Antimicrobial Target Discovery With Metabolic Network Models

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

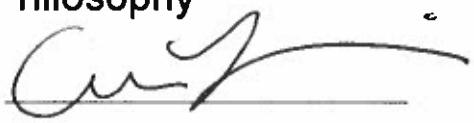
by

Anna S Blazier

May 2019

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: 

This Dissertation has been read and approved by the examining committee:

Advisor: Jason Papin

Committee Member: Philip Bourne

Committee Member: Alison Criss

Committee Member: Kevin Janes

Committee Member: Shayn Peirce-Cottler

Committee Member: _____

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

May 2019

Abstract

There is an urgent need to discover new therapeutic targets to treat infections. The rise in antimicrobial resistance exacerbates this need. If left unmitigated, antimicrobial resistance is estimated to claim over 10 million lives worldwide by 2050. However, despite this rise in resistance, very few new antibiotics have been brought to market. In fact, over the past thirty years, there has been a lack of discovery of new antibiotic classes as a whole.

There are several reasons for this lack of antibiotic discovery, many of which are scientific bottlenecks. For instance, current target identification platforms require extensive screening and downstream follow-up experiments that are very time-consuming. Frequently, all that work leads to low success rates because they identify targets of unknown function, requiring even more experiments, or they identify targets that actually promote resistance. Additionally, current platforms use whole bacterial population approaches and fail to capture heterogeneous subpopulations with unique susceptibilities.

Metabolic network modeling is emerging as a powerful tool for antimicrobial target discovery to overcome these limitations. Genome-scale metabolic network reconstructions (or GENREs) serve as knowledge-bases for everything we know to-date about the metabolism of an organism. These reconstructions are also tools that allow us to study the genotype-to-phenotype relationship within a cell. Ultimately, using these models, we can

probe the capability of an organism in different environmental conditions. Importantly, we can use these models to identify essential processes for different organism objectives, such as growth or the production of metabolites of interest. By identifying these essential processes, we can suggest potential therapeutic targets.

In this work, I demonstrate that antimicrobial target discovery with metabolic network modeling overcomes challenges associated with current target identification platforms. Specifically, I show that metabolic network models (1) enable high-throughput target discovery, (2) delineate targets of known function, (3) determine targets that may mitigate resistance, and (4) identify targets for heterogeneous subpopulations. To do this, I applied a metabolic network model of the Gram-negative, multi-antimicrobial resistant pathogen *Pseudomonas aeruginosa* to antimicrobial target discovery in three different applications. In the first, I use the model to probe the interrelationship between growth and the synthesis of metabolites important for infection known as virulence factors (Chapter 2). In the second application, I reconcile conflicting high-throughput *in vitro* gene essentiality datasets and demonstrate the utility of contextualizing and interpreting these datasets with the model (Chapter 3). Finally, in the third, I generate condition-specific metabolic network models to identify targets for a specific subpopulation of bacteria, called persister cells, that is known to tolerate antimicrobial treatment (Chapter 4).

Together, this research demonstrates the unique ability of metabolic network modeling to facilitate the drug discovery pipeline and identify antibacterial targets that would be impossible to delineate without the use of computational models.

Acknowledgments

My graduate experience has been a formative one. I owe my success to my amazing support system who has encouraged me along the way and made me a better researcher, colleague, and friend.

To my advisor, Jason Papin – I have been very fortunate to have an advisor who encourages independent thinking, who pushes students to "just try it", who approaches science with optimism, and who is committed to developing the individual, not just the researcher. Thank you, Jason, for your endless support and your positivity.

To my various mentors – Shirley Masand, thank you for showing me that there's room for laughter in between experiments. Alejandro Amezquita and Aline Metris, thank you for your mentorship during my Unilever internship and support throughout my persister project. Glynis Kolling, thank you for all your help with experiments. Your jovial approach to science is refreshing and something I hope to emulate. To my thesis committee, thank you for your guidance throughout my PhD. Your feedback has made me a better researcher and a better communicator.

To my labmates, past and present – I am so fortunate to have worked in a lab where I can call my labmates my friends. Thank you all for the laughs, the criticism sandwiches, and the much-needed distractions. I will miss our lab lunches, movie nights, and dominating all the sys-bio competitions with you all!

To my sister, Karen Blazier – You have always been the best at balancing praising my accomplishments and keeping me humble. Thank you for humoring me when I talk science and for being a supportive outlet.

To my parents, Linda and Stephen Blazier - Thank you for being my biggest cheerleaders. I am eternally grateful to you both for everything that you have done and continue to do for me. Daddy, thank you for teaching me what it means to be an engineer. Your genuine curiosity about how the world works inspires me to be a better scientist everyday. Ma, thank you for encouraging me to pursue interests and passions outside of engineering. Your dedication to service and helping others inspires me to think beyond myself and focus my work on bettering society.

To my cat, Oliver – meow, meow, meow! MEOW! Kiss, kiss, kiss, kiss, kiss!

To my husband, Joshua - I cannot imagine life without you by my side. You are my favorite person to brainstorm with, my favorite person to vent to, and my favorite person to brag about. You are my best friend. Thank you for your encouragement, your patience, your advice, your laughs, and your love. I am so excited for our next adventure!

To my son, Samuel - With you in the world, the stars twinkle brighter, the birds sing louder, the cookies taste sweeter, and the hugs feel softer. Being your mother is the best and most rewarding job in the world. I hope to always inspire your curiosity, foster your creativity, and encourage your passions.

Contents

Abstract ..... **i**

Acknowledgments ..... **iii**

List of Figures ..... **x**

List of Tables ..... **xii**

Chapter 1. Background and Significance **1**

1.1 Introduction 1

1.2 Challenges of current target identification platforms 3

1.3 Metabolic network modeling is a useful tool 4

1.4 *Pseudomonas aeruginosa* is a model organism 5

1.5 Dissertation preview 6

1.6. References 9

Chapter 2. Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis **12**

2.1 Abstract 13

2.2 Introduction 14

2.3 Results 16

2.3.1 Metabolic network reconstruction of *P. aeruginosa* 16

2.3.2 Model Validation 19

2.3.3 Virulence associations of Tn-seq-based essential genes 21

2.3.4 Modeling VF production capabilities 25

2.3.5 Core set of growth-essential genes impact VFs 25

2.3.6 Interconnectivity of synthesis of pyoverdine and biomass 31

2.3.7 Experimental evaluation of pyoverdine mutants 32

2.4 Discussion 36

2.5 Methods ... 39
2.5.1 Metabolic network reconstruction ... 39
2.5.2 Model validation ... 43
2.5.3 Screen and database assembly ... 44
2.5.4 Prediction of virulence-related production versus growth ... 46
2.5.5 Network visualization ... 46
2.5.6 Strains and growth conditions ... 46
2.5.7 Pyoverdine assay ... 47
2.5.8 Data availability ... 47
2.6 Acknowledgments ... 47
2.7 Author Contributions ... 47
2.8 Conflicts of Interest ... 48
2.9 References ... 48
2.10 Supplementary Information ... 54

Chapter 3. Reconciling high-throughput gene essentiality data with metabolic network reconstructions ... 56
3.2 Introduction ... 58
3.3 Results ... 59
3.3.1 Comparison of candidate essential genes reveals variability across transposon mutagenesis screens ... 59
3.3.2 Contextualization of gene essentiality datasets using genome-scale metabolic network reconstructions ... 67
3.3.3 Quantitative evaluation of the impact of media formulation on condition-independent essential gene identification ... 79
3.4 Discussion ... 84
3.5 Methods ... 87
3.5.1 Data sources ... 87
3.5.2 Generation of candidate essential gene lists ... 88
3.5.3 Comparison of candidate essential gene lists ... 89
3.5.4 Re-analysis of transposon sequencing datasets ... 89
3.5.5 Model gene essentiality predictions ... 90
3.5.6 Subsystem assignment of consensus essential and nonessential genes ... 90

3.5.7 Flux sampling in LB and sputum 91
3.5.8 Media formulation impact on essentiality 91
3.5.9 Code and data availability 92
3.5.10 Computational resources 92
3.6 Acknowledgments 92
3.7 Author Contributions 93
3.8 Conflicts of Interests 93
3.9 References 93
3.10 Supplementary Information 97

Chapter 4. Metabolic network analysis of *Pseudomonas aeruginosa* persister cells .. 99
4.1 Abstract 100
4.2 Introduction 101
4.3 Results 103
4.2.1 Persister cells enable *Pseudomonas aeruginosa* to tolerate treatment with benzisothiazolinone 103
4.2.2 The transcriptional state of persister cells is distinct from untreated cells 104
4.2.3 The metabolite footprint of persister samples is different from untreated and dead samples 108
4.2.4 Condition-specific modeling of the persister and untreated states 115
4.2.5 Metabolic network analysis suggests targets for inhibiting persister viability 116
4.4 Discussion 121
4.5 Materials and Methods 124
4.5.1 Bacterial strains and growth conditions 124
4.5.2 Minimal inhibitory concentration assay 124
4.5.3 Persister isolation and re-growth 125
4.5.4 RNA-sequencing 126
4.5.5 Metabolomics profiling 126
4.5.6 Condition-specific model generation 127
4.5.7 Model prediction of persister targets 128
4.6 Acknowledgments 129
4.7 Author Contributions 129

4.8 Conflicts of Interest 129
4.9 References 130

Chapter 5. Reflections and Future Directions 134
5.1 Introduction 134
5.2 Contributions 134
5.2.1 High-throughput target discovery 135
5.2.2 Targets with known function 136
5.2.3 Targets that mitigate resistance 137
5.2.4 Targets of heterogeneous populations 138
5.2.5 *Pseudomonas aeruginosa* reconstructions 138
5.3 Future work 139
5.3.1 Target follow-through 140
5.3.2 Biomass objective function 141
5.3.3 Ensemble integration of omics data 144
5.3.4 Expansion to other organisms 145
5.4 My Perspective 147
5.5 References 148

List of Figures

Figure 1.1. The rise of resistance. 1
Figure 1.2. Number of antibacterial new drug applications approvals versus year intervals. 2
Figure 1.3. Antimicrobial target discovery with metabolic network models. 7

Figure 2.1. Network model characteristics. 17
Figure 2.2. Network model characteristics for mPAO1. 17
Figure 2.3. Comparison of experimental and computational single substrate source utilization for *P. aeruginosa* PA14. 19
Figure 2.4. Comparison of experimental and computational single substrate source utilization for *P. aeruginosa* PAO1. 20
Figure 2.5. Analysis of essential gene predictions compared to in vitro essentiality. 21
Figure 2.6. Visualization of experimental virulence-linked essential genes. 24
Figure 2.7. Genes essential for VF synthesis versus growth in SCFM. 26
Figure 2.8. VF synthesis and growth interconnectivity. 28
Figure 2.9. Pyoverdine synthesis capabilities in vitro on LB. 34

Figure 3.1. Comparison of candidate essential genes from transposon mutagenesis screens reveals variability. 61
Figure 3.2. Comparison of candidate essential genes from PAO1 LB transposon mutagenesis screens reveals variability across screens. 64
Figure 3.3. Comparison of candidate essential genes from LB transposon mutagenesis screens reveals variability across screens. 66
Figure 3.4. Contextualization of gene essentiality datasets using genome-scale metabolic network reconstructions. 69
Figure 3.5. Distribution of PAO1 consensus essential and nonessential genes across model subsystems. 73
Figure 3.6. Computational assessment of the impact of number of minimal media conditions considered on condition-independent essentiality. 80

Figure 3.7. Computational assessment of the impact of LB media composition on condition-independent essentiality. 82

Figure 4.1. Experimental and computational systems biology approach for characterizing the metabolism of *Pseudomonas aeruginosa* PA14 persister 102
Figure 4.2. Killing of *Pseudomonas aeruginosa* PA14 by benzisothiazolinone (BIT) over time. 103
Figure 4.3. RNA-sequencing analysis of *P. aeruginosa* persister cells. 105
Figure 4.4. Overlap of upregulated (A) and downregulated (B) genes across all four comparisons. 107
Figure 4.5. Metabolomics profiling of culture supernatants. 109
Figure 4.6. Metabolites that significantly change in abundance over time common to the untreated, persister, and dead conditions 111
Figure 4.7. Metabolites that significantly change in abundance over time unique to the untreated (A), persister (B), and dead (C) conditions. 113
Figure 4.8. Integrated metabolites for condition-specific models. 116
Figure 4.9. Time-kill assays for *P. aeruginosa* mutants. 118

Figure 5.1. Impact of biomass metabolite presence on gene essentiality predictions 143
Figure 5.2. Killing of *Burkholderia cenocepacia* by industrial antimicrobials. 146

List of Tables

Table 2.1. Small molecules associated with virulence accounted for in mPA14 and mPAO1. 18
Table 2.2. Comparison of experimental essentiality screens. 22

Table 3.1. Detailed description of in vitro transposon mutagenesis screens. 60
Table 3.2. Characteristics of the in vitro transposon mutagenesis screens. 60
Table 3.3. Percent accuracy between model predictions of essentiality and in vitro identified essential genes. 70
Table 3.4. Consensus metabolic essential and non-essential genes for PAO1 and PA14 media conditions with more than two screens. 72
Table 3.5. Discrepancies between model predicted essential genes and in vitro identified consensus nonessential genes for PAO1. 76

Chapter 1

Background and Significance

1.1 Introduction

There is an urgent need to discover new therapeutic targets to treat infections. The rise in antimicrobial resistance exacerbates this need. Resistance has been observed for nearly all antibiotics that have been developed (Figure 1.1) [1].

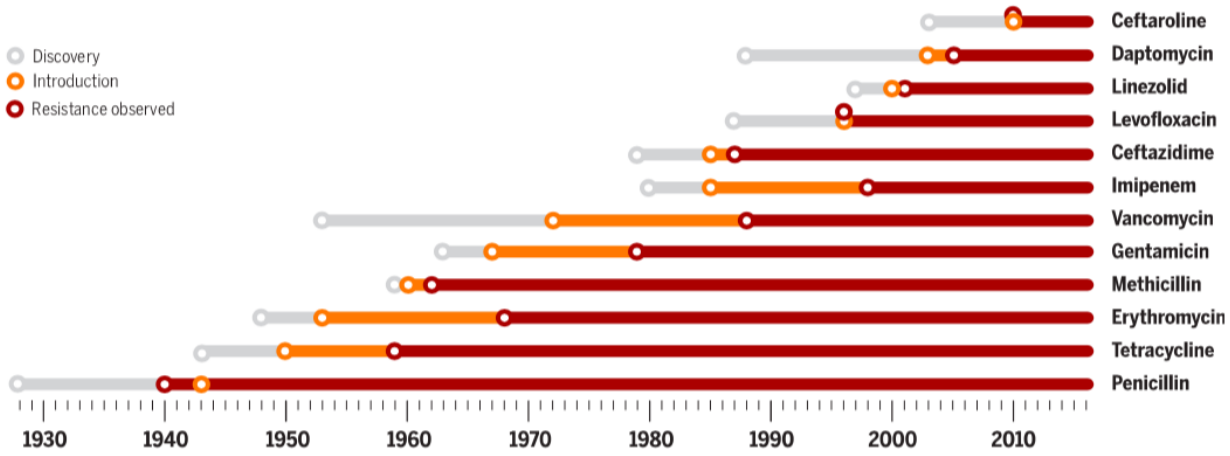


Figure 1.1. The rise of resistance.
Reprinted from [2], with permission from the American Association for the Advancement of Science.

Concerningly, multidrug resistant pathogens are becoming more common, giving rise to life-threatening infections that cannot be treated with our current antibiotic arsenal [3]. If left unmitigated, antimicrobial resistance is estimated to claim over 10 million lives worldwide by 2050 [4].

Despite this rise in resistance, very few new antibiotics have been brought to market (Figure 1.2) [1,5].

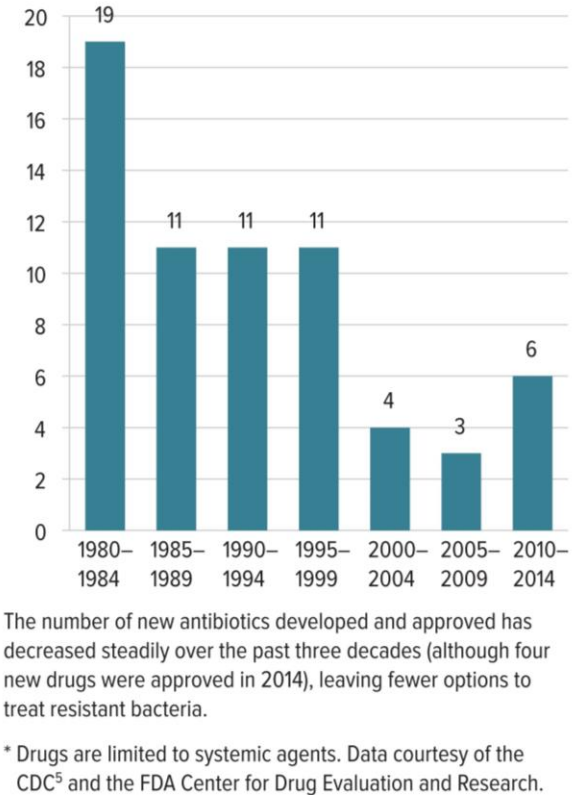


Figure 1.2. Number of antibacterial new drug applications approvals versus year intervals.
Reprinted from [1], with permission from MediMedia Managed Markets.

In fact, over the past 30 years, there has been a decline in the discovery and approval of new antibiotic classes as a whole [6,7]. Notably, in the past three years, two new classes of antibiotics have been discovered by academic groups but have not yet been brought to market [8,9]. Unfortunately, both new classes are only active against Gram-positive bacteria. Discovery and development of new antibiotics effective against the more problematic, multidrug resistant Gram-negative bacteria are lacking.

This dearth in antibiotic discovery can be attributed to several factors including regulatory hurdles, reduced economic incentives as well as scientific bottlenecks [1,7]. For example, the cost for Phase III Clinical Trials for an antibiotic (~$70 million) is prohibitive, considering that antibiotics are typically used for short periods and are often curative [1,5]. Furthermore, new antibiotics are typically prescribed as "last-line" defenses, used only after other antibiotics have failed. Additionally, the current approach of high-throughput screening of chemical libraries for antimicrobial activity is oftentimes unsuccessful because these compounds lack physicochemical properties unique to antibiotics, such as being less lipophilic [7]. Together, these hurdles have stalled the discovery and development of new antibiotic classes.

1.2 Challenges of current target identification platforms

Challenges associated with current antimicrobial target identification platforms contribute to the lack of discovery of new antibiotics. For instance, current target identification platforms require extensive screening and downstream follow-up experiments that are very time-consuming and resource intensive [10]. Oftentimes, these screens lead to low success rates because they identify targets of unknown function, requiring even more experiments to fully characterize their function and determine their 3D structure. Additionally, promising targets identified by these screens are regularly found to actually promote resistance rather than mitigate it [6,10,11]. Furthermore, current platforms use whole bacterial population approaches and fail to capture heterogeneous populations with unique susceptibilities. Finally, current strategies focus on targets essential for growth *in vitro*, failing to consider targets essential for growth *in vivo [11]*. Altogether,

these challenges have hindered the success of current target identification platforms to discover novel antimicrobial targets.

1.3 Metabolic network modeling is a useful tool

Metabolic network modeling is emerging as a powerful tool for antimicrobial target discovery to overcome these limitations [12]. Briefly, genome-scale metabolic network reconstructions (or GENREs) serve as knowledgebases for everything we know to-date about the metabolism of an organism or cell-type. Importantly, these reconstructions are also tools that allow us to study the genotype-to-phenotype relationship within a cell.

Assembled in part from annotated genomes as well as biochemical, genetic, and cell phenotype data, metabolic networks contain curated information about the known metabolites and metabolic reactions of a cell type [13]. An important aspect of these networks is that they account for the gene, protein, reaction relationships as well as physicochemical and thermodynamic constraints. By converting them into a mathematical model, we can use them to study the metabolic capabilities of an organism in a variety of conditions. Because of the size of these genome-scale reconstructions, which typically account for over a thousand genes, proteins, and reactions, computational tools are necessary for their analysis. One such tool is flux balance analysis (FBA), which is a constraint-based modeling approach. Using FBA, we can probe the capability of the model to convert inputs, such as carbon sources, into metabolites of interest through what is called an "objective function". For most cases, this objective function is a biomass reaction, serving as an approximation of growth. However, it can also be metabolites of interest, such as small molecule virulence factors or biomarkers. These reconstructions also serve as a framework

for contextualizing high-throughput, conflicting datasets. Through the integration of omics datasets with these reconstructions, condition-specific models can be generated, enabling the analysis of context-dependent metabolism [14].

Ultimately, using these models, we can probe the capability of an organism in different environmental conditions. Importantly, we can use these models to identify essential processes for different organism objectives, such as growth or the production of metabolites of interest. By identifying these essential processes, we can suggest potential therapeutic targets.

1.4 *Pseudomonas aeruginosa* is a model organism

In this work, I have applied metabolic network modeling to the bacterium *Pseudomonas aeruginosa*. *Pseudomonas aeruginosa* is a model organism for the identification of targets using GENREs because of its metabolic versatility, which allows it to thrive in diverse environments, such as in the soil [15], on medical devices like catheters [16], on human tissues as in the case of the lungs of cystic fibrosis patients [17], and in personal care products like shampoos [18]. Because of this versatility, *Pseudomonas aeruginosa* is a problem both in the clinic and in industry.

In the clinic, it is an opportunistic pathogen. It is the leading cause of hospital acquired infections and is best known for chronically infecting immunocompromised patients [19]. Furthermore, *P. aeruginosa* is notoriously multidrug resistant, leading to severe and recalcitrant infections with adverse clinical outcomes [20]. Given this threat to public health, the World Health Organization recently labelled *P. aeruginosa* as a “Priority 1: Critical”

pathogen [21], highlighting the urgent need for new antibiotics to treat *P. aeruginosa* infections.

In industry, *P. aeruginosa* is known to routinely contaminate personal care products and exhibit resistance to preservative treatment [22]. The presence of *P. aeruginosa* and other microbial pathogens in personal care products is a health risk for consumers [23,24]. Additionally, contamination by *P. aeruginosa* can lead to spoilage of the product through breakdown of the formulation, resulting in economic losses for the manufacturer [25]. To prevent microbiological contamination of personal care products, preservation systems are used; however, industry isolates of *P. aeruginosa* have exhibited resistance to preservative treatment [26], underscoring the need to identify new antimicrobial targets to increase the safety and utility of personal care products.

Because of its troublesome activity in both the clinic and in industry, there is a pressing need to discover new antimicrobial targets to treat *P. aeruginosa* infections and contamination. The metabolic versatility of *P. aeruginosa* makes it a prime candidate for antimicrobial target discovery with metabolic network modeling.

1.5 Dissertation preview

In this work, I demonstrate that antimicrobial target discovery with metabolic network modeling overcomes challenges associated with current target identification platforms. Specifically, I show that metabolic network models:

1. Enable high-throughput target discovery
2. Delineate targets of known function

3. Determine targets that mitigate resistance
4. Identify targets for heterogeneous subpopulations

To do this, I applied a metabolic network model of *Pseudomonas aeruginosa* to antimicrobial target discovery in three different applications (Figure 1.3).

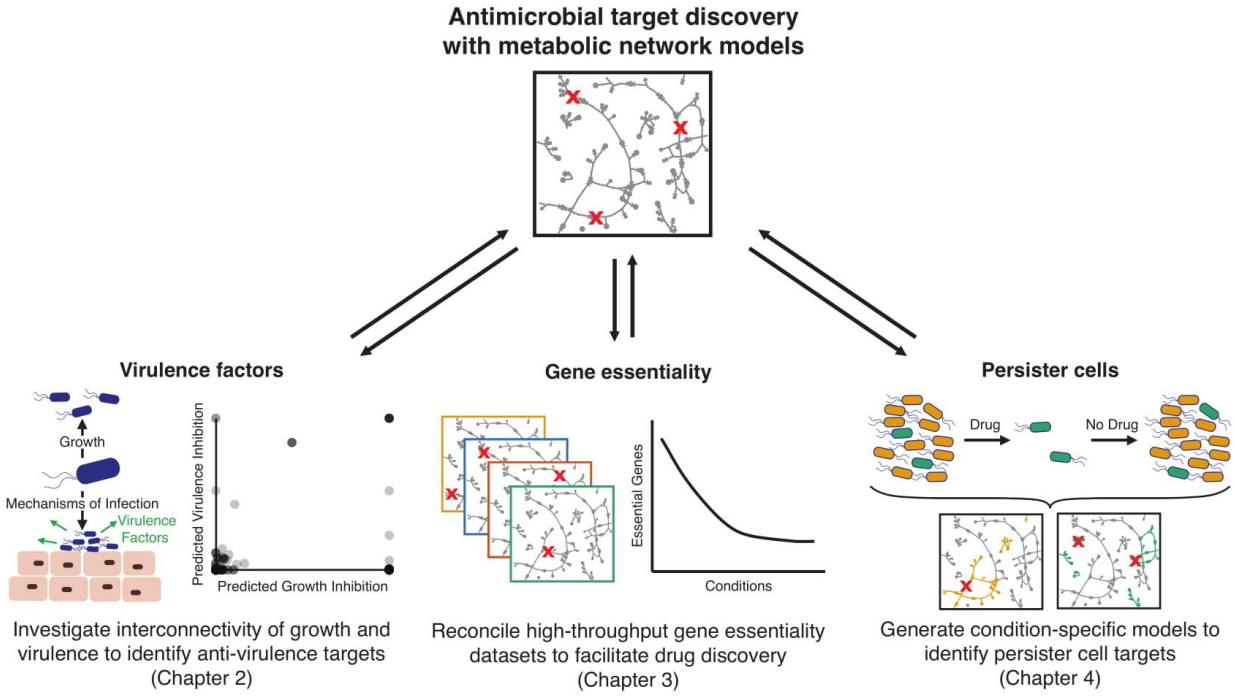


Figure 1.3. Antimicrobial target discovery with metabolic network models.
Graphical abstract for dissertation research.

In the first, I use the model to probe the interrelationship between growth and the synthesis of metabolites important for infection known as virulence factors (Chapter 2). Virulence factors are pathogen-produced molecules that promote the establishment of a pathogen within a host and enhance its potential to cause disease. These compounds can range from single metabolites, such as siderophores involved in iron-scavenging, to macromolecular structures, such as flagella involved in bacterial motility. In this work, I focused on small molecule virulence factors because of the tractability of reconstructing

their metabolic synthesis. There has been recent interest in targeting virulence as an alternative approach. It is thought that, by targeting virulence, cells might be less likely to develop resistance because virulence is not an essential cell process. However, there has not been much research into the relationship between virulence factor and growth. Given this uncertainty, I applied genome-scale metabolic network reconstructions to aid in teasing out the interconnectivity between growth in virulence.

In the second application, I reconcile conflicting high-throughput *in vitro* gene essentiality datasets and demonstrate the utility of contextualizing and interpreting these datasets with the model (Chapter 3). In this work, I define a gene as essential if disruption of the gene through transposon mutagenesis results in the inability of the mutant to grow. Genes essential for growth are promising antimicrobial targets. Transposon mutagenesis screens are the state-of-the-art for experimentally identifying genes essential for growth. However, these screens are subject to variability and are challenging to interpret, impeding our ability to identify therapeutic targets. Given these challenges, I applied genome-scale metabolic network reconstructions to reconcile essentiality data and assist in the antibiotic discovery pipeline.

Finally, in the third, I generate condition-specific metabolic network models to identify targets for a specific subpopulation of bacteria, called persister cells, that is known to tolerate antimicrobial treatment (Chapter 4). Traditionally, persister cells are thought to evade antimicrobial treatment due to a reduced metabolic state. However, the metabolism of persister cells is not well understood. Given this ambiguity, I built a metabolic network model of persister cell metabolism to aid in characterizing their metabolic state and suggest targets of persister viability.

Together, this research demonstrates the unique ability of metabolic network modeling to facilitate the antimicrobial pipeline. I conclude this dissertation with a discussion of how this work has contributed to the fields of antimicrobial target discovery and metabolic network modeling as well as point to areas of promising future research.

1.6. References

1. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P T. 2015;40: 277–283.
2. Kupferschmidt K. Resistance fighters. Science. 2016;352: 758–761.
3. Nikaido H. Multidrug resistance in bacteria. Annu Rev Biochem. 2009;78: 119–146.
4. Organization WH, Others. Antimicrobial resistance: global report on surveillance. World Health Organization; 2014.
5. Cooper MA, Shlaes D. Fix the antibiotics pipeline. Nature. 2011;472: 32.
6. Silver LL. Challenges of antibacterial discovery. Clin Microbiol Rev. 2011;24: 71–109.
7. Trusts PC. A Scientific Roadmap for Antibiotic Discovery. The Pew Charitable Trusts Philadelphia; 2016.
8. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, et al. A new antibiotic kills pathogens without detectable resistance. doi:10.1038/nature14098
9. Hover BM, Kim S-H, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nat Microbiol. 2018;3: 415–422.
10. Wright GD. Something old, something new: revisiting natural products in antibiotic drug discovery. Can J Microbiol. 2014;60: 147–154.
11. Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. Nature. 2016;529: 336–343.
12. Dunphy LJ, Papin JA. Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. Curr Opin Biotechnol. 2018;51: 70–79.
13. Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin J a. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol. Elsevier Ltd; 2012;20: 113–123.

14. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Frontiers in Computational Physiology And Medicine. 3. Available: http://www.frontiersin.org/Computational_Physiology_and_Medicine/10.3389/fphys.2012.00299/abstract

15. Green SK, Schroth MN, Cho JJ, Kominos SK, Vitanza-jack VB. Agricultural plants and soil as a reservoir for Pseudomonas aeruginosa. Appl Microbiol. 1974;28: 987–991.

16. Mittal R, Aggarwal S, Sharma S, Chhibber S, Harjai K. Urinary tract infections caused by Pseudomonas aeruginosa: a minireview. J Infect Public Health. 2009;2: 101–111.

17. Gellatly SL, Hancock REW. Pseudomonas aeruginosa: new insights into pathogenesis and host defenses. Pathog Dis. 2013;67: 159–173.

18. Neza E, Centini M. Microbiologically Contaminated and Over-Preserved Cosmetic Products According Rapex 2008–2014. Cosmet Toiletries. Multidisciplinary Digital Publishing Institute; 2016;3: 3.

19. Moradali MF, Ghods S, Rehm BHA. Pseudomonas aeruginosa Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. Front Cell Infect Microbiol. 2017;7: 39.

20. Aloush V, Navon-Venezia S, Seigman-Igra Y, Cabili S, Carmeli Y. Multidrug-resistant Pseudomonas aeruginosa: risk factors and clinical impact. Antimicrob Agents Chemother. 2006;50: 43–48.

21. World Health Organization. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. Geneva: World Health Organization. 2017;

22. Green AE, Amézquita A, Le Marc Y, Bull MJ, Connor TR, Mahenthiralingam E. The consistent differential expression of genetic pathways following exposure of an industrial Pseudomonas aeruginosa strain to preservatives and a laundry detergent formulation. FEMS Microbiol Lett. Oxford University Press; 2018;365: fny062.

23. Becks VE, Lorenzoni NM. Pseudomonas aeruginosa outbreak in a neonatal intensive care unit: a possible link to contaminated hand lotion. Am J Infect Control. 1995;23: 396–398.

24. Behravan J, Bazzaz F, Malaekeh P. Survey of bacteriological contamination of cosmetic creams in Iran (2000). Int J Dermatol. 2005;44: 482–485.

25. Smart R, Spooner DF. Microbiological spoilage in pharmaceuticals and cosmetics. J Soc Cosmet Chem. Citeseer; 1972;23: 721–737.

26. Weiser R. The resistance of Pseudomonas aeruginosa to preservatives used in industrial formulations [Internet]. phd, Cardiff University. 2015. Available: http://orca.cf.ac.uk/id/eprint/84672

Chapter 2

Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis

Anna S Blazier[2]†, Jennifer A Bartell[1,2]†, Phillip Yen[2], Juliane C Thøgersen[3], Lars Jelsbak[3], Joanna B Goldberg[4,5], Jason A Papin[2]

[1]Center for Biosustainability, Technical University of Denmark, 2970 Hørsholm, Denmark
[2]Biomedical Engineering, University of Virginia, Charlottesville, 22908 Virginia, USA
[3]Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[4]Department of Pediatrics, Division of Pulmonology, Allergy/Immunology, Cystic Fibrosis and Sleep, Children's Healthcare of Atlanta, Atlanta, GA, 30322, USA
[5]Emory + Children's Center for Cystic Fibrosis Research, Emory University and Children's Healthcare of Atlanta, Atlanta, GA, 30322, USA

† These authors contributed equally to this article.

The text included in this chapter has been published here:

Bartell JA, Blazier AS, Yen P, Thøgersen JC, Jelsbak L, Goldberg JB, Papin JA. (2017). Reconstruction of the metabolic network of Pseudomonas aeruginosa to interrogate virulence factor synthesis. *Nature Communications*. 8(14631). DOI: 10.1038/ncomms14631.

2.1 Abstract

Virulence-linked pathways in opportunistic pathogens are potential therapeutic targets that may be associated with less potential for resistance than targets in growth-essential pathways. However, efficacy of virulence-linked targets may be affected by the contribution of 'virulence' genes to metabolism. We evaluate the complex interrelationships between growth and virulence-linked pathways using a new genome-scale metabolic network reconstruction of *P. aeruginosa* strain PA14 and an updated, expanded reconstruction of *P. aeruginosa* strain PAO1. The PA14 reconstruction accounts for the activity of 112 virulence-linked genes and virulence factor synthesis pathways that produce 17 unique compounds. We integrate 8 published genome-scale mutant screens to validate gene essentiality predictions in rich media, contextualize intra-screen discrepancies, and evaluate virulence-linked gene distribution across essentiality datasets. Computational screening further elucidates interconnectivity between inhibition of virulence factor synthesis and growth. Successful validation of selected gene perturbations using PA14 transposon mutants demonstrates the utility of model-driven screening of therapeutic targets.

2.2 Introduction

There is a need for new drugs that effectively inhibit microbial infection while avoiding the development of resistance. Traditional antibiotics that kill bacteria by targeting growth-essential functions actively select for antibiotic-resistant mutants that overtake the infection. This growth-based selection promotes the rapid development of resistance and consequently exacerbates infections [1], resulting in substantial patient morbidity, mortality, and healthcare costs [2]. Inhibiting mechanisms of infection by targeting the synthesis of virulence factors and virulence-linked genes may be a promising new therapeutic strategy that avoids growth-based target selection, improves patient outcomes, and mitigates the spread of resistance [1,3,4]. However, genes that contribute to growth and genes that contribute to virulence are not necessarily distinct actors in an organism's genetic network; understanding the impact of genes on each pathogen directive (growth versus virulence) is critical to therapy design and prediction of resistance development.

Virulence-linked genes contribute to survival and fitness within a host. Many of these genes encode the synthesis pathways of virulence factors (VFs). Here, we focus on pathogen-produced small molecule virulence factors that enable adaptation to the host environment and enhance infection potential through such activities as iron sequestration and bacterial communication [5,6]. In targeting the synthesis of these metabolites, resistance may develop more slowly because of weakened selection pressure versus traditional targets that directly impact growth-essential catabolism of substrates or cell wall construction and repair [3]. However, our understanding of the role of virulence-linked genes is evolving [7] – significant links between virulence and pathogen metabolism are now emerging. For example, antibiotic pigments called phenazines enable opportunistic bacteria to combat the effects of

immune cell oxidative bursts, but these pigments may also induce rewiring of redox-linked pathways within the pathogen [8]. Furthermore, the production of virulence-linked compounds relies on essential components of central metabolism that connect substrate catabolism to VF synthesis pathways. A clear division between therapeutic targets impacting growth and virulence is therefore not straightforward [7]. We need to map the interconnectivity of these systems to identify genes that contribute to either or both systems, determine their function and essentiality in a clinically relevant environment, and estimate the impact of their inhibition on virulence versus growth.

To study the relationship between VF synthesis and growth from a systems level perspective, we used genome-scale metabolic network models (GEMs). Assembled from annotated genomic data, GEMs are mathematical frameworks that incorporate biochemical, genetic, and cell phenotypic data and account for hundreds to thousands of gene-protein-reaction (GPR) relationships and reaction stoichiometry and directionality [9]; they have been used to predict novel drug targets that inhibit growth [10] as well as probe the capability of an organism to synthesize various metabolites [11], including VFs [12].

Here, we present a new GEM of *Pseudomonas aeruginosa* strain PA14 (iPau1129) as well as an updated GEM of reference strain *P. aeruginosa* PAO1 (iPae1146). *P. aeruginosa* is a Gram-negative opportunistic pathogen capable of developing multi-drug antibiotic resistance, hospital-acquired infections [13–15]*, and* infections in cystic fibrosis patient lungs, burn wounds, and immunocompromised individuals. We validate our GEMs using substrate utilization data and gene essentiality screens from transposon mutant libraries and use six previously-published transposon sequencing (Tn-seq) screens to evaluate essential virulence-linked genes [16–18]. To study the relationship between VF production

and growth, we compare the effect of *in silico* gene knockouts on synthesis of biomass versus 17 VFs and identified genes uniquely critical for VF production, genes solely important for the synthesis of biomass, as well as genes involved in both VF production and biomass production. A case study of the VF pyoverdine shows the utility of GEMs in probing network dependencies that offer novel insights into links between virulence and metabolism that may enhance design of cycled or combination drug therapies as well as reduce the development of resistance.

2.3 Results

2.3.1 Metabolic network reconstruction of P. aeruginosa

Here, we present an updated GEM of *P. aeruginosa* strain PAO1 (iPae1146) as well as a new GEM of *P. aeruginosa* strain PA14 (iPau1129) (for ease of reference in this study, we refer to these reconstructions as mPAO1 and mPA14, respectively). The network reconstruction process began with previous *P. aeruginosa* PAO1 GEMs [19,20]. We implemented a more detailed biomass equation, incorporated new biological information, and curated the model against carbon source utilization and gene essentiality data (see below). We also assigned potential roles to 59 and 44 genes annotated as hypothetical proteins in PAO1 and PA14 genome annotations from the Pseudomonas Genome Database (PGD), respectively. In conclusion, the new GEM mPA14 accounts for the function of 1,129 genes, 1,495 reactions, and 1,286 metabolites while the updated GEM mPAO1 accounts for the function of 1,146 genes, 1,493 reactions, and 1,284 metabolites (Figure 2.1A).

A.

Organism		PAO1		PA14
Genome size		6.3 Mbp		6.5 Mbp
GC content		66.60%		66.30%
CDS		5716		5892

Model	iMO1056	iMO1086	mPAO1	mPA14
Genes	1056	1086	1146	1129
Unique EC identifiers	502	NA	674	674
Metabolites	760	1021	1284	1286
Cytoplasmic	760	916	1083	1085
Extracellular	NA	105	201	201
Reactions	883	1031	1493	1495
Metabolic	750	898	1083	1085
Transport	133	133	238	238
Exchange	NA	105	172	172
Gene-protein-reaction associations				
Single gene	839	839	800	796
Multi-gene			471	476
None	44	47	49	50

B.

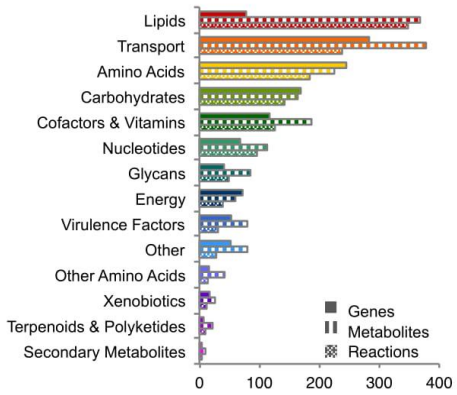


Figure 2.1. Network model characteristics.
(A) Properties of the updated PAO1 model as compared to previously published GEMs for *P. aeruginosa*, iMO1056 and iMO1096, as well as properties of the new PA14 model. (B) The number of genes, metabolites, and reactions in mPA14 grouped into functional categories as defined by KEGG [62]. For the distribution of genes, metabolites, and reactions in mPAO1, see Figure 2.2.

The distribution of genes, metabolites and reactions in mPA14 across a variety of KEGG functional categories is shown in Figure 2.1B (for the distribution of mPAO1, see Figure 2.2).

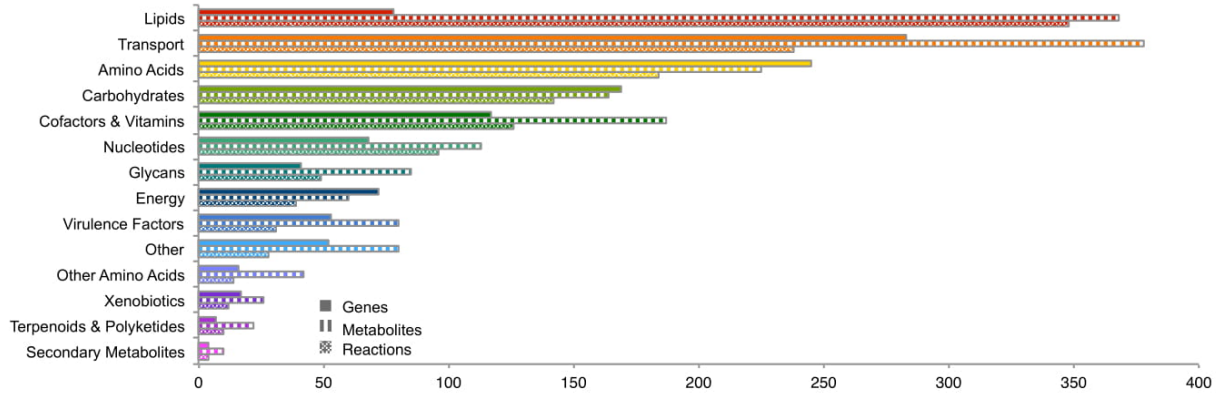


Figure 2.2. Network model characteristics for mPAO1.
The number of genes, metabolites, and reactions in mPA14 grouped into functional categories as defined by KEGG [62].

During curation, we specifically accounted for the synthesis pathways of several small molecule VFs. *P. aeruginosa* produces an array of VFs which can be grouped into several categories including exopolysaccharides, lipopolysaccharides, phenazines, quorum sensing

signal molecules, siderophores, and surfactant [21,22]. Table 2.1 lists the compounds that can be synthesized by mPA14 – the six italicized factors are new to mPAO1 and mPA14 compared to previous GEMs and are based on experimental evidence.

Category	Virulence Factor	Metabolite ID
Exopolysaccharide	Alginate	cpd17074[c]
Lipopolysaccharide	A-band-O-antigen	cpd17056[c]
	B-band-O-antigen	cpd17057[c]
	Lipid A	cpd17066[c]
Phenazines	Chorismate	cpd00216[c]
	1-Carboxyphenazine	*cpd17083[c]*
	Pyocyanin	cpd01206[c]
Quorum Sensing Signal Molecules	Acyl-homoserine-lactone (AHL)	cpd17082[c] and cpd08635[c]
	cis-2-Decenoic acid (DSF)	*cJB00127[c]*
	2-heptyl-4-quinolone (HHQ)	*cpd17078[c]*
	Pseudomonas quinolone signal (PQS)	cpd17085[c]
Siderophores	**Dihydroaeruginoic acid (Dha)**	**cJB00126[c]**
	Pyochelin	*cpd08828[c]*
	Pyoverdine	*cPY00164[c]*
	Salicylate	*cpd00599[c]*
Surfactant	Rhamnolipid	cpd17081[c] and cpd17080[c]

Table 2.1. Small molecules associated with virulence accounted for in mPA14 and mPAO1.
The six italicized factors are new additions to mPA14 and mPAO1 as compared to previous GEMs, and bolded dihydroaeruginoic acid is a recently identified PA14-specific VF included only in mPA14 [23]. The metabolite ID is the compound reference ID used in our models.

Bolded dihydroaeruginoic acid is a recently identified PA14-specific VF included only in mPA14 [23]. Additionally, we evaluated a list of 454 genes linked to virulence of PAO1, PA14, or both in the Virulence Factor Annotations tool from the recently updated PGD to identify model genes that are associated with virulence. Only 123 of these virulence-linked genes were annotated as part of a BRITE metabolic pathway by KEGG, and 49 of the 454 genes were annotated as hypothetical proteins. Using KEGG and PseudoCAP annotations (functional system annotations developed by the Pseudomonas Genome Database) as well as literature on VF synthesis, we focused on accounting for genes relevant to metabolism and virulence-linked synthesis pathways. Ultimately, there are 112 and 108 virulence-linked genes incorporated into mPAO1 and mPA14, respectively [24].

2.3.2 Model Validation

We used two data types to curate and validate the models: a carbon source utilization dataset and a published gene essentiality dataset. We generated the substrate utilization dataset using BIOLOG phenotype microarrays, which indicated whether PAO1 and PA14 were able to grow on particular carbon sources. We then compared these results to model predictions of biomass production (an approximation of growth) on different minimal media. After extensive transport reaction curation and refinement of metabolic pathways, mPA14 and mPAO1 account for 91 and 93 carbon sources and predict utilization with accuracies of 81% and 80%, respectively (Figure 2.3 and Figure 2.4, respectively).

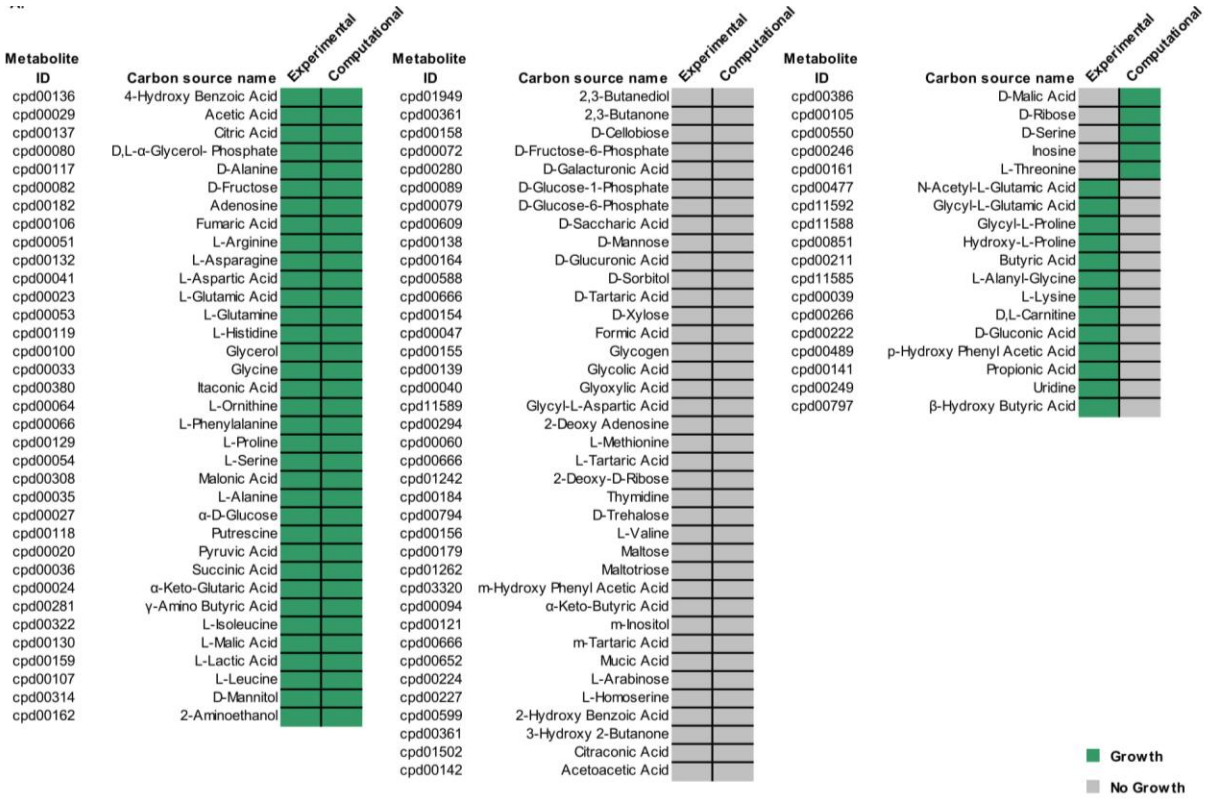


Figure 2.3. Comparison of experimental and computational single substrate source utilization for *P. aeruginosa* PA14.

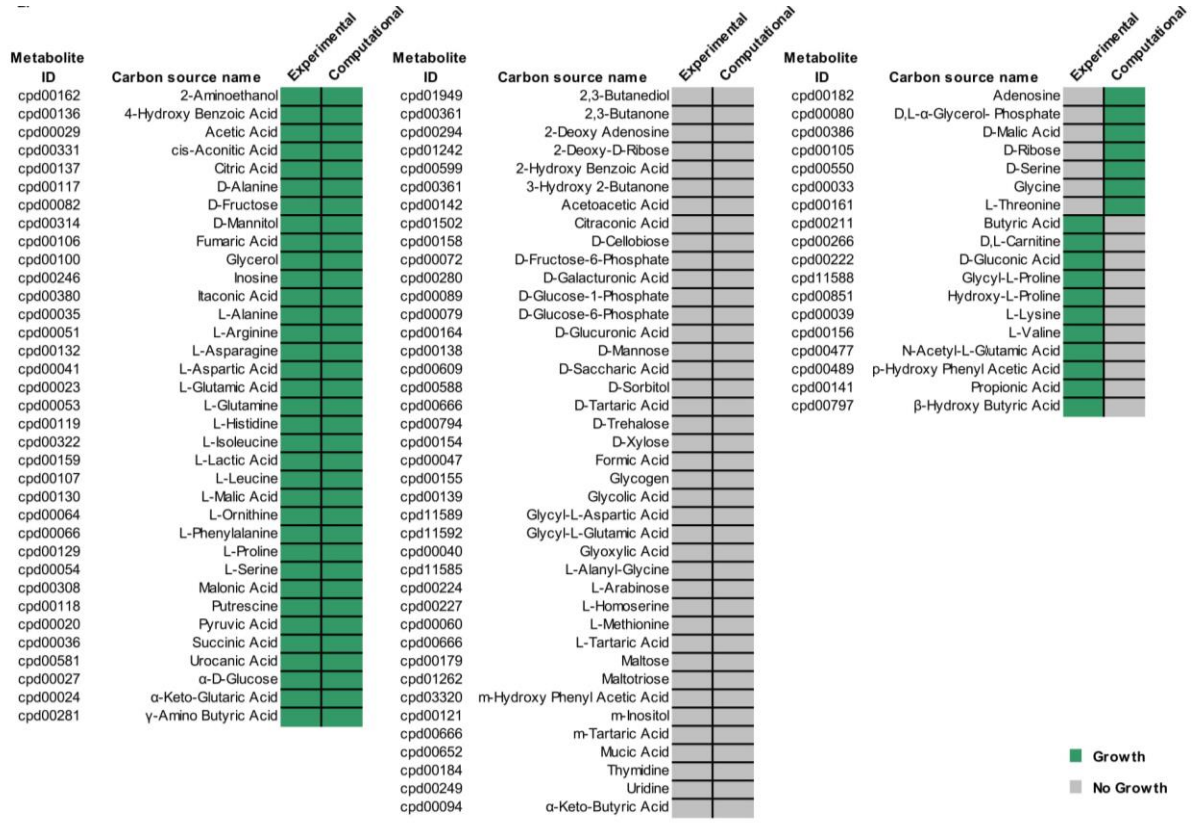


Figure 2.4. Comparison of experimental and computational single substrate source utilization for *P. aeruginosa* PAO1.

For the gene essentiality validation, we used a published dataset comprised of the overlap between the essential PAO1 genes identified in Jacobs et al. [25] and the essential PA14 genes identified in Liberati et al. [26] by creation of transposon insertion mutant libraries in a Luria-Bertani (LB) media background. We used this overlap dataset for our curation efforts instead of the individual libraries because there is a higher confidence in which genes are essential since they were not disrupted in either of the PAO1 and PA14 screens and the libraries are validated, publicly available, and created with established approaches in rich media. To compare this overlap dataset with our model predictions, we performed *in silico* single gene knockouts in our models and measured the subsequent

effects on biomass production. Ultimately, both mPA14 and mPAO1 can be used to predict gene essentiality with an accuracy of 91% (Figure 2.5).

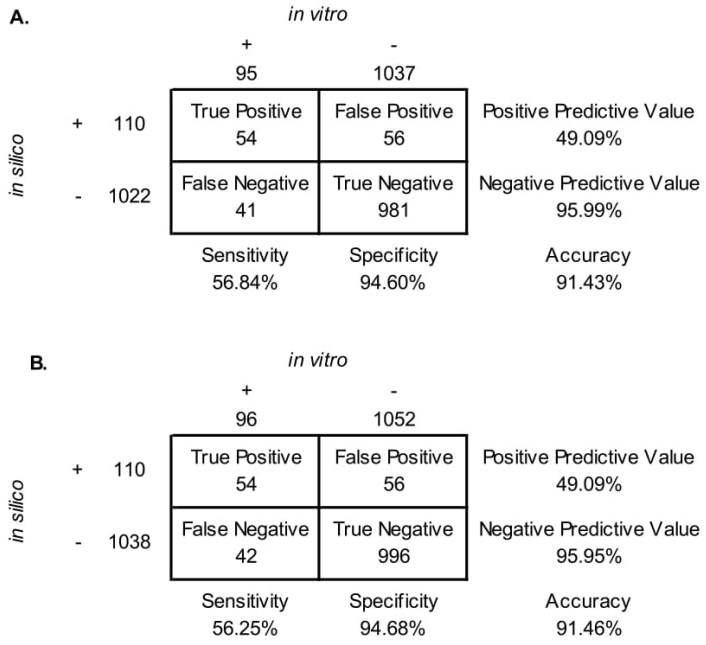


Figure 2.5. Analysis of essential gene predictions compared to in vitro essentiality. Overlap of essential genes derived from the PAO1 single transposon mutant library and PA14 single transposon mutant library were used. Both mPA14 (Panel A) and mPAO1 (Panel B) predict gene essentiality with an accuracy of 91%.

2.3.3 Virulence associations of Tn-seq-based essential genes

The recent advent of Tn-seq high-throughput screening has enabled the rapid evaluation of the fitness of a transposon insertion mutant in a given condition. These screens contribute important information regarding bacterial survival in specific contexts. Given our interest in studying the relationship between growth and virulence, we sought to determine how many essential genes were also virulence-linked in recent *P. aeruginosa* Tn-seq screens.

We obtained data from published Tn-seq screens for PAO1 and PA14 in several culturing conditions and identified essential genes for each individual screen. We then compared these individual essential gene lists to a list of either PAO1 or PA14 virulence-

linked genes from the PGD to identify virulence-linked essential genes for each screen (Table 2.2).

Strain	PA14		PAO1					
Reference	Pier	Whiteley	Manoil	Whiteley	Manoil	Whiteley	Manoil	Whiteley
Media	LB	Sputum	LB	LB	Sputum	Sputum	Pyruvate	Succinate
# Essential genes	634	510	201	336	224	445	182	641
# VF-linked essential genes	49	25	21	20	30	41	27	54
%VF-linked essential genes	7.73	4.90	10.45	5.95	13.39	9.21	14.84	8.42

Table 2.2. Comparison of experimental essentiality screens.
The number of essential genes, virulence-linked essential genes (VF-linked essential genes), and the proportion of essential genes that are virulence-linked for all eight of the transposon mutant screens analyzed. The data is assembled from studies by the Pier lab in 2013 [16] and the Whiteley lab [17] and Manoil lab [18] in 2015.

The number of virulence-linked essential genes identified across the screens ranged from 20 in the Whiteley PAO1 LB dataset [17] to 54 in the Whiteley PAO1 succinate dataset [17]. Furthermore, the proportion of essential genes that were also virulence-linked varied across the screens, ranging from 4.9% in the Whiteley PA14 sputum dataset [17] to 14.8% in the Manoil PAO1 pyruvate dataset [18]. This variability in the percentage of virulence-linked essential genes may stem from the variability in transposon insertion coverage of the individual screens. While some screens identified over 600 essential genes [16,17], other screens identified less than 200 essential genes [18] in the same media.

The moderate number of virulence-linked genes present in the Tn-seq screens can partially be explained by the lack of host selection pressure in the generation of the mutant libraries and the imperfect replication of *in vivo* growth conditions in *in vitro* studies. Tn-seq screens performed in infection models have demonstrated that mutants unable to synthesize certain VFs are unable to colonize the infection site [27], suggesting that virulence-linked genes may be essential in some contexts, while elsewhere (such as in liquid culture) they are unnecessary for bacterial fitness. However, using only Tn-seq screens that differ by growth

media rather than host selection pressure to contextualize virulence-linked gene essentiality still shows that some virulence-linked genes have important, potentially non-virulence related, functions. This analysis indicates that these genes may play a more focused metabolic role in the development of infection or are capable of dual functions linked to both virulence and growth.

To evaluate the potential overlap of virulence-linked genes with growth activity, we first used mPA14 as a framework to compare sets of growth essential genes and virulence-linked genes that have been curated as functionally relevant to metabolic activity using the Whiteley PA14 sputum screen [17]. Figure 2.6 shows the model reactions linked to 205 genes required for growth of PA14 in sputum (blue), and the 108 PA14 virulence-linked genes from the PGD (red).

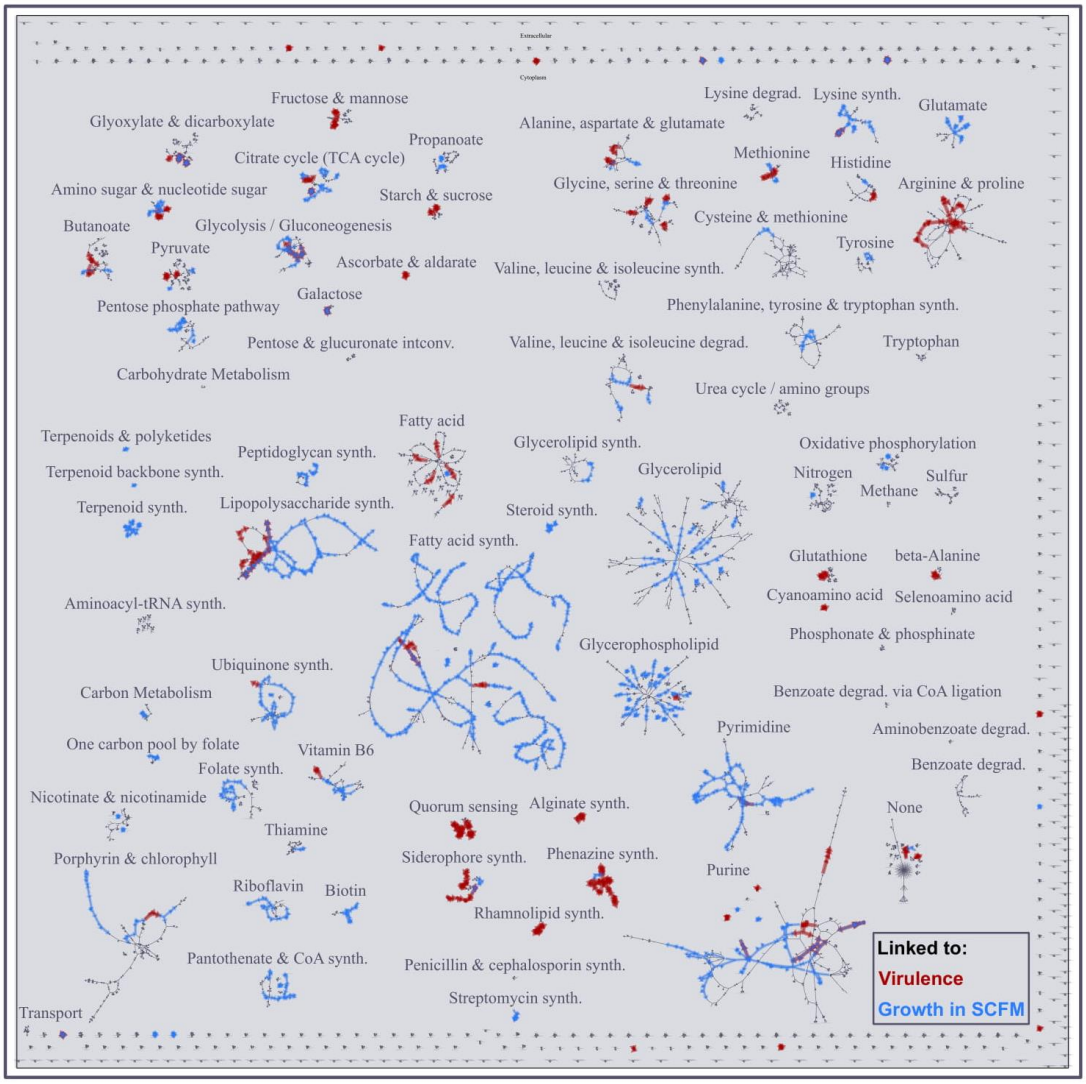


Figure 2.6. Visualization of experimental virulence-linked essential genes.
Distribution of virulence-linked genes and growth-essential genes from experiments in CF sputum visualized across all mPA14 reactions (gray) using MetDraw. Reactions associated with virulence-linked genes (as defined by the Pseudomonas Genome Database) are highlighted in red, and reactions associated with genes essential to growth in synthetic cystic fibrosis sputum are shown in blue. Purple reactions are associated with both virulence and growth essentiality. All reactions and metabolites are labeled with unique identifiers referenced in the model, visible at high magnification, and text-searchable.

The overlap between reactions associated with required genes and virulence-linked genes, totaling 21 reactions (11 genes) are linked to a broad array of systems and present at high density in central metabolic pathways, amino acids, lipids, and nucleotide metabolism

(overlap reactions in purple). Intriguingly, many reactions associated only with growth or virulence group together in the same pathways, which may indicate functional connections even if specific genes are not shared between the distinct gene sets. This analysis supports the need for a mechanistic evaluation of virulence-linked genes in the context of growth.

2.3.4 Modeling VF production capabilities

While infection-based Tn-seq screens have demonstrated that mutants incapable of VF synthesis lack the ability to infect, it is challenging to discern whether this occurs due to the inhibited gene's essentiality for the expression of virulence-linked compound(s), essentiality for growth, or essentiality for both [7,28]. To address this gap in knowledge, we employed genome-scale metabolic network modeling. We implemented a medium that mimics the lung of cystic fibrosis patients (Synthetic Cystic Fibrosis Medium, SCFM) in order to more closely model *in vivo* conditions [29]. The ability of *P. aeruginosa* to maintain decades-long infections in the lungs of cystic fibrosis patients may be due to both its metabolic adaptability and deployment of an array of VFs, such that pathway interconnectivity may proffer unique metabolic benefits as well as enable resistance to treatment [30]. Using an *in silico* SCFM medium, we performed *in silico* single-gene knockouts and assessed the levels of growth inhibition and VF synthesis inhibition by normalizing the resulting biomass flux and VF flux to wildtype production levels. By repeating this analysis for all 17 VFs in our model, we quantitatively compared the broad effects of simple genetic perturbations on the production of different VFs versus growth.

2.3.5 Core set of growth-essential genes impact VFs

To study the role of genes critical to both growth and VF synthesis, we compared the 116 genes predicted by mPA14 as essential for growth on SCFM to the genes essential for synthesis of VFs and found that 46 of the growth-essential genes are also essential for the production of at least one VF. These 46 genes critical to both biomass production and virulence are listed in Figure 2.7 with their PseudoCAP category and function and a heatmap showing the affected VFs.

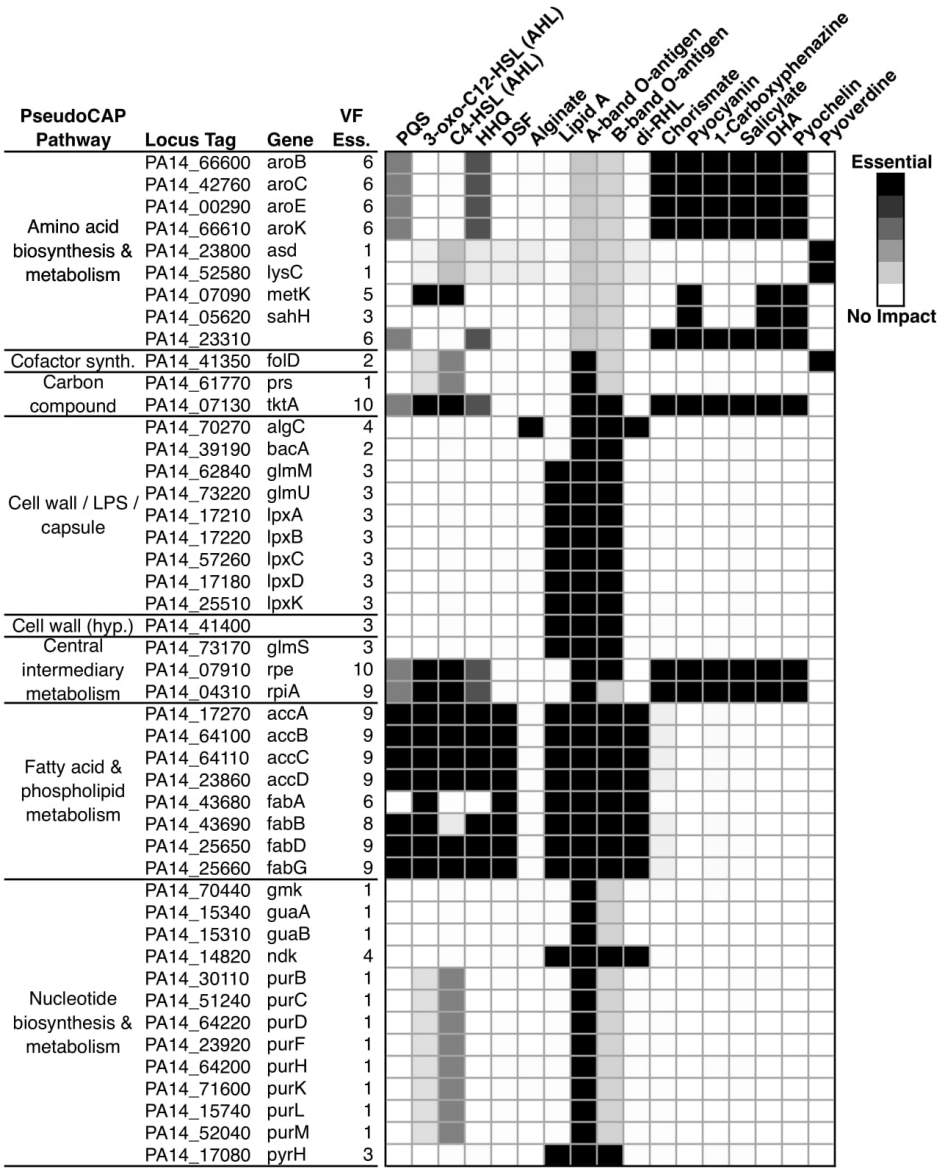


Figure 2.7. Genes essential for VF synthesis versus growth in SCFM.

The table lists the 46 genes essential for growth and production of at least one VF. Pathway assignment via PseudoCAP annotation and tabulated count of VFs for which the gene is essential are also included. Impact of a given gene's deletion is shown as white indicating 0% inhibition and black indicating 100% inhibition.

The PseudoCAP category critical for the largest number of VFs is fatty acid and phospholipid metabolism, with 7 genes predicted to be essential for the production of at least 8 VFs in addition to biomass production. Additionally, several *aro* operon genes contributing to aromatic amino acid synthesis are essential for the production of 6 VFs in the phenazine and siderophore families, while an array of genes involved in purine metabolism fully inhibit only the production of A-band-O-antigen. Ultimately, this analysis provides a novel list of genes ranked by their impact on virulence pathways in addition to growth inhibition, which may assist the design of therapeutics with broad impact on metabolic processes.

We then expanded our analysis to all genes in our model, plotting inhibition of each VF versus growth (Figure 2.8).

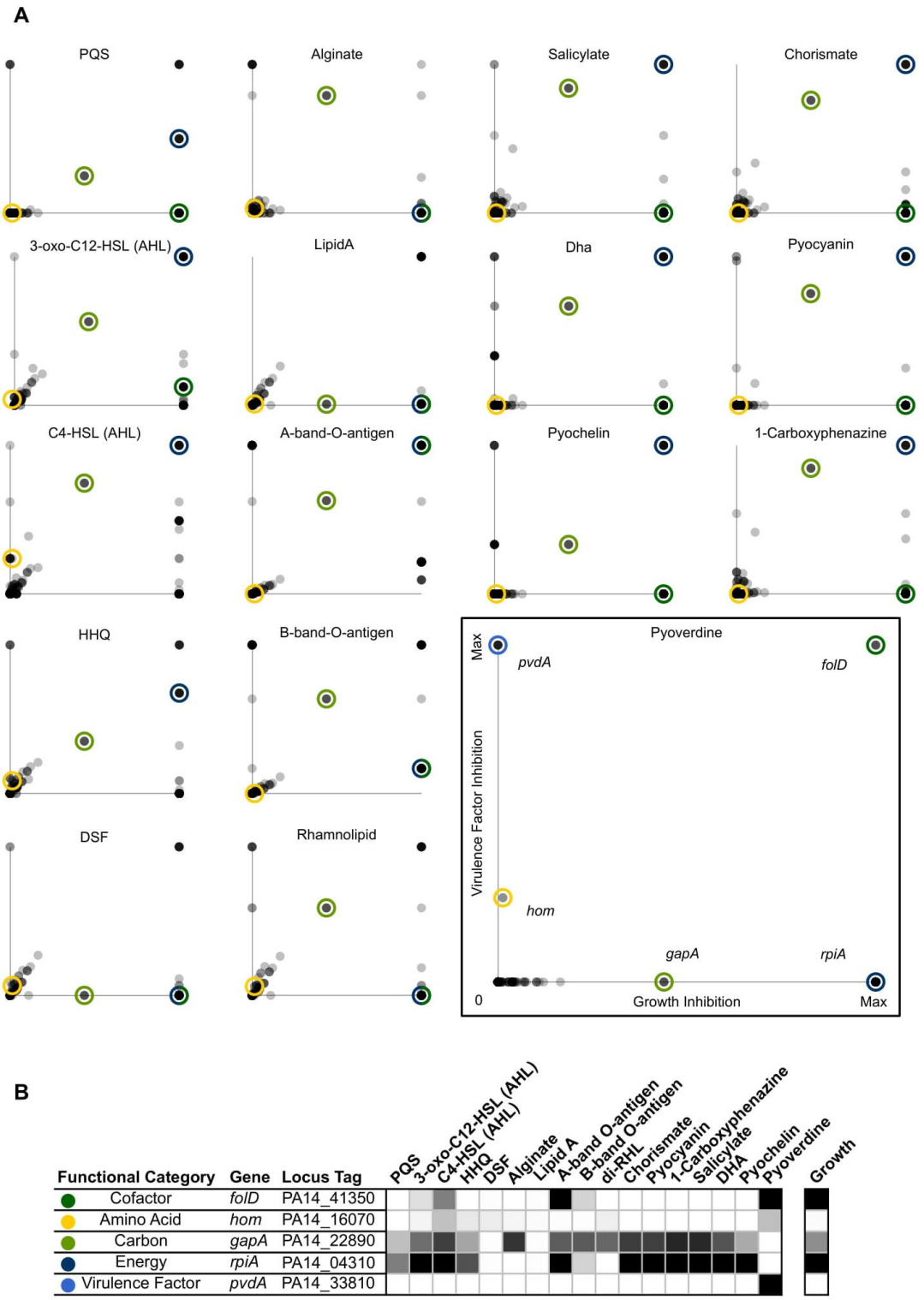


Figure 2.8. VF synthesis and growth interconnectivity.
(A) *In silico* gene knockouts were performed and the subsequent levels of biomass and VF production for each VF in the model were measured. The amount of growth inhibition was calculated by normalizing the knockout biomass production to the wild-type biomass production. Likewise, the amount of VF synthesis inhibition was calculated by normalizing

the mutant level of VF production to the wild-type level of VF production. Each point indicates the growth inhibition (x-axis) and VF inhibition (y-axis) relative to wild-type for a given *in silico* knockout. All data points are transparent such that a high density of data points results in an increase in color intensity. Colored circles are used to indicate genes of interest as labeled in the pyoverdine example with yellow representing genes involved in amino acid metabolism, green carbohydrate metabolism, dark blue energy metabolism, and light blue VF metabolism. (B) We highlight genes representing unique subtypes of impact on pyoverdine synthesis versus growth in a quantitative way that enables easy comparison of the activity of these genes across all VFs, with white indicating 0% inhibition and black indicating 100% inhibition.

Each point in the resulting plots indicates the level of growth inhibition (x-axis) and VF inhibition (y-axis) relative to wild-type for a given *in silico* knockout. All data points are transparent such that a high density of data points results in an increase in color intensity. Thus, the color intensity at the origin of the plots indicates a high number of gene deletions that have no effect on production of either biomass or the indicated VF. Data points in the upper right corner of each plot represent genes essential to both VF production and growth, while data points arrayed between axes indicate the degree of biased impact on growth versus VF production by a given knockout.

This analysis enabled the identification of non-obvious relationships between growth and VF production. Unsurprisingly, most gene knockouts resulted in marginal or no growth defects, as indicated by data point clusters near the origin along the x-axis. This result was mirrored for VF synthesis, with most gene knockouts also resulting in marginal or no VF production defects. We hypothesized that VF synthesis would be less robust to perturbation as compared to growth because these compounds rely on the catabolism of growth substrates prior to VF anabolism. We instead see that for several VFs, many genes essential for growth only partially inhibit synthesis when disrupted. The number of genes essential solely for the production of a given VF varies considerably, and is not always correlated with

the complexity of the synthesis pathway or final compound. These results highlight critical differences in the degree of interconnectivity of VF synthesis and biomass production across the VFs, which we can evaluate mechanistically through the use of our computational model.

VFs that are less sensitive to genetic perturbations than biomass production may have a high degree of redundancy in their synthesis pathways. For example, relatively few genes impact lipid A, chorismate, and 1-carboxyphenazine production without also impacting growth and no gene is essential solely for the production of the respective VF. Instead, genes that are essential for VF production are also essential for growth, thus indicating the high level of integration of VF synthesis with the overall metabolism of *P. aeruginosa*. While this integration of VF and biomass synthesis is expected for lipid A given its presence in the biomass reaction in the model as an essential component, this was a surprising result for chorismate and 1-carboxyphenazine as we do not consider these essential components. Upon closer network inspection, we find alternative pathways for the production of chorismate and multiple isozymes for the synthesis of 1-carboxyphenazine. Both instances highlight redundancies in the network that reduce the occurrence of predicted essential genes unique to these two VFs.

Interestingly, B-band-O-antigen and A-band-O-antigen demonstrate the other extreme – all growth essential genes also impact the synthesis of both O-antigens to some extent. This case highlights the dependency of the production of these O-antigens on some of the biomass components themselves, namely lipid A. Since the O-antigens rely on the production of lipid A, all the genes that inhibit the synthesis of lipid A (and, thus, biomass), also inhibit the synthesis of the O-antigens. This result indicates the importance of biomass function formulation; here, we retain a standard list of components for consistency with

other models, but more targeted analyses may improve upon addition and removal of less integral biomass components.

In contrast to these VFs, there were several VFs that displayed much less interconnectivity with the rest of the network. For example, relatively few genes maximally inhibit both growth and alginate production compared to the other VFs in the model. Most of the genes that are essential for alginate production have no impact on growth when the associated function is removed from the model, thus indicating that the synthesis pathway for alginate is less highly integrated into the metabolism of *P. aeruginosa*. VFs like alginate may be more peripheral to the general metabolic function of *P. aeruginosa* due to specialization. While i*n vivo* studies have highlighted the importance of these metabolites in maintaining infections [31–34], here we systematically demonstrate pathway independence from essential metabolic function.

2.3.6 Interconnectivity of synthesis of pyoverdine and biomass

In addition to studying the interconnectivity of individual VF synthesis pathways, this analysis also enabled the investigation of the role of individual genes. While the disruption of some gene functions results in a similar response across the VFs, other gene function disruptions produce a highly varied response, suggesting that these genes play a unique role in the synthesis of each VF (example genes circled in Figure 2.8). Using the VF pyoverdine as a reference, when the function of the gene *hom*, encoding for homoserine dehydrogenase, is removed, there is a very slight impact on growth and only marginal impact on each of the VFs, with pyoverdine synthesis demonstrating the most inhibition as a result of a *hom* knockout. Similarly, while a functional disruption of *pvdA*, which encodes for l-ornithine N5-

oxygenase, maximally inhibits pyoverdine production, it has no impact on synthesis of the other VFs accounted for in mPA14.

In contrast, other simulated gene knockouts have more varied impact on VF inhibition. Functional disruption of *gapA*, which encodes for glyceraldehyde-3-phosphate dehydrogenase, has varied impact on the synthesis of VFs in the network, illustrated by preservation of pyoverdine production but near maximal impact on salicylate production. Similarly, while functional disruption of *rpiA*, which encodes for ribose-5-phosphate isomerase A, again does not inhibit pyoverdine production, it does inhibit production of the AHLs incorporated into mPA14. Likewise, functionally disrupting *folD*, which encodes for 5,10-methylene-tetrahydrofolate cyclohydrolase, maximally inhibits pyoverdine production and has no impact on PQS production. Thus, we can tease out the role of different genes on the synthesis of different VFs, with some simulated gene knockouts demonstrating consistent levels of inhibition across all of the VFs and others demonstrating varied levels of inhibition.

2.3.7 Experimental evaluation of pyoverdine mutants

We chose to extend our investigation of the inhibition of pyoverdine synthesis because of the important role it has in iron scavenging and the tractability of experiments measuring pyoverdine production. In fluorescent Pseudomonads, pyoverdine is the main siderophore, a molecule that solubilizes iron for use by essential metabolic processes. It has been implicated in bacterial interactions in biofilms, it is essential for burn wound colonization, and it is upregulated in initial CF lung colonization [35–37]. Pyoverdine is also considered a 'public good' compound that is produced by select members of a community to

benefit the whole. Thus, the inhibition of pyoverdine synthesis within the small group of producer cells may affect the whole community while reducing the possibility of acquisition and spread of resistance genes [38,39].

To interrogate the relationship between growth and pyoverdine synthesis, we identified gene function disruptions with varied impact on pyoverdine synthesis and growth as shown by the circled points of Figure 2.8. We chose *pvdA* because it was predicted to be essential for pyoverdine production but not growth. Conversely, *rpiA* was chosen because it was predicted to be essential for growth but not for pyoverdine production. We chose *folD* because our model predicted it to be essential for both growth and pyoverdine production, and *hom* and *gapA* because of their predicted sub-inhibitory effects on pyoverdine production and growth, respectively. We then investigated the accuracy of these predictions with literature and experiments using available transposon mutants.

Mutants for both *folD* and *rpiA* were not present in the PA14 genome-wide transposon mutant library [26], suggesting that both these genes are indeed essential for growth of *P. aeruginosa*. Involved in the folate biosynthetic pathway, *folD* plays a critical upstream role in the synthesis of several compounds such as thymidine, purines and various amino acids. Studies have investigated *folD* as a potential therapeutic target to kill a variety of pathogens including *P. aeruginosa [40–42]*. Also important for purine synthesis, *rpiA* plays a critical role in the pentose phosphate pathway, converting D-Ribulose-5 to D-Ribose-5. Due to their growth essentiality, it is not feasible to study their role in VF synthesis experimentally – we instead use our computational model to offer unique insight. While mPA14 predicts that *rpiA* is not important in pyoverdine synthesis via a simulated knockout, it does predict that *folD* plays a crucial role, as evidenced by a simulated knockout resulting

in total inhibition of pyoverdine synthesis. An analysis of pyoverdine synthesis precursors that cannot be produced after an *in silico folD* knockout in mPA14 highlights N5-formyl-N5-hydroxy-L-ornithine as the missing metabolite. This metabolite is not included in the much longer list of missing metabolites including purines that prevent biomass formation by the model. Thus, while *folD* may be essential for growth because of its role in purine synthesis, it appears to be essential for pyoverdine synthesis because of its role in amino acid metabolism. Understanding the metabolic interconnectivity of these genes provides insight into their potential impact on multiple systems if targeted therapeutically and we are able to determine the role of growth essential genes in VF synthesis which would otherwise be intractable.

Using transposon mutants of *pvdA*, *hom*, and *gapA* from the PA14 genome-wide transposon mutant library [26], we performed absorbance-based assays of pyoverdine production and growth in SCFM as described in the methods. The extent of growth and pyoverdine production (normalized to growth) for wild-type PA14 and each mutant strain are shown in Figure 2.9.

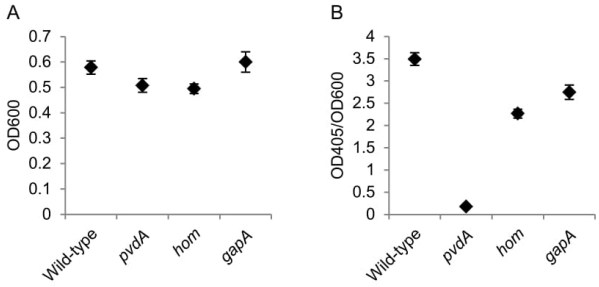


Figure 2.9. Pyoverdine synthesis capabilities in vitro on LB.
PA14 wild-type and *pvdA*, *hom*, and *gapA* PA14 mutants were grown to stationary phase in LB and the supernatants isolated. The OD405 of each condition's supernatant was then measured as a proxy for pyoverdine levels. The OD405 was divided by the OD600 of the culture in order to normalize for growth. Error bars indicate s.d. among five biological replicates.

As expected, the *pvdA* mutant showed markedly decreased pyoverdine production compared to wild-type, as it is an established pyoverdine assay control directly connected to the pyoverdine synthesis pathway. Interestingly, the *pvdA* mutant resulted in a minor growth defect relative to wild-type. This result could indicate that the lack of pyoverdine, and thus lack of access to iron, somewhat hindered the growth of the *pvdA* mutant. The *hom* mutant displayed a marginal growth defect and an approximately 1.5-fold decrease in pyoverdine production relative to wild-type. These results are consistent with model predictions that a *hom* knockout would result in slight growth inhibition and significant, but not total, inhibition of pyoverdine synthesis. Homoserine dehydrogenase, the gene product of *hom*, catalyzes the reaction converting L-homoserine to L-aspartate 4-semialdehyde which then gets converted to L-2,4-diaminobutryate, an important precursor of pyoverdine [43]. Thus, a mutated *hom* limits the production of L-aspartate 4-semialdehyde, creating a bottleneck in pyoverdine synthesis. Because, L-aspartate 4-semialdehyde is not a growth-essential metabolite, both *in vitro* and *in silico*, targeting it may specifically prevent pyoverdine production without strong growth-based resistance selection. Unlike the *pvdA* and *hom* mutants, the *gapA* mutant did not exhibit a growth defect, disagreeing with our model prediction of an approximately 50% reduction in growth. Additionally, while we predicted that a *gapA* knockout would not impact pyoverdine production, we observed that the *gapA* mutant did indeed reduce pyoverdine synthesis, albeit to a lesser extent than the *pvdA* and *hom* mutants. *gapA* encodes for the enzyme glyceraldehyde-3-phosphate dehydrogenase, which catalyzes the conversion of glyceraldehyde-3-phosphate to 1,3,-biphospho-D-glycerate, a central reaction in glycolysis. The discrepancy between our model predictions and experimental results for both growth and pyoverdine synthesis identifies a “gap” in our

knowledge regarding the function of glyceraldehyde-3-phosphate dehydrogenase in the overall metabolic network. Thus, we experimentally tested model predictions regarding genes in both growth and VF synthesis to tease out the role of genes upstream of pyoverdine synthesis and identify "gaps" in our current understanding of *P. aeruginosa* metabolism.

2.4 Discussion

We utilized a novel approach to systematically evaluate the contribution of metabolic genes to the synthesis of factors critical to virulence as well as growth at genome-scale using metabolic models. A new GEM for *P. aeruginosa* PA14 and an updated GEM for strain PAO1 were curated using single transposon insertion mutant data, virulence-linked gene databases, substrate utilization data, updated genome annotations, and recent literature. Using our PA14 model, we contextualized the PGD database of virulence-linked genes that were identified as essential for growth in Tn-seq screens, and then identified a core set of metabolic genes that were necessary for both growth and the synthesis of at least one VF. Subsequent analyses mapped the metabolic interconnections between growth and the synthesis of individual VFs, using a case study of pyoverdine to demonstrate model utility in teasing out the role of individual genes with regards to both growth and VF production. Our work enhances understanding of relationships between VF synthesis and growth, which is challenging to elucidate with experimental approaches. By quantifying the impact of genetic targets on growth versus virulence using a mechanistic model, we contribute novel insights for the design of therapeutic strategies that account for potential resistance development.

The rapid fitness screening enabled by transposon insertion mutants have produced valuable insights into gene function in different environments [44], but the genome-scale datasets can be difficult to interpret. Signature-tagged mutagenesis screens in infection

models have identified virulence-linked genes, but require a highly accurate replication of growth conditions *in vitro* for a baseline of effective comparison with infection [45,46]. Recently, Tn-seq has been used to measure *in vivo* gene fitness [44], but transposon insertion coverage, interpretation of Tn-seq results and essential gene identification are difficult to replicate across studies. Our model provides important nuance when considering the true definition of an 'essential' gene and survival fitness in varied environmental conditions; genes can be classified in a quantitative manner instead of a binary 'essential/nonessential' categorization. Thus, the high degree of variability in the number and roles of genes identified as critical for fitness even in different Tn-seq studies of the same growth environment can be elucidated when paired with mechanistic modeling. While we focused our efforts on mapping the distribution of virulence-linked genes in the datasets in an effort to understand their impact without the pressure of survival in a host, there is a rich opportunity to expand this comparison to the distribution of all metabolic genes in future work.

Our study provides an important expansion of genes to consider during study of VF synthesis during adaptation. We identify 46 genes as critical for the production of up to 10 of the 17 assessed VFs as well as biomass, which represent a novel core set of metabolic functions integral to the development of infection by *P. aeruginosa*. Long term adaptation may result in altered virulence capabilities due to accumulated mutations in these genes as well as regulatory genes and genes linked directly to product synthesis [47]. When we expand our analysis to all model genes, we can then group genes by their functional impact on growth, VF synthesis, or a combination of roles emphasizing a higher degree of connection. This novel analysis provides testable hypotheses regarding the contribution of

a gene to a given synthesis task. While we focus on experimental assessment of uniquely categorized genes in our pyoverdine analysis, this approach can be expanded to the other VFs. The analysis also demonstrates the flexibility of *P. aeruginosa* metabolism in a substrate-rich environment; the impact of competition for resources and the role of auxotrophy in evolving strains can be compared by predictions of differing optimal growth and production levels. Thus, further mapping of gene function may be enhanced by repeating the study in less complex and varied growth conditions.

Therapeutically targeting virulence-related pathways is an approach attracting much attention from a field struggling to find effective treatment for drug-resistant pathogens [1,48]. Quorum sensing inhibitors have been investigated through small molecule screening for a range of pathogens including *P. aeruginosa [49–51]* partially as a means of broad-spectrum anti-virulence treatment. A recent study showed that gallium-based quenching of extracellular siderophore activity successfully inhibited infection of caterpillars by *P. aeruginosa* while avoiding the development of resistance [52]. While inhibiting siderophore synthesis may increase resistance incidence in comparison to quenching, it will also enable pathogen-specific targeting of iron sequestration and avoid other systemic side effects (e.g., radiation with respect to gallium). While these VFs are regulated by quorum sensing molecules, related signaling networks are complex; more direct routes of inhibition provide an efficient avenue for precision treatment. This study provides curated sets of potential targets for diminishing or preventing the production of a large array of VFs. Reducing experimental costs and time to identify targets while simultaneously elucidating the underlying mechanisms by which targets inhibit infection are major contributions of our models to effective development of new therapies.

Our quantitative analysis of metabolic gene contribution to both growth and virulence is the most comprehensive genome-scale computational screen to date of virulence-related metabolism. Concerns regarding resistance to growth-targeting antibiotics in the context of multi-drug treatments may benefit from incorporation of new therapeutics that target VF synthesis [53–55]. However, new proposals regarding sequential cycling of drugs with different mechanisms of action in an attempt to avoid drug resistance may favor drugs that inhibit VF production and growth simultaneously to maximize impact [56,57]. Our mechanistic modeling approach allows us to predict the graded contribution of a given target gene to growth versus virulence systems to aid in these treatment designs. Ultimately, our updated models are valuable tools for quantitatively assessing relationships that would be challenging to interrogate experimentally at genome-scale. Our experimental validation of model predictions indicates that our approach provides testable hypotheses of gene function that can be used to elucidate critical interactions that may inform development of "resistance-resistant" therapeutics.

2.5 Methods

2.5.1 Metabolic network reconstruction

Previously published iterations of *P. aeruginosa* PAO1 GEMs iMO1056 [19] and iMO1086 [20] were both used as resources during reconstruction efforts. iMO1056 was created using field-standard syntax consistent with many models in the BiGG database, while iMO1086 was built using the ToBiN platform which is not currently available [19,20]. Since these original models were published, the modelSEED has become a favored draft reconstruction resource, and offers a comprehensive database of balanced reactions and

metabolites referenced from KEGG and MetaCyc from which hundreds of draft models have been created for use within the modeling community [58]. In light of this, we used a draft conversion of iMO1056 to model SEED format as the starting point for our GEM update to enable consistency with our past *P. aeruginosa* models, improve annotation of model reactions and metabolites (KEGG IDs, E.C. numbers, pathway assignments) and enable easy comparison with a large collection of models created and curated by other groups [12,59–61]. Because the conversion was an automated step performed by the modelSEED in an early iteration of the SEED database, manual curation was performed to add additional species-specific reactions that did not successfully convert from the original iMO1056 model or were present in iMO1086 as well as to correct conversion errors in reaction stoichiometry, directionality, and gene-protein-reaction (GPR) assignments. Further updates to SEED reactions and metabolite names using the modelSEED database were implemented to ensure consistency, and a KEGG subsystem assignment was added to each reaction when possible [62].

The genomic contents of *P. aeruginosa* PAO1 and *P. aeruginosa* PA14 and two closely related pathogens from the *Burkholderia cepacia* complex were compared to assist development of new, reconciled GEMs for each strain from previously built models. *P. aeruginosa* PA14 is a primary clinical isolate that is used as a model strain due to its substantial virulence in a variety of hosts, while *P. aeruginosa* PAO1, a wound isolate, is the main reference strain of this species [48]. We used *Burkholderia* species specifically because of their similarities to *Pseudomonas* as opportunistic Gram-negatives that also chronically infect cystic fibrosis patients and share similar virulence mechanisms. We also previously built and extensively curated GEMs for these species in modelSEED syntax as described

further below, making them useful resources. *P. aeruginosa* PA14, *P. aeruginosa* PAO1, *Burkholderia cenocepacia* J2315, and *Burkholderia multivorans* ATCC17616 were compared using genome-scale reciprocal BLASTP with an E-value cutoff of 0.01 with no low-complexity filter using CLC Main Workbench (CLC bio, Aarhus, Denmark). Hits with E-values below 1E-40 were considered high confidence hits and automatically matched. Genes with hits that received a higher E-value score were manually evaluated based on predicted function, gene descriptions, and PseudoCAP category (custom system/pathway annotations) on the Pseudomonas Genome Database (PGD) [24] before inclusion in the models in the few cases they were employed. There is a distinction, however, between confident gene matches between the organisms used, and utilization of genes annotated as hypothetical proteins in all species. We propose functions for a number of hypothetical proteins in the models, some of which are missing a specific functional annotation only in PA14 versus PAO1, and some of which are hypothetical proteins in both strains which we have utilized in the models based on functional domain associations and other predictions provided by PGD annotations, literature, and manual curation based on BLAST results against other species. Many of these hypothetical proteins are implemented in transport reactions, fatty acid and lipid pathways, and VF pathways. A table of these low-confidence gene assignments and hypothetical proteins to which we assigned functions in the model is provided in Supplementary Data 1.

Updated, strain-specific biomass formulas were created using a field standard approach that approximates the biomass composition by accounting for DNA, RNA, protein, cell wall components, lipids, and organism-specific compounds whose production is required for growth [9,12]. This effort expanded the number of components considered necessary for growth according to an improved biomass formulation and an updated search

of literature pertaining to *Pseudomonas* species. Additional *Pseudomonas*-specific requirements, such as preference for ubiquinone-9 versus ubiquinone-8 as a key cofactor in respiration [63], were implemented. More specific lipids were implemented using recent studies from literature and as enabled by the expanded lipid reactions used in modelSEED draft reconstructions. Specifically, while iMO1056 and iMO1086 accounted for simple representations of cardiolipin, phosphatidylethanolamine, phosphatidylglycerol, and phosphatidylserine, we implemented the specific saturated, unsaturated, and cyclopropane fatty acids making up the full lipid profile of *P. aeruginosa* as described in literature [64–67]. Details of the new biomass formulations can be found in Supplementary Datasets 7-8.

To fill gaps and improve predictions, additional model components were first derived from iMO1086 and recently published GEMs of *Burkholderia* species. We built on prior curation efforts while maintaining consistent modelSEED syntax to enable future cross-species comparisons and community modeling. If SEED reactions in *Burkholderia* models were not present in the new *Pseudomonas* SEED model, the high confidence BLASTP results were used in conjunction with the PGD and Burkholderia Genome Database [68] and literature to evaluate addition of these reactions. Many of the new reactions were added to increase the number of Biolog carbon sources accounted for in the *Pseudomonas* models (from only the PM1 substrate set to both PM1 and PM2a substrate sets); this effort was guided by previous work we performed for the highly catabolically flexible *Burkholderia [12]*. We also implemented new VF synthesis pathways using similar *Burkholderia* pathways as a guide. Other new reactions were added to expand lipid metabolism pathways using literature regarding *Pseudomonas*-specific lipid composition and the increased specificity of SEED reactions in this subsystem. Reactions implemented in the other well-curated SEED

model available during our build work, *B. subtilis* iBsu1103, as well as reactions included in the MetaCyc and MetRxn databases were also used as resources [59,69,70]. PAO1 and PA14 genes categorized as linked to virulence via data from experimental studies incorporated into the PGD v3 [24] were specifically evaluated for inclusion in the models to expand clinically-relevant functional prediction ability (Supplementary Data 2 and Methods – Screen and Database Assembly).

2.5.2 Model validation

Models were validated using new, comprehensive assessments of experimental data from genome-scale transposon libraries and carbon utilization screening. Similar data had been used with prior models, but unexpected discrepancies identified in comparisons between PAO1 and PA14 measurements motivated careful re-assessment of data sets and experimental confirmation of results.

Gene essentiality predictions were performed by *in silico* deletions of single genes while optimizing for production of biomass using flux balance analysis (FBA) via the COBRA Toolbox [71]. Predicted essential genes were compared with a list of genes that were not successfully targeted by transposon insertions in both genome-scale transposon insertion libraries of *P. aeruginosa* PAO1 [25] and *P. aeruginosa* PA14 [26]. By using genes lacking transposon insertions in both studies, which used different transposon systems and resulted in differing levels of insertion rate and genome coverage, we increased our confidence that these genes were truly essential for growth in rich media for *P. aeruginosa* strains. Curation with essentiality data resulted in improved prediction accuracy of gene essentiality via curated GPR relationships as well as the addition of new components to the biomass formula.

Single carbon source catabolic ability of the strains was predicted by providing a single carbon source and salts to the model via exchange constraints and optimizing for biomass production using FBA [12]. Carbon utilization data were compiled from literature for both PAO1 and PA14, but discrepancies between studies motivated us to perform our own growth screens for both strains using Biolog phenotype arrays PM1 and PM2a. Growth curve screens were performed in triplicate using a microplate reader with shaking at 37°C for 48 hours. Curves were evaluated to identify substrates enabling growth versus no growth [12]. Results guided specific curation of catabolic pathways and expansion of transport systems included in the model to improve prediction accuracy.

2.5.3 Screen and database assembly

Information on virulence-linked genes was compiled from the Pseudomonas Genome Database [24] (current as of February 2016) using the Annotations by Category tool that provides Virulence Factor Annotation lists for several strains. We used the lists for PAO1 and PA14, which provided 427 and 208 genes, respectively, which were culled by the PGD from experimental screens in many different infection models, the Virulence Factor Database, and the Victors database as indicated in Supplementary Data 2. The bias towards PAO1 is partly due to more screens and studies performed for PAO1 versus PA14 in the literature; however, 419 of these genes are present in both genomes. We assumed that many of the genes identified as virulence-linked in PAO1 could also be virulence-linked in PA14; however, virulence-linked genes truly active in only one strain would be of interest to track in future work; these genes must then have alternate functions in addition to a role in virulence. Nevertheless, building on the above assumption, we created a combined list of genes

associated with virulence that included any genes noted in either list which were present in both genomes to which we then added strain-specific virulence genes. The resulting lists included 432 and 441 plausible virulence-linked genes for PAO1 and PA14, respectively; the strain in which each gene was originally classified as a VF is also indicated as well as its presence in each model in Supplementary Data 2.

For the Tn-seq-based essential gene analysis, we obtained gene essentiality data from eight recently published Tn-seq screens for PA14 and PAO1 in a variety of culturing conditions. These screens are listed in Table 2.2 and are identified by the name of the paper's senior author (Pier [16], Whiteley [17], and Manoil [18]), strain, and media condition. For the PA14 Pier dataset [16], we used the essential genes identified in Table S1 of the original manuscript. Similarly, for the PAO1 and PA14 Whiteley datasets [17], we used the essential genes identified in Dataset S1 and Dataset S3 of the original manuscript, respectively. For the PAO1 Manoil datasets [18], we curated the "General essential genes" identified in Dataset S1 of the original manuscript to determine essential genes for each of the three media conditions studied: LB, sputum, and pyruvate. Specifically, we applied a cutoff such that if a mutant for a particular gene failed to be generated in at least one of the independent transposon mutant pools for a particular media condition, that gene was deemed essential for that media condition. This approach does not take into account the location of the transposon insertion and, thus, may miss some essential genes. In the end, we obtained eight unique lists of essential genes for either PAO1 or PA14 in different media conditions based on the Pier, Whiteley, and Manoil datasets. Once we obtained these lists of the essential genes identified in each screen, we compared them individually to the list of virulence-linked genes from the PGD database for either PA14 or PAO1 as appropriate. Genes that were in both a

particular screen's essential list and the virulence-linked list were categorized as virulence-linked essential genes for that particular screen.

2.5.4 Prediction of virulence-related production versus growth

VF production capacity was first evaluated by optimizing the flux through an artificial 'demand' reaction for each virulence-related metabolite. Single gene deletions were implemented by identifying reactions for which a given gene was essential via the model's Boolean relationships and then constraining the flux through each of these reactions to zero. The effect of each of these deletions was evaluated by predicting production levels of each VF and biomass separately; resulting production levels lower than 0.001 were categorized as completely inhibitory (i.e., the deleted gene is essential for production of that component). Production levels were normalized by maximum possible production of a component under wild-type conditions for comparison within VFs.

2.5.5 Network visualization

mPA14 was visualized using a command line implementation of MetDraw [72] that enables color overlay which was then edited in Inkscape (https://inkscape.org/en/).

2.5.6 Strains and growth conditions

Wild-type strains of *P. aeruginosa* PAO1 and PA14 and PA14 single gene knock-out mutants from the PA14 non-redundant genome-scale transposon library [26] were grown in Luria-Bertani (LB) media supplemented with 15 ug/ml gentamycin as necessary at 37°C with aeration for liquid cultures.

2.5.7 Pyoverdine assay

To measure pyoverdine production, strains were grown in synthetic cystic fibrosis media [29] for 24 hours in 50 ml flasks and the absorbance of culture supernatants was measured at 405 nm according to a previously published protocol [73]. All measurements were normalized to culture density as determined by the absorbance of the bacterial culture at 600 nm.

2.5.8 Data availability

The new metabolic network reconstructions for *P. aeruginosa* PAO1 and PA14, iPae1146 and iPau1129, respectively, are provided in spreadsheet format (Supplementary Data 3-4) that includes curation notes and SBML file format (Supplementary Data 5-6).

2.6 Acknowledgments

We thank Heath Damron, Mariette Barbier and Glynis Kolling for discussions about virulence factor assays as well as Edik Blais for discussions regarding modeling. Support for this project was provided by the National Institute of General Medical Sciences (GM088244 and GM108501) and the Whitaker Foundation.

2.7 Author Contributions

J.B., A. B., and J.P. conceived and designed the study. J.B. and A.B. reconstructed and curated the models with assistance from P.Y. and J.T. J.B. and A.B. conducted all simulations and analyses. A.B. performed the experiments. J.G. and L.J. provided commentary and revision for the manuscript written by J.B., A.B., and J.P.

2.8 Conflicts of Interest

We have no conflicts of interest to declare.

2.9 References

1. Allen RC, Popat R, Diggle SP, Brown SP. Targeting virulence: can we make evolution-proof drugs? Nat Rev Microbiol. Nature Publishing Group; 2014;12: 300–308.

2. Center for Disease Control. Antibiotic resistance threats in the United States, 2013. Centres for Disease Control and Prevention, US Department of Health and Human Services; 2013.

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

4. Rasko D., Sperandio V. Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov. Nature Publishing Group; 2010;9: 117–128.

5. Konings AF, Martin LW, Sharples KJ, Roddam LF, Latham R, Reid DW, et al. Pseudomonas aeruginosa uses multiple pathways to acquire iron during chronic infection in cystic fibrosis lungs. Infect Immun. 2013;81: 2697–2704.

6. Wilder CN, Diggle SP, Schuster M. Cooperation and cheating in Pseudomonas aeruginosa: the roles of the las, rhl and pqs quorum-sensing systems. ISME J. 2011;5: 1332–1343.

7. de Lorenzo V. Pseudomonas aeruginosa: the making of a pathogen. Environ Microbiol. 2015;17: 1–3.

8. Price-Whelan A, Dietrich LEP, Newman DK. Pyocyanin alters redox homeostasis and carbon flux through central metabolic pathways in Pseudomonas aeruginosa PA14. J Bacteriol. 2007;189: 6372–6381.

9. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. Nature Publishing Group; 2010;5: 93–121.

10. Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin J a. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol. Elsevier Ltd; 2012;20: 113–123.

11. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6: 1290–1307.

12. Bartell JA, Yen P, Varga JJ, Goldberg JB, Papin JA. Comparative metabolic systems analysis of pathogenic Burkholderia. J Bacteriol. 2014;196: 210–226.

13. Rodríguez-Rojas A, Oliver A, Blázquez J. Intrinsic and environmental mutagenesis drive diversification and persistence of Pseudomonas aeruginosa in chronic lung infections. J Infect Dis. 2012;205: 121–127.

14. Curcio D. Multidrug-resistant Gram-negative bacterial infections: are you ready for the challenge? Curr Clin Pharmacol. 2014;9: 27–38.

15. Falkinham JO 3rd, Hilborn ED, Arduino MJ, Pruden A, Edwards MA. Epidemiology and Ecology of Opportunistic Premise Plumbing Pathogens: Legionella pneumophila, Mycobacterium avium, and Pseudomonas aeruginosa. Environ Health Perspect. 2015;123: 749–758.

16. Skurnik D, Roux D, Aschard H, Cattoir V, Yoder-Himes D, Lory S, et al. A Comprehensive Analysis of In Vitro and In Vivo Genetic Fitness of Pseudomonas aeruginosa Using High-Throughput Sequencing of Transposon Libraries. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003582

17. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proceedings of the National Academy of Sciences. 2015; 201419677.

18. Lee S a., Gallagher L a., Thongdee M, Staudinger BJ, Lippman S, Singh PK, et al. General and condition-specific essential functions of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences. 2015; 201422186.

19. Oberhardt M a., Puchałka J, Fryer KE, Martins dos Santos V a. P, Papin J a. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. J Bacteriol. 2008;190: 2790–2803.

20. Oberhardt M a., Puchałka J, Martins dos Santos V a. P, Papin J a. Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. PLoS Comput Biol. 2011;7: e1001116.

21. Ballok AE, O'Toole GA. Pouring Salt on a Wound: Pseudomonas aeruginosa Virulence Factors Alter Na+ and Cl− Flux in the Lung. J Bacteriol. American Society for Microbiology Journals; 2013;195: 4013–4019.

22. Williams P, Cámara M. Quorum sensing and environmental adaptation in Pseudomonas aeruginosa: a tale of regulatory networks and multifunctional signal molecules. Curr Opin Microbiol. 2009;12: 182–191.

23. Maspoli A, Wenner N, Mislin GLA, Reimmann C. Functional analysis of pyochelin-

/enantiopyochelin-related genes from a pathogenicity island of Pseudomonas aeruginosa strain PA14. Biometals. 2014;27: 559–573.

24. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. Nucleic Acids Res. 2016;44: D646–53.

25. Jacobs M a., Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, et al. Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 2003;100: 14339–14344.

26. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, et al. An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A. 2006;103: 2833–2838.

27. Feinbaum RL, Urbach JM, Liberati NT, Djonovic S, Adonizio A, Carvunis A-R, et al. Genome-wide identification of Pseudomonas aeruginosa virulence-related genes using a Caenorhabditis elegans infection model. PLoS Pathog. 2012;8: e1002813.

28. Hilker R, Munder A, Klockgether J, Moran Losada P, Chouvarine P, Cramer N, et al. Interclonal gradient of virulence in the Pseudomonas aeruginosa pangenome from disease and environment. Environ Microbiol. 2014;17: 29–46.

29. Palmer KL, Aye LM, Whiteley M. Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum. J Bacteriol. 2007;189: 8079–8087.

30. Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Høiby N, et al. Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nat Rev Microbiol. Nature Publishing Group; 2012;10: 841–851.

31. Uehlinger S, Schwager S, Bernier SP, Riedel K, Nguyen DT, Sokol PA, et al. Identification of specific and universal virulence factors in Burkholderia cenocepacia strains by using multiple infection hosts. Infect Immun. 2009;77: 4102–4110.

32. Byrd MS, Pang B, Hong W, Waligora EA, Juneau RA, Armbruster CE, et al. Direct evaluation of Pseudomonas aeruginosa biofilm mediators in a chronic infection model. Infect Immun. 2011;79: 3087–3095.

33. Turner KH, Everett J, Trivedi U, Rumbaugh KP, Whiteley M. Requirements for Pseudomonas aeruginosa Acute Burn and Chronic Surgical Wound Infection. PLoS Genet. 2014;10. doi:10.1371/journal.pgen.1004518

34. Hunter RC, Klepac-Ceraj V, Lorenzi MM, Grotzinger H, Martin TR, Newman DK. Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. Am J Respir Cell Mol Biol. 2012;47: 738–745.

35. Yeterian E, Martin LW, Guillon L, Journet L, Lamont IL, Schalk IJ. Synthesis of the siderophore pyoverdine in Pseudomonas aeruginosa involves a periplasmic maturation. Amino Acids. 2010;38: 1447–1459.

36. Yang L, Nilsson M, Gjermansen M, Givskov M, Tolker-Nielsen T. Pyoverdine and PQS mediated subpopulation interactions involved in Pseudomonas aeruginosa biofilm formation. Mol Microbiol. 2009;74: 1380–1392.

37. Meyer JM, Neely A, Stintzi A, Georges C, Holder IA. Pyoverdin is essential for virulence of Pseudomonas aeruginosa. Infect Immun. 1996;64: 518–523.

38. Pepper JW. Defeating pathogen drug resistance: guidance from evolutionary theory. Evolution. 2008;62: 3185–3191.

39. Pepper JW. Drugs that target pathogen public goods are robust against evolved drug resistance. Evol Appl. 2012;5: 757–761.

40. Eadsforth TC, Gardiner M, Maluf FV, McElroy S, James D, Frearson J, et al. Assessment of Pseudomonas aeruginosa N5,N10-methylenetetrahydrofolate dehydrogenase-cyclohydrolase as a potential antibacterial drug target. PLoS One. 2012;7: e35973.

41. Eadsforth TC, Maluf FV, Hunter WN. Acinetobacter baumannii FolD ligand complexes--potent inhibitors of folate metabolism and a re-evaluation of the structure of LY 374571. FEBS J. Wiley Online Library; 2012;279: 4350–4360.

42. Eadsforth TC, Pinto A, Luciani R, Tamborini L, Cullia G, De Micheli C, et al. Characterization of 2,4-Diamino-6-oxo-1,6-dihydropyrimidin-5-yl Ureido Based Inhibitors of Trypanosoma brucei FolD and Testing for Antiparasitic Activity. J Med Chem. 2015;58: 7938–7948.

43. Vandenende CS, Vlasschaert M, Seah SYK. Functional characterization of an aminotransferase required for pyovierdine siderophore biosynthesis in Pseudomonas aeruginosa PAO1. J Bacteriol. 2004;186: 5596–5602.

44. Van Opijnen T, Camilli A. Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. Nat Rev Microbiol. Nature Publishing Group; 2013;11: 435–442.

45. Saenz HL, Dehio C. Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Curr Opin Microbiol. 2005;8: 612–619.

46. Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. Nat Rev Genet. 2006;7: 929–939.

47. Andersen SB, Marvig RL, Molin S, Krogh Johansen H, Griffin AS. Long-term social

dynamics drive loss of function in pathogenic bacteria. Proceedings of the National Academy of Sciences. 2015;112: 201508324.

48. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, et al. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. Genome Biol. 2006;7: R90.

49. Fletcher MP, Diggle SP, Cámara M, Williams P. Biosensor-based assays for PQS, HHQ and related 2-alkyl-4-quinolone quorum sensing signal molecules. Nat Protoc. 2007;2: 1254–1262.

50. Bjarnsholt T, van Gennip M, Jakobsen TH, Christensen LD, Jensen PØ, Givskov M. In vitro screens for quorum sensing inhibitors and in vivo confirmation of their effect. Nat Protoc. 2010;5: 282–293.

51. Valentine CD, Zhang H, Phuan P-W, Nguyen J, Verkman AS, Haggie PM. Small molecule screen yields inhibitors of pseudomonas homoserine lactone-induced host responses. Cell Microbiol. Wiley Online Library; 2014;16: 1–14.

52. Ross-Gillespie A, Weigert M, Brown SP, Kümmerli R. Gallium-mediated siderophore quenching as an evolutionarily robust antibacterial treatment. Evolution, medicine, and public health. 2014;2014: 18–29.

53. Chait R, Craney A, Kishony R. Antibiotic interactions that select against resistance. Nature. 2007;446: 668–671.

54. Torella JP, Chait R, Kishony R. Optimal drug synergy in antimicrobial treatments. PLoS Comput Biol. 2010;6: e1000796.

55. Ejim L, Farha MA, Falconer SB, Wildenhain J, Coombes BK, Tyers M, et al. Combinations of antibiotics and nonantibiotic drugs enhance antimicrobial efficacy. Nat Chem Biol. 2011;7: 348–350.

56. Imamovic L, Sommer MO a. Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. Sci Transl Med. 2013;5: 204ra132.

57. Kim S, Lieberman TD, Kishony R. Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. Proceedings of the National Academy of Sciences. 2014;111: 14494–14499.

58. Henry CS, DeJongh M, Best A a., Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. Nature Publishing Group; 2010;28: 977–982.

59. Henry CS, Zinner JF, Cohoon MP, Stevens RL. i Bsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. Genome Biol. BioMed Central;

2009;10: R69.

60. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, et al. Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun. 2011;2: 589.

61. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. Methods Mol Biol. 2013;985: 17–45.

62. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40: D109–14.

63. Yamada Y, Takinami-Nakamura H, Tahara Y, Oyaizu H, Komagata K. The ubiquinone systems in the strains of Pseudomonas species. J Gen Appl Microbiol. Applied Microbiology, Molecular and Cellular Biosciences Research Foundation; 1982;28: 7–12.

64. Zhu K, Choi K-H, Schweizer HP, Rock CO, Zhang Y-M. Two aerobic pathways for the formation of unsaturated fatty acids in Pseudomonas aeruginosa. Mol Microbiol. 2006;60: 260–273.

65. Hoffmann N, Steinbüchel A, Rehm BHA. The Pseudomonas aeruginosa phaG gene product is involved in the synthesis of polyhydroxyalkanoic acid consisting of medium-chain-length constituents from non-related carbon sources. FEMS Microbiol Lett. Blackwell Publishing Ltd Oxford, UK; 2000;184: 253–259.

66. Chao J, Wolfaardt GM, Arts MT. Characterization of Pseudomonas aeruginosa fatty acid profiles in biofilms and batch planktonic cultures. Can J Microbiol. 2010;56: 1028–1039.

67. Ramos J-L, Filloux A. Pseudomonas: Volume 5: A Model System in Biology. Springer Science & Business Media; 2007.

68. Winsor GL, Khaira B, Van Rossum T, Lo R, Whiteside MD, Brinkman FSL. The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. Bioinformatics. 2008;24: 2803–2804.

69. Caspi R, Altman T, Dale JM, Dreher K, Fulcher C a., Gilham F, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2010;38: D473–9.

70. Kumar A, Suthers PF, Maranas CD. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics. 2012;13: 6.

71. Becker S a., Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc. 2007;2: 727–738.

72. Jensen PA, Papin JA. MetDraw: automated visualization of genome-scale metabolic network reconstructions and high-throughput data. Bioinformatics. 2014;30: 1327–1328.

73. Imperi F, Tiburzi F, Visca P. Molecular basis of pyoverdine siderophore recycling in Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 2009;106: 20440–20445.

2.10 Supplementary Information

The supplemental files listed below can be found at the following website:

https://doi.org/10.18130/V3/Xf0AVD

Supplementary Data 1- Curation of putative proteins added to mPA14 and mPAO1.

We present information on every gene included in the models with hypothetical, putative, and probable annotations in the Pseudomonas Genome Database as well as genes which we explicitly use for an alternate function than that annotated. We show how these genes were implemented functionally in the model and include notes and any relevant literature supporting our curation decision.

Supplementary Data 2 – Integration of virulence-linked genes from the Pseudomonas Genome Database.

We present our reconciliation of the lists of virluence-linked genes from PAO1 versus PA14 from the Pseudomonas Genome Database. The Virulence-linked genes spreadsheet includes mainly data provided by the PGD for each gene including database or study source and infection model. We organized this information such that genes matched between strains are paired. We used both BLASTP, manual evaluation of similar operons, and literature in creating these pairings – genes missing in one strain genome are noted as 'not available'. The summary count information provides insight into the bias of virulence-linked annotations towards PAO1 versus PA14 despite nearly all genes being duplicated in each genome.

Supplementary Data 3 – Genome-scale metabolic network model of Pseudomonas aeruginosa PA14, iPau1129, in spreadsheet format.

Our model accounts for the function of 1129 genes and is provided in spreadsheet format consistent with usage of COBRA Toolbox 2.0 using SEED notation for reactions and metabolites.

Supplementary Data 4 – Genome-scale metabolic network model of Pseudomonas aeruginosa PAO1, iPae1146, in spreadsheet format.

Our model accounts for the function of 1146 genes and is provided in spreadsheet format consistent with usage of COBRA Toolbox 2.0 using SEED notation for reactions and metabolites.

Supplementary Data 5 – Genome-scale metabolic network model of Pseudomonas aeruginosa PA14, iPau1129, in SBML format.

Our model accounts for the function of 1129 genes and is provided in SBML format consistent with usage of COBRA Toolbox 2.0 using SEED notation for reactions and metabolites.

Supplementary Data 6 – Genome-scale metabolic network model of Pseudomonas aeruginosa PAO1, iPae1146, in SBML format.

Our model accounts for the function of 1146 genes and is provided in SBML format consistent with usage of COBRA Toolbox 2.0 using SEED notation for reactions and metabolites.

Supplementary Data 7 – Biomass formulation for genome-scale metabolic network model of Pseudomonas aeruginosa PA14, iPau1129.

Supplementary Data 8 – Biomass formulation for genome-scale metabolic network model of Pseudomonas aeruginosa PAO1, iPae1146.

Chapter 3

Reconciling high-throughput gene essentiality data with metabolic network reconstructions

Anna S. Blazier[1], Jason A. Papin[1,2,3]
[1]Biomedical Engineering
[2]Medicine, Infectious Diseases & International Health
[3]Biochemistry & Molecular Genetics
University of Virginia; Charlottesville, Virginia, 22908; USA

The text included in this chapter has been published here:

Blazier AS, Papin JA. (2018). Reconciling high-through gene essentiality data with metabolic network reconstructions. *BioRxiv*. DOI: 10.1101/415448.

At the writing of this dissertation, this manuscript is currently under revision at *PLOS Computational Biology*.

3.1 Abstract

The identification of genes essential for bacterial growth and survival represents a promising strategy for the discovery of antimicrobial targets. Essential genes can be identified on a genome-scale using transposon mutagenesis approaches; however, variability between screens and challenges with interpretation of essentiality data hinder the identification of both condition-independent and condition-dependent essential genes. To illustrate the scope of these challenges, we perform a large-scale comparison of multiple published *Pseudomonas aeruginosa* gene essentiality datasets, revealing substantial differences between the screens. We then contextualize essentiality using genome-scale metabolic network reconstructions and demonstrate the utility of this approach in providing functional explanations for essentiality and reconciling differences between screens. Genome-scale metabolic network reconstructions also enable a high-throughput, quantitative analysis to assess the impact of media conditions on the identification of condition-independent essential genes. Our computational model-driven analysis provides mechanistic insight into essentiality and contributes novel insights for design of future gene essentiality screens and the identification of core metabolic processes.

3.2 Introduction

With the rise of antibiotic resistance, there is a growing need to discover new therapeutic targets to treat bacterial infections. One attractive strategy is to target genes that are essential for growth and survival [1–4]. Discovery of such genes has been a long-standing interest, and advances in transposon mutagenesis combined with high-throughput sequencing have enabled their identification on a genome-scale. Transposon mutagenesis screens have been used to discriminate between *in vivo* and *in vitro* essential genes [1,5], discover genes uniquely required at different infection sites [6], and assess the impact of co-infection on gene essentiality status [7]. However, nuanced differences in experimental methods and data analysis can lead to variable essentiality calls between screens and hamper the identification of essential genes with high-confidence [8,9]. Additionally, a central challenge of these screens is in interpreting why a gene is or is not essential in a given condition, hindering the identification of promising drug targets.

These data are often used to validate and curate genome-scale metabolic network reconstructions (GENREs) [10,11]. GENREs are knowledgebases that capture the genotype-to-phenotype relationship by accounting for all the known metabolic genes and associated reactions within an organism of interest. These reconstructions can be converted into mathematical models and subsequently used to probe the metabolic capabilities of an organism or cell type in a wide range of conditions. GENREs of human pathogens have been used to discover novel drug targets [12], determine metabolic constraints on the development of antibiotic resistance [13], and identify metabolic determinants of virulence [14]. Importantly, GENREs can be used to assess gene essentiality by simulating gene

knockouts. Through *in silico* gene essentiality analysis, GENREs can be useful in the systematic comparison of gene essentiality datasets.

Here, we perform the first large-scale, comprehensive comparison and reconciliation of multiple gene essentiality screens and contextualize these datasets using genome-scale metabolic network reconstructions. We apply this framework to the Gram-negative, multi-drug resistant pathogen *Pseudomonas aeruginosa*, using several published transposon mutagenesis screens performed in various media conditions and the recently published GENREs for strains PAO1 and PA14. We demonstrate the utility of interpreting transposon mutagenesis screens with GENREs by providing functional explanations for essentiality, resolving differences between the screens, and highlighting gaps in our knowledge of *P. aeruginosa* metabolism. Finally, we perform a high-throughput, quantitative analysis to assess the impact of media conditions on identification of core essential genes. This work demonstrates how genome-scale metabolic network reconstructions can help interpret gene essentiality data and guide future experiments to further enable the identification of essential genes with high-confidence.

3.3 Results

3.3.1 Comparison of candidate essential genes reveals variability across transposon mutagenesis screens

We obtained data from several published transposon mutagenesis screens for *P. aeruginosa* strains PAO1 and PA14 in various media conditions and determined candidate essential genes for each screen as described in Methods (Table 3.1) [15–19].

Screen	Strain	Media	Number of Essential Genes	Publication of analyzed datasets	Initial publication of dataset	Publication of initial insertion library	Experimental overview to generate initial insertion library	Analysis overview
PAO1.LB.913	PAO1	LB	913	Jacobs et al., 2003	Jacobs et al., 2003	Jacobs et al., 2003	Random transposon insertion mutagenesis with ISphoA and ISlacZ	Manual tally of recovered genes
PAO1.LB.201	PAO1	LB	201	Lee et al., 2015	Lee et al., 2015	Lee et al., 2015	Tn-seq circle method with Tn5-based transposon T8	Number of transposon insertions per gene compared to a normal distribution
PAO1.LB.335	PAO1	LB	335	Turner et al., 2015	Gallagher et al., 2011	Gallagher et al., 2011	Tn-seq circle method with Tn5-based transposon T8	Number of transposon insertions per gene compared to Monte Carlo simulated data
PAO1.Sputum.224	PAO1	Sputum	224	Lee et al., 2015	Lee et al., 2015	Lee et al., 2015	Tn-seq circle method with Tn5-based transposon T8	Number of transposon insertions per gene compared to a normal distribution
PAO1.Sputum.405	PAO1	Sputum	405	Turner et al., 2015	Turner et al., 2015	Gallagher et al., 2011	Tn-seq circle method with Tn5-based transposon T8	Number of transposon insertions per gene compared to Monte Carlo simulated data
PAO1.Pyruvate.179	PAO1	Pyruvate minimal media	179	Lee et al., 2015	Lee et al., 2015	Lee et al., 2015	Tn-seq circle method with Tn5-based transposon T8	Number of transposon insertions per gene compared to a normal distribution
PAO1.Succinate.640	PAO1	Succinate minimal media	640	Turner et al., 2015	Turner et al., 2014	Gallagher et al., 2011	PCR-based Tn-seq method with Tn5-based transposon T8	Number of transposon insertions per gene compared to Monte Carlo simulated data
PA14.LB.1544	PA14	LB	1544	Liberati et al., 2006	Liberati et al., 2006	Liberati et al., 2006	Random transposon insertion mutagenesis with mariner family transposon	Manual tally of recovered genes
PA14.LB.634	PA14	LB	634	Skurnik et al., 2013	Skurnik et al., 2013	Skurnik et al., 2013	INSeq method with mariner farmily transposon	Fold change between reads per kilobase per million reads
PA14.Sputum.510	PA14	Sputum	510	Turner et al., 2015	Turner et al., 2015	Turner et al., 2015	PCR-based Tn-seq method with Tn5-based transposon T8	Number of transposon insertions per gene compared to Monte Carlo simulated data

Table 3.1. Detailed description of in vitro transposon mutagenesis screens.

Briefly, where available, we used the published essential gene lists identified by the authors of the screen. Otherwise, we defined genes as essential in a particular screen if the corresponding mutant did not appear in that screen, suggesting that a mutation in the corresponding gene resulted in a non-viable mutant. Candidate essential gene lists ranged in size from 179 to 913 for PAO1 and from 510 to 1544 for PA14, suggesting substantial variability between the screens (Table 3.2, Supplementary Data 1, Supplementary Data 2).

Screen	Strain	Media	Number of Essential Genes	Publication
PAO1.LB.913	PAO1	LB	913	Jacobs et al., 2003
PAO1.LB.201	PAO1	LB	201	Lee et al., 2015
PAO1.LB.335	PAO1	LB	335	Turner et al., 2015
PAO1.Sputum.224	PAO1	Sputum	224	Lee et al., 2015
PAO1.Sputum.405	PAO1	Sputum	405	Turner et al., 2015
PAO1.Pyruvate.179	PAO1	Pyruvate minimal media	179	Lee et al., 2015
PAO1.Succinate.640	PAO1	Succinate minimal media	640	Turner et al., 2015
PA14.LB.1544	PA14	LB	1544	Liberati et al., 2006
PA14.LB.634	PA14	LB	634	Skurnik et al., 2013
PA14.Sputum.510	PA14	Sputum	510	Turner et al., 2015

Table 3.2. Characteristics of the in vitro transposon mutagenesis screens.

To investigate the similarity between the different candidate essential gene lists for the two strains, we performed hierarchical clustering with complete linkage on the dissimilarity between the candidate essential gene lists, as measured by Jaccard distance (Figure 3.1A and 3.1C).

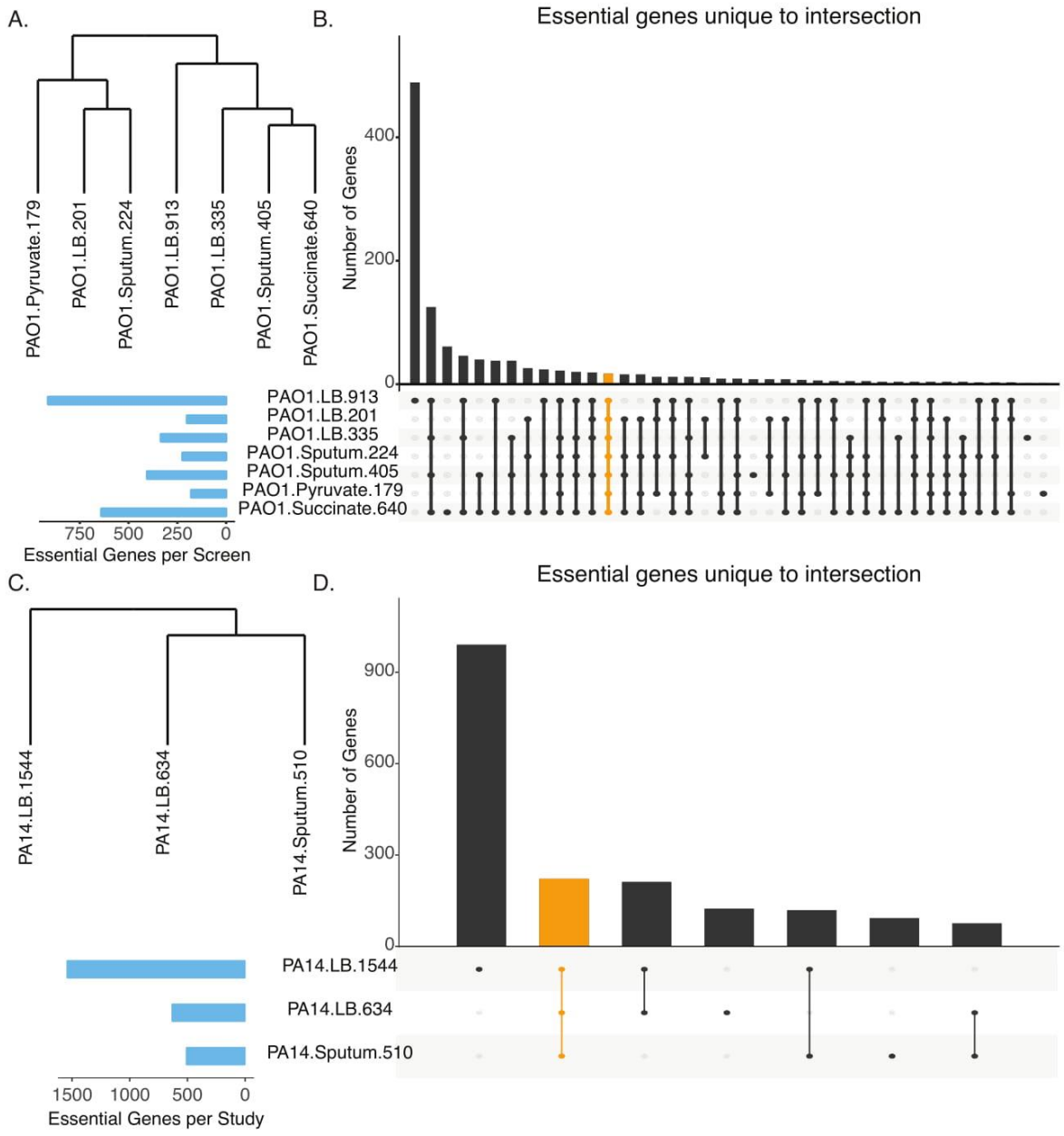


Figure 3.1. Comparison of candidate essential genes from transposon mutagenesis screens reveals variability.

(A and C). Hierarchical clustering of candidate essential gene lists from transposon mutagenesis screens for PAO1 and PA14, respectively. (B and D). Overlap analysis of candidate essential gene lists for transposon mutagenesis screens for PAO1 and PA14, respectively. Blue bars indicate the total number of candidate essential genes identified in each screen. Black bars indicate the number of candidate essential genes unique to the intersection given by the filled-in dots. The orange bar indicates the overlap for all screens for either PAO1 (Panel B) or PA14 (Panel D). For the relationship between the overlap analysis and Venn diagrams, see Figures 3.2 and 3.3.

Interestingly, the screens clustered by publication rather than by media condition for both strains. As an example from the PAO1 screens, rather than clustering by lysogeny broth (LB) media, sputum media, pyruvate minimal media, and succinate minimal media, all three of the screens from the Lee et al. publication clustered together, all three of the screens analyzed in the Turner et al. publication clustered together, and the Jacobs et al. transposon mutant library clustered independently. This result suggests that experimental technique and downstream data analysis play a large role in determining essential gene calls, motivating the importance of comparing several screens to identify consensus essential gene lists, or genes identified as essential across multiple screens.

We then measured the overlap of the candidate essential gene lists to calculate how many genes were shared across all the screens as well as those unique to particular sets of screens, defined as intersections (Figure 3.1B and 3.1D). For both strains, the candidate essential genes unique to the transposon mutant libraries (i.e., PAO1.LB.913 and PA14.LB.1544) accounted for the largest grouping, reflecting the disproportionately large size of both screens' candidate essential gene lists relative to the transposon sequencing screens. Approximately 63% and 54% of the essential genes were unique to the PA14.LB.1544 and PAO1.LB.913 screens, respectively. While genes were uniquely essential for PAO1 on individual LB screens, there were no genes uniquely essential to all three LB

screens; rather, the genes identified as commonly essential in all three LB screens were also identified in one or more of the sputum, pyruvate and succinate screens. This trend also held for the PAO1 sputum screens; however, 61 genes were uniquely identified in the succinate minimal media screen and two genes were uniquely identified in the pyruvate minimal media screen, perhaps reflecting the more stringent conditions of the minimal media screens relative to the more rich conditions of the LB and sputum screens.

This analysis revealed substantial differences in the overlap of the candidate essential genes across the screens. Using the number of intersections as an indicator of variability, comparison of the PAO1 screens resulted in more than 30 intersections, while comparison of the PA14 screens resulted in seven, highlighting the discrepancies between the screens for both *P. aeruginosa* strains. This heterogeneity across the screens could be attributed to a number of factors such as screening approach (e.g., individually mapped mutants versus transposon sequencing), library complexity, metrics of essentiality, data analysis, and the media conditions tested. To investigate the possibility that these discrepancies were completely due to data analysis alone and not experimental differences, we re-analyzed the sequencing data for the PAO1 transposon sequencing screens performed on LB where sequencing data was publicly available using the same analytical pipeline (Figure 3.2)[18,20].

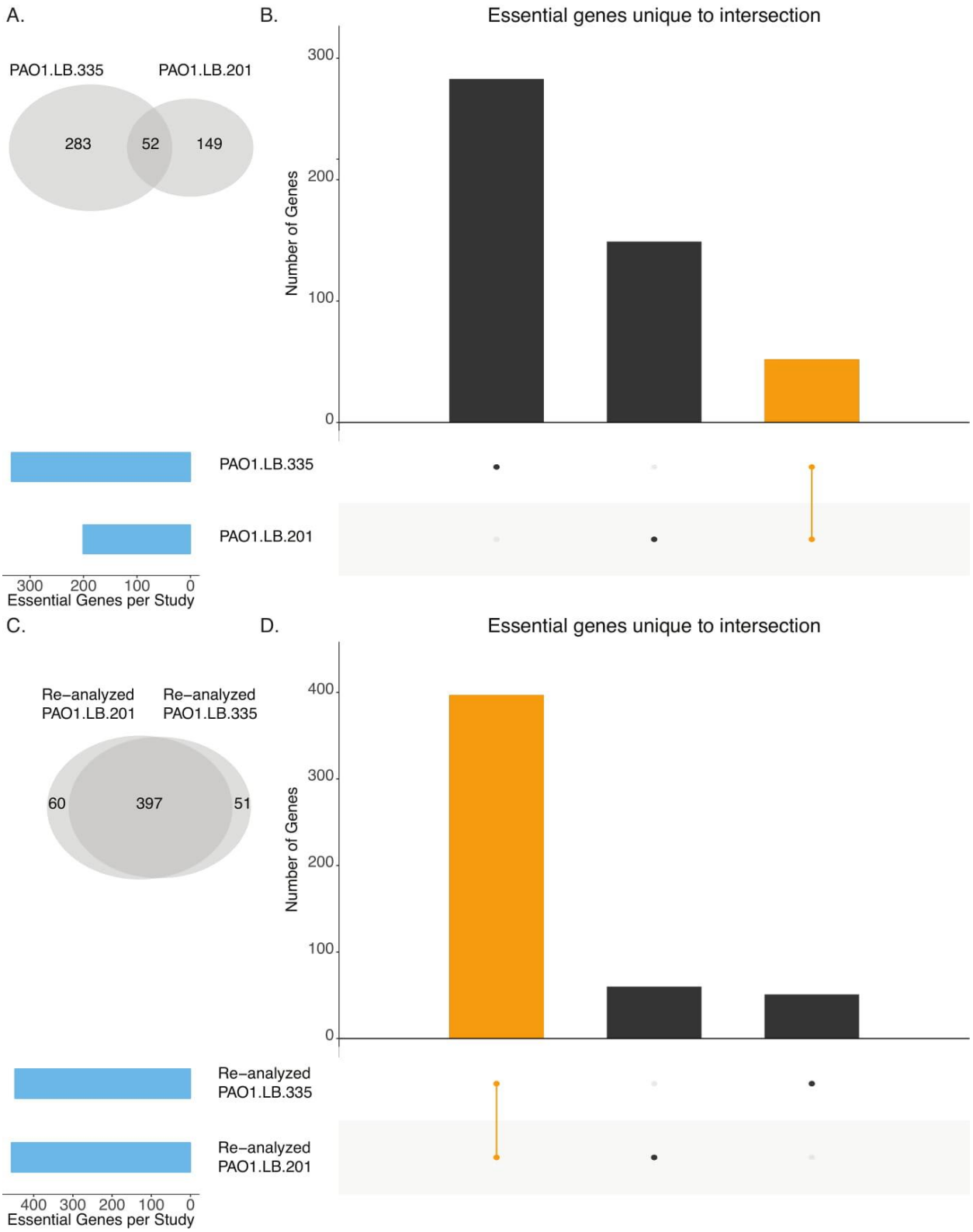


Figure 3.2. Comparison of candidate essential genes from PAO1 LB transposon mutagenesis screens reveals variability across screens.

(A and C). Venn diagrams of original (Panel A) and re-analyzed (Panel C) candidate essential gene lists from PAO1 transposon mutagenesis screens performed on LB. (B and D). Overlap analysis of original (Panel B) and re-analyzed (Panel D) candidate essential gene lists for PAO1 transposon mutagenesis screens performed on LB. Blue bars indicate the total number of candidate essential genes identified in each screen. Black bars indicate the number of candidate essential genes unique to the intersection given by the filled-in dots. The orange bar indicates the overlap of both screens.

As expected, when the same analysis pipeline was applied to the two screens, there was an increase in the number of commonly essential genes compared to the overlap between the published results. However, there were still genes that were identified as uniquely essential to each screen. These results suggest that differences in data processing alone do not account for the observed variability between the screens but that experimental differences, such as library complexity, number of replicates, and read depth, likely also contribute.

To determine potential core essential genes (i.e., genes that are essential regardless of media or other conditions), we measured the number of genes that were shared by all of the screens for either PAO1 or PA14. Surprisingly, only 17 genes were shared by all PAO1 screens while 192 genes were shared by all PA14 screens. These numbers of core essential genes are lower than expected, particularly for strain PAO1. Typically, essential genes are thought to number a few hundred for the average bacterial genome [21]. We reasoned that this unexpectedly low number of observed core essential genes might be due to the variety of media conditions across the PAO1 screens, so we repeated our analysis focusing only on the LB media screens for both PA14 and PAO1 (Figure 3.3).

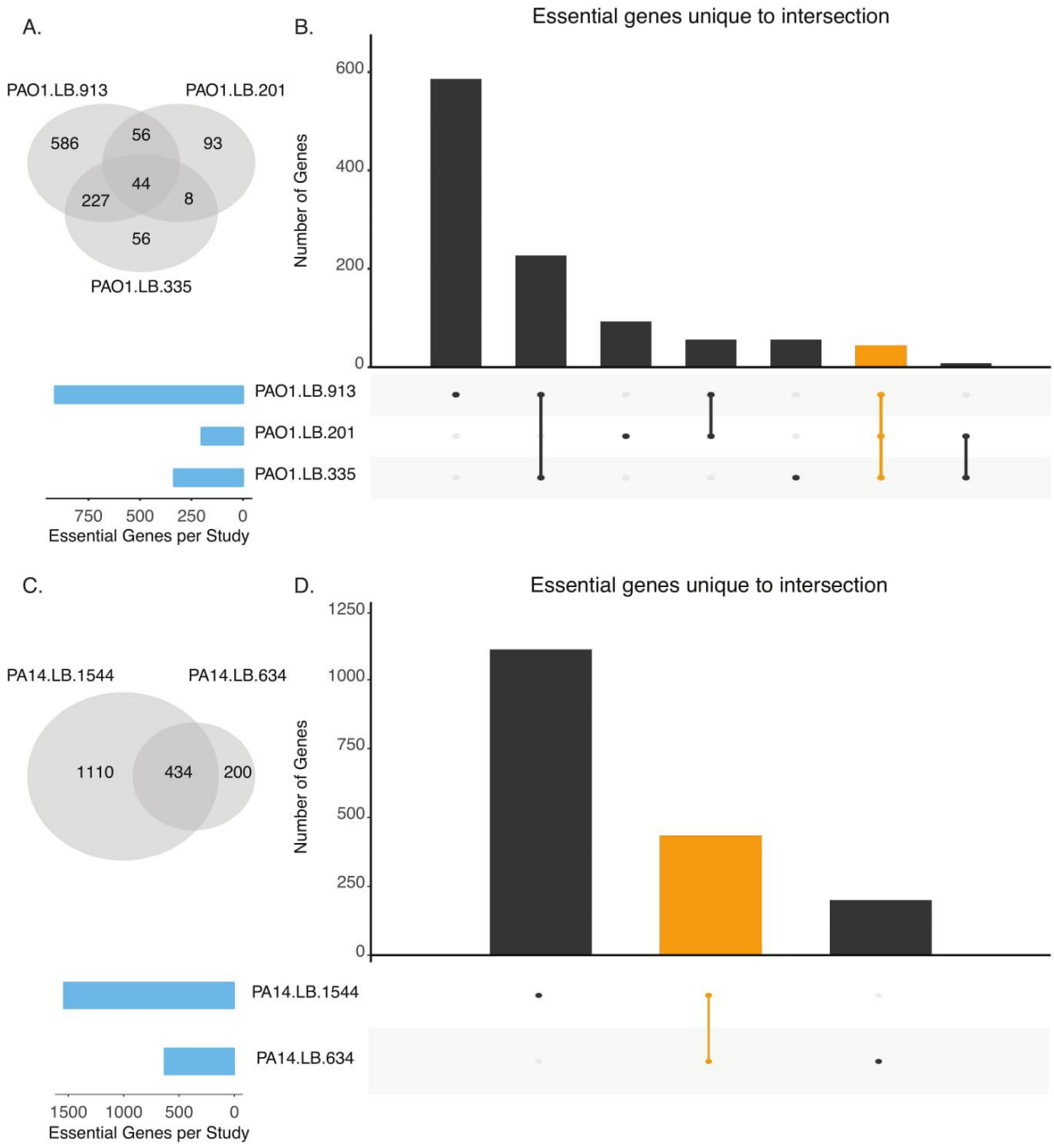


Figure 3.3. Comparison of candidate essential genes from LB transposon mutagenesis screens reveals variability across screens.
(A and C). Venn diagram of candidate essential genes lists for transposon mutagenesis screens performed on LB for PAO1 and PA14, respectively. (B and D). Overlap analysis of candidate essential gene lists for transposon mutagenesis screens performed on LB for PAO1 and PA14, respectively. Blue bars indicate the total number of candidate essential genes identified in each screen. Black bars indicate the number of candidate essential genes unique to the intersection given by the filled-in dots. The orange bar indicates the overlap for all

screens for either PAO1 (Panel B) or PA14 (Panel D). The black and orange bars correspond to the intersections identified in the venn diagrams in panels A and C.

Interestingly, the trends remained the same, with 434 genes shared across both PA14 LB media screens and only 44 genes shared across all PAO1 LB media screens. Overall, the PA14 screens had higher numbers of essential genes compared to those for PAO1, with all the PA14 screens having at least 400 essential genes. In contrast, there were four PAO1 screens with less than 350 essential genes. Together, these differences suggest greater variability for transposon mutagenesis in PAO1 compared to PA14. Strain-specific differences in essentiality have been reported previously but are underappreciated [22]. This result adds to the growing literature emphasizing how the genetic background of the strain analyzed may impact the identification of essential genes. Nevertheless, the identified core essential genes point to genes that may potentially be indispensable for bacterial growth and survival regardless of condition.

Taken together, results from this comparison revealed vast differences between the candidate essential gene lists across screens, even for those from the same media condition. These differences may be due to a number of factors such as experimental screening approach, library complexity, read depth, and downstream data analysis. Ultimately, this variability complicates the discovery of essential genes with high-confidence.

3.3.2 Contextualization of gene essentiality datasets using genome-scale metabolic network reconstructions

A central challenge of transposon mutagenesis screens lies in the interpretation of why a gene is or is not essential in a given condition. Here, we demonstrate the utility of genome-scale metabolic network reconstructions to contextualize gene essentiality and

provide mechanistic explanations for the essentiality status of metabolic genes. To do this, we compared the *in vitro* candidate essential gene lists to predicted essential genes from the PAO1 and PA14 GENREs [23]. These GENREs were previously shown to predict gene essentiality with an accuracy of 91% [23]. For both models, we simulated *in silico* gene knockouts under media conditions that approximated those used in the *in vitro* screens and assessed the resulting impact on biomass synthesis as an approximation for growth (Supplementary Data 3, Supplementary Data 4). Genes were predicted to be essential if biomass production for the associated mutant model was below a standard threshold. Predicted essential gene lists for both the PAO1 and PA14 models under the different media conditions were compared to the candidate essential gene lists for each of the experimental screens and the matching accuracy between model predictions and the *in vitro* screens was assessed (Figure 3.4A, Table 3.3).

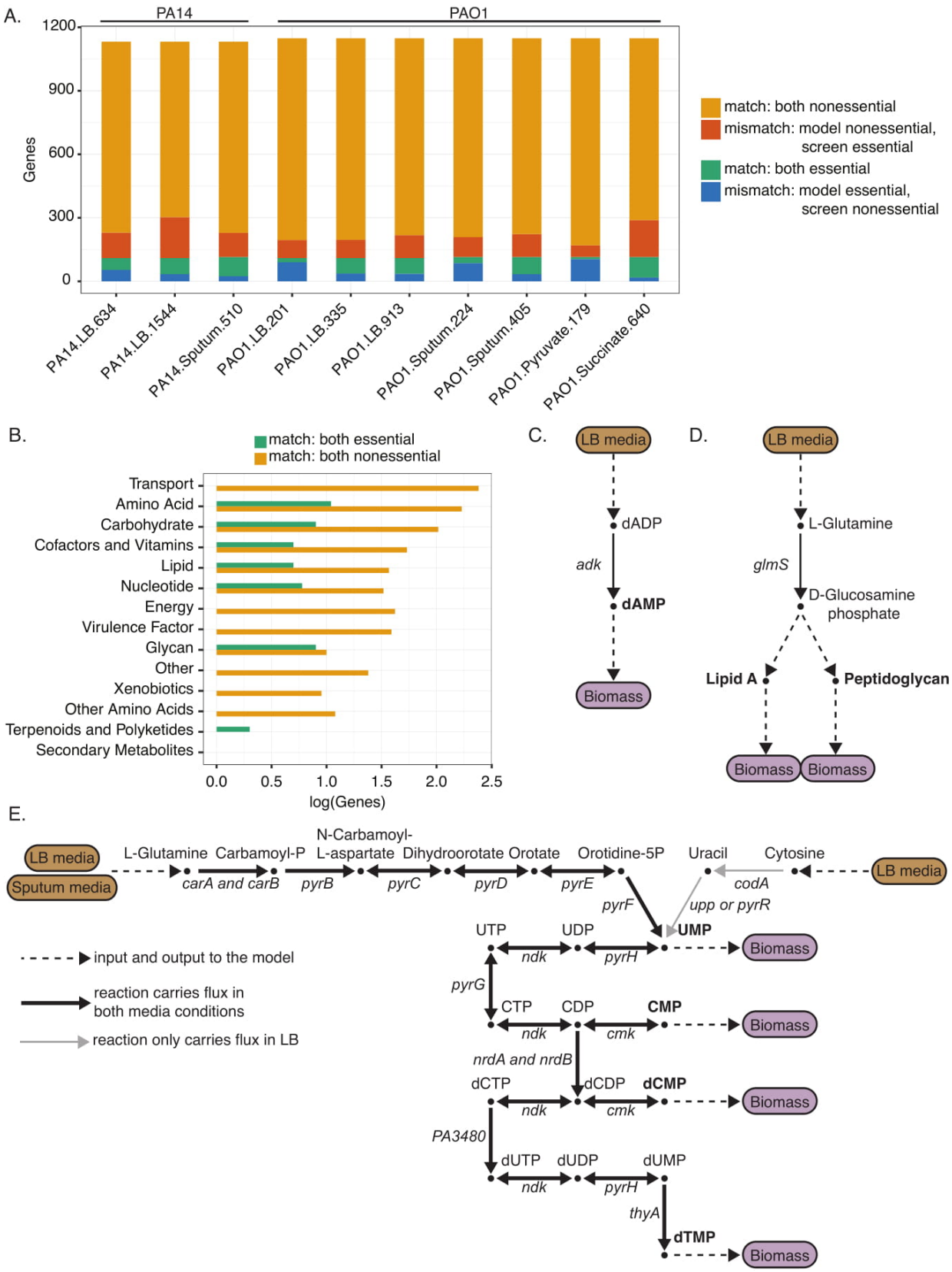


Figure 3.4. Contextualization of gene essentiality datasets using genome-scale metabolic network reconstructions.

(A). Comparison of model essentiality predictions to *in vitro* essentiality screens. *In silico* gene knockouts were performed for both PA14 and PAO1 genome-scale metabolic network reconstructions to predict essential genes. Model-predicted essential genes were compared to the candidate essential genes for each *in vitro* screen. The bars show the result of this comparison, with orange indicating the number of genes for which both the model and experimental screen identified the gene as nonessential (match: both nonessential), red indicating the number of genes for which the model identified the gene as nonessential whereas the screen identified the gene as essential (mismatch: model-nonessential, screen-essential), green indicating the number of genes for which both the model and experimental screen identified the gene as essential (match: essential), and blue indicating the number of genes for which the model identified the gene as essential whereas the screen identified the gene as nonessential (mismatch: model-essential, screen-nonessential). (B). Functional subsystems for PA14 consensus essential and nonessential genes that were also correctly predicted to be essential or nonessential in the PA14 GENRE. Consensus essential and nonessential genes were identified for PA14 by comparing all three LB screens and determining genes essential or nonessential in all three screens. (C and D). Metabolic pathways demonstrating essentiality for the consensus essential genes *adk* and *glmS*, respectively. Dashed lines represent inputs and outputs of the pathway, or, as in D, multiple steps. Brown boxes indicate media inputs, while purple boxes indicate biomass outputs. Metabolites are labeled beside the nodes, with bold metabolites indicating biomass components. Genes associated with the specific reaction are indicated. (E). Flux activity in pyrimidine metabolism under both sputum and LB media conditions. Consensus LB essential genes were compared to consensus sputum essential genes for PAO1. The PAO1 GENRE was used to explain differences in essentiality between the two media-types. Black lines indicate that the reaction is capable of carrying flux under both sputum and LB conditions, while the gray lines indicate that the reaction does not carry flux in sputum conditions but does in LB conditions. Brown boxes are media inputs, purple boxes are biomass outputs. Metabolites are labeled above the nodes, with bold metabolites indicating biomass components. Many of these metabolites are involved in many reactions beyond pyrimidine metabolism. Gene-protein-reaction relationships are indicated in italics beside each reaction edge.

Screen	% Accuracy
PAO1.LB.913	87.46
PAO1.LB.201	84.76
PAO1.LB.335	89.29
PAO1.Sputum.224	84.32
PAO1.Sputum.405	87.63
PAO1.Pyruvate.179	86.07
PAO1.Succinate.640	83.28
PA14.LB.1544	79.95
PA14.LB.634	84.63
PA14.Sputum.510	87.81

Table 3.3. Percent accuracy between model predictions of essentiality and in vitro identified essential genes.

As expected, most genes were identified as nonessential by both the screens and the models. These nonessential genes likely encode redundant features in the metabolic network, such as isozymes or alternative pathways, or are involved in accessory metabolism, such as the production of small molecule virulence factors. Interestingly, the number of screen-essential genes predicted as nonessential was significantly larger than the number of screen-nonessential genes predicted as essential ($p < 0.01$, as measured by Wilcoxon signed-rank test). We hypothesize that the reason for this difference is due to the increased likelihood of an *in vitro* screen missing a gene, potentially due to gene length or transposition cold spots [16], and subsequently incorrectly identifying it as essential.

This analysis can help to provide specific functional explanations for essentiality. Where there is agreement between the model predictions and *in vitro* screens, we can use the network to explain why a gene is or is not essential. Similarly, we can analyze the network to explain why a gene may be essential in one media condition versus another. A mismatch denotes some discrepancy between the model predictions and the experimental results. These mismatches may point to a gap in the model, indicating that it is missing some relevant biological information. Alternatively, the mismatches may be due to experimental variability such as differences in environmental conditions or technique.

To begin contextualizing the gene essentiality datasets using the GENREs, we focused on metabolic genes that were identified as essential or as nonessential in all LB screens for either PAO1 or PA14 (which we termed "consensus essential genes" and "consensus nonessential genes", respectively) (Table 3.4, Supplementary Data 5, Supplementary Data 6).

Strain	Media	Consensus Essential Genes	Consensus Non-essential Genes	Original Model Predicted Consensus Essential Genes	Original Model Predicted Consensus Non-essential Genes	Original Model Accuracy (%)	Updated Model Predicted Consensus Essential Genes	Updated Model Predicted Consensus Non-essential Genes	Updated Model Accuracy (%)
PAO1	LB	15	863	7	843	96.81	7	848	97.38
PAO1	Sputum	67	903	24	874	92.58	24	878	92.99
PA14	LB	113	800	45	777	90.03	45	781	90.47

Table 3.4. Consensus metabolic essential and non-essential genes for PAO1 and PA14 media conditions with more than two screens.

Consensus essential genes have a greater likelihood of being truly essential rather than experimental artifacts since they were identified as such in multiple independent screens. We then compared these lists of consensus essential genes and consensus nonessential genes to the model predictions of essentiality in LB media.

From this comparison, we found 45 of 113 consensus essential genes predicted to be essential by the PA14 model and 777 of 800 consensus nonessential genes predicted to be nonessential by the PA14 model. For PAO1, we found seven of 15 consensus essential genes predicted to be essential by the PAO1 model and 843 of 863 consensus nonessential genes predicted as nonessential by the PAO1 model (Table 3.4). The low number of consensus essential genes for PAO1 reflects the high variability between screens, as highlighted in Figures 3.1 and 3.2.

We then used the models to delineate subsystem assignments for the model-predicted consensus essential and nonessential genes (Figure 3.4B for PA14 and Figure 3.5 for PAO1).

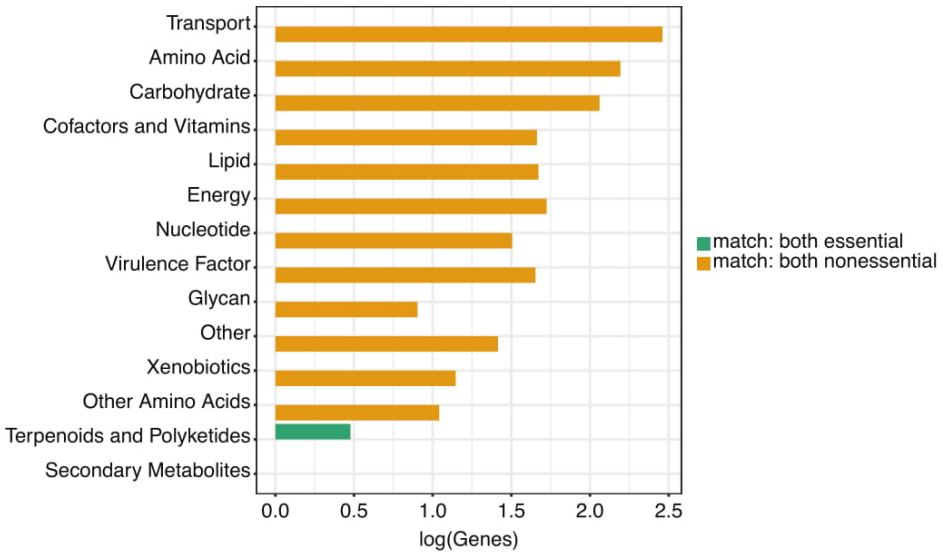


Figure 3.5. Distribution of PAO1 consensus essential and nonessential genes across model subsystems.
Functional subsystems for PAO1 consensus essential and nonessential genes that were also identified to be essential or nonessential in the PAO1 genome-scale metabolic network model. Consensus essential and nonessential genes were identified for PAO1 by comparing all three LB screens and identifying those genes which were either essential or nonessential in all three screens.

As expected, the consensus nonessential genes spanned most subsystems within the network, likely due to redundancy in the network as well as the presence of accessory metabolic functions that are not critical for biomass production. In contrast, for PA14, the consensus essential genes were limited to seven of the 14 subsystems within the network (note that this trend does not hold for PAO1 because there were very few consensus essential genes to consider). These seven subsystems capture metabolic pathways that are critical for bacterial growth and survival. For instance, lipid metabolism is essential for building and maintaining cell membranes, while carbohydrate metabolism is critical for ATP generation. None of the genes involved in transport were consensus essential genes. Because we only considered screens performed in LB media, transport of individual important metabolites, such as a specific carbon sources, was not a limiting factor given the abundant availability of

such compounds in rich media conditions. However, we would expect that if we considered screens performed under minimal media conditions, relevant transport genes would be essential for bacterial growth.

Because these consensus essential genes were also predicted to be essential by the model, we can use the network to provide functional reasons for essentiality. For example, both the model and screens identified the gene *adk*, encoding adenylate kinase, as essential. Using the model, we determined that when *adk* is not functional, the conversion of deoxyadenosine diphosphate (dADP) to deoxyadenosine monophosphate (dAMP) cannot proceed, impacting the cell's ability to synthesize DNA and ultimately produce biomass (Figure 3.4C). The model can also tease out less obvious relationships. For instance, both the model and the screens identified *glmS*, encoding glucosamine-fructose-6-phosphate aminotransferase, as essential. Using the model, we found that when *glmS* is not functional, the conversion of L-Glutamine to D-Glucosamine phosphate cannot proceed. D-Glucosamine phosphate is an essential precursor to both Lipid A, a component of the endotoxin lipopolysaccharide, and peptidoglycan, which forms the cell wall (Figure 3.4D). For each of the model-predicted consensus essential genes, we identified which biomass components could not be synthesized when the gene was removed from the model (Supplementary Data 7 and Supplementary Data 8). Further analysis is necessary to tease out the metabolic pathways that prevent synthesis of these biomass metabolites; however, from the examples above it is evident that GENREs can provide both obvious and non-obvious functional explanations for essentiality, streamlining the interpretation of transposon mutagenesis screens.

In addition to identifying consensus essential and nonessential genes that were in agreement with the models, we also uncovered discrepancies between model predictions and experimental results. For PAO1 and PA14, respectively, there were 8 and 68 consensus essential genes that the models predicted to be nonessential and 20 and 23 consensus nonessential genes that the models predicted to be essential. These mismatches between model predictions and experimental results provide insight into gaps in our understanding of *P. aeruginosa* metabolism.

In the case where a consensus essential gene was predicted to be non-essential by the model, this result indicates that the model has some additional functionality that is not available *in vitro*. This result could be an inaccuracy of the network reconstruction or it could be a result of using a non-condition-specific network where the model has access to all possible reactions in the network. Because cells undergo varying states of regulation, gene essentiality can be modulated as a result. Thus, profiling data such as transcriptomics could be integrated into the network reconstruction to generate a condition-specific model to improve model predictions under specified conditions [24,25].

In contrast, in the case where a consensus nonessential gene was predicted to be essential, this result indicates that the model is missing key functionality, pointing to areas of potential model curation. Using this list of discrepancies to guide curation (Table 3.5), we performed an extensive literature review and found several suggested changes to the metabolic network reconstruction (Supplementary Data 9).

PAO1 Locus Tag	Name	Function	Subsystem
PA0265	*davD*	Glutaric semialdehyde dehydrogenase	Carbohydrate
PA0546	*metK*	Methionine adenosyltransferase	Amino Acid
PA0581	*ygiH*	Glycerol-3-phosphate acyltransferase	Lipid
PA1758	*pabB*	Para-aminobenzoate synthase component I	Cofactors and Vitamins
PA1806	*fabI*	NADH-dependent enoyl-ACP reductase	Lipid
PA1959	*bacA*	Bacitracin resistance protein	Glycan
PA2165	*glgA*	Probable glycogen synthase	Carbohydrate
PA2964	*pabC*	4-Amino-4-deoxychorismate lyase	Cofactors and Vitamins
PA2969	*plsX*	Fatty acid biosynthesis protein PlsX	Lipid
PA3164		Frameshift 3-phosphoshikimate-carboxyvinyltransferase prephenate dehydrogenase	Amino Acid
PA3296	*phoA*	Alkaline phosphatase	Cofactors and Vitamins
PA3333	*fabH2*	3-Oxoacyl-[acyl-carrier-protein] synthase III	Lipid
PA3633	*ygbP*	4-Diphosphocytidyl-2-C-methylerythritol synthase	Lipid
PA3659	*dapC*	Succinyldiaminopimelate transaminase	Amino Acid
PA3686	*adk*	Adenylate kinase	Nucleotide
PA4050	*pgpA*	Phosphatidylglycerophosphatase A	Lipid
PA4693	*pssA*	Phosphatidylserine synthase	Lipid
PA4770	*lldP*	L-lactate permease	Transport
PA5322	*algC*	Phosphomannomutase AlgC	Carbohydrate
PA5357	*ubiC*	Chorismate pyruvate lyase	Cofactors and Vitamins

Table 3.5. Discrepancies between model predicted essential genes and in vitro identified consensus nonessential genes for PAO1.

For instance, we incorrectly predicted as essential the gene *fabI* (PA1806), which is linked to triclosan resistance; however, a recent study discovered an isozyme of *fabI* in PAO1 called *fabV* (PA2950) [26]. To account for this new information, we suggest changing the gene-protein-reaction (GPR) relationship for the 28 reactions governed by *fabI* to be "*fabI* OR *fabV*", making *fabI* no longer essential in the model. Additionally, our model incorrectly predicted the genes *ygiH* (PA0581) and *plsX* (PA2969) to be essential due to a GPR formulation of "*ygiH* AND *plsX*" for several reactions in glycerolipid metabolism. Literature

evidence suggests that the gene-product of *plsB* (PA3673) is also able to catalyze these reactions. Specifically, the gene-products of both *plsB* and the *ygiH*/*plsX* system are able to carry out the acylation of glycerol-3-phosphate from an acyl carrier protein whereas only the gene-product of *plsB* is able to carry out this reaction for acyl-CoA thioesters [27,28]. This experimental evidence motivates changing the GPRs for 16 reactions in glycerolipid metabolism.

In addition to changes in the GPR formulation for specific reactions, we also identified a potential change to the biomass reaction. Two PAO1 genes, *glgA* (PA2165) and *algC* (PA5322), are incorrectly predicted as essential for the synthesis of glycogen, a biomass component. Glycogen is not an essential metabolite for *P. aeruginosa* growth; however, it is very important for energy storage, which is why it was initially included in the biomass reaction [29]. Removal of glycogen from the biomass equation would make *glgA* and *algC* accurate predictions as nonessential genes in PAO1. Implementing these proposed changes in the PAO1 and PA14 GENREs resulted in enhanced predictive capability of the models (Supplementary Data 10, Supplementary Data 11, Table 3.4). The updated PAO1 model predicted consensus gene essentiality status in LB media with an accuracy of 97.4% compared to 96.8% for the original model. Meanwhile, the updated PA14 model predicted consensus gene essentiality status in LB media with an accuracy of 90.5% compared to 90.0% for the original mode. It is worth noting that, although these changes to the reconstructions were made to address essentiality discrepancies in LB media conditions, they also improved the PAO1 model predictive capabilities for consensus genes in sputum media, increasing accuracy from 92.6% to 93.0%.

While we identified several changes to the model to improve predictions, there were several genes for which we could find no literature evidence to change their predicted essentiality status. These genes highlight gaps in our current knowledge and understanding of *Pseudomonas* metabolism and indicate areas of future research. Identification of these knowledge gaps is not possible without the reconciliation of experimental data with model predictions. Ultimately, this analysis demonstrates the utility of integrating data with GENREs to improve gene annotation and suggest areas of future research.

In addition to contextualizing essentiality for a given media condition, we also used the model to explain why certain metabolic genes are essential in one media-type versus another. We compared consensus LB essential genes to consensus sputum essential genes for PAO1 and identified the essential genes that were either shared by both conditions or unique to one condition versus the other. Overall, 18 genes were commonly essential, while 92 genes were uniquely essential in sputum and 26 genes were uniquely essential in LB, indicating the presence of condition-dependent essential genes.

We then focused our analysis just on those genes that were also present in the PAO1 model and compared these lists to model predictions. We found four genes that both the model and the screens indicated as uniquely essential in sputum but not in LB. Interestingly, all four of these genes (*pyrB*, *pyrC*, *pyrD*, and *pyrF*) are involved in pyrimidine metabolism. Applying flux sampling [30] to the PAO1 metabolic network model, we investigated why these four genes were uniquely essential in sputum but not in LB (Figure 3.4E). The pyrimidine metabolic pathway directly leads to the synthesis of several key biomass precursors (UMP, CMP, dCMP and dTMP), making it an essential subsystem within the network. Under LB media conditions, there are two inputs into the pathway, one through L-

Glutamine and the other through Cytosine. However, in sputum media conditions, L-Glutamine is the only input into the pathway. Because of this reduction in the number of available substrates in sputum media, the steps for L-Glutamine breakdown must be active to synthesize the biomass precursors. Thus, the genes responsible for catalyzing this breakdown are essential in sputum media conditions. In contrast, because there are two LB substrates that feed into pyrimidine metabolism, if a gene involved in the breakdown of one of the substrates is not functional the other substrate is still accessible, thus making the deletion of that gene nonessential.

As stated above, further constraining the model with profiling data from both media conditions would help to further contextualize differences in the essentiality results by modulating the availability of certain reactions. Nevertheless, as demonstrated here, the metabolic network reconstruction can be a useful tool for providing functional explanations for why certain genes are essential in one condition versus another.

3.3.3 Quantitative evaluation of the impact of media formulation on condition-independent essential gene identification

Given the variability in the number of candidate essential genes across the screens, we were interested in using the models to quantitatively evaluate the impact of media conditions on essentiality. We first focused our analysis on how the number of considered minimal media conditions impacts the number of condition-independent essential genes identified, or the number of genes found as essential in every condition. To do this, we simulated growth of the PA14 model on 42 different minimal media and performed *in silico* gene knockouts, identifying the genes essential for biomass production on each media condition (Figure 3.6A).

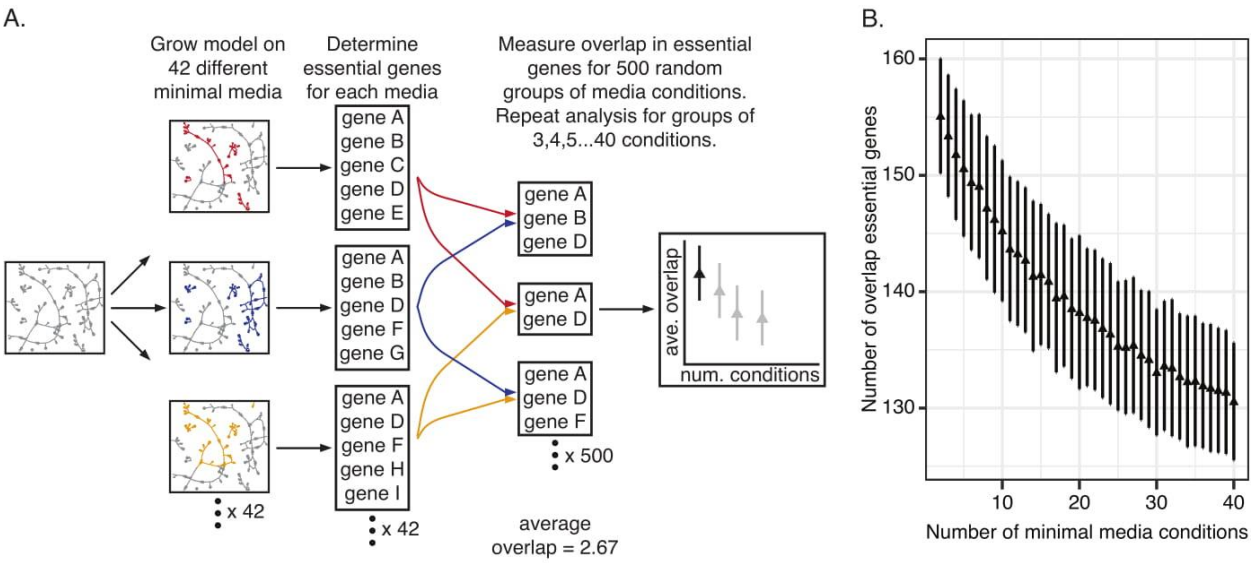


Figure 3.6. Computational assessment of the impact of number of minimal media conditions considered on condition-independent essentiality.
(A). Pipeline for computational assessment of the impact of minimal media composition on condition-independent essentiality. The base PA14 model is grown on 42 different minimal media. For each minimal media condition, the *in silico* essential genes are identified, resulting in 42 essential gene lists. Initially, pairwise comparisons are made between minimal media essential gene lists to identify the shared essential genes. Specifically, the essential gene lists from two randomly selected minimal media conditions are compared to determine the overlap between the two gene lists. This random selection of two minimal media conditions to compare is repeated 500 times. The average number of overlap genes for all 500 comparisons is calculated as well as the standard deviation. Ultimately, this random selection of groups of minimal media conditions to compare is repeated for groups of three minimal media conditions, groups of four, and so on, up to groups of 40 minimal media conditions. (B). Impact of minimal media differences on the identification of condition-independent essential genes. Each data point represents the mean from 500 comparisons. Error bars indicate standard deviation.

We then randomly selected groups of minimal media conditions and compared their essential gene lists to determine the commonly essential genes, defined as the overlap. We performed this random selection of minimal media conditions for group sizes ranging from two to 40 minimal media conditions considered. For each group size, we randomly selected minimal media conditions 500 times. As expected, the more media conditions considered, the smaller the overlap of essential genes (Figure 3.6B). This relationship between the number of media conditions considered and the size of the overlap is best characterized by

an exponential decay, with the size of the overlap eventually converging on 131 genes as 40 conditions are considered. This result suggests that to identify a core set of condition-independent essential genes, dozens of minimal media screens need to be compared. However, variability between the screens, as indicated by the error bars, could still confound interpretation, necessitating the comparison of replicates and potentially even more screens to truly identify condition-independent essential genes with high confidence.

We next assessed how modifications to a rich media, like LB, impact gene essentiality. LB is a complex media with known batch-to-batch variability [31,32], motivating this analysis of how differences in LB composition can alter essentiality. Given the challenge of modeling concentration, here the simulations focus on the presence or absence of metabolites in LB media. Specifically, we randomly selected carbon source components from LB media in sets of varying sizes, ranging from two to 21 LB media components considered. We then used these sets as the model media conditions and performed *in silico* gene knockouts to identify essential genes for biomass production on each LB media formulation (Figure 3.7A).

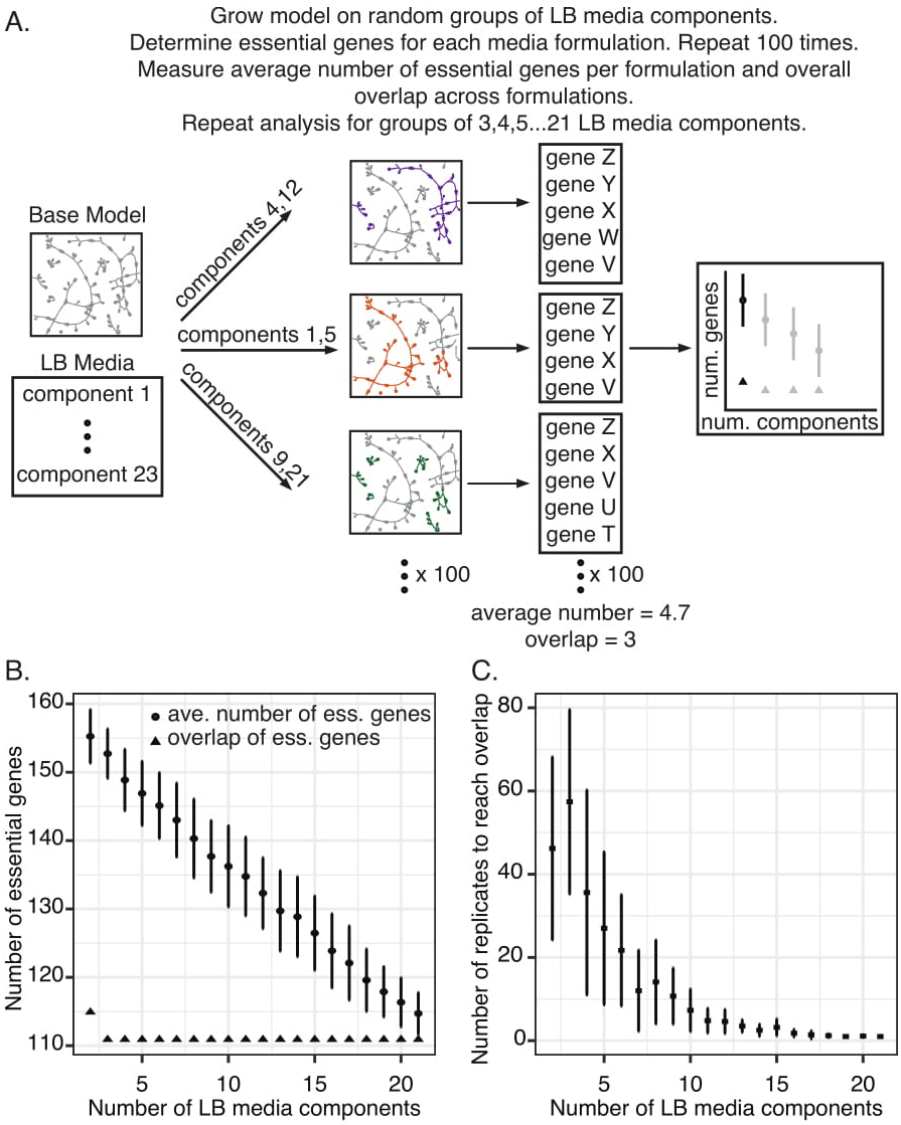


Figure 3.7. Computational assessment of the impact of LB media composition on condition-independent essentiality.

(A). Pipeline for computational assessment of the impact of LB media formulation on condition-independent essentiality. The PA14 model is grown on different media formulations consisting of random groups of LB components. For instance, two random LB components are selected out of a pool of 23 LB components. The model is grown on these randomly selected pairs and the essential genes for growth on this media formulation are identified. This analysis is repeated 100 times for 100 pairs of LB media components. The average number of essential genes for growth on these random pairs across 100 different formulations is calculated as well as the standard deviation. Additionally, the essential genes common to all 100 different formulations is determined. Ultimately, this random selection of groups of LB media components to support growth of the model and essential gene identification is repeated for groups of three LB components, groups of four, and so on, to groups of 21 LB media components. (B) Impact of LB media formulation on the identification of condition-independent essential genes. Circles represent the average number of essential genes identified in different LB media formulations across 100 comparisons. Triangles

represent the shared essential genes (i.e., the overlap) across all 100 comparisons. Error bars indicate standard deviation. (C) Number of replicates needed to converge on shared essential genes in different LB formulations. The pipeline outlined in Panel A was repeated 10 independent times, with 100 replicates per set size. For each iteration, the number of replicates needed to recapture the 111 overlapping genes was calculated. Each data point represents the average number of replicates from the 10 runs. Error bars indicate standard deviation.

For each set size, we randomly selected LB components 100 times and calculated the average number of essential genes identified as well as the number of shared essential genes across all 100 sets. As the number of LB media components increases, we found that the size of the essential gene lists decreases linearly (Figure 3.7B). If we were to consider even more media components beyond the scope of LB, we predict that this linear relationship would eventually plateau due to limitations in the metabolic network. This result suggests that a media richer than LB may be necessary to identify a core set of condition-independent essential genes.

Interestingly, we found that as more complex LB media formulations are considered, the number of shared essential genes across 100 simulations quickly converges on 111. Indeed, only three LB media components were needed to achieve this overlap. Thus, even though the average size of essential gene lists is larger for less complex media formulations, the overlap of these larger essential gene lists still results in the same overlap as more complex media formulations, suggesting that changes in complex media formulation have minimal impact on determining a core set of essential genes.

However, for this analysis, we had compared 100 random media formulations for each set size, potentially masking the impact of media changes on essentiality. To identify how many LB media formulations need to be compared to converge on this overlap value, we re-ran this analysis 10 times and, for each iteration, determined the number of samples,

or replicates, needed to recapture the 111 overlapping genes (Figure 3.7C). In more complex media formulations, relatively few comparisons are needed to identify the 111 overlapping essential genes. However, as fewer LB media components are considered, more comparisons need to be made. For example, in the case of formulations consisting of only three LB media components, nearly 60 comparisons are needed to converge on the 111 overlap essential genes. Thus, as the media formulation diverges from true LB due to batch-to-batch variability, more comparisons are necessary to converge on a core set of essential genes.

Taken together, these computational analyses define the scope that is needed to identify condition-independent essential genes. These results suggest that both the number of media conditions and the number of replicates analyzed can impact our ability to determine condition-independent essential genes.

3.4 Discussion

The identification of both condition-dependent and condition-independent essential genes has been a long-standing interest [33,34]. Determination of these essential processes can aid in the discovery of novel antibacterial targets as well as the discovery of minimal genomes required to sustain life [7,35]. In this study, we performed a large-scale comparison of multiple gene essentiality datasets and contextualized essential genes using genome-scale metabolic network reconstructions. We applied this approach to several *P. aeruginosa* transposon mutagenesis screens performed on multiple media conditions and demonstrated the utility of GENREs in providing functional explanations for essentiality and resolving differences between screens. Finally, using the *P. aeruginosa* GENRE, we performed a high-throughput, quantitative analysis to determine how media conditions impact the

identification of condition-independent essential genes. The resulting insights would be challenging to develop without the use of a computational model of *P. aeruginosa* metabolism. Our work enables the elucidation of mechanistic explanations for essentiality, which is challenging to determine experimentally. Ultimately, this approach serves as a framework for future contextualization of gene essentiality data and can be applied to any cell type for which such data is available. Additionally, by quantifying the impact of media conditions on the identification of condition-independent essential genes, we contribute novel insights for design of future gene essentiality screens and identification of core metabolic processes.

Recent advances in deep-sequencing technologies combined with transposon mutagenesis have enabled high-throughput determination of candidate essential genes for a variety of bacterial species in a wide range of environmental conditions [36]. While researchers have demonstrated reasonable reproducibility within a given study [37], variability across studies has been suggested but not assessed on a large-scale [1,38]. Our comparison of multiple *P. aeruginosa* transposon mutagenesis screens revealed substantial variability in candidate essential genes within and across media conditions, particularly for strain PAO1. Numerous factors may contribute to this lack of overlap between the screens, such as differences in transposon insertion library complexity, differences in data analysis and statistical determination of essentiality, as well as environmental variability between the screens [8,9]. Factors such as these lead to discrepancies between screens and complicate our ability to identify high-confidence sets of condition-dependent and condition-independent essential genes.

Focusing on one of these factors, we used the metabolic model of *P. aeruginosa* strain PA14 to quantitatively assess how media formulation impacts the identification of condition-independent essential genes. While previous *in vitro* studies have surveyed conditional essentiality in numerous environmental conditions, these screens used an already established mutant library for each media-type [39]. In this work, we computationally generated *de novo* mutant libraries for individual media conditions, eliminating any bias from starting with an established mutant library. Ultimately, we found that to determine a high-confidence set of core essential genes for minimal media conditions, more than 40 minimal media formulations need to be compared. We extended this analysis to consider how differences in rich media formulations impact gene essentiality and found that as rich media formulations diverge, as many as 60 replicates are needed to identify condition-independent essential genes with high-confidence. Taken together, these computational results suggest a rich opportunity for a large-scale experimental effort to identify with high confidence condition-independent essential genes. These insights would be impossible to garner without computational modeling due to the sheer number of comparisons made.

In addition to variability between datasets, a central difficulty of performing gene essentiality screens lies in the interpretation of why a gene is essential in a given condition. Oftentimes, laborious follow-up experiments are necessary to investigate the role of a gene in a given condition using lower-throughput approaches [36]. Here, we presented a strategy for contextualizing gene essentiality data using genome-scale metabolic network reconstructions. We demonstrated the utility of this approach by providing functional reasons for essentiality for consensus LB media essential genes. For these genes, we determined which specific components of biomass could not be synthesized when the gene

was knocked out. Additionally, by analyzing the network structure and flux patterns, we used the model to explain why certain genes are essential in one condition versus another. Our computational approach provides testable hypotheses regarding the functional role of a gene in synthesizing biomass in a given environmental condition, streamlining downstream follow-up experiments. In future work, profiling data could be integrated with the metabolic networks to further enhance the utility of these models in contextualizing gene essentiality [24]. Additionally, integration of transcriptional regulatory networks with the GENREs would further expand the number of genes considered [40].

In summary, genome-scale metabolic network reconstructions can guide the design of gene essentiality screens and help to interpret their results. The identification of both condition-independent and condition-dependent essential genes is vital for the discovery of novel therapeutic strategies and mechanistic modeling streamlines the ability to identify these genes. This framework can be applied to numerous other organisms of both clinical and industrial relevance.

3.5 Methods

3.5.1 Data sources

Transposon insertion library datasets were downloaded from the original publication for each screen where available. Screens were renamed following this pattern: *Strain.Media.NumEssentials*, where *Strain* indicated whether the screen was for strain PAO1 or PA14, *Media* indicated which media condition the screen was performed on, and *NumEssentials* indicated the number of essential genes identified for the given strain on the given media condition. Specifically, for the PAO1.LB.201, PAO1.Sputum.224, and

PAO1.Pyruvate.179 datasets, Dataset_S01 was downloaded from [19]. For the PAO1.LB.335, PAO1.Sputum.405, and PAO1.Succinate.640 datasets, Dataset_S01 was downloaded from [18]. For the PA14.LB.634 dataset, Table S1 was downloaded from [17]. For the PA14.Sputum.510 dataset, Dataset_S04 was downloaded from [18]. For the PAO1.LB.913 dataset, PA_two_allele_library5.xlsx was downloaded from the Manoil Laboratory website (http://www.gs.washington.edu/labs/manoil/libraryindex.htm). For the PA14.LB.1544 dataset, NRSetFile_v5_061004.xls was downloaded from the PA14 Transposon Insertion Mutant Library website (http://pa14.mgh.harvard.edu/cgi-bin/pa14/downloads.cgi).

The PAO1 and PA14 genome-scale metabolic network reconstructions were downloaded from the Papin Laboratory website (http://www.bme.virginia.edu/csbl/Downloads1-pseudomonas.html).

3.5.2 Generation of candidate essential gene lists

Candidate essential genes were determined for each screen as follows. For PAO1.LB.201, we considered genes to be essential if they were not disrupted in all six of the Tn-seq runs on LB in the original dataset. For PAO1.Sputum.224, we considered genes to be essential if they were not disrupted in all four of the Tn-seq runs on sputum in the original dataset. For PAO1.Pyruvate.179, we considered genes to be essential if they were not disrupted in all three of the Tn-seq screens on Pyruvate minimal media in the original dataset. For PAO1.LB.335, PAO1.Sputum.405, and PAO1.Succinate.640, we used the genes that were labeled as essential in the original dataset. For PAO1.LB.913, the mutants listed in the transposon insertion library were compared to a list of all known genes in the PAO1 genome. Genes in the PAO1 genome that were not in the mutant library list were considered

to be essential. For PA14.LB.634, we used the genes listed as essential in the original dataset. For PA14.BHI.424 and PA14.Sputum.510, we used the genes that were labeled as essential in the original dataset. For PA14.LB.1544, the mutants listed in the transposon insertion library were compared to a list of all known genes in the PA14 genome. Genes in the PA14 genome that were not in the mutant library list were considered to be essential.

3.5.3 Comparison of candidate essential gene lists

Hierarchical clustering with complete linkage was performed on the candidate essential gene lists for the PA14 and PAO1 screens and visualized with a dendrogram. The overlap between the datasets was visualized using the R-package, UpsetR [41].

3.5.4 Re-analysis of transposon sequencing datasets

PAO1.LB.335 sequencing data were downloaded from NCBI SRA under the accession number SRX031647. PAO1.LB.201 sequencing data were downloaded from NCBI SRA under the accession number PRJNA273663. Data were analyzed using methods adapted from [18,20]. Briefly, reads were mapped to the PAO1 reference genome (GCA_000006765.1 ASM676v1 assembly downloaded from NCBI) using bowtie2 v.2.3.4.1. Open reading frame assignments were modified where 10% of the 3' end of every gene was removed in order to disregard insertions that may not interrupt gene function. Aligned reads were mapped to genes and we removed the 50 most abundant sites to account for potential PCR amplification bias. We applied weighted LOESS smoothing to correct for genome position-dependent effects. One-hundred random datasets were generated by randomizing insertion locations. Previous analysis showed that results begin to converge after 50 random datasets [18]. We compared the random datasets to the experimental datasets with a negative binomial test in

DESeq2. We corrected for multiple testing by adjusting the p-value with the Benjamini-Hochberg method. We used the mclust package in R to test whether a gene was 'reduced' or 'unchanged'. Genes were called 'essential' if they were assigned to the 'reduced' category by mclust with an adjusted p-value <0.05 and uncertainty <0.1.

3.5.5 Model gene essentiality predictions

In silico gene essentiality screens were performed in relevant media conditions using the PAO1 and PA14 genome-scale metabolic network reconstructions [23]. Specifically, media formulations were computationally approximated for LB, sputum, pyruvate minimal media, and succinate minimal media for the PAO1 simulations and LB and sputum for the PA14 simulations. Systematically, genes were deleted from the models one-by-one and the resulting impact on biomass production was assessed. If biomass production for the associated mutant model was below 0.0001 h^{-1}, a standard threshold, the knocked-out gene was predicted to be essential [23]. For each *in silico* predicted essential gene, we determined which biomass components specifically could not be synthesized using the COBRA toolbox function, biomassPrecursorCheck() [42]. Statistical significance for the comparison of the "mismatch: model nonessential, screen essential" category and the "mismatch: model essential, screen nonessential" category was assessed using the Wilcoxon signed-rank test.

3.5.6 Subsystem assignment of consensus essential and nonessential genes

For each of the consensus essential and nonessential genes that were also present in the PAO1 and PA14 models, we determined which subsystems they participated in using an in-house script (see Supplementary Information). Briefly, we first converted model subsystems to broad subsystems based on KEGG functional categories [43]. We then

identified the reactions associated with the gene of interest and used the broad subsystem of this reaction to indicate the subsystem assignment for the gene of interest. Where there was more than one reaction connected to a gene, we used the reaction associated with the first instance of the gene in the network for subsystem assignment.

3.5.7 Flux sampling in LB and sputum

The impact of media conditions on flux through pyrimidine metabolism in the PAO1 metabolic network reconstruction was assessed using the flux sampling algorithm optGpSampler [30]. Briefly, optGpSampler samples the solution space of genome-scale metabolic networks using the Artificial Centering Hit-and-Run algorithm and returns a distribution of possible flux values for reactions of interest. Three-thousand flux samples were collected for each simulation, using one thread and a step-size of one. Maximization of biomass synthesis was set as the objective function. Flux sampling simulations were performed for PAO1 grown in LB media and sputum media. The median flux values for every reaction in pyrimidine metabolism were compared between the LB and sputum simulations to determine whether flux was higher, lower, or unchanged in sputum versus LB.

3.5.8 Media formulation impact on essentiality

The impact of media formulation on gene essentiality predictions was assessed using the PA14 genome-scale metabolic network reconstruction. For the minimal media analysis, the PA14 model was grown on 42 different minimal media and *in silico* essential genes were identified as described above. We then randomly selected groups of minimal media conditions of varying sizes, ranging from two to 41 minimal media conditions considered, and found the intersection of the group's predicted essential gene lists, or the genes that

were identified as essential in every condition considered within that group. For each group size, we randomly selected minimal media conditions 500 times.

For the LB media analysis, we randomly selected components from LB media in sets of varying sizes, ranging from two to 21 LB media components considered, used these sets as the model media conditions, and identified *in silico* essential genes as above. For each set size, we randomly selected LB components 100 times and calculated the average total number of essential genes identified and the intersection of the essential genes across all 100 sets. To determine how many LB media formulations needed to be compared to converge on this intersection, we re-ran this LB media formulation analysis 10 times and, for each iteration, determined the number of samples needed to achieve the size of the overlap if all 100 samples were considered at each set size

3.5.9 Code and data availability

Code and files necessary to recreate figures and data can be found here: https://github.com/ablazier/gene-essentiality

3.5.10 Computational resources

The COBRA Toolbox 2.0.5 [42], the Gurobi 6.5 solver, and MATLAB R2016a were used for model simulations. optGPSampler1.1 was used for flux sampling simulations [30]. Bowtie2 v.2.3.4.1 [44] and Samtools v.1.3.1 [45] were used for transposon sequencing analysis. R 3.3.3 was used for all other analyses and figure generation.

3.6 Acknowledgments

The authors would like to thank Jennifer Bartell and Joanna Goldberg for early discussions related to the project, Laura Dunphy, Kevin Janes, and Phillip Yen for thoughtful comments on the manuscript, and Sean Leonard for assistance in analysis of the transposon sequencing raw datasets. Support for this project was provided by The Wagner Fellowship and Unilever.

3.7 Author Contributions

A.S.B. and J.A.P. conceived and designed the study. A.S.B. completed all analyses. A.S.B. and J.A.P. wrote and edited the manuscript.

3.8 Conflicts of Interests

The authors declare no conflicts of interest.

3.9 References

1. Umland TC, Schultz LW, MacDonald U, Beanan JM, Olson R, Russo TA. In vivo-validated essential genes identified in Acinetobacter baumannii by using human ascites overlap poorly with essential genes detected on laboratory media. MBio. 2012;3. doi:10.1128/mBio.00113-12

2. Le Breton Y, Belew AT, Valdes KM, Islam E, Curry P, Tettelin H, et al. Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes. Sci Rep. 2015;5: 9838.

3. Gallagher LA, Shendure J, Manoil C. Genome-Scale Identification of Resistance Functions in Pseudomonas aeruginosa Using Tn-seq. 2011; doi:10.1128/mBio.00315-10.Editor

4. Moule MG, Hemsley CM, Seet Q. Genome-Wide Saturation Mutagenesis of Burkholderia pseudomallei. MBio. 2014;5: 1–9.

5. van Opijnen T, Camilli A. A fine scale phenotype–genotype virulence map of a bacterial pathogen. Genome Res. 2012;22: 2541–2551.

6. Turner KH, Everett J, Trivedi U, Rumbaugh KP, Whiteley M. Requirements for Pseudomonas aeruginosa Acute Burn and Chronic Surgical Wound Infection. PLoS

Genet. 2014;10. doi:10.1371/journal.pgen.1004518

7. Ibberson CB, Stacy A, Fleming D, Dees JL, Rumbaugh K, Gilmore MS, et al. Co-infecting microorganisms dramatically alter pathogen gene essentiality during polymicrobial infection. Nature Microbiology. Nature Publishing Group; 2017;2: 1–6.

8. Chao MC, Abel S, Davis BM, Waldor MK. The design and analysis of transposon insertion sequencing experiments. Nat Rev Microbiol. Nature Publishing Group; 2016;14: 119–128.

9. Grenov AI, Gerdes SY. Modeling competitive outgrowth of mutant populations: why do essentiality screens yield divergent results? Methods Mol Biol. 2008;416: 361–367.

10. Burger BT, Imam S, Scarborough MJ, Noguera DR, Donohue TJ. Combining genome-scale experimental and computational methods to identify essential genes in Rhodobacter sphaeroides. mSystems. 2017;2: 1–18.

11. Broddrick JT, Rubin BE, Welkie DG, Du N, Mih N, Diamond S, et al. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. Proceedings of the National Academy of Sciences. 2016;113: E8344–E8353.

12. Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin J a. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol. Elsevier Ltd; 2012;20: 113–123.

13. Zampieri M, Enke T, Chubukov V, Ricci V, Piddock L, Sauer U. Metabolic constraints on the evolution of antibiotic resistance. 2017; 1–14.

14. Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BØ. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proceedings of the National Academy of Sciences. 2016; 201523199.

15. Jacobs M a., Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, et al. Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 2003;100: 14339–14344.

16. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, et al. An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A. 2006;103: 2833–2838.

17. Skurnik D, Roux D, Aschard H, Cattoir V, Yoder-Himes D, Lory S, et al. A Comprehensive Analysis of In Vitro and In Vivo Genetic Fitness of Pseudomonas aeruginosa Using High-Throughput Sequencing of Transposon Libraries. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003582

18. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proceedings of the National Academy of Sciences. 2015; 201419677.

19. Lee S a., Gallagher L a., Thongdee M, Staudinger BJ, Lippman S, Singh PK, et al. General and condition-specific essential functions of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences. 2015; 201422186.

20. Powell JE, Leonard SP, Kwong WK, Engel P, Moran NA. Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. Proc Natl Acad Sci U S A. 2016;113: 13887–13892.

21. Juhas M, Eberl L, Glass JI. Essence of life: essential genes of minimal genomes. Trends Cell Biol. 2011;21: 562–568.

22. Juhas M. Pseudomonas aeruginosa essentials: An update on investigation of essential genes. Microbiology. 2015;161: 2053–2060.

23. Bartell JA, Blazier AS, Yen P, Thøgersen JC, Jelsbak L, Goldberg JB, et al. Reconstruction of the metabolic network of Pseudomonas aeruginosa to interrogate virulence factor synthesis. Nat Commun. 2017;8. doi:10.1038/ncomms14631

24. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes Escherichia coli traits_supplement. Nat Biotechnol. 2017;35: 904–908.

25. Ghosh S, Baloni P, Mukherjee S, Anand P, Chandra N. A multi-level multi-scale approach to study essential genes in Mycobacterium tuberculosis. BMC Syst Biol. 2013;7: 132.

26. Zhu L, Lin J, Ma J, Cronan JE, Wang H. Triclosan resistance of Pseudomonas aeruginosa PAO1 is due to FabV, a triclosan-resistant enoyl-acyl carrier protein reductase. Antimicrob Agents Chemother. 2010;54: 689–698.

27. Lu Y-J, Zhang Y-M, Grimes KD, Qi J, Lee RE, Rock CO. Acyl-phosphates initiate membrane phospholipid synthesis in Gram-positive pathogens. Mol Cell. 2006;23: 765–772.

28. Kondakova T, D'Heygère F, Feuilloley MJ, Orange N, Heipieper HJ, Duclairoir Poc C. Glycerophospholipid synthesis and functions in Pseudomonas. Chem Phys Lipids. 2015;190: 27–42.

29. Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, et al. Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev. 2010;34: 952–985.

30. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. PLoS One.

2014;9: e86587.

31. Sridhar S, Steele-Mortimer O. Inherent Variability of Growth Media Impacts the Ability of Salmonella Typhimurium to Interact with Host Cells. PLoS One. 2016;11: e0157043.

32. Sezonov G, Joseleau-Petit D, D'Ari R. Escherichia coli physiology in Luria-Bertani broth. J Bacteriol. 2007;189: 8746–8749.

33. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential Bacillus subtilis genes. Proceedings of the National Academy of Sciences. 2003;100: 4678–4683.

34. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in mycobacteria. Proceedings of the National Academy of Sciences. 2001;98: 12712–12717.

35. Hutchison C a. III, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, et al. Global Transposon Mutagenesis and a Minimal *Mycoplasma* Genome. Science. 1999;286: 2165–2169.

36. Van Opijnen T, Camilli A. Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. Nat Rev Microbiol. Nature Publishing Group; 2013;11: 435–442.

37. Fu Y, Waldor MK, Mekalanos JJ. Tn-seq analysis of vibrio cholerae intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. Cell Host Microbe. Elsevier Inc.; 2013;14: 652–663.

38. Osterman AL, Gerdes SY. Comparative approach to analysis of gene essentiality. Methods Mol Biol. 2008;416: 459–466.

39. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, et al. Phenotypic landscape of a bacterial cell. Cell. 2011;144: 143–156.

40. Wang Z, Danziger SA, Heavner BD, Ma S, Smith JJ, Li S, et al. Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast. PLoS Comput Biol. 2017;13: e1005489.

41. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33: 2938–2940.

42. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6: 1290–1307.

43. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and

interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40: D109–14.

44. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359.

45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.

3.10 Supplementary Information

The supplemental files listed below can be found at the following website:

https://doi.org/10.18130/V3/Xf0AVD

Supplementary Data 1 - PAO1 candidate essential genes for *in vitro* screens
Candidate essential genes lists for each PAO1 transposon mutagenesis screen. Candidate essential genes are marked with a '1', while non-essential genes are marked with a '0'.

Supplementary Data 2 - PA14 candidate essential genes for *in vitro* screens
Candidate essential genes lists for each PA14 transposon mutagenesis screen. Candidate essential genes are marked with a '1', while non-essential genes are marked with a '0'.

Supplementary Data 3 - PAO1 model predicted essential genes for *in silico* screens
Model predicted essential genes lists for PAO1 growth simulated on LB media, Sputum media, Pyruvate minimal media, and Succinate minimal media. Model predicted essential genes are marked with a '1', while non-essential genes are marked with a '0'.

Supplementary Data 4 - PA14 model predicted essential genes for *in silico* screens
Model predicted essential genes lists for PA14 growth simulated on LB media and Sputum media. Model predicted essential genes are marked with a '1', while non-essential genes are marked with a '0'.

Supplementary Data 5 - PAO1 consensus metabolic essential/non-essential genes
Lists of consensus metabolic essential and non-essential genes for PAO1 on LB media and Sputum media.

Supplementary Data 6 - PA14 consensus metabolic essential/non-essential genes
Lists of consensus metabolic essential and non-essential genes for PA14 on LB media.

Supplementary Data 7 - Biomass precursors for PAO1 model predicted consensus essential genes

List of biomass precursors that cannot be synthesized when PAO1 model predicted consensus essential genes are removed from the model.

Supplementary Data 8 - Biomass precursors for PA14 model predicted consensus essential genes

List of biomass precursors that cannot be synthesized when PA14 model predicted consensus essential genes are removed from the model.

Supplementary Data 9 - Proposed model changes

Table of proposed model changes based on discrepancies between model predictions and consensus metabolic non-essential genes for PAO1 on LB.

Supplementary Data 10 - PAO1 model predicted essential genes for *in silico* screens for the updated PAO1 model

Model predicted essential genes lists for PAO1 growth simulated on LB media and Sputum media. Model predicted essential genes are marked with a '1', while non-essential genes are marked with a '0'.

Supplementary Data 11 - PA14 model predicted essential genes for *in silico* screens for the updated PA14 model

Model predicted essential genes lists for PA14 growth simulated on LB media. Model predicted essential genes are marked with a '1', while non-essential genes are marked with a '0'.

Chapter 4

Metabolic network analysis of *Pseudomonas aeruginosa* persister cells

Anna S. Blazier[1], Glynis Kolling[1], Alejandro Amezquita[2], Aline Metris[2], Jason A. Papin[1,3,4]

[1]Biomedical Engineering
[3]Medicine, Infectious Diseases & International Health
[4]Biochemistry & Molecular Genetics
University of Virginia; Charlottesville, Virginia, 22908; USA
[2]Saftey and Environment Assurance Centre
Unilever; Sharnbrook, Bedford MK44 1LQ; UK

4.1 Abstract

Persister cells are transient phenotypic variants in a bacterial population that are able to tolerate antimicrobial treatment. These cells have been implicated in the recalcitrant nature of chronic infections and in the resistance to disinfection and preservation systems that can lead to contamination of industrial products. While the tolerant nature of persister cells is classically associated with a reduced metabolic state, the characteristics of persister cell metabolism are not well understood. In this work, we perform an experimental and computational systems-level analysis to characterize the metabolic state of persister cells. We collected samples from *Pseudomonas aeruginosa* untreated and persister samples for transcriptomics and metabolomics profiling, revealing a distinct metabolic repertoire in persister cells marked by an increase in central metabolism activity. To aid in their analysis, we integrated these datasets with a *P. aeruginosa* genome-scale metabolic network reconstruction (GENRE), generating condition-specific models of the persister and untreated states. We then used this model of persister cell metabolism to propose targets of persister cell viability. Experimental testing of model predictions suggested that persister cells are robust to single gene deletions and that combinatorial targeting strategies may be necessary to completely inhibit the persister phenotype.

4.2 Introduction

Bacterial persistence is a transient, phenotypic state marked by the ability to withstand antimicrobial treatment [1]. When a culture is exposed to high doses of antimicrobial, a majority of the cells rapidly die; however, a subpopulation, known as persister cells, remain and tolerate treatment. Upon removal of the antimicrobial pressure, these persister cells can then switch into a normal phenotype and repopulate the culture [2]. Because of their transient, tolerant nature, persister cells have been implicated in the recalcitrant nature of chronic infections and in the resistance to disinfection [3,4]. Furthermore, the presence of persister cells may contribute to the development of resistance [5]. Despite their clinical and industrial relevance, relatively little is known about persister cell formation and maintenance across different bacterial species [6].

Traditionally, persister cells are thought to tolerate antimicrobial treatment due to a reduced metabolic state [7]. However, the specifics of persister cell metabolism are poorly understood. While some studies demonstrate that persistence is linked to arrested cell growth [8,9], others indicate that dormancy is not a prerequisite for persistence [10,11]. Furthermore, studies have shown that inactivation of the tricarboxylic acid cycle can differentially modulate persister levels depending on the organism being studied [12,13]. Nevertheless, the metabolism of persister cells has been shown to be important for their ability to tolerate antimicrobial treatment [14–17]. While individual metabolic genes have been linked to the persister phenotype [18], robust analyses of persister cell metabolism are lacking [11]. A systems level analysis of persister cell metabolism will enhance our understanding of the persister cell metabolic program and will enable the identification of promising targeting strategies to eliminate the persister phenotype.

In this work, we perform a systems-level analysis to quantitatively characterize the metabolic state of persister cells (Figure 4.1).

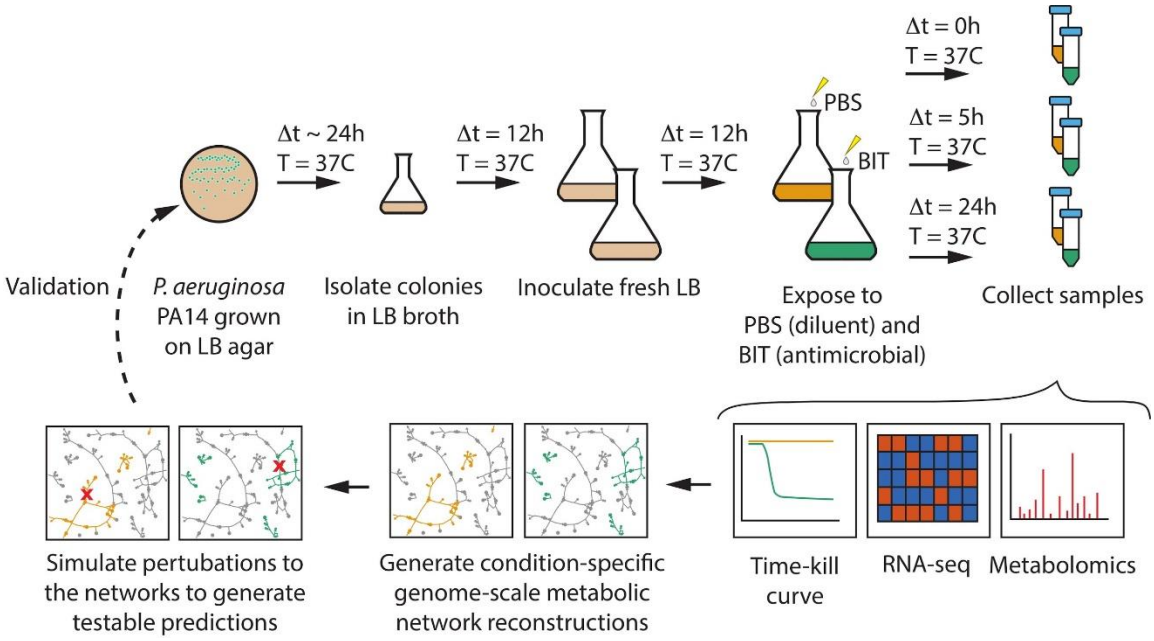


Figure 4.1. Experimental and computational systems biology approach for characterizing the metabolism of *Pseudomonas aeruginosa* PA14 persister.

We collected transcriptomics and metabolomics data from persister cells of the Gram-negative, opportunistic pathogen *Pseudomonas aeruginosa* exposed to the antimicrobial benzisothiazolinone (BIT). To facilitate the systems-level analysis of persister cell metabolism, we employed genome-scale metabolic network reconstructions (GENREs), which are computational modeling frameworks that capture the genotype-to-phenotype relationship and enable the study of metabolic capabilities of an organism of interest. Through the integration of transcriptomics and metabolomics datasets with a *P. aeruginosa* GENRE [19], we generated condition-specific metabolic models of the persister and untreated states. We then used these models to suggest targets of persister cell metabolism to inhibit persister cell viability.

4.3 Results

4.2.1 Persister cells enable Pseudomonas aeruginosa *to tolerate treatment with benzisothiazolinone*

To determine whether *P. aeruginosa* is able to tolerate treatment with BIT due to the presence of persister cells, we performed a time-kill assay (Figure 4.2A).

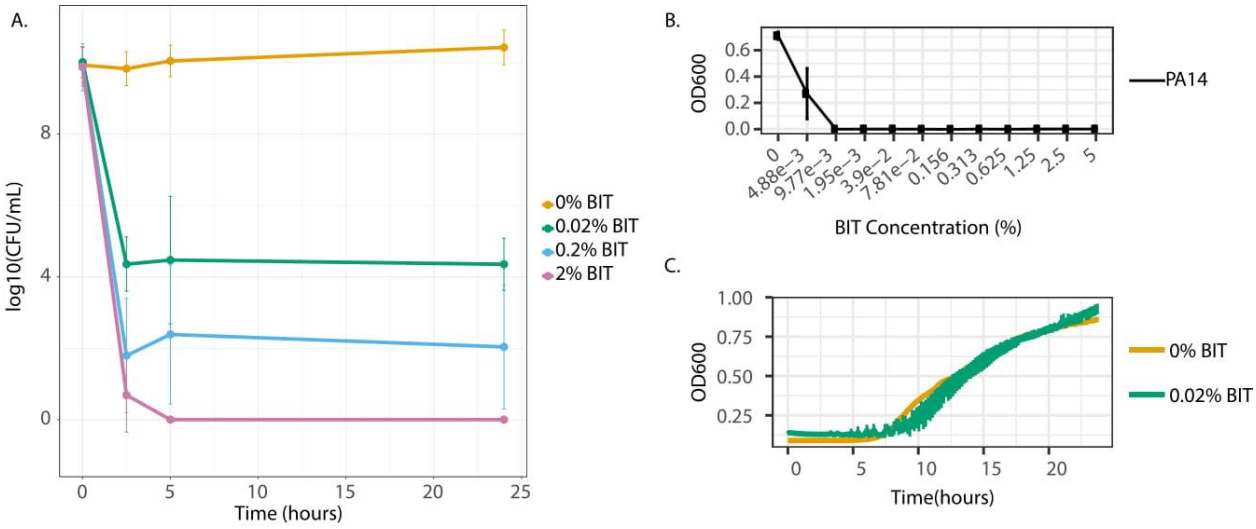


Figure 4.2. Killing of *Pseudomonas aeruginosa* PA14 by benzisothiazolinone (BIT) over time.
(A). Time-kill curve of *Pseudomonas aeruginosa* cultures exposed to the antimicrobial benzisothiazolinone (BIT). Stationary phase cultures of *P. aeruginosa* PA14 were exposed to various concentrations of BIT for 24 hours. The number of colony forming units (CFUs) in each culture were counted immediately before exposure at 0 hours and 2.5, 5, and 24 hours after exposure. Untreated cultures exposed to PBS are given by the yellow line and labelled as 0% BIT. The green line indicates cultures exposed to 0.02% BIT (10X the minimal inhibitory concentration, MIC). The blue line indicates cultures exposed to 0.2% BIT (100X the MIC). The pink line indicates cultures exposed to 2% BIT (1000X the MIC). (B). Dose response curve of *P. aeruginosa* treated with BIT. Overnight cultures of PA14 were diluted and exposed to various concentrations of BIT. Growth was measured 24-hours after inoculation via optical density (OD600) measurements. (C). Re-growth of untreated and 0.02% BIT treated *P. aeruginosa* samples. Untreated and 0.02% BIT treated samples were diluted to the same starting CFU concentration in LB media. Growth was measured for 24-hours using optical density (OD600) measurements.

We exposed stationary-phase cultures of *P. aeruginosa* to various concentrations of BIT corresponding to 0x, 10x, 100x, and 1000x the minimal inhibitory concentration (MIC) of BIT against *P. aeruginosa* (Figure 4.2B). We measured the number of colony forming units

(CFUs) in the culture immediately before exposure, 2.5, 5, and 24 hours after exposure. The time-kill curves of the 10x and 100x cultures exhibited biphasic behavior marked by an initial high rate of killing followed by a plateau in killing. This biphasic behavior is indicative of the presence of persister cells in the culture. In contrast, the 1000x MIC culture exhibited complete killing by 5 hours after exposure. These results indicate that at concentrations as high as 100x the minimal inhibitory concentration, *P. aeruginosa* is able to tolerate treatment with BIT due to the presence of persister cells.

To further confirm that the cells that survived treatment are indeed persister cells, we measured the re-growth of the untreated and 10X MIC (0.02% BIT) conditions (Figure 4.2C). When the samples were diluted to the same starting CFU concentration, there was not an appreciable difference in the growth profiles of the two samples. These results further confirm that persister cells enable *P. aeruginosa* to tolerate treatment with the antimicrobial BIT.

4.2.2 The transcriptional state of persister cells is distinct from untreated cells

We were then interested in comparing the transcriptional state of *P. aeruginosa* persister cells in the presence of BIT to untreated cells. We collected samples for RNA-sequencing from the 0x and 10x MIC conditions (henceforth referred to as the untreated and persister conditions, respectively) immediately before exposure (i.e., time 0) as well as 5 and 24 hours after exposure. Differential expression analysis of the RNA-sequencing data revealed that persister cells are in a distinct transcriptional state from untreated cells (Figure 4.3).

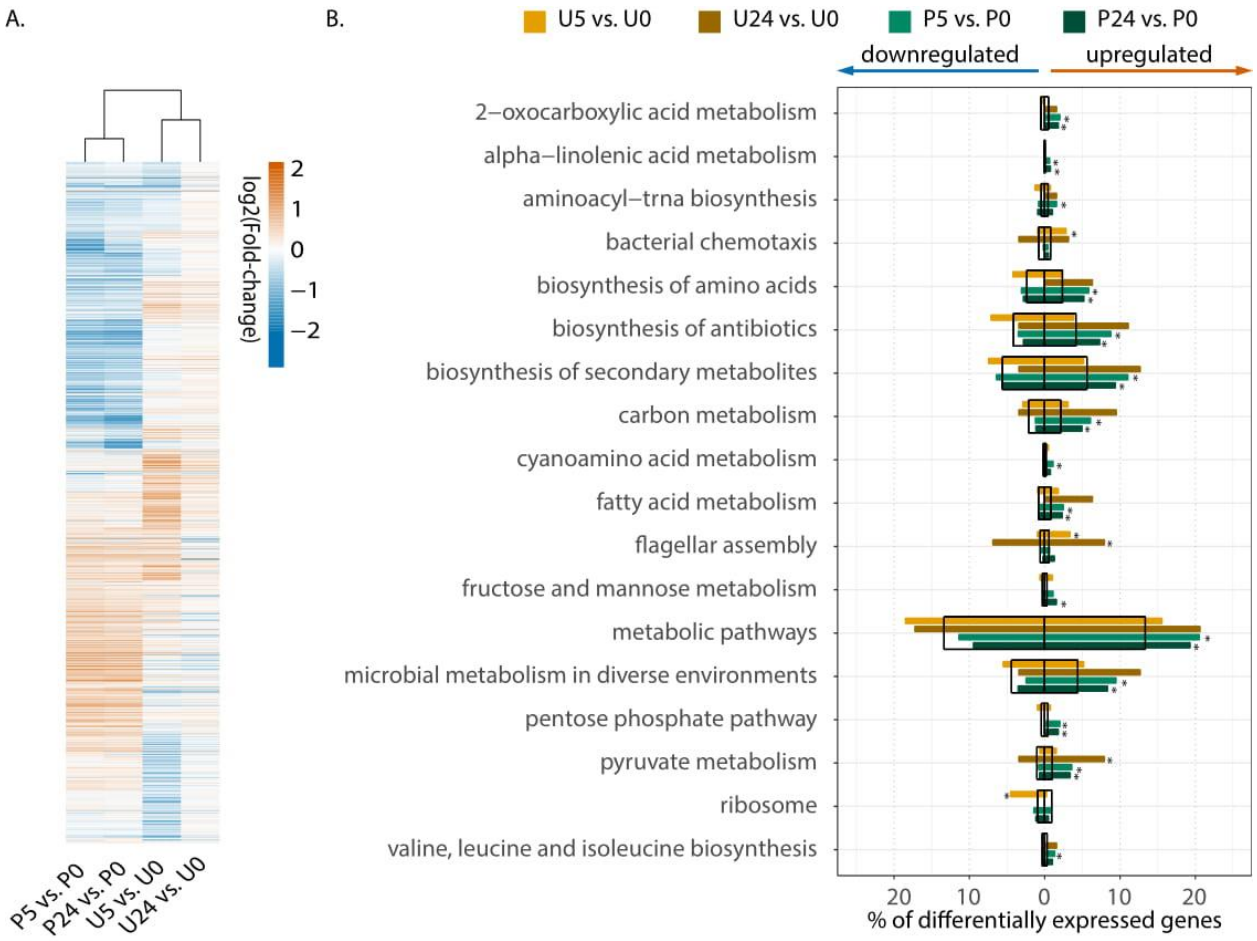


Figure 4.3. RNA-sequencing analysis of *P. aeruginosa* persister cells.
Samples were collected for RNA-sequencing from the untreated and persister conditions (0 and 10X the MIC, respectively) immediately before exposure and 5 and 24 hours exposure. (A). Differential expression analysis was performed by comparing the 5 and 24 hour samples to their respective time 0 samples. The heatmap shows the log2(fold-change) for genes significantly differentially regulated in at least one comparison. (B). Gene set enrichment analysis based on KEGG pathway classification was performed on the significantly up- and down-regulated genes for each comparison. Black boxes represent the fraction of the entire genome represented in the categories.

When comparing the 5 and 24 hour samples of the persister and untreated conditions to their respective time 0 samples, the two persister conditions clustered separately from the untreated conditions (Figure 4.3A). Indeed, the persister conditions are marked by distinct regions of downregulation and upregulation compared to the untreated conditions.

We then assessed the overlap between the significantly upregulated genes for the persister and untreated conditions. The two persister conditions had the largest number of shared upregulated genes followed by genes uniquely upregulated in the untreated condition at 5 hours after exposure (Figure 4.4A).

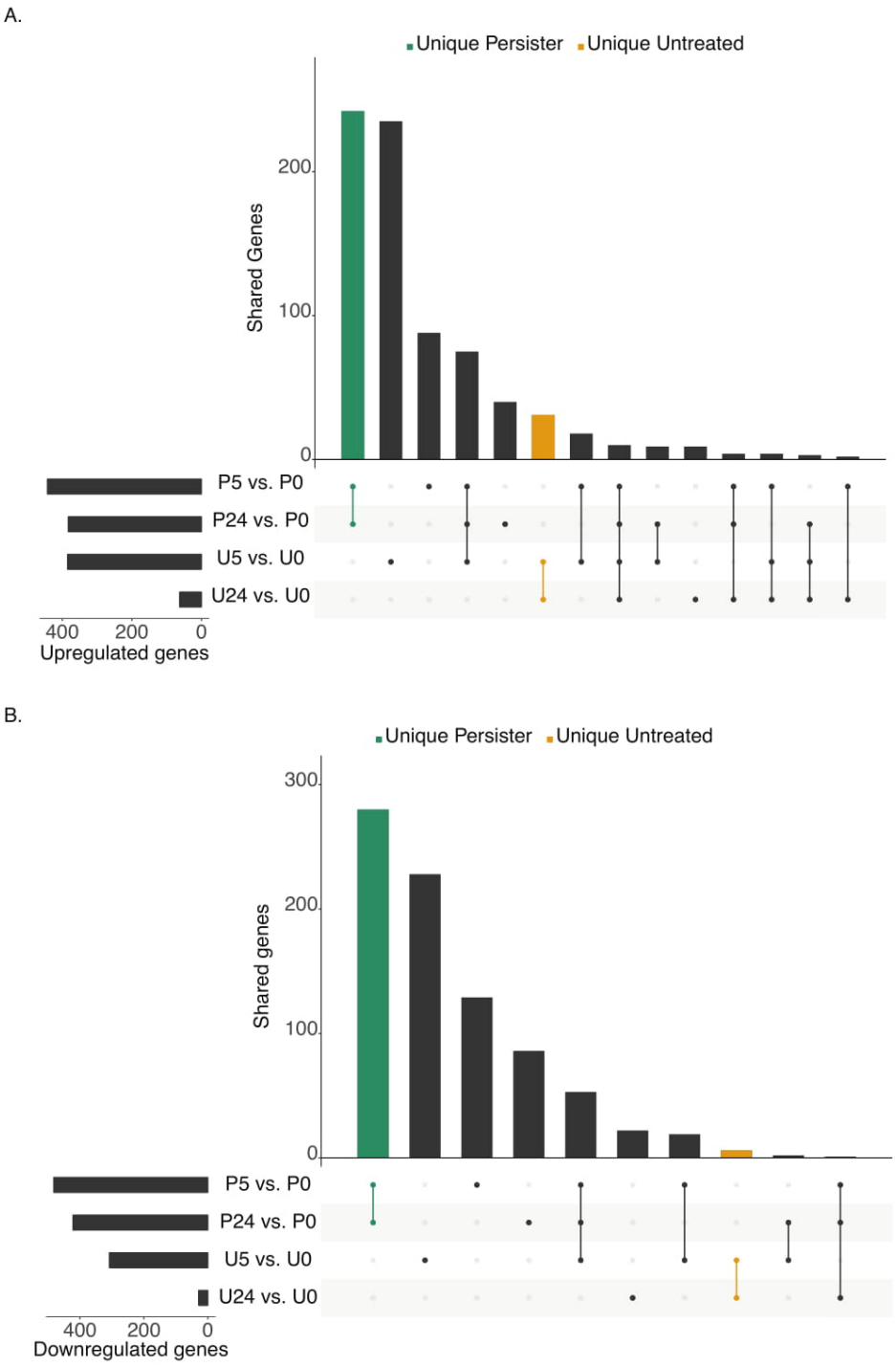


Figure 4.4. Overlap of upregulated (A) and downregulated (B) genes across all four comparisons.

The same comparison was done for the downregulated genes across the conditions. Similarly to the upregulated genes, the two persister conditions had the largest number of shared downregulated genes (Figure 4.4B). Overall, there were 522 genes commonly differentially expressed in the two persister conditions but not in the untreated conditions.

To get a sense for systems level changes in the expression data, we performed gene set enrichment analysis based on KEGG pathway classification systems (Figure 4.3B). Interestingly, we found that the upregulated genes in the persister conditions were significantly enriched for several metabolic pathways, such as those involved in carbon metabolism and the biosynthesis of secondary metabolites. In contrast, the upregulated genes in the untreated samples were enriched for motility pathways, such as flagellar assembly for both timepoints and chemotaxis for the untreated at 5 hours after exposure sample. Pyruvate metabolism was the only metabolic pathway significantly enriched in the untreated sample 24 hours after exposure. No metabolic pathways were significantly enriched for the untreated sample 5 hours after exposure. These enrichment results suggest an overall shift in the metabolic state of persister cells compared to untreated samples. Furthermore, because the enriched metabolic pathways were observed in the upregulated genes of the persister conditions, this suggests that persister cells are in a more metabolically active state than previously thought. Interestingly, few pathways were significantly enriched for the downregulated genes. This lack of enrichment suggests a widespread decrease in functionality across cellular subsystems for all conditions.

4.2.3 The metabolite footprint of persister samples is different from untreated and dead samples

To probe the metabolism of *P. aeruginosa* persister cells further, we collected samples for metabolomics profiling of culture supernatants. Similarly to the RNA-sequencing analysis, we collected samples from the persister and untreated conditions immediately before exposure and 5 and 24 hours after exposure. Additionally, we collected samples from the 1000X MIC condition, henceforth referred to as the dead condition. We included this dead condition to filter metabolites associated with dying from the persister condition. Analysis of the relative abundance of metabolites across the samples revealed that the persister metabolite footprint is distinct from the untreated and dead conditions (Figure 4.5A).

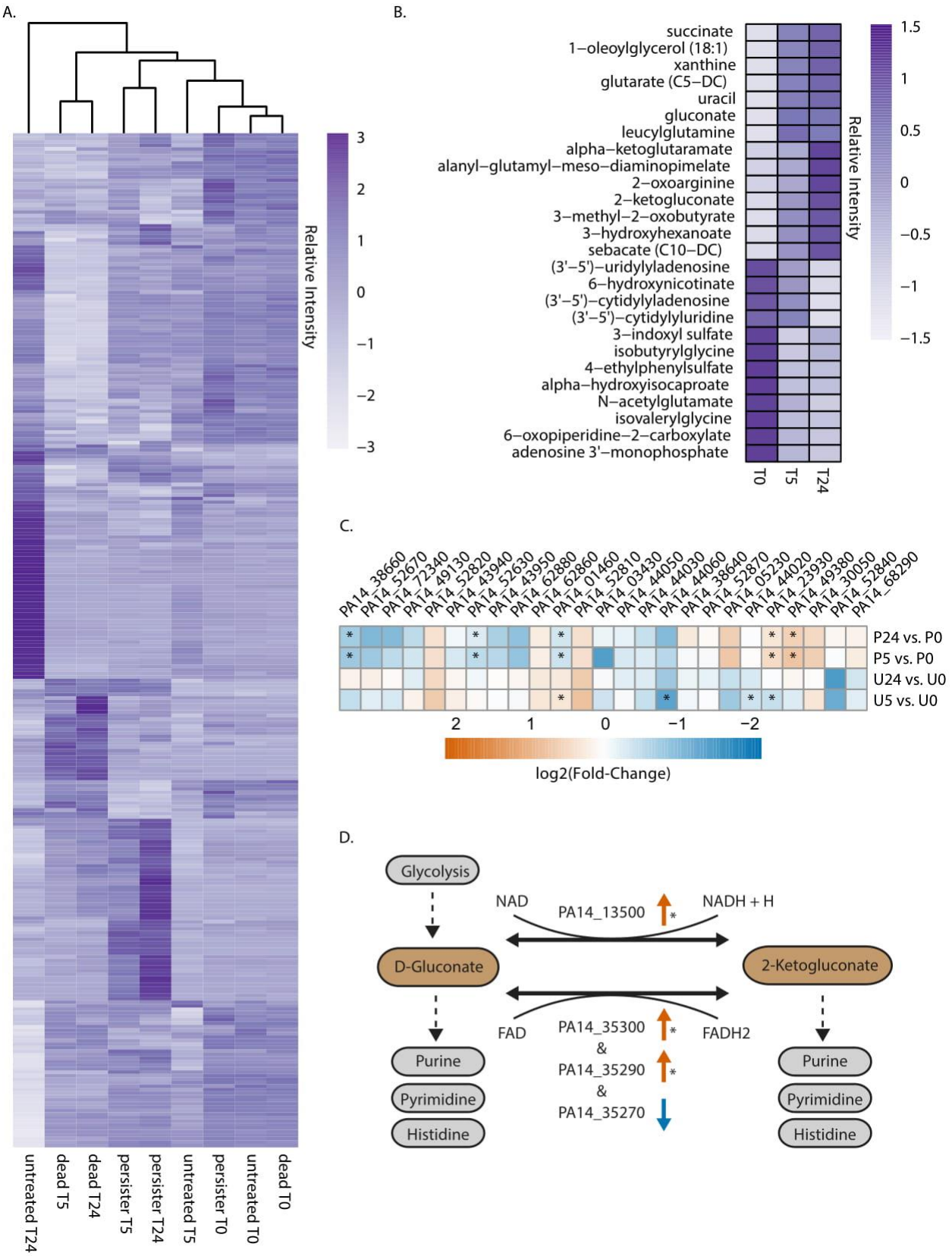


Figure 4.5. Metabolomics profiling of culture supernatants.

Samples were collected for metabolomics profiling from the culture supernatants of the untreated, persister, and dead conditions (0, 10X, and 1000X the MIC, respectively) immediately before exposure and 5 and 24 hours exposure. (A). Metabolites that significantly changed in abundance over time for each condition were identified. The heatmap shows the relative abundance for a metabolite in a given sample compared to the other samples. (B). Metabolites that changed in abundance over time only in the persister condition were identified. The heatmap shows the relative abundance for a metabolite in a given timepoint compared to the other timepoints. (C) Heatmap of the log2(fold-change) for genes involved in succinate metabolism. Asterisks denote significant differential expression ($p < 0.01$). (D) Metabolism of D-Gluconate and 2-Ketogluconate. Brown boxes refer to metabolites while grey boxes refer to other metabolic pathways. Orange arrows indicate upregulation in persister conditions relative to time 0 samples. Blue arrows indicate downregulation in persister conditions relative to time 0 samples. Asterisks denote significant differential expression ($p < 0.01$).

The two BIT-exposed persister conditions clustered separately from the other conditions. Furthermore, certain metabolites were markedly more abundant in the persister samples relative to the other samples. For example, 20 metabolites were significantly more abundant in the persister condition at both 5 and 24 hours after exposure compared to the other conditions at those respective time points.

In contrast, 60 metabolites significantly changed in abundance for every condition (Figure 4.6).

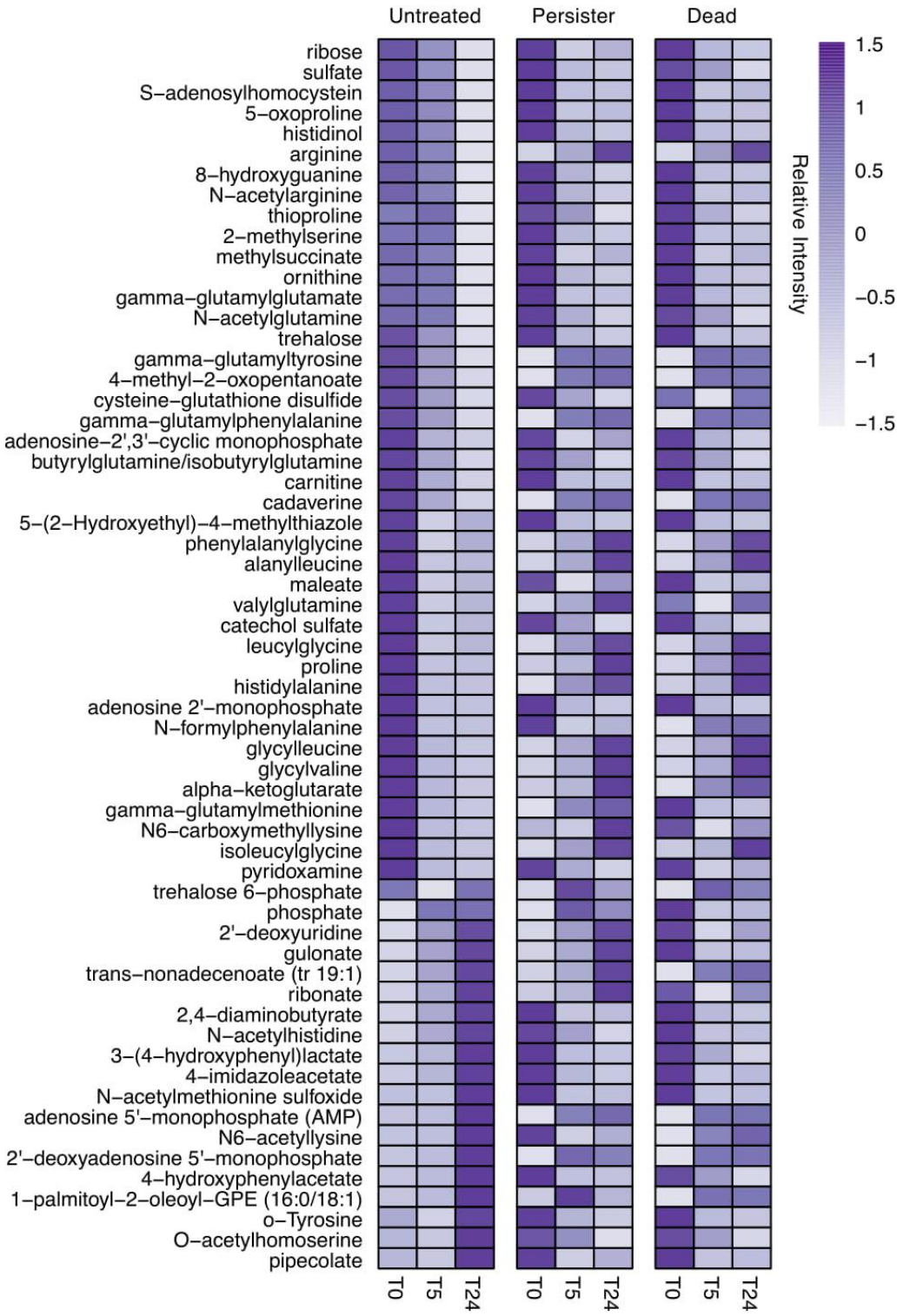


Figure 4.6. Metabolites that significantly change in abundance over time common to the untreated, persister, and dead conditions.

Interestingly, while each of these 60 metabolites changed in abundance for every condition, it varied whether they were produced or consumed across the conditions. For example, while ribose decreased in abundance for each condition, arginine only decreased in abundance for the untreated condition but increased in abundance for the persister and dead conditions. The increase in abundance of arginine in the persister and dead samples is likely due to the prevalence of dying cells in both conditions. Other metabolites varied in abundance trajectories across all three conditions. For example, N6-acetyllysine increased rapidly in abundance for the dead condition, slowly for the untreated condition, and instead decreased in abundance for the persister condition. Lysine acetylation has been associated with cell-cell signaling, transcriptional regulation, and cell survival pathways in mammalian systems [20]. Mounting evidence suggests that lysine acetylation in prokaryotes plays a similar role [20,21]. Thus, this difference in N6-acetyllysine abundance between the persister and untreated and dead conditions may indicate a difference in regulatory activity.

Additionally, we identified metabolites that uniquely varied across time for a given condition (Figure 4.5B and 4.7).

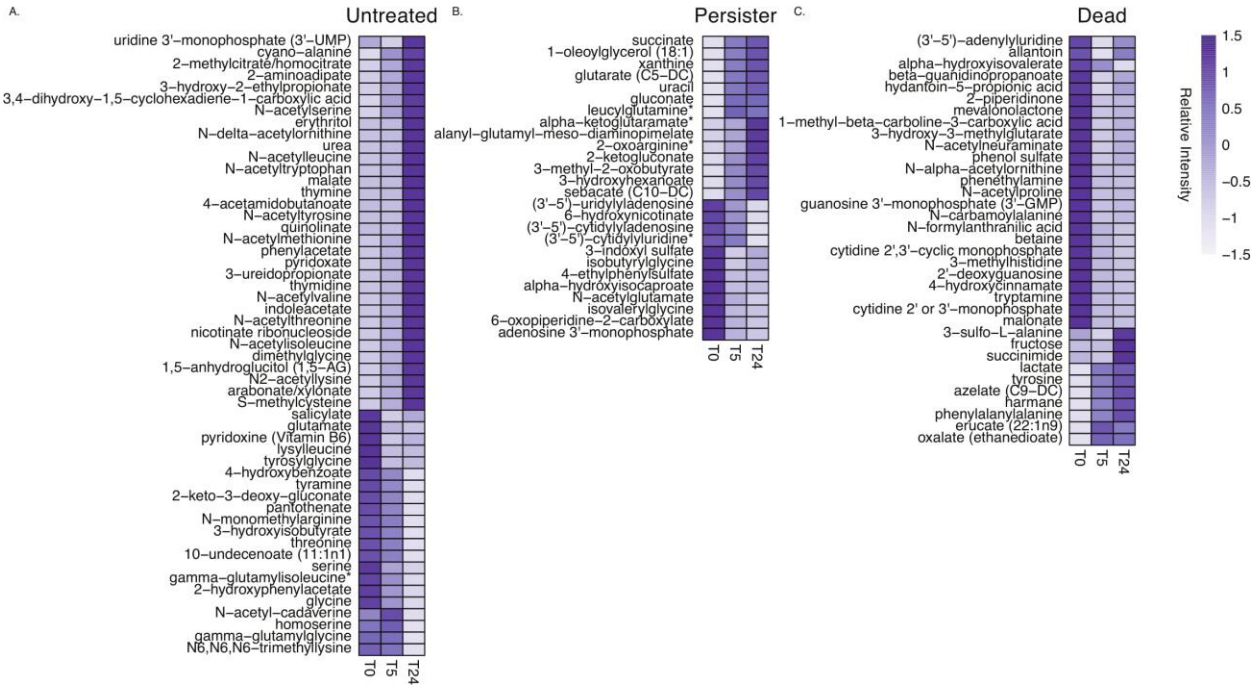


Figure 4.7. Metabolites that significantly change in abundance over time unique to the untreated (A), persister (B), and dead (C) conditions.

For example, succinate increased in abundance over time for the persister condition but not for the untreated and dead conditions. Filtering the data this way allowed us to determine metabolites uniquely associated with the untreated state, those associated with dying, and those associated with the persister phenotype. Overall, 53 metabolites were uniquely associated with the untreated condition, 35 with dying, and 26 with the persister phenotype.

Transcriptomics analysis revealed that genes associated with the production and consumption of these unique persister metabolites were differentially regulated. For example, the metabolomics data suggested that succinate production is uniquely associated with the persister phenotype. The RNA-sequencing data also indicated a difference in succinate metabolism, with unique differential regulation of succinate-associated genes across the conditions (Figure 4.5C). Succinate is a key intermediate in the tricarboxylic acid (TCA) cycle. Previous studies have implicated the TCA cycle in persister cell formation and

viability. Indeed, one study showed that accumulation of fumarate, the product of succinate oxidation, enhanced persister formation [22]. Other studies showed that mutants of TCA genes, such as *sdh* and *sucB*, had significantly reduced persister levels [13,22,23]. Interestingly, other studies have found conflicting roles for the TCA cycle in persister formation and viability. For example, one study showed that introducing succinate sensitized stationary-phase *P. aeruginosa* cells to tobramycin [24]. However, given the overproduction of succinate in our persister samples, our results suggest that succinate actually serves as a biomarker of the persister phenotype, not a potentiator for antimicrobial activity. Indeed, the previous study showed that carbon-source induced potentiation was specific to tobramycin and not applicable to other antibiotics. Additionally, another study found that inhibitors of the TCA cycle actually enhanced persister formation in Gram-positive bacteria [12]. Altogether, these results indicate that the TCA cycle plays an important, albeit heterogeneous, role in persister viability depending on the stressor.

Additionally, both the metabolomics data and transcriptomics data indicated enhanced gluconate metabolism in persister conditions (Figure 4.5D). Both gluconate and 2-ketogluconate were uniquely produced in the persister condition. Additionally, genes associated with the conversion of the two metabolites were significantly upregulated in the persister samples but not the untreated samples. Both gluconate and 2-ketogluconate are important precursors of purine and pyrimidine metabolism; however, the role of these two metabolites is understudied. Interestingly, previous research implicated gluconate accumulation with multi-drug resistant clinical isolates [25]. Together with our results, gluconate accumulation appears to be associated with antimicrobial inefficacy and may

serve as a good biomarker for resistance. Future work will be necessary to determine if targeting gluconate metabolism is a potential antimicrobial strategy.

4.2.4 Condition-specific modeling of the persister and untreated states

Finally, to aid in the interpretation of the metabolomics and transcriptomics datasets, we generated condition-specific models of the persister and untreated metabolic states. To build the models, we integrated the transcriptomics and metabolomics datasets with a previously published genome-scale metabolic network reconstruction (GENRE) of *Pseudomonas aeruginosa* strain PA14 [19]. We first integrated the RNA-sequencing dataset with the GENRE using the MADE algorithm [26], which modifies gene functionality in the model based on significant differential expression. Then, we forced production of timepoint and condition-specific metabolites based on the metabolomics data. Because the structure of the condition-specific models changed depending on the parameters used during the integration of the RNA-sequencing data, we repeated the integration process for multiple parameter sets. Upon integration of the gene expression and metabolomics datasets with the PA14 metabolic model, we generated condition-specific models for the metabolism of the persister and untreated states at 5 and 24 hours after exposure to BIT. The resulting models differed in their functionality. For instance, the untreated models for the 24-hour timepoint contained the most metabolites with forced production based on the metabolomics data, followed by the persister models for the 5-hour timepoint (Figure 4.8).

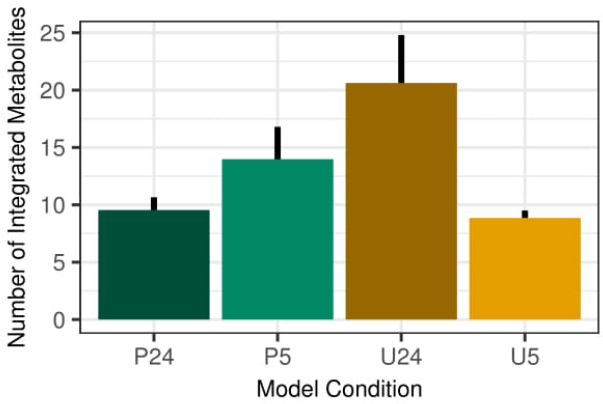


Figure 4.8. Integrated metabolites for condition-specific models.
The number of metabolites with forced production based on the metabolomics data for the condition-specific models.

These differences in metabolite production highlight unique functionality of the networks and point to potential vulnerabilities of the individual conditions.

4.2.5 Metabolic network analysis suggests targets for inhibiting persister viability

Using these condition-specific metabolic models, we sought to identify genes that, when inhibited, interrupted persister cell viability. A central challenge to generating these predictions is using an appropriate objective function that captures the persister state. For our analyses, we explored two potential objective functions: minimal biomass production as an approximation for growth and minimal ATP production. We chose to investigate biomass production because it is a standard, well-studied objective function in the field of metabolic modeling. However, because persister cells are traditionally thought to be in a dormant metabolic state, biomass production may not be the best suited objective function. To complement our analysis, we also investigated ATP production, which has previously been shown to be associated with persistence [16]. While these objective functions are reasonable

starting points, they may be limited in their applicability because the overall goal of persister metabolism is unclear.

Given this uncertainty, for both objective functions, we implemented three different strategies to identify potential gene targets. In the first strategy, we performed *in silico* single gene knockouts on the models and identified essential genes. We repeated this simulation for models generated with different MADE parameters and tallied the number of times a gene was predicted as essential across the different iterations. In the second strategy, we used OptGpSampler to perform flux sampling on the networks [27], providing information on the distribution of fluxes each reaction in the network can carry. We then used this flux sampling information to determine reactions that had a greater likelihood of carrying flux in the persister models compared to the untreated models. Once these reactions were identified, we found the genes associated with them. In the third strategy, we used the flux sampling data to identify reactions strongly correlated with essential reactions in the models. Once these reactions were identified, we found the genes associated with them. We then combined the predictions generated from these three strategies to generate a list of candidate target genes. To arrive at a set of genes for experimental testing, we filtered this list of candidate target genes based on several factors such subsystem diversity, availability of transposon mutants in the PA14 mutant library [28], and the presence of enzymatic subunits.

Ultimately, we experimentally tested 20 mutants that were predicted to be critical for the biomass and ATP objective functions analyzed. Following the same experimental procedure as before, we exposed the mutants to 10X the MIC of BIT and performed a time-kill assay (Figure 4.9).

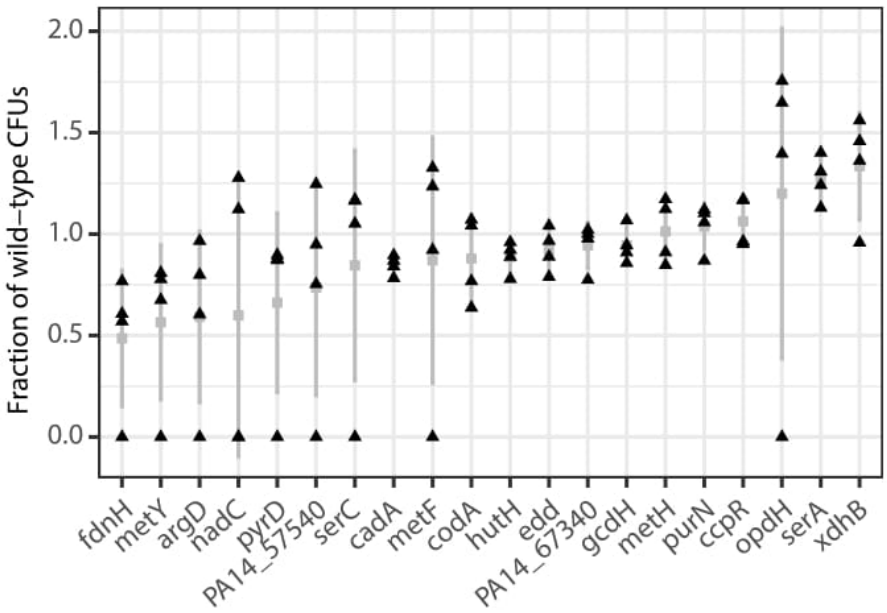


Figure 4.9. Time-kill assays for P. aeruginosa mutants.
Mutants of genes predicted to be critical for persister viability were exposed to 10X the MIC of BIT (0.02% BIT) for 24 hours. The number of CFUs in the culture at 24 hours after exposure were normalized to the number of CFUs in the culture immediately before exposure. The normalized 24 hour CFU counts were then compared to wild-type for each mutant.

Of 20 mutants tested, one mutant, fdnH, exhibited a significant, yet incomplete, reduction in the number of CFUs after 24 hours of BIT exposure ($p < 0.05$). The gene product of fdnH, nitrate-inducible formate dehydrogenase subunit beta, is involved in the breakdown of formate, an electron acceptor [29]. In the *P. aeruginosa* metabolic network, there are two enzymes that breakdown formate, formate dehydrogenase and 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase. Interestingly, the gene for 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase is downregulated in expression in the persister samples, suggesting that the only route for formate breakdown is through formate dehydrogenase. Thus, when a gene encoding one of the formate dehydrogenase subunits is mutated, the breakdown of formate cannot proceed in the persister state. Ultimately, as indicated by the

metabolic model, this prevents persister cells from being able to generate the metabolite 5-10-Methylenetetrahydrofolate, which is important for the synthesis of thymidine and various amino acids.

The discrepancies between the model predictions and the experimental results can provide insight into gaps in our understanding of *P. aeruginosa* persister cell metabolism. For example, the models predicted that xanthine dehydrogenase is essential for the persister state but not the untreated state. However, when we tested a mutant for one of the subunits of xanthine dehydrogenase (*xdhB*), there was no impact on persister cell viability. Based on the metabolomics data, xanthine production was uniquely associated with the persister state at both 5 and 24 hours after exposure to BIT. Because of this evidence, we forced production of xanthine in the persister models. In the base *P. aeruginosa* metabolic network, there are two routes to xanthine production, one through xanthine dehydrogenase and the other through guanine deaminase. However, the reaction catalyzed by guanine deaminase is blocked. As a result, xanthine production is forced through the reaction catalyzed by xanthine dehydrogenase, ultimately making the genes encoding the subunits of xanthine dehydrogenase essential in the persister models. However, the results from the time-kill assay for the *xdhB* mutant suggest that both routes for xanthine synthesis are accessible *in vitro*. Because of the unique production of xanthine in the persister state, we hypothesize that a combination targeting approach that removed the functionality of both xanthine dehydrogenase and guanine deaminase would impact persister viability.

Additionally, the models predicted the gene *purN* to be uniquely critical for the persister state. The gene product of *purN* is phosphoribosylglycinamide formyltransferase, which is involved in purine metabolism and important for the synthesis of several biomass

precursors. In *P. aeruginosa*, there is an isozyme for phosphoribosylglycinamide formyltransferase encoded by the gene *purT*. According to the RNA-sequencing data, purT is significantly downregulated in the persister conditions at both time points ($p < 10^{-10}$), suggesting that the preferred route for phosphoribosylglycinamide formyltransferase generation is through *purN*. Because a mutant for *purN* did not exhibit a reduction in the persister state, this implies that the mutant shifted phosphoribosylglycinamide formyltransferase generation to *purT*. Based on these results, we hypothesize that a double knockout of *purN* and *purT* would inhibit persister viability. Alternatively, a compound that targets the enzymatic activity of phosphoribosylglycinamide formyltransferase, such as Pemetrexed, may inhibit persister viability [30].

Overall, the incomplete killing of the *P. aeruginosa* mutants and the variability in the data suggest that the persister phenotype is robust to single gene deletions. This result is consistent with literature. Two previously published screens of *P. aeruginosa* mutant libraries attempted to identify genes associated with persistence. One of these screens tested 5000 mutants and found four mutants with a reduced persister phenotype [31]. The other tested 4,411 mutants and found 118 mutants with a reduced persister phenotype [6]. Of the mutants identified, none of them completely eliminated the persister state, potentially due to a difference in the expression states between the wildtype and mutant strains. Additionally, mutants of interest did not decrease the level of persistence to the same extent across different antimicrobials. Altogether, our data and data from literature suggest that persistence is complex and robust to single gene deletions. To more completely inhibit the persister state, we propose that multi-gene targeting will be necessary. Through analysis of

the transcriptomics and metabolomics dataset with the metabolic network reconstructions of *P. aeruginosa*, we were able to suggest combination targeting strategies.

4.4 Discussion

Traditionally, persister cells are thought to be able to tolerate antimicrobial treatment due to a reduced metabolic state. However, the extent of persister cell metabolism is not well understood. In this study, we employed both experimental and computational systems biology techniques to better characterize the metabolism of persister cells and suggest vulnerabilities in persister cells metabolism. We applied this approach to persister cells isolated from *Pseudomonas aeruginosa* cultures exposed to the industrial antimicrobial benzisothiazolinone. We collected both transcriptomics and metabolomics data on persister cell samples, which indicated that persister cells are in a distinct metabolic state relative to untreated samples marked by increased activity in central metabolism. We then integrated the transcriptomics and metabolomics datasets with a genome-scale metabolic network reconstruction of *P. aeruginosa* to generate a condition-specific model of persister cell metabolism. Using this model and information from the experimental data, we suggested targets for inhibiting the viability of persister cells. The resulting insights about vulnerabilities in persister cell metabolism would be challenging to develop without the use of a computational model of *P. aeruginosa* metabolism. Ultimately, this experimental and computational approach enabled the characterization of persister cell metabolic activity and the elucidation of potential targets that may interrupt persister cell functionality.

Both the transcriptomics and metabolomics datasets indicate that persister cells are not in a dormant state. Rather, these two datasets suggest that persister cells are active in

central metabolic processes. For example, gene set enrichment analysis on the RNA-sequencing data revealed that upregulated genes in the persister conditions are significantly enriched in carbon metabolism, pyruvate metabolism, and the biosynthesis of amino acids. These results contrast the enrichment results for the untreated conditions, which were enriched for motility pathways. Furthermore, analysis of the metabolomics data revealed certain metabolites uniquely associated with the persister state. For example, gluconate and 2-oxoarginine, intermediates in amino acid metabolism, were uniquely produced in the persister condition. Additionally, both the transcriptomics and metabolomics data indicated heightened metabolism of succinate, a key metabolite of the TCA cycle, and gluconate, a key metabolite in the pentose phosphate pathway. Together, these data add to the growing body of evidence that the persister state is a metabolically active state [11,13,32,33]. However, because no single pathway was uniquely associated with the persister state, this points to the heterogeneity and complexity of the persister phenotype.

There is a large effort to identify genetic determinants of persistence. Several studies have performed high-throughput screens of mutant libraries to identify genes with altered persister levels [6,31,34,35]. A limitation of these screens is that they fail to capture the biphasic rates of killing of a bacterial culture by an antimicrobial, which is indicative of the presence of persister cells. Time-kill assays traditionally used to isolate persister cells are, by nature, not amenable to high-throughput screening approaches. To overcome this limitation, we sought to expedite the identification of genetic determinants of persistence using condition-specific genome-scale metabolic network reconstructions for the persister and untreated states. Using these models, we predicted targets critical for persister biomass production and ATP production. Upon experimentally testing these model predictions, we

found that the persister phenotype is overall robust to single gene deletions, consistent with previous findings [6,31]. Most mutants had no impact on persister levels. Even more so, the mutants that did have reduced persister levels did not exhibit a complete loss of persistence. These results add to the growing body of literature that there is not a single persistence mechanism [32,36,37]. Rather, persistence may be mediated through a variety of routes within a cell and these routes may differ depending on the stressor applied.

To this end, we employed the persister condition-specific model to suggest combinatorial targeting strategies that may inhibit this complex phenotype. For example, based on the metabolomics data, xanthine production is uniquely associated with the persister phenotype. Using the persister model, we found two routes for xanthine production in *P. aeruginosa* through xanthine dehydrogenase and guanine deaminase. While inhibition of xanthine dehydrogenase alone did not reduce the persister phenotype, we hypothesize that combinatorially targeting both enzymes will inhibit the persister state. Future work will be necessary to test these multi-target approaches. However, a recent study showed that single-drug tolerant persister cells could be eliminated by groups of three antibiotics, demonstrating the potential of combination strategies [38]. The identification of combinatorial targeting approaches would be challenging to delineate without the use of computational models.

In summary, persister cell metabolism is distinct from untreated cells and may be more metabolically active than previously appreciated. Characterization of the persister metabolic state enabled the identification of processes unique to the persister phenotype. These processes may be crucial to persister cell functionality. Further work needs to be performed to determine if these attributes are common to persister cells generated by other

stressors as well as persister cells generated in other bacteria. Finally, computational modeling of persister cell metabolism enabled the identification of potential targets that may interrupt persister cell functionality. Further studies, including combinatorial gene knockout strategies, will need to be performed to validate these computational predictions. In generating condition-specific models of persister and untreated cell metabolism, we provide the community with tools to further probe the metabolic capabilities of *P. aeruginosa* persister cells. Additionally, the transcriptomics and metabolomics datasets serve as invaluable resources to study the complex state of persistence from a systems perspective.

4.5 Materials and Methods

4.5.1 Bacterial strains and growth conditions

Wild-type *Pseudomonas aeruginosa* PA14 and PA14 single gene knock-out mutants from the PA14 non-redundant genome-scale transposon library [28] were grown in LB media supplemented with 15 ug/mL gentamicin as necessary at 37 and 125 RPM for liquid cultures. Desired concentrations of BIT were made by diluting the antimicrobial in PBS.

4.5.2 Minimal inhibitory concentration assay

A frozen stock of PA14 was streaked on an LB agar plate and grown overnight. Individual colonies were inoculated into 10 mL of LB and grown overnight with aeration. The culture was then diluted to an OD600 of 0.001 (approximately 10^6 CFU/mL) and inoculated into a 96-well plate containing varying concentrations of BIT. The plate was incubated overnight in a plastic container to prevent evaporation at 37 C and 125 RPM. Approximately 16-hours after exposure, the OD600 was measured using a plate reader (Tecan Infinite M200 Pro) as an approximation for growth. Background subtraction was

performed and the minimal inhibitory concentration was defined as the lowest concentration that prevented growth of PA14 (i.e., OD600 < 0.1).

4.5.3 Persister isolation and re-growth

A frozen stock of PA14 wild-type or PA14 mutants was streaked on an LB agar plate and grown overnight. Individual colonies were inoculated into a 50 mL flask containing 10 mL of LB and grown overnight with aeration. The overnight cultures were diluted in fresh LB in 500 mL flasks to an OD600 of 0.01 for a total volume of 50 mL and grown for 12 hours to reach stationary phase. The cultures were then subsequently exposed to BIT by removing 5 mL of the culture and adding 5mL of antimicrobial at various concentrations (PBS for the untreated condition, 0.2% BIT for the 10X MIC condition, 2% BIT for the 100X MIC condition, and 20% BIT for the 1000X MIC condition). The flasks were returned to the incubator for 24 hours. One microlitre samples were collected to determine the number of viable cells in the culture immediately before exposure (0 hr), 2.5, 5, and 24 hours after exposure. Samples were washed in PBS twice followed by resuspension in 1 mL of PBS. The samples were then serially diluted in PBS and 10 uL was plated for each dilution on LB agar to measure the number of colony forming units (CFU).

To verify the re-growth of the persister cells, 1 mL samples were collected from the untreated and 10X MIC conditions. Samples were washed in PBS twice and resuspended in 1 mL PBS. The 10X MIC condition was diluted 1:100 and the untreated condition was diluted to approximately the same CFU concentration as the diluted 10X MIC condition (~10^3 CFU/mL). The diluted cultures were plated in a 96-well plate and the OD600 was measured using a plate reader (Tecan Infinite M200 Pro) for 24 hours.

4.5.4 RNA-sequencing

One millilitre samples for RNA-sequencing were collected immediately before exposure, 5 and 24 hours after exposure from the untreated and 10X MIC conditions. Samples were mixed with 2 mL of RNAprotect (Qiagen) to stabilize RNA and frozen at -80 C. Samples were sent to Genewiz for further processing. Ribosomal RNA was depleted using the Ribo-Zero rRNA removal kit (Illumina). Libraries were sequenced on an Illumina HiSeq, 2x150bp configuration.

Reads were aligned to the reference *P. aeruginosa* PA14 genome (NC_008463.1) using the BWA aligner with the MEM algorithm with default parameters [39]. Aligned reads were tallied for each gene using FeatureCounts [40]. Differential expression was then determined with DESeq2 [41]. Gene set enrichment analysis was performed using the procedure outlined in [42].

4.5.5 Metabolomics profiling

Six millilitre samples for supernatant metabolomics profiling were collected immediately before exposure, 5 and 24 hours after exposure from the untreated, 10X MIC, and 1000X MIC conditions. The samples were centrifuged for 5 minutes at 6000 RCF. The resulting supernatant was filter sterilized in 1 mL aliquots and frozen at -20 C. The samples were shipped to Metabolon (Durham, NC) for metabolomics data collection. Briefly, all methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. Sample extract was dried and reconstituted in solvents compatible

to each of the four methods described below. One aliquot was analyzed using acidic positive ion conditions optimized for hydrophilic compounds with a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 μm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was analyzed using acidic positive ion conditions optimized for hydrophobic compounds with the same C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions with a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slightly between methods but covered 70-1000 m/z. Library matches for each compound were checked for each sample and corrected if necessary. Peaks were quantified using area under the curve. Missing values were imputed with half the minimum. Data was log2-transformed and centered around the mean. Metabolites significantly changing over time were identified using a Welch one-way ANOVA with a Games-Howell post-hoc test. Metabolites significantly changing across the samples were identified using a one-way ANOVA with a Tukey post-hoc test.

4.5.6 Condition-specific model generation

Transcriptomics data was integrated with the PA14 genome-scale metabolic network reconstruction [19] using the algorithm MADE [26] implemented in the Tiger Toolbox [43]. Briefly, MADE modulates gene functionality in the model based on statistically significant gene expression.

Metabolomics data was integrated after RNA-sequencing data integration. Using the metabolomics data, metabolites uniquely produced in a given condition for each timepoint were identified. Where feasible, the production of these metabolites was forced in the RNA-sequencing integrated models.

4.5.7 Model prediction of persister targets

Three strategies were implemented to identify potential gene targets. In the first, we performed *in silico* single gene knockouts on the condition-specific models and identified genes essential for minimal flux through the objective function (0.0001 h^{-1}). We repeated this simulation for models generated with different MADE parameters and tallied the number of times a gene was predicted as essential across the different iterations. The MADE parameters altered were the adjusted p-value threshold and the log2(fold-change) cut-offs.

In the second strategy, we used OptGpSampler to perform flux sampling on the networks generated with one set of MADE parameters[27]. We used this flux sampling information to determine reactions that had a greater likelihood of carrying flux in the persister models compared to the untreated models. Mood's median test was used to determine reactions with a significantly different median in the persister conditions compared to the untreated conditions ($p < 0.05$). Once these reactions were identified, we found the genes associated with them.

In the third strategy, we used the flux sampling data collected above to identify reactions strongly correlated with essential reactions in the models. Spearman's correlation coefficient was used to find correlated reactions ($p < 0.05$). Once these reactions were identified, we found the genes associated with them.

We repeated these analyses for minimal biomass production and minimal ATP production as the objective functions.

4.6 Acknowledgments

The authors would like to thank Jennifer Bartell, Laura Dunphy, and Phillip Yen for discussions related to the project. The authors would also like to thank Stephen Turner and Kris Rawls for assistance with RNA-sequencing data analysis. The authors would also like to thank Julia Hiser for contributions to other on-going persistence studies in the lab that provided insight for this project. Support for this project was provided by The Wagner Fellowship, the Sture G. Olsson Fellowship, and Unilever.

4.7 Author Contributions

A. B, and J.P. conceived and designed the study with thoughtful insights from G.K., A.A., and A.M. A.B. performed all experiments with assistance from G.K. A.B. performed all data analysis and modeling simulations with insight from A.M. and J.P. . A.B wrote the manuscript and J.P. provided feedback.

4.8 Conflicts of Interest

The authors declare no conflicts of interest.

4.9 References

1. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. Science. 2004;305: 1622–1625.

2. Keren I, Kaldalu N, Spoering A, Wang Y, Lewis K. Persister cells and tolerance to antimicrobials. FEMS Microbiol Lett. 2004;230: 13–18.

3. Fisher RA, Gollan B, Helaine S. Persistent bacterial infections and persister cells. Nat Rev Microbiol. Nature Publishing Group; 2017;15: 453–464.

4. Cogan NG, Brown J, Darres K, Petty K. Optimal Control Strategies for Disinfection of Bacterial Populations with Persister/Susceptible Dynamics. Antimicrob Agents Chemother. American Society for Microbiology Journals; 2012; AAC.00675–12.

5. Levin-reisman I, Ronin I, Gefen O, Braniss I, Shoresh N, Balaban NQ. Antibiotic tolerance facilitates the evolution of resistance. 2017; 1–10.

6. Cameron DR, Shan Y, Zalis EA, Isabella V, Lewis K. A Genetic Determinant of Persister Cell Formation in Bacterial Pathogens. J Bacteriol. 2018;200. doi:10.1128/JB.00303-18

7. Wood TK, Knabel SJ, Kwan BW. Bacterial persister cell formation and dormancy. Appl Environ Microbiol. 2013;79: 7116–7121.

8. Kwan BW, Valenta JA, Benedik MJ, Wood TK. Arrested protein synthesis increases persister-like cell formation. Antimicrob Agents Chemother. 2013;57: 1468–1473.

9. Shah D, Zhang Z, Khodursky A, Kaldalu N, Kurg K, Lewis K. Persisters: a distinct physiological state of E. coli. BMC Microbiol. 2006;6: 53.

10. Orman M a., Brynildsen MP. Dormancy is not necessary or sufficient for bacterial persistence. Antimicrob Agents Chemother. 2013;57: 3230–3239.

11. Radzikowski JL, Vedelaar S, Siegel D, Ortega ÁD, Schmidt A, Heinemann M. Bacterial persistence is an active σS stress response to metabolic flux limitation. Mol Syst Biol. 2016;12: 882.

12. Wang Y, Bojer MS, George SE, Wang Z, Jensen PR, Wolz C, et al. Inactivation of TCA cycle enhances Staphylococcus aureus persister cell formation in stationary phase. Sci Rep. 2018;8: 10849.

13. Ma C, Sim S, Shi W, Du L, Xing D, Zhang Y. Energy production genes sucB and ubiF are involved in persister survival and tolerance to multiple antibiotics and stresses in Escherichia coli. FEMS Microbiol Lett. 2010;303: 33–40.

14. Pu Y, Zhao Z, Li Y, Zou J, Ma Q, Zhao Y, et al. Enhanced Efflux Activity Facilitates Drug Tolerance in Dormant Bacterial Cells. Mol Cell. 2016;62: 284–294.

15. Amato SM, Orman MA, Brynildsen MP. Metabolic Control of Persister Formation in Escherichia coli. Mol Cell. Elsevier Inc.; 2013; 1–13.

16. Conlon BP, Rowe SE, Gandt AB, Nuxoll AS, Donegan NP, Zalis EA, et al. Persister formation in Staphylococcus aureus is associated with ATP depletion. Nature Microbiology. 2016;1: 16051.

17. Allison KR, Brynildsen MP, Collins JJ. Metabolite-enabled eradication of bacterial persisters by aminoglycosides. Nature. Nature Publishing Group; 2011;473: 216–220.

18. Amato SM, Fazen CH, Henry TC, Mok WWK, Orman M a., Sandvik EL, et al. The role of metabolism in bacterial persistence. Front Microbiol. 2014;5: 1–9.

19. Bartell JA, Blazier AS, Yen P, Thøgersen JC, Jelsbak L, Goldberg JB, et al. Reconstruction of the metabolic network of Pseudomonas aeruginosa to interrogate virulence factor synthesis. Nat Commun. 2017;8. doi:10.1038/ncomms14631

20. Zhang J, Sprung R, Pei J, Tan X, Kim S, Zhu H, et al. Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. Mol Cell Proteomics. 2009;8: 215–225.

21. Thao S, Chen C-S, Zhu H, Escalante-Semerena JC. Nε- lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity. PLoS One. Public Library of Science; 2010;5: e15123.

22. Kim J-S, Cho D-H, Heo P, Jung S-C, Park M, Oh E-J, et al. Fumarate-Mediated Persistence of Escherichia coli against Antibiotics. Antimicrob Agents Chemother. 2016;60: 2232–2240.

23. Orman M a., Brynildsen MP. Inhibition of stationary phase respiration impairs persister formation in E. coli. Nat Commun. Nature Publishing Group; 2015;6: 7983.

24. Meylan S, Porter CBM, Yang JH, Belenky P, Gutierrez A, Lobritz MA, et al. Carbon Sources Tune Antibiotic Susceptibility in Pseudomonas aeruginosa via Tricarboxylic Acid Cycle Control. Cell Chemical Biology. Elsevier Ltd.; 2017; 1–12.

25. Behrends V, Bell TJ, Liebeke M, Cordes-Blauert A, Ashraf SN, Nair C, et al. Metabolite profiling to characterize disease-related bacteria: gluconate excretion by Pseudomonas aeruginosa mutants and clinical isolates from cystic fibrosis patients. J Biol Chem. ASBMB; 2013; jbc–M112.

26. Jensen PA, Papin JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics. 2011;27: 541–547.

27. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. PLoS One. 2014;9: e86587.

28. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, et al. An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A. 2006;103: 2833–2838.

29. Kirkpatrick C, Maurer LM, Oyelakin NE, Yoncheva YN, Maurer R, Slonczewski JL. Acetate and formate stress: opposite responses in the proteome of Escherichia coli. J Bacteriol. 2001;183: 6466–6477.

30. Morfouace M, Shelat A, Jacus M, Freeman BB 3rd, Turner D, Robinson S, et al. Pemetrexed and gemcitabine as combination therapy for the treatment of Group3 medulloblastoma. Cancer Cell. 2014;25: 516–529.

31. De Groote VN, Verstraeten N, Fauvart M, Kint CI, Verbeeck AM, Beullens S, et al. Novel persistence genes in Pseudomonas aeruginosa identified by high-throughput screening. FEMS Microbiol Lett. 2009;297: 73–79.

32. Radzikowski JL, Schramke H, Heinemann M. Bacterial persistence from a system-level perspective. Curr Opin Biotechnol. Elsevier Ltd; 2017;46: 98–105.

33. Mok WWK, Park JO, Rabinowitz JD, Brynildsen MP. RNA Futile Cycling in Model Persisters Derived from MazF Accumulation. MBio. 2015;6: e01588–15.

34. Lewis. Genetic Basis of Persister Tolerance to Aminoglycosides in Escherichia coli. 2015;6: 1–10.

35. Held K, Gasper J, Morgan S, Siehnel R, Singh P, Manoil C. Determinants of Extreme β-Lactam Tolerance in the Burkholderia pseudomallei Complex. Antimicrob Agents Chemother. 2018;62. doi:10.1128/AAC.00068-18

36. Goormaghtigh F, Fraikin N, Putrinš M, Hallaert T, Hauryliuk V, Garcia-Pino A, et al. Reassessing the Role of Type II Toxin-Antitoxin Systems in Formation of Escherichia coli Type II Persister Cells. MBio. 2018;9. doi:10.1128/mBio.00640-18

37. Amato SM, Brynildsen MP. Persister Heterogeneity Arising from a Single Metabolic Stress. Curr Biol. Elsevier; 2015; 1–9.

38. Tang T, Wu Y, Lin H, Li Y, Zuo H, Gao Q, et al. The drug tolerant persisters of Riemerella anatipestifer can be eradicated by a combination of two or three antibiotics. BMC Microbiol. 2018;18: 137.

39. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler

transform. Bioinformatics. 2009;25: 1754–1760.

40. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30: 923–930.

41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

42. Rossi E, Falcone M, Molin S, Johansen HK. High-resolution in situ transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-phenotypes in cystic fibrosis lungs. Nat Commun. 2018;9: 3459.

43. Jensen P a., Lutz K a., Papin J a. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst Biol. BioMed Central Ltd; 2011;5: 147.

Chapter 5

Reflections and Future Directions

5.1 Introduction

Over the course of my dissertation work, I demonstrated that metabolic network models are valuable tools for antimicrobial target discovery. More specifically, using metabolic network modeling, I revealed the interconnectivity of virulence factor synthesis and growth, I reconciled conflicting high-throughput essentiality datasets and suggested mechanisms for essentiality, and I uncovered the complex and robust phenotype of the antimicrobial tolerant bacterial subpopulation known as persister cells. Having completed this work, I now turn to discussing key contributions my research has made to my field. Additionally, I present areas of future research to further extend and improve upon my findings. Finally, I close with my perspective on the role of metabolic network modeling in the future.

5.2 Contributions

Through completion of these Aims, I demonstrated that metabolic network modeling can overcome challenges associated with current target identification platforms. First and foremost, metabolic network modeling is a high-throughput approach for target discovery. In each of the Aims, I used the models to identify potential antimicrobial targets. Secondly,

because of the mechanistic nature of these models, the targets delineated by these models all have known function. Thirdly, through this approach, I was able to determine targets that may mitigate resistance. Finally, through condition-specific modeling, I was able to capture heterogeneous populations and identify targets unique to certain sub-populations. Additionally, through all of this work, I have developed tools and analyses that benefit the metabolic network modeling community.

5.2.1 High-throughput target discovery

A major challenge of current target identification platforms is that they require extensive screening and downstream follow-up experiments that are very time-consuming. In this work, I showed that metabolic network modeling is a valuable high-throughput approach for target discovery. In each of the Aims, I used metabolic network modeling to generate lists of high-confidence, computationally predicted antimicrobial targets. More specifically, in my virulence project (Chapter 2), I predicted targets that inhibited the synthesis of various virulence factors without inhibiting the ability of *P. aeruginosa* to grow. In my gene essentiality project (Chapter 3), I reconciled conflicting experimental essentiality datasets with the *P. aeruginosa* metabolic network model and predicted high-confidence targets essential for growth both computationally and in several experimental datasets. In my persister project (Chapter 4), I predicted combinatorial targeting strategies that inhibit the viability of the complex persister phenotype. Through each of these projects, I demonstrated that metabolic network modeling streamlines the target discovery pipeline by enabling efficient, cost-effective, and directed target discovery.

5.2.2 Targets with known function

One of the significant bottlenecks of current target identification platforms is that they oftentimes identify targets of unknown function. To make these targets druggable, laborious and time-consuming experiments need to be performed to characterize their function. These follow-up experiments ultimately delay the antimicrobial target discovery pipeline. In contrast, because of the mechanistic nature of metabolic network models, all the targets I identified in my work have known function. Furthermore, in each of my projects, I used the model to provide functional reasons for why certain targets are critical for various *P. aeruginosa* processes. Specifically, in my essentiality project, I demonstrated the utility of the model in providing non-obvious mechanistic explanations for essentiality for growth essential genes. For example, with the network, I determined that when the gene *glmS* is not functional, *P. aeruginosa* cannot synthesize metabolites critical for Lipid A and peptidoglycan production, which are important components of the Gram-negative cell wall. Additionally, through simulations on the network, I determined why certain genes involved in pyrimidine metabolism are essential in sputum media conditions but not in LB media conditions. In my virulence project, I used the model to help explain why certain genes are essential for either growth, virulence factor production or both. For instance, through analysis of the network, we uncovered that while *folD* is essential for growth because of its role in purine synthesis, it is essential for the synthesis of the virulence factor pyoverdine because of its role in amino acid synthesis. In my persister project, I used condition-specific modeling of the persister state to provide functional reasons for why certain genes are uniquely critical to the persister phenotype. For instance, we found that, while the base network has two genes involved in the breakdown of formate, which is an important precursor for thymidine, serine, and

methionine, in the persister state, only one of these genes, *fdnH*, is functional. By identifying targets with known function, metabolic network modeling streamlines the antimicrobial discovery pipeline. Furthermore, the mechanistic nature of these models helps to focus downstream experiments to further characterize the identified targets.

5.2.3 Targets that mitigate resistance

A second significant bottleneck of current target identification platforms is that they oftentimes identify targets that actually promote the development of resistance rather than mitigate it. Through my research, I was able to propose targets that may actually dampen the development of resistance. In my virulence factor project, by identifying targets that uniquely inhibit virulence factor synthesis, we avoid targeting processes essential for growth, potentially lessening the resistance development. For example, both with the model and experimentally, I identified that when the gene *hom* is knocked out, pyoverdine synthesis is significantly interrupted with only a marginal impact on growth. These results suggest that inhibiting *hom* may be a promising anti-virulence strategy that targets mechanisms necessary for infection without inhibiting growth essential processes. In the persister project, by identifying targets that inhibit this tolerant bacterial population, we prevent re-growth of the bacterial culture, ultimately giving the cells less time to evolve resistance mechanisms. For example, we propose that a combinatorial targeting strategy that inhibits both purN and purT, two genes involved in purine metabolism and critical for the synthesis of several biomass precursors, will inhibit persister cell viability. Based on our RNA-sequencing data, *purT* was significantly downregulated in the persister conditions while *purN* was not, suggesting that combinatorial perturbations to this pathway will have a

detrimental impact on persister cells. These examples demonstrate the use of metabolic network models to suggest targets that may mitigate the development of resistance, overcoming one of the central limitations of current target identification strategies.

5.2.4 Targets of heterogeneous populations

Another limitation of current target identification platforms is that they rely on whole bacterial population approaches and fail to capture heterogeneous subpopulations with unique susceptibilities. Through my persister project, I demonstrate that condition-specific metabolic network models are useful tools for identifying targets for distinct subpopulations within bacterial cultures. Specifically, through the integration of transcriptomics and metabolomics datasets, I generated models for the antimicrobial tolerant subpopulation known as persister cells. Using these models, I suggested targets that uniquely inhibit the persister state. For example, through analysis of the networks and the experimental data, I suggest that targeting both xanthine dehydrogenase and guanine deaminase may uniquely inhibit the persister cell phenotype by inhibiting production of xanthine, a metabolite uniquely produced in the persister state. Identification of these combinatorial targeting strategies to inhibit heterogeneous subpopulations, would be very challenging to determine experimentally. Condition-specific metabolic network models are an ideal framework for streamlining combinatorial antimicrobial target discovery.

5.2.5 Pseudomonas aeruginosa reconstructions

In addition to facilitating the antimicrobial target discovery pipeline, as part of this work, I have developed tools and analyses that benefit the metabolic network modeling community. For instance, I and co-authors built a new reconstruction for *Pseudomonas*

aeruginosa strain PA14 and an updated reconstruction for strain PAO1. These reconstructions serve as valuable resources for the metabolic network modeling communities. First, as a knowledgebase, these reconstructions contain everything we know to-date about *P. aeruginosa* metabolism. Second, as a tool, these reconstructions can be converted into mathematical models to enable simulation and perturbation of the *P. aeruginosa* metabolic network for discovery of emergent behavior. Ultimately, these reconstructions enable discovery and a deeper understanding of *P. aeruginosa* metabolism. Already, the new and updated *P. aeruginosa* reconstructions have been applied for a variety of applications beyond the scope of this dissertation. For example, Angharad Green from Cardiff University generated isolate specific models, using the PA14 model as a base, to study resistance to preservative treatment [1]. Additionally, Laura Dunphy from the University of Virginia analyzed carbon source utilization data from antibiotic resistant strains in conjunction with the PA14 model to suggest mechanistic reasons for differential catabolism across the strains [2]. I am excited to see how these models are used in the future to assist in antimicrobial target discovery

5.3 Future work

Insights from my dissertation work point to exciting avenues for future research. For example, the computationally predicted essential genes identified in my work are promising candidate targets. Further work needs to be done to fully tease out their potential as antimicrobial targets. Additionally, through my research, I discovered that both the model objective function and the parameters used in omics integration with the models can have a large impact on model predictions. In the future, it will be interesting to perform a robust

analysis on both of these aspects to more completely analyze their impact on model predictive capability. Finally, my work demonstrated that metabolic network models are promising tools for antimicrobial target discovery. In the future, it will be beneficial to apply this approach to other microbes and other contexts.

5.3.1 Target follow-through

While metabolic network modeling is a useful tool for generating curated lists of promising targets in a high-throughput, efficient, and cost-effective manner, it is necessary to further evaluate the candidacy of these targets with downstream experiments and analyses. For example, an immediate next step could be to investigate the druggability of these targets. With the database DrugBank, we can determine whether any of these targets are currently targeted by drugs. As an example, in my persister project, I proposed that targeting both the enzymes xanthine dehydrogenase and guanine deaminase might inhibit the persister phenotype by preventing xanthine production. Using DrugBank, I found that both enzymes are targeted by FDA approved compounds. For example, xanthine dehydrogenase is targeted by Allopurinol and Febuxostat, two drugs used to treat hyperuricemia associated with gout. Guanine deaminase has been shown to be targeted by Imidazole, which has medicinal activity ranging from anticancer to antibacterial. Because compounds exist that target these enzymes, this indicates that these enzymes are druggable, promoting their candidacy as promising targets. A future experiment exposing *P. aeruginosa* to inhibitors of both xanthine dehydrogenase and guanine deaminase to see if there is a reduction in the persister phenotype would motivate their candidacy even further. In

filtering the targets on druggability, this will lessen the number of time-consuming and resource intensive experiments necessary to validate the targets.

5.3.2 Biomass objective function

The identification of promising antimicrobial targets with metabolic network models is dependent on an appropriate objective function. However, it can be challenging to know what objective function to use and when. For instance, the same bacteria growing in different environments might depend on different objectives. Even subpopulations within a larger culture might differ in their objectives.

In my dissertation work, I explored some alternative objectives; however, I think a more robust analysis of different potential objective functions could greatly benefit antimicrobial target discovery with metabolic network models. For example, in my persister project, a central challenge of the modeling work was determining an appropriate objective for the persister state. We explored several possibilities: minimal biomass production, minimal ATP production, minimal redox potential, minimization of internal fluxes, as well as various combinations of the aforementioned. Each of these objective functions had limitations. For instance, while biomass production is perhaps the most established objective function in the metabolic modeling field, it might not be appropriate for the persister state which is traditionally thought to be a state of dormancy and non-growth. Additionally, integration of the RNA-sequencing data prevented the maximization of a minimal redox potential in some of the condition-specific models. Ultimately, because of the scope of the project, I limited my target identification analysis to the minimal biomass production and minimal ATP production objectives. In the future, it would be interesting to

explore other objective functions, such as optimizing the model based on the metabolomics data, and see if these improve the predictive capabilities of the condition-specific models.

Additionally, in my essentiality project, I identified that subtle modifications to the objective function can have an impact on the model predictions. Specifically, I proposed changes to the *P. aeruginosa* biomass reaction based on discrepancies between model predictions and the *in vitro* consensus essential genes. For instance, the *P. aeruginosa* model incorrectly identified the genes *glgA* and *algC* as essential because of their role in glycogen metabolism. While glycogen is certainly an important metabolite for energy storage, it is not essential for *P. aeruginosa* growth. Given the experimental data and literature evidence, I proposed removing glycogen from the biomass reaction. This result demonstrated the importance of contextualizing objective functions with experimental data.

Based on these results, I was interested to see how modifications to the biomass reaction can impact model predictions. Working with Patrick Gelbach, a previous undergraduate student in the Papin lab, we performed a robust analysis on how the presence or absence of individual metabolites in the biomass objective function can impact different model predictions like gene essentiality, carbon source utilization, and internal flux distributions (Figure 5.1).

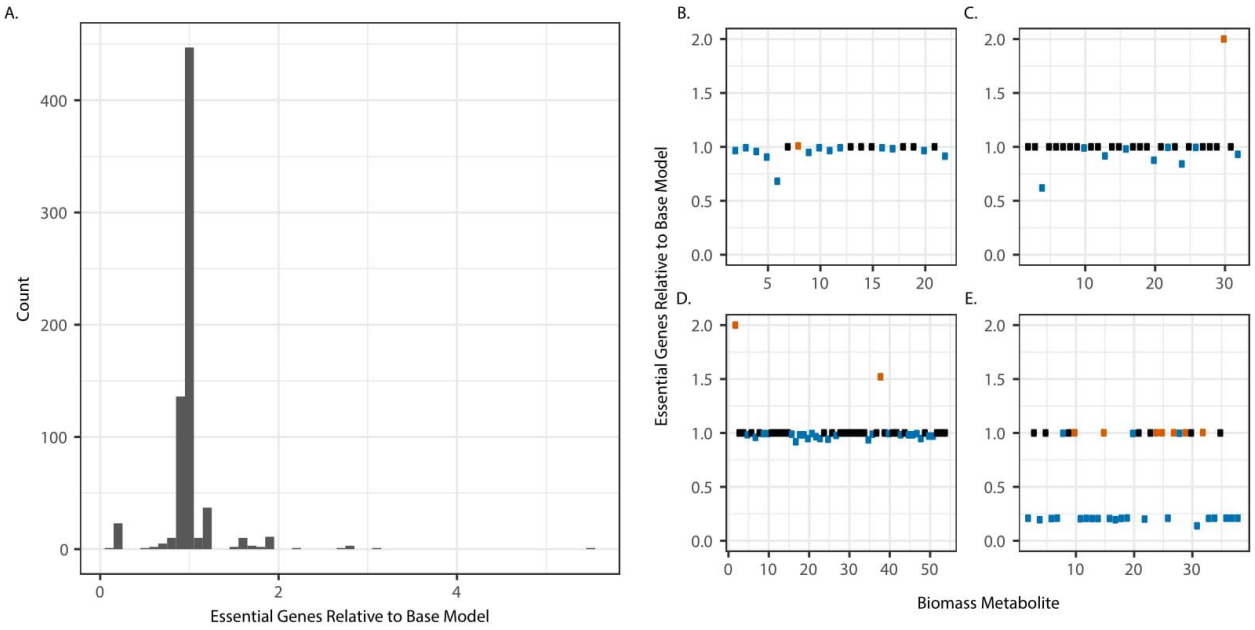


Figure 5.1. Impact of biomass metabolite presence on gene essentiality predictions. Biomass metabolites were removed from biomass reactions one-by-one for 50 genome-scale metabolic network reconstructions. *In silico* gene essentiality analysis was performed on the resulting models. The number of essential genes was identified for each model and normalized relative to the base model. (A). The distribution of the normalized essential gene counts across the modified models. Only models with altered essentiality counts are displayed because a majority of the modified models (~2000) had no impact on essentiality predictions. (B-E). Examples of how removal of single biomass metabolites from the biomass reaction can impact essentiality calls for individual models. Orange indicates that removal of that particular biomass metabolite from the biomass reaction resulted in an increase in the number of essential genes identified. Blue indicates a decrease in the number of essential genes identified. Black indicates no change in the number of essential genes identified. Panel B is model iNF518 [3]. Panel C is model iBsu1103 [4]. Panel D is model iTZ478 [5]. Panel E is model iOD907 [6].

To do this, we analyzed the biomass reactions of approximately 50 models and one-by-one removed metabolites from the equation and measured the resulting impact on the aforementioned predictions. This analysis revealed that certain metabolites, such as ATP and isoleucine, are more influential on model predictions. Additionally, while most changes to model predictions were involved in the pathways closely related to the metabolite removed from the biomass reaction, there were some non-obvious changes to the model predictions which will be interesting to investigate further. In the future, we plan to compare the

predictions of these models to experimental data with the goal of identifying modifications to the biomass objective function that improve model accuracy. Ultimately, I think this type of analysis will be very beneficial for the metabolic network modeling community in designing appropriate objective functions for different microbes in different environments.

5.3.3 Ensemble integration of omics data

Condition-specific models for antimicrobial target discovery are only as useful as the experimental data and integration methods used to generate them. Through my persister project, I revealed that the network of condition-specific models is highly dependent on the parameters used in the integration of the transcriptomics and metabolomics datasets. For instance, the p-value threshold used in the integration of the RNA-sequencing data can have a large impact on resulting functionality of genes in the condition-specific models. Furthermore, condition-specific models built with different integration parameters generated different predictions. To investigate this, I performed a sensitivity analysis on the integration of my persister RNA-sequencing data with the *P. aeruginosa* metabolic network model on gene essentiality predictions. Specifically, I evaluated the impact of changing the p-value and log2(fold-change) threshold parameters in the RNA-sequencing integration on the essentiality status of each gene in the *P. aeruginosa* metabolic network. I found that while the essentiality status of some of the genes in the model never changes (i.e., they are either always essential or always non-essential), the essentiality status of several genes in the model changed depending on the integration parameters used. Because we can never be certain of the right integration parameters to use, this suggests that, to overcome this

uncertainty, we need to consider the predictions generated by ensembles, or groups, of condition-specific models generated with different integration parameters.

The idea of using ensembles of metabolic network models for improved predictive capability is not new. For example, ensembles have been shown to be useful to address uncertainty in gap-filling of draft reconstructions [7]. However, the application of the ensemble approach to condition-specific modeling, as introduced in my persister project, is novel. In the future, I think a more detailed analysis on the use of ensembles of condition-specific models will be beneficial for the metabolic modeling community. For instance, it will be useful to know how large do ensembles of condition-specific models need to be to improve predictive capability? What integration parameters have the largest impact on predictive capability? And what type of predictions are most impacted by varying integration parameters? Additionally, it will be interesting to extend this sensitivity analysis to other methods for integrating the different omics datasets. For example, do ensembles of condition-specific models generated with multiple methods perform better than ensembles of condition-specific models generated with a single method? Ultimately, I think ensembles of condition-specific models have the potential to improve metabolic network modeling utility not just in antimicrobial target discovery applications but in the field as a whole.

5.3.4 Expansion to other organisms

Finally, based on the promising results from my dissertation work, it will be interesting to investigate the applicability of metabolic network modeling to assist in antimicrobial target discovery for other microorganisms. For example, metabolic network models might be invaluable in their ability to help identify both broad-spectrum antibiotics,

those that target multiple groups of pathogens, and narrow-spectrum antibiotics, those that target a select few. Through metabolic network modeling, we can efficiently identify targets that are common to groups of bacteria or unique to specific pathogens in a high-throughput manner. Additionally, metabolic network modeling may be useful in suggesting antimicrobial targeting strategies for pathogens that operate in communities, such as *P. aeruginosa*, *Burkholderia cenocepacia*, and *Staphylococcus aureus* in cystic fibrosis infections. Through community modeling, we might be able to identify perturbations that result in an imbalance and weakness in the community structure.

In my own work, I have already begun to explore the utility of metabolic network modeling in antimicrobial target discovery for other microbes. For instance, working with Julia Hiser, a previous undergraduate student in the Papin lab, we found that persister cells enable the Gram-negative, opportunistic pathogen and industrial contaminant *B. cenocepacia* to tolerate treatment with the industrial antimicrobials benzisothiazolinone and benzyl alcohol (Figure 5.2).

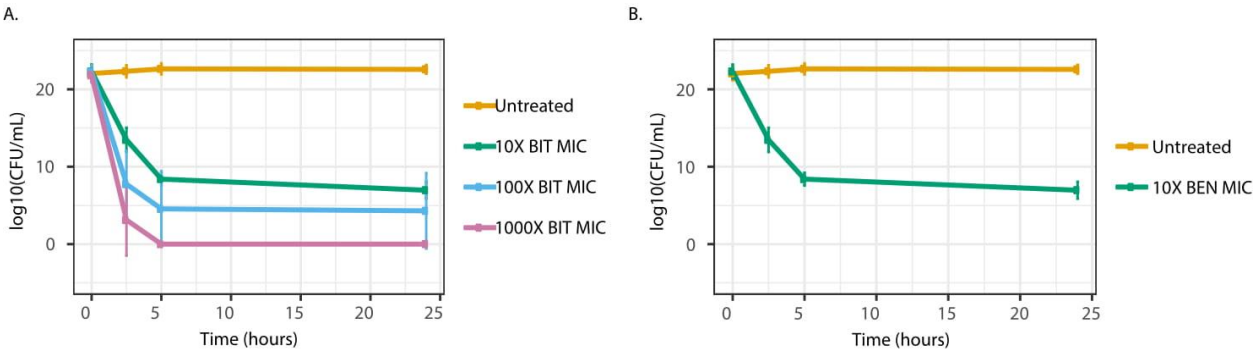


Figure 5.2. Killing of *Burkholderia cenocepacia* by industrial antimicrobials.
(A). Time-kill curve of *Burkholderia cenocepacia* cultures exposed to the antimicrobial benzisothiazolinone (BIT). Stationary phase cultures of *B. cenocepacia* were exposed to various concentrations of BIT for 24 hours. The number of colony forming units (CFUs) in each culture were counted immediately before exposure at 0 hours and 2.5, 5, and 24 hours after exposure. Untreated cultures exposed to PBS are given by the yellow line and labelled as 0% BIT. The green line indicates cultures exposed to 0.0006% BIT (10X the minimal inhibitory concentration, MIC). The blue line indicates cultures exposed to 0.006% BIT (100X

the MIC). The pink line indicates cultures exposed to 0.06% BIT (1000X the MIC). (B). Time-kill curve of *Burkholderia cenocepacia* cultures exposed to the antimicrobial benzyl alcohol (BEN). Stationary phase cultures of *B. cenocepacia* were exposed to various concentrations of BEN for 24 hours. The number of colony forming units (CFUs) in each culture were counted immediately before exposure at 0 hours and 2.5, 5, and 24 hours after exposure. Untreated cultures exposed to PBS are given by the yellow line and labelled as 0% BIT. The green line indicates cultures exposed to 0.03% BEN (10X the MIC). Error bars are standard deviation from three experiments.

We are currently in the process of collecting samples for RNA-sequencing of the *B. cenocepacia* persister cells to integrate with the *B. cenocepacia* metabolic network model, iPY1537 [8], to generate a model of *B. cenocepacia* persister cell metabolism. Comparison of this model with the *P. aeruginosa* persister cell model may identify persister traits present in both pathogens as well as unique to the individual bacteria. Furthermore, it will be interesting to see if the *B. cenocepacia* persister state is just as complex as the *P. aeruginosa* persister state, requiring a combinatorial targeting approach for complete inhibition.

5.4 My Perspective

There is an urgent need to discover new antimicrobial targets to treat infections and mitigate the development of antibiotic resistance. My graduate work has demonstrated that metabolic network models are promising tools to assist in antimicrobial target discovery because of their ability to delineate targets of known function, determine targets that may mitigate resistance, and identify targets for heterogeneous bacterial subpopulations with unique susceptibilities in a high-throughput manner. My essentiality work showed the utility of reconciling conflicting datasets with metabolic network models. Additionally, in this project, I present an approach for interpreting *in vitro* essentiality datasets with these *in silico* mechanistic models. My virulence work was the most comprehensive, genome-scale analysis to uncover the complex interrelationships between virulence factor synthesis and

growth in a quantitative and high-throughput manner. My persister work relied on novel modeling approaches to reveal that the persister state is robust and, through condition-specific modeling, suggested combinatorial targeting strategies that would otherwise be challenging to discern without the use of computational tools. Each of these projects has laid a strong foundation for the discovery of novel antimicrobial targets as well as enhanced the field of metabolic modeling through the creation of tools and novel simulation approaches.

Over time, the utility of these models will only be enhanced as additional informative data is collected to curate them and more creative tools are developed for novel analyses and simulations. For example, as personalized medicine becomes more prevalent, the development of patient-specific genome-scale metabolic network reconstructions may be realized [9]. These personalized GENREs will enable targeted antimicrobial therapeutic strategies rather than the traditional resistance-promoting, broad-spectrum approach [10]. As my dissertation shows, metabolic network modeling is a promising tool to facilitate current antimicrobial target identification platforms as well as propel next-generation drug discovery forward.

5.5 References

1. Green A. Investigating the genetic basis of preservative resistance in an industrial Pseudomonas aeruginosa strain. Mahenthiralingam E, editor. PhD, Cardiff University. 2017.

2. Dunphy LJ, Yen P, Papin JA. Network analysis reveals differential metabolic functionality in antibiotic-resistant Pseudomonas aeruginosa [Internet]. 2018. doi:10.1101/303289

3. Flahaut NAL, Wiersma A, van de Bunt B, Martens DE, Schaap PJ, Sijtsma L, et al. Genome-scale metabolic model for Lactococcus lactis MG1363 and its application to the analysis of flavor formation. Appl Microbiol Biotechnol. 2013;97: 8729–8739.

4. Henry CS, Zinner JF, Cohoon MP, Stevens RL. i Bsu1103: a new genome-scale metabolic

model of Bacillus subtilis based on SEED annotations. Genome Biol. BioMed Central; 2009;10: R69.

5. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, et al. Three-dimensional structural view of the central metabolic network of Thermotoga maritima. Science. 2009;325: 1544–1549.

6. Dias O, Pereira R, Gombert AK, Ferreira EC, Rocha I. iOD907, the first genome-scale metabolic model for the milk yeast Kluyveromyces lactis. Biotechnol J. Wiley Online Library; 2014;9: 776–790.

7. Biggs MB, Papin JA. Managing uncertainty in metabolic network structure and improving predictions using EnsembleFBA. PLoS Comput Biol. 2017;13: 1–25.

8. Bartell JA, Yen P, Varga JJ, Goldberg JB, Papin JA. Comparative metabolic systems analysis of pathogenic Burkholderia. J Bacteriol. 2014;196: 210–226.

9. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol. EMBO Press; 2014; 10: 721.

10. Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. Nature. 2016; 529: 336-343.