completely abolish the biological functions of their targets. In addition, the sub-inhibition of a good gene target should not induce the formation of biofilms before antimicrobial agents totally inhibit its biological functions.

- Step 3: The trend to form biofilms for the mutant of essential gene m is quantified by comparing the distribution of fluxes of each biofilm-associated reaction, represented by $\mu_{v_n}^{gene\ m}$ and $f_{v_n}^{gene\ m}$, to its counterpart $\mu_{v_n}^{wild}$ and $f_{v_n}^{wild}$ obtained for the wild-type strain.

The Kolmogorov – Smirnov (K-S) test approach [54] is first applied to determine whether the distribution of fluxes of each biofilm-associated reactions (i.e., $f_{v_n}^{gene\ m}$ with $n = 1, 2, \dots$,

$num_{\text{biofilm-reactions}}$), is of statistically significant change from the corresponding distribution obtained for the wild-type strain (i.e., $f_{v_n}^{\text{wild}}$, $n = 1, 2, \dots, num_{\text{biofilm-reactions}}$). $num_{\text{biofilm-reactions}}$ refers to the total number of biofilm-associated reactions. The two sample K-S test routine *kstest2* from MATLAB is used to perform K-S test, at a 5% significance level. If the two distributions $f_{v_n}^{\text{gene } m}$ and $f_{v_n}^{\text{wild}}$ are different, $KS(f_{v_n}^{\text{gene } m}, f_{v_n}^{\text{wild}})$ is equal to one, otherwise, $KS(f_{v_n}^{\text{gene } m}, f_{v_n}^{\text{wild}})$ is set to zero. The difference between the two distributions $f_{v_n}^{\text{gene } m}$ and $f_{v_n}^{\text{wild}}$ is further quantified by the following equation

$$Activity_Var_{v_n}^{\text{gene } m} = \frac{\mu_{v_n}^{\text{gene } m} - \mu_{v_n}^{\text{wild}}}{\mu_{v_n}^{\text{wild}}} \times KS(f_{v_n}^{\text{gene } m}, f_{v_n}^{\text{wild}})$$

where $Activity_Var_{v_n}^{\text{gene } m}$ is the relative fold change of the activity (i.e., the flux) of the biofilm-associated reaction n upon the mutation of the essential gene m . Equation (2) is applied to all biofilm-associated reactions. An activity-variation vector is then constructed as shown in Equation (3) to describe the change in the distributions of fluxes across all biofilm-associated reactions upon the mutation of gene m .

$$\begin{aligned} \text{Vector_activity_var}^{\text{gene } m} = \\ [Activity_Var_{v_1}^{\text{gene } m} \quad Activity_Var_{v_2}^{\text{gene } m} \quad \dots \\ Activity_Var_{v_{num_{\text{biofilm-reactions}}}}^{\text{gene } m}] \end{aligned}$$

where $\text{Vector_activity_var}^{\text{gene } m}$ is referred to as the relative activity change profile for the gene m mutant in the following text. When Equation (3) is applied to all genes that are essential for planktonic growth, a matrix given by Equation (4) is constructed, which shows in each column the relative activity changes of one biofilm-associated reaction i for the mutants of all essential genes, and which shows in each row the relative activity changes in all biofilm-associated reactions upon the mutation of one essential gene.

$$\text{Matrix_activity_var} = \begin{bmatrix} Activity_Var_{v_1}^{\text{gene } 1} & Activity_Var_{v_2}^{\text{gene } 1} & \dots & Activity_Var_{v_{num_{\text{biofilm-reactions}}}}^{\text{gene } 1} \\ Activity_Var_{v_1}^{\text{gene } 2} & Activity_Var_{v_2}^{\text{gene } 2} & \dots & Activity_Var_{v_{num_{\text{biofilm-reactions}}}}^{\text{gene } 2} \\ \vdots & \vdots & \ddots & \vdots \\ Activity_Var_{v_1}^{\text{gene } num_{\text{gene}}} & Activity_Var_{v_2}^{\text{gene } num_{\text{gene}}} & \dots & Activity_Var_{v_{num_{\text{biofilm-reactions}}}}^{\text{gene } num_{\text{gene}}} \end{bmatrix}$$

where num_{gene} is the total number of essential genes for the planktonic pathogens. It can be seen from Equation (4) that the activity-variation vector $\text{Vector_activity_var}^{\text{gene } m}$ given in Equation (3) constitutes one row of **Matrix_activity_var**.

- Step 4: Potential gene targets are identified from the essential planktonic-growth genes by hierarchically categorizing them into different clusters based upon the capability of their mutants to form biofilms.

One hypothesis to test in this work is that the mutation of an essential gene might induce the formation of biofilms. $\text{Vector_activity_var}^{\text{gene } m}$ shown in Equation (3) describes the relative change of activity levels of all biofilm-associated reactions for the gene m mutant. $\text{Vector_activity_var}^{\text{gene } m}$ is thus used to represent the trend of biofilm formation upon the mutation of gene m . In particular, if certain elements in

$\text{Vector_activity_var}^{\text{gene } m}$, are of values larger than three, it means a significant amount of fluxes are re-distributed in the biofilm-associated reactions. The mutant thus exhibits high potential to switch from the planktonic state to the biofilm growth mode. In this work, the essential genes are categorized into different clusters based upon the capability of their mutants to form biofilms. The clustering result is then used to determine the gene targets to treat biofilm-associated pathogens. Specifically two relative activity change profiles are used to quantify the similarity of the impact of the mutation of two genes on biofilm formation via Equation (5).

$Similarity_{j,k}$

$$\begin{aligned} &= \frac{(\text{Vector_activity_var}^{\text{gene } i})^T \times \text{Vector_activity_var}^{\text{gene } k}}{\|\text{Vector_activity_var}^{\text{gene } i}\| \|\text{Vector_activity_var}^{\text{gene } k}\|} \\ &\times \frac{2}{\frac{\text{Vector_activity_var}^{\text{gene } i}}{\text{Vector_activity_var}^{\text{gene } k}} + \frac{\text{Vector_activity_var}^{\text{gene } k}}{\text{Vector_activity_var}^{\text{gene } i}}} \end{aligned}$$

where $Similarity_{i,k}$ defines a similarity measure of the biofilm formation capability of the mutants of genes i and k . If the similarity measure is equal to unity, the mutations of the two genes have the same impact on the formation of biofilms.

$\text{Vector_activity_var}^{\text{gene } i}$ is the average relative activity change over all biofilm-associated reactions upon the mutation of gene i . The first term on the right-hand side of Equation (5) reflects the similarity of the shapes of activity change profiles contained in $\text{Vector_activity_var}^{\text{gene } i}$ and $\text{Vector_activity_var}^{\text{gene } k}$. If these two activity change profiles are parallel, e.g., $\text{Vector_activity_var}^{\text{gene } k}$ equal to $3 \times \text{Vector_activity_var}^{\text{gene } i}$, the first term is of a value of one. For this case, if only the first term is used as the measurement of similarity, the mutation of gene k is determined to have the same impact on biofilm formation as that of gene i . However, this is not true as the activity change magnitude in the mutant of gene k is three times of that for the gene i mutant. The second term on the right-hand side of Equation (5) is thus used to take the magnitude of activity change into account in the evaluation of similarity. If the activity change magnitudes of two profiles are quite different, the denominator of the second term is large. This reduces the similarity value of the two activity change profiles. Only two activity change profiles with similar shapes and comparable magnitudes are of a large similarity value.

Based upon the quantified similarity, the genes essential for planktonic growth can be hierarchically categorized into different clusters via the hierarchical clustering routine *dendrogram* from MATLAB. Those essential planktonic-growth genes upon which the mutations lead to a low enhancement in the activity levels across biofilm-associated reactions are identified and regarded as good gene targets to treat biofilm-associated pathogens.

- Step 6: The metabolic reactions that facilitate biofilm formation in the single essential mutants are determined by identifying biofilm-associated reactions whose activity levels surge in most mutants. Each column in **Matrix_activity_var** shown in Equation (4) lists the relative change of the activity of one biofilm-associated reaction upon the mutation of all essential genes. Those biofilm-associated reactions whose activity levels significantly increase for most single-mutants account for the underlying metabolic mechanisms that facilitate biofilm formation in these mutants. In this work,

the average relative activity change for a biofilm-associated reaction, $\overline{Activity}_{v_i}$, is quantified by Equation (6). Based upon the value of $\overline{Activity}_{v_i}$ the biofilm-associated reactions are categorized into two types, one containing reactions with minor activity increase and the other containing reactions with large activity increase for most single-mutants. Two types were preferred here, as the simulation result showed that the relative flux activity changes of a biofilm-associated reaction over most mutants can be generally characterized by either minor changes or significant ones. The *k*-mean clustering routine in MATLAB is used to perform the clustering operation.

$$\overline{Activity}_{v_i} = \frac{\sum_{m=1}^{num_{gene}} |Activity_Var_{v_i}^{gene\ m}|}{num_{gene}}$$

References

- Lew DP, Waldvogel FA (2004) Osteomyelitis. *Lancet* 364(9431):369–379.
- Gjødskøl K, Christensen JJ, Karlsmark T, Jørgensen B, Klein BM, et al. (2006) Multiple bacterial species reside in chronic wounds: a longitudinal study. *Int Wound J* 3(3):225–231.
- James GA, Swogger E, Wolcott R, Pulcini E, Secor P, et al. (2008) Biofilms in chronic wounds. *Wound Repair Regen* 16(1):37–44.
- Singh PK, Schaefer AL, Parsek MR, Moninger TO, Welsh MJ, et al. (2000) Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* 407(6805):762–764.
- Mah TF, O'Toole GA (2001) Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol* 9(1):34–39.
- Nguyen D, Joshi-Datar A, Lepine F, Bauerle E, Olakanmi O, et al. (2011) Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science* 334(6058):982–986.
- Stoodley P, Sauer K, Davies DG, Costerton JW (2002) Biofilms as complex differentiated communities. *Annu Rev Microbiol* 56:187–209.
- Jayaraman A, Wood TK (2008) Bacterial quorum sensing: signals, circuits, and implications for biofilms and disease. *Annu Rev Biomed Eng* 10:145–167.
- Senadheera D, Cvikovitch DG (2008) Quorum sensing and biofilm formation by *Streptococcus mutans*. *Adv Exp Med Biol* 631:178–188.
- Jones MB, Jani R, Ren D, Wood TK, Blaser MJ (2005) Inhibition of *Bacillus anthracis* growth and virulence-gene expression by inhibitors of quorum-sensing. *J Infect Dis* 191(11):1881–1888.
- Ueda A, Wood TK (2009) Connecting quorum sensing, c-di-GMP, pel polysaccharide, and biofilm formation in *Pseudomonas aeruginosa* through tyrosine phosphatase TpbA (PA3885). *PLoS Pathog* 5(6):e1000483.
- Edwards JS, Palsson BO (2000) Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol Prog* 16(6):927–939.
- Heinemann M, Kümmel A, Ruinatscha R, Panke S (2005) In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnol Bioeng* 92(7):850–864.
- Jamshidi N, Palsson BO (2007) Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol* 1:26.
- Kim HU, Kim TY, Lee SY (2010) Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen *Acinetobacter baumannii* AYE. *Mol Biosyst* 6(2):339–348.
- Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA (2008) Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J Bacteriol* 190(8):2790–2803.
- Sigurdsson G, Fleming RM, Heinken A, Thiele I (2012) A systems biology approach to drug targets in *Pseudomonas aeruginosa* biofilm. *PLoS One* 7(4):e34337.
- Müsken M, Di Fiore S, Dötsch A, Fischer R, Häussler S (2010) Genetic determinants of *Pseudomonas aeruginosa* biofilm establishment. *Microbiology* 156(Pt 2):431–441.
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2(3):727–738.
- Kaufman DE, Smith RL (1998) Direction choice for accelerated convergence in hit-and-run sampling. *Operations Res* 46: 84–96.
- Cao B, Wang H, Sun H, Zhu Y, Chen M (2004) Risk factors and clinical outcomes of nosocomial multi-drug resistant *Pseudomonas aeruginosa* infections. *J Hosp Infect* 57(2):112–118.
- Oberhardt MA, Goldberg JB, Hogardt M, Papin JA (2010) Metabolic network analysis of *Pseudomonas aeruginosa* during chronic cystic fibrosis lung infection. *J Bacteriol* 192(20):5534–5548.
- Schlag S, Nerz C, Birkenstock TA, Altenberend F, Götz F (2007) Inhibition of staphylococcal biofilm formation by nitrite. *J Bacteriol* 189(21):7911–7919.
- Wolfe AJ (2005) The Acetate switch. *Microbiol Mol Biol Rev* 69(1): 12–50.
- Gaupp R, Schlag S, Liebeck M, Lalk M, Götz F (2010) Advantage of upregulation of succinate dehydrogenase in *Staphylococcus aureus* biofilms. *J Bacteriol* 192(9):2385–2394.
- Kroukamp O, Wolfaardt GM (2009) CO₂ production as an indicator of biofilm metabolism. *Appl Environ Microbiol* 75(13):4391–4397.
- Hancock V, Dahl M, Klemm P (2010) Abolition of biofilm formation in urinary tract *Escherichia coli* and *Klebsiella* isolates by metal interference through competition for fur. *Appl Environ Microbiol* 76(12):3836–3841.
- Hancock V, Ferrières L, Klemm P (2008) The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious *Escherichia coli* in human urine. *Microbiology* 154(Pt 1):167–175.
- Wu Y, Outten FW (2009) IscR controls iron-dependent biofilm formation in *Escherichia coli* by regulating type I fimbria expression. *J Bacteriol* 191(4):1248–1257.
- Itzek A, Zheng L, Chen Z, Merritt J, Kreth J (2011) Hydrogen peroxide-dependent DNA release and transfer of antibiotic resistance genes in *Streptococcus gordonii*. *J Bacteriol* 193(24):6912–6922.
- Lu CD, Yang Z, Li W (2004) Transcriptome analysis of the ArgR regulon in *Pseudomonas aeruginosa*. *J Bacteriol* 186:3855–3861.
- Chandra H, Basir SF, Gupta M, Banerjee N (2010) Glutamine synthetase encoded by glnA-1 is necessary for cell wall resistance and pathogenicity of *Mycobacterium bovis*. *Microbiology* 156(Pt 12):3669–3677.
- Garavaglia M, Rossi E, Landini P (2012) The pyrimidine nucleotide biosynthetic pathway modulates production of biofilm determinants in *Escherichia coli*. *PLoS One* 7(2):e31252.
- Ueda A, Attila C, Whiteley M, Wood TK (2009) Uracil influences quorum sensing and biofilm formation in *Pseudomonas aeruginosa* and fluorouracil is an antagonist. *Microb Biotechnol* 2: 62–74.
- Serra DO, Lücking G, Weiland F, Schulz S, Görg A, et al. (2008) Proteome approaches combined with Fourier transform infrared spectroscopy revealed a distinctive biofilm physiology in *Bordetella pertussis*. *Proteomics* 8(23–24):4995–5010.
- Burton E, Gawande PV, Yakandawala N, LoVetri K, Zhanel GG, et al. (2006) Antibiofilm activity of GlnU enzyme inhibitors against catheter-associated uropathogens. *Antimicrob Agents Chemother* 50(5):1835–1840.
- Shiba Y, Miyagawa H, Nagahama H, Matsumoto K, Kondo D, et al. (2012) Exploring the relationship between lipoprotein mislocalization and activation of the Rcs signal transduction system in *Escherichia coli*. *Microbiology* 158(Pt 5):1238–1248.
- Rathsam C, Eaton RE, Simpson CL, Browne GV, Valova VA, et al. (2005) Two-dimensional fluorescence difference gel electrophoretic analysis of *Streptococcus mutans* biofilms. *J Proteome Res* 4(6):2161–2173.
- Yoshida A, Tomita T, Kuzuyama T, Nishiyama M (2007) Purification, crystallization and preliminary X-ray analysis of the regulatory subunit of aspartate kinase from *Thermus thermophilus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 63(Pt 2): 96–98.
- Mueller RS, McDougald D, Cusumano D, Sodhi N, Kjelleberg S, et al. (2007) *Vibrio cholerae* strains possess multiple strategies for abiotic and biotic surface colonization. *J Bacteriol* 189(14):5348–5360.
- Kopriva S, Büchert T, Fritz G, Suter M, Benda R, et al. (2002) The presence of an iron-sulfur cluster in adenosine 5'-phosphosulfate reductase separates organisms utilizing adenosine 5'-phosphosulfate and phosphoadenosine 5'-phosphosulfate for sulfate assimilation. *J Biol Chem* 277(24):21786–21791.

Supporting Information

Table S1 The distribution of the essential genes from Cluster 1 of Figure 2 in biological subsystems. (DOCX)

Author Contributions

Conceived and designed the experiments: ZX ZH. Performed the experiments: TKW. Analyzed the data: ZX XF TKW ZH. Contributed reagents/materials/analysis tools: ZX ZH. Wrote the paper: ZX XF TKW ZH.

42. ewBonofiglio L, García E, Mollerach M (2005) Biochemical characterization of the pneumococcal glucose 1-phosphate uridylyltransferase (GalU) essential for capsule biosynthesis. *Curr Microbiol* 51(4):217–221.
43. Markaryan A, Zaborina O, Punj V, Chakrabarty AM (2001) Adenylate kinase as a virulence factor of *Pseudomonas aeruginosa*. *J Bacteriol* 183(11):3345–3352.
44. Prüss BM, Verma K, Samanta P, Sule P, Kumar S, et al. (2010) Environmental and genetic factors that contribute to *Escherichia coli* K-12 biofilm formation. *Arch Microbiol* (9):715–728.
45. Beenken KE, Dunman PM, McAleese F, Macapagal D, Murphy E, et al. (2004) Global gene expression in *Staphylococcus aureus* biofilms. *J Bacteriol* 186(14):4665–4684.
46. Liu J, Ma X, Wang Y, Liu F, Qiao J, et al. (2011) Depressed biofilm production in *Bacillus amyloliquefaciens* C06 causes c-Polyglutamic acid (c-PGA) overproduction. *Curr Microbiol* 62(1):235–241.
47. Mai-Prochnow A, Lucas-Elio P, Egan S, Thomas T, Webb JS, et al. (2008) Hydrogen peroxide linked to lysine oxidase activity facilitates biofilm differentiation and dispersal in several gram-negative bacteria. *J Bacteriol* 190(15):5493–5501.
48. Monds RD, Newell PD, Gross RH, O'Toole GA (2007) Phosphate-dependent modulation of c-di-GMP levels regulates *Pseudomonas fluorescens* Pf0-1 biofilm formation by controlling secretion of the adhesin LapA. *Mol Microbiol* 63(3):656–679.
49. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, et al. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci U S A* 103(8): 2833–2838.
50. Zhang X, Garcia-Contreras R, Wood TK (2007). YcfR (BhsA) Influences *Escherichia coli* biofilm formation through stress response and surface hydrophobicity. *J Bacteriol* 189(8):3051–3062.
51. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6(9):1290–1307.
52. Makhorin A. GLPK (GNU Linear Programming Kit) User's Guide. [Online]. Available: <http://www.gnu.org/software/glpk/glpk.html>. Accessed 2013 Jan 28.
53. Schellenberger J, Palsson BO (2009) Use of randomized sampling for analysis of metabolic networks. *J Biol Chem* 284(9):5457–5461.
54. Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* 46(253): 68–78.