Comparative systems analysis of opportunistic Gram-negative pathogens

A Dissertation

presented to the faculty of the School of Engineering and Applied Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

by

Jennifer Anne Gardner Bartell

May 2015

Department of Biomedical Engineering

UNIVERSITY of VIRGINIA

APPROVAL SHEET

This dissertation is in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biomedical Engineering

Jennifer Anne Gardner Bartell
Author

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

Jason Papin, Ph.D.
Dissertation Advisor
Department of Biomedical Engineering

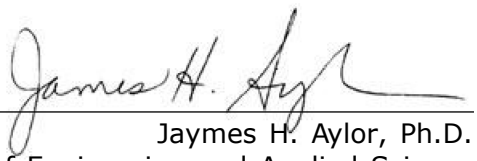
Shayn Peirce-Cottler, Ph.D.
Committee Chair
Department of Biomedical Engineering

Erik Hewlett, M.D.
Committee Member, Division of Infectious Diseases and International Health

Joanna Goldberg, Ph.D.
Committee Member
Department of Pediatrics, Emory University

Jeff Saucerman, Ph.D.
Committee Member
Department of Biomedical Engineering

Accepted for the School of Engineering and Applied Science:



Jaymes H. Aylor, Ph.D.
Dean, School of Engineering and Applied Science
May 2015

Dedication

To my committee and secondary advisors – thank you so much for your steady guidance and support. It has been so helpful to know that you have all been rooting for me to succeed while you helped to shape my path forward. I appreciate your reliability in always being willing to meet with me at short notice, answer questions, and have critical but supportive discussions about my projects. Your advice, additional training and unique perspectives have been invaluable over the years.

To my labmates past and present – thanks for playing your role as both colleagues working in partnership towards the same goals and reliable friends so well. You've made my graduate years so much more positive and enjoyable than they would be without you. Thanks for all the work chats, the coffee breaks, the happy hours, and the evenings out sprinkled with both jokes and camaraderie as fellow students and nerdy conversations about project details. I will miss working with all of you.

To my friends, my second family – I have tried to convey how much your friendship has meant to me as I've gotten older and learned the value of reliable and supportive relationships, but I think I have still failed in expressing it as well as I would like. Thanks for the listening, the advice, the easy hang outs, the shared rants, the growth together rather than apart, the ridiculous dance parties, and the wine Skypes so essential to continued close friendships even as we spread across the world.

To my advisor Jason – you have been an incredible mentor; I am so lucky that a position opened in your lab nearly five years ago that you thought I could fill. You have let me grow at my own pace, start too many projects and analyses at one time, and then guided me back towards the promising ones, and I have nearly always left your office feeling more encouraged and positive than when I went in. You have been a kind and incredibly supportive listener through both professional and personal struggles, and allowed me to address each in turn as I needed. Your optimism has balanced my pessimism, and you have shown me that it is possible to handle the crazy demands of professorship with grace, humor, balance, and a large grain of salt. Thank you so much for your mentorship and friendship.

To my family – thank you for your unflagging support over the years that I have pursued this dream. You know more than anyone what it means to me to finally be at the end of my time at UVa with a successful PhD behind me and excitedly planning the next steps of my career. You were the ones to never have a shred of doubt in my abilities from the very beginning and always encourage my next steps with full belief in my future success. You have always, always been my staunchest, loudest, and most faithful and unquestioning fans and perpetually available listeners in this long process, and I thank you for that and love you very much.

Abstract

Persistent bacterial infections are a rapidly growing concern worldwide; combatting these drug-resistant infections is hampered by a limited understanding of multifactorial pathogen evolution. Pathogens must adapt to novel nutrient restrictions, stresses induced by the host environment, competition with other microbes, and therapeutic interventions such as antibiotic treatment. The lung infections of cystic fibrosis (CF) patients are an excellent model of long-term pathogen evolution. *Pseudomonas aeruginosa* and species of the *Burkholderia cepacia* complex (Bcc) are considered the most problematic CF pathogens, known for their dominance over other pathogens, ability to induce infections lasting decades, and resistance to antibiotics (factors which also contribute to their role in serious hospital-acquired infections).

With the goal of providing novel insights into potential new therapeutic targets to combat rising drug resistance, I investigate the metabolic flexibility, virulence capability, and adaptive metabolic rewiring of these pathogens using comparative systems analyses via the framework of genome-scale metabolic models. I have built two new genome-scale metabolic reconstructions of Bcc species, updated and reconciled an existing model of *P. aeruginosa* PAO1 and also created a new model of *P. aeruginosa* PA14. I couple these models with constraint-based analysis techniques and experimental phenotype screening to validate predictions and pursue model-generated hypotheses.

My work has resulted in the prediction of specific mechanistic causes for differential antibiotic resistance and capacity for virulence between *B. cenocepacia* and *B. multivorans* via quantitative examinations of genetic redundancy and predicted activity of secondary metabolite production pathways. I extended my comparative analysis approaches to the study of decades-long evolution in *P. aeruginosa* clinical isolates from chronic CF infections, using a novel integration of single nucleotide polymorphism- and expression-based constraints to create isolate-specific metabolic models representing early and late stage adaptation. I identified network rewiring of redox metabolism as a potentially important contributor to the successful persistence of the late stage strains. The *P. aeruginosa* models have also been used to evaluate genes and enzymes essential to the production of known virulence factors and tradeoffs between virulence factor production and bacterial growth, providing a parallel avenue of treatment to enhance current antibiotic approaches. My analysis of these predictions identifies novel therapeutic targets for inhibiting virulence factor production alone or in addition to growth, of which a subset are experimentally evaluated using *in vitro* gene knockouts. In summary, I use an integrated computational and experimental framework to conduct a comparative systems analysis of important drug-resistant pathogens, contributing novel insights into their adaptive metabolic capabilities and proposing new therapeutic approaches.

Table of Contents

List of Figures and Tables

Chapter 1: Clinical applications of microbial metabolic reconstructions

Systems biology in biomedical research

Biomedical research has now been in the 'era of genomics' for two decades since the first bacterial genome was sequenced in 1995, and improving technology has been opening new doors in rapidly advancing subfields such as functional genomics, epigenomics, and metagenomics (Preidis and Versalovic 2009; Davies 2013). As the price and ease of genome sequencing drops exponentially, the amount of 'omics' data generation increases exponentially. Omics assays are now significant components in characterization of non-communicable diseases such as cancer (Kristensen et al. 2014) and metabolic syndrome (Dumas et al. 2014), pathogens inducing infectious disease, and interactions between humans and environmental factors ranging from toxins and pollutants (Kyrtopoulos 2013; Zeise et al. 2013; Vrijheid 2014) to microbiota (Preidis and Versalovic 2009). However, the full promise of these research efforts has not been realized. The key limitation in pursuit of novel discoveries is not data generation, but analyses that comprehensively integrate and interpret these data to identify meaningful trends and underlying mechanisms that can be translated into impactful clinical solutions (Ritchie et al. 2015).

The impediment of omics integration and interpretation has contributed to the parallel, rapid development of systems biology research approaches (Hood et al. 2004; Ritchie et al. 2015). Traditional molecular biology approaches use a reductionist strategy to define the function of single genes or pathways of interest and then branch outward to more complex analyses of related components. Systems biology uses the tremendous amount of biological information which has been produced during decades of the reductionist approach as a starting point to holistically evaluate complex interactions between an array of components within a given system or systems. Relying on mathematical and statistical modeling approaches, systems biologists and bioinformaticians parse important trends from collections of big data acquired by sequencing, high throughput screening of transcriptomics, proteomics, and metabolomics, and other measurement-dense assays. It is an inherently interdisciplinary field that has made great strides in the

discovery of emergent properties of molecular networks from a systems perspective.

A common target of systems biology analyses is the study of biochemical networks through elucidation of their structure, function, and flexibility. The ability to computationally map the full range of potential activity within these networks is an invaluable tool when designing more expensive and time-consuming experimental studies. It enables researchers to target specific questions and interpret the results of discovery-based experimental screens from a mechanistic standpoint. A broad array of systems modeling approaches has been used in this effort: dynamic versus steady state, single-scale versus multi-scale, broadly comprehensive versus tightly focused and parameterized (Wolkenhauer 2014). Each approach has been applied to elucidation of the key biological networks of transcription, translation, regulation, and metabolism with varying degrees of success. The perpetual need for more advanced and robust methods of big data interpretation continues to motivate the expansion of systems biology applications in biomedical research.

One of the first systems approaches to large-scale network analysis was rooted in the most well defined and conserved of biological networks – metabolism. In this review, we briefly describe the development of metabolic systems modeling specifically via genome-scale metabolic reconstructions along with model construction and analysis techniques. We then explore the growing development of clinical applications of these techniques, and show how multiple applications can be relevant in the study of infections by a class of particularly troublesome pathogens, opportunistic Gram-negative bacteria.

Systems analysis via metabolic modeling

A series of studies published in the early 1990s describe the application of linear systems optimization techniques to the prediction of metabolic reaction fluxes and production rates in *Escherichia coli*, hybridomas, and red blood cells using initial networks on the order of tens of reactions (Savinell and Palsson 1992a).

Interestingly, this novel biological network optimization approach was originally developed for bioprocess engineering; for example, rates of monoclonal antibody production by hybridomas were predicted in an attempt to optimize production levels. Further development paired with the arrival of the first sequenced microbial genome, *Haemophilus influenzae,* resulted in the publication of a revolutionary new tool, the genome-scale metabolic reconstruction, in 1999 (Edwards and Palsson 1999).

Genome-scale metabolic reconstructions are, on a basic level, curated lists of balanced chemical reactions enabled by an organism's particular repertoire of metabolic enzymes (Haggart et al. 2011). Constructed using information from an organism's annotated genome, they account for relationships among hundreds of metabolites, genes, proteins, and reactions. These relationships can be mathematically represented via a stoichiometric matrix, Boolean gene-to-protein relationships (GPRs), and reaction bounds that enforce physicochemical constraints on reaction reversibility and substrate uptake. Exchange reactions enable the *in silico* replication of nutrients available in the environment for catabolization. Using linear optimization techniques such as flux balance analysis (FBA), reaction flux patterns are predicted that enable optimal flux through a set objective function. This function is often designed as a 'growth'-predictive reaction that accounts for the production of all required biomass components for a microbe. With FBA, we can predict (1) the ability of an organism to use substrates from its environment to grow and adapt to nutrient restrictions, (2) the activity of metabolic pathways of interest, and (3) the production of important compounds and byproducts that range from ATP to microbial virulence factors (Oberhardt et al. 2009).

Construction & analysis approaches

Detailed explanations of the basic principles and protocols for building and analyzing metabolic reconstructions have been reviewed thoroughly in the literature (Feist et al. 2009; Orth et al. 2010). Chapter 2 of this thesis provides such an explanation as well. Here, we will instead mention that while all reconstructions

require manual curation of their contents for high quality predictions, new tools such as the modelSEED, KBASE, and RAVEN enable the automated creation of draft reconstructions for microbes with sequenced genomes that can be curated even by researchers new to the field (Henry et al. 2010; Agren et al. 2013) (Department of Energy Systems Biology Knowledgebase (KBase), https://kbase.us). An array of optimization toolboxes for reconstruction analysis have been written for Matlab, Python, R, and even standalone GUIs such as Optflux that cater to a range of coding aptitudes (Rocha et al. 2010; Schellenberger et al. 2011; Ebrahim et al. 2013). Increasing access to these automated and semi-automated techniques has substantially expanded the number of available reconstructions.

The expansion of semi-automated reconstruction has been accompanied by an expansion of analysis techniques and applications. Over 100 unique algorithms have been developed to evaluate different aspects of metabolism via constraint-based modeling (Lewis et al. 2012). For example, model constraints can be formulated from 'omics' data such as gene expression levels and proteomics; resulting model predictions can be used to evaluate the genome-scale functional impact of significant expression changes of particular genes and proteins (Blazier and Papin 2012). Iterative cycles of integration, prediction, experimental validation, and model refinement often also contribute many refinements to the organism's genome annotation through the filling of network gaps and assignment and testing of putative functions of orphan ORFs (Blais et al. 2013). Thus, metabolic reconstructions can be incorporated into research approaches as a novel tool for contextualizing 'omics' data and be continually expanded and refined as new experimental data is incorporated. Enhancements in automation of omics-based constraint development provide an excellent starting place for integration and interpretation of data characterizing microbial metabolic activity in a wide array of contexts (Blazier and Papin 2012).

Growth of clinical applications

The rapid developments in metabolic modeling have corresponded with an explosion of publications in the last 15 years. As new techniques are developed, subfields have sprung up within the metabolic modeling community that apply analysis techniques to unique scientific concerns. To evaluate the growth of these subfields and identify current trends within our research area, we used field-specific search terms to select and download the titles and abstracts of 1147 publications from PubMed spanning from very preliminary mathematical models of metabolic pathways published in 1979 to the current state-of-the-art as of February 2015. Publication descriptions (titles and abstracts grouped together) were parsed computationally in Matlab to count the iterations of each unique word within the total collection of text and track changes in these unique word counts by publication date (a small set of common words was removed and stemming pre-processing was applied to minimally reduce the dataset). Figure 1.1A shows the final counts of key terms repeated at least 10 times between 2000 and 2015 in the text data set to highlight important areas of investigation. The terms have been divided into appropriate subcategories in a qualitative fashion to show the relative activity within proposed subfields. The trends summarized in the figure give an indication of the relative impact of various developing subfields.

As the sorting of terms into different categories is an imperfect and qualitative process (‘engineer’ could obviously be sorted into multiple categories), terms of particular interest have been bolded. The category ‘Multi-field applications’ is used to resolve this by incorporating terms that are of interest to multiple subfields. Terms of expected high impact include ‘reconstruction’, ‘engineer’, ‘disease’, and ‘biofuels’. Perhaps less expected are the high counts of the terms ‘plant’, ‘community’, and ‘stress.’ Even more interesting is the relatively low counts and number of terms included in the ‘Biotech’ subfield; this could be due to term sorting judgements, but also perhaps the consistency of approach versus high diversity of applications within this field.

A

Network biology	1420	Multi-field Applications	1322	Biomedicine	2183	Biotech	1395	Omics Analysis	551	Agriculture & Biofuels	791	Community & Ecology	648
reconstruction	**575**	comparison	130	human	271	**engineer**	**433**	13c	159	energy	247	environmental	137
annotation	**134**	**evolution**	**126**	drug	236	synthesis	187	**transcriptomics**	**52**	**plant**	**226**	**environment**	**133**
Seed	133	**community**	**75**	cancer	150	scale	124	**omics**	**49**	**biofuels**	**61**	**community**	**75**
Kegg	123	physiology	69	**disease**	**143**	**fermentation**	**99**	**genotype-phenotype**	**47**	biofuel	41	**microbiota**	**59**
transport	108	essentiality	67	mouse	122	**batch**	**80**	**proteomic**	**44**	aragem	34	fixation	53
metacyc	74	recombinant	64	**pathogen**	**87**	**industrial**	**68**	high-throughput	36	photosynthesis	33	sporulation	49
physiology	69	**stress**	**61**	**host**	**73**	fed-batch	64	proteome	35	**cyanobacteria**	**26**	population	42
annotate	59	comparative	58	**biomarkers**	**72**	fermentative	56	transcriptome	32	photosynthetic	24	microbiome	36
topological	**39**	synthetic	55	tissue	72	mixture	54	transcriptomic	28	chlamydomonas	20	ecological	18
refinement	39	**adaptive**	**54**	treatment	71	byproduct	51	metabolomics	27	algal	19	envirome	17
feedback	23	adaptation	51	tuberculosis	70	biosynthetic	41	proteomics	26	microalgae	18	ecosystem	16
bigg	17	redox	51	mitochondrial	67	biotechnology	40	fluxomics	16	rice	16	co-culture	13
biocyc	15	mechanistic	49	microbiota	59	synthesize	38			maize	16		
orthologs	12	nutritional	45	liver	53	bioremediation	29			cyanobacterial	10		
		biosynthetic	41	tumor	50	biotechnological	19						
		evolutionary	41	melanoma	47	chemostat	12						
		cobra	38	brain	45								
		adapt	34	**infection**	**44**								
		condition-specific	33	tissue-specific	36								
		reconciliation	**30**	**antibiotic**	**35**								
		curated	27	virulence	34								
		multiscale	25	multi-tissue	32								
		feedback	23	gut	26								
		knock-out	23	dosage	26								
		biofilm	15	antimicrobial	25								
		context-specific	13	cardiac	24								
		antimetabolite	12	diabetic	24								
		organism-specific	12	adipocyte	23								
				hepatic	22								
				patient	18								
				lung	16								
				mitochondrion	16								
				myocardial	16								
				respirative	16								
				warburg	15								
				tissue/organ	14								
				antibacterial	13								
				hepatocytes	10								
				immune	10								

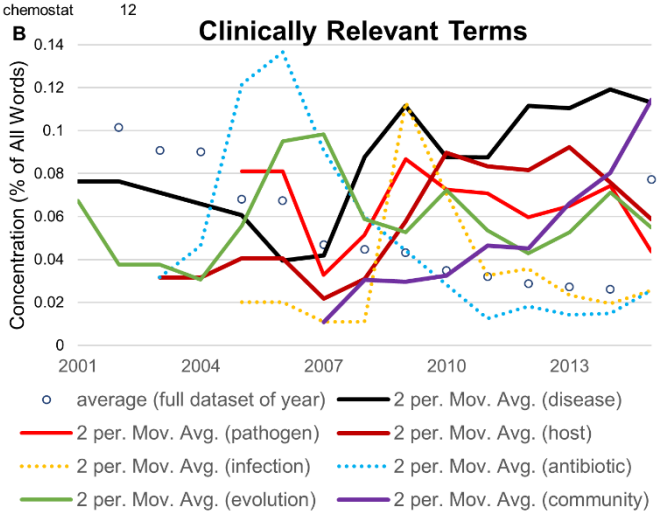


FIG 1.1. Examination of term usage in metabolic modeling publications. (A) Total counts of term usage from 1979 to 2015. (B) Changing concentration of clinically relevant terms from 2001 to 2015 compared to average word concentration.

We acknowledge this qualitative categorizing is biased by our ability to identify clinically relevant terms from the large data set which likely contributes to the extended Biomedicine column in contrast to other categories, but the diversification of disease states of interest is obvious. Several terms were chosen from this subfield for more detailed temporal tracking. Figure 1.1B shows averaged trendlines of the changing concentration of these terms by publication year in comparison to the average concentration of any single word in that year's data set. As the number of publications per year grows, the full dataset average concentration decreases. In contrast, many of the clinically-related terms maintain or increase in average concentration. In fact, Figure 1.1B indicates a transition in 2007 for most terms to above average concentration, indicating a substantial growth of interest in

developing biomedical applications within the metabolic modeling field. Surprisingly, 'infection' and 'antibiotic' both spike dramatically and then fall to lower than average use. Interest appears to have then grown particularly fast in the study of pathogens and their hosts and disease applications. The most dramatic increase is seen in use of the word 'community'; while this also incorporates instances of references to the modeling community as a whole, it also aligns with the general growth of interest in microbial communities.

In summary, biomedical applications of metabolic reconstruction are diversifying with time while gaining the interest of new generations of metabolic modelers. In the following sections, we detail the application and impact of metabolic reconstructions on several different clinically-relevant research areas. We focus on the microbial applications of these subfields in contrast to work focused specifically on human metabolism. While a few targeted reviews in this area have been published, I wanted to present a broader summary of the several subfields which interweave in the following chapters of this dissertation.

Drug targeting

One of the most well-developed biomedical applications of metabolic modeling is the identification of new therapeutic targets. Metabolic reconstructions allow modelers to predict genes and enzymes integral to the production of metabolites critical for cell survival and growth. The systems perspective provided by the models allows for efficient proposal of targets for experimental testing and provides insight into the degree of pathways and systems impacted by a particular target's inhibition *in silico*. These factors contribute to prioritization of targets when one is designing experimental validation studies.

Computational approaches range from single and double gene-deletion studies of essential genes in rich media (Lee et al. 2009) to complex analyses of correlated reaction activity that provides insight into potential drug synergies (Jamshidi and Palsson 2007). While some work has been performed to identify cancer drug

targets (Folger et al. 2011), many more studies have proposed new targets for antibacterial therapeutics. The first review of reconstruction-based strategies for antimicrobial drug targeting was presented by Chavali et al in 2012 (Chavali et al. 2012). Within this subfield, noted successes include the identification of a new malarial therapeutic target, nicotinate nucleotide adenylyltransferase, using a model of *Plasmodium falciparum* and validation via a small molecule inhibitor (Plata et al. 2010).

Comparative genomics

Metabolic reconstructions are also an excellent framework for comparative genomics; in contrast to many other approaches based heavily on bioinformatics, the metabolic models enable immediate prediction of the functional impacts of differing genomic content on a systems level. Given the dramatic increase in sequenced bacterial genomes and growing collections of clinical isolates of pathogens, this approach is a valuable tool in evaluating subtle differences between highly related species. The models enable automated comparison of network content and metabolic capacity through their standardized organization of pathways and GPRs, but truly effective comparisons can only be performed if model syntax and pathway structure are consistent among distinct models.

This requirement has resulted in the development of manual (Oberhardt et al. 2011) and then semi-automated approaches (Damiani et al. 2015.) to model reconciliation. The improvements in semi-automated model construction using the same tool such as the modelSEED offer an alternate path to model consistency. Studies have then been performed to evaluate differences in metabolic capacity between pathogen and non-pathogen of the same genus (Oberhardt et al. 2011), graded virulence between two related pathogens (Bartell et al. 2014), and the functional impact of SNPs in a collection of clinical isolates of the same species (Lee et al. 2009; Monk et al. 2013).

Adaptive evolution

The relationship between genome-scale metabolic modeling and adaptive evolution is a fundamental one. The predictions produced by linear optimization techniques are interpreted as the ideal theoretical representation of fully adapted metabolic performance of a task such as growth in a given condition. However, inherent tradeoffs in an organism's optimization towards one particular state versus retaining suboptimal but flexible metabolic capabilities in many different environments means that predictions often overestimate actual experimental measurements (Ibarra et al. 2002).

From the perspective of bioprocess applications, the ability to predict the potential flux solutions that result in optimal performance can be used to improve production efficiency via genetic engineering (Poblete-Castro et al. 2013). In the study of adaptive evolution during infection of a host, metabolic modeling can be used to evaluate the potential effects of selective stresses on an organism adapting to a new environment. Many microbes undergo a shift in their metabolic activity to enable successful infection; this shift can encompass development of niche specializations, reduction in superfluous metabolic abilities, and adaptive response to host defenses over infectious periods that can range from weeks to decades of evolution within the host (Folkesson et al. 2012; Proctor et al. 2014). Biologists and bioinformaticists can track the accumulation of mutations via sequencing of a continuum of clinical isolates collected over the period of infection (Marvig et al. 2014), but the functional impacts of SNPs and genomic rearrangements including insertions and deletions have systemic effects that are difficult to fully evaluate. Metabolic modeling offers a framework to systematically study the repercussions of these adaptive changes, comparing the original environmental strain to strains at representative stages of evolution.

These snapshots along a continuum of adaptation are created by collecting omics data via gene expression arrays, RNAseq, or proteomics that encapsulates the metabolic activity of each isolate. Differential expression levels among isolates can be converted into model constraints for each stage, providing isolate-specific

metabolic models for comparison (Blazier and Papin 2012). The first iteration of this approach was used in a study of adaptation in *Pseudomonas aeruginosa* isolates causing chronic lung infections in cystic fibrosis patients (Oberhardt et al. 2010) using constraints based on simple differential expression cut-offs. More sophisticated automated algorithms for consistently applying constraints representing expression states have since been developed (Jensen et al. 2011; Machado and Herrgård 2014) and will contribute to expansion of adaptive evolution studies.

Engineering the host environment

Early investigation of microbial infections often targeted specific mechanisms of virulence such as production of toxins and mechanisms of invasion. While these factors are important to successful colonization, there is growing interest in how organisms utilize compounds available in the environment around them to survive (Wessel et al. 2013; Staib and Fuchs 2014). Researchers are investigating the treatment potential of engineering an infection-prone environment to inhibit or prevent colonization and growth, or alternatively encourage the growth of healthy commensal bacteria. Compounds available to bacteria inhabiting airways or the human gut could be manipulated in several ways, including altered diet and nutrition (David et al. 2014) and therapeutic delivery of anti-metabolites(Rangel-vega et al. 2015; Singh et al. 2015).

Metabolic reconstructions offer promise as a source of hypotheses on how to block essential nutrient exchanges, but actual studies are somewhat limited. One of the earliest modeling studies in this area was an effort by Klitgord et al. to identify environmental conditions that promoted the development of commensal microbial communities (Klitgord et al. 2010), though this study was not developed with regards to any specific clinical application. Another study used coupled reconstructions of *Plasmodium falciparum* and a human erythrocyte to carefully examine nutrient uptake by *P. falciparum* in this parasitic relationship (Huthmacher et al. 2010); while the authors used this information to predict enzymes to be

targeted therapeutically, these analyses would also be relevant in designing application of anti-metabolites. The most relevant study is likely one by Saulnier et al., which examined the abilities of 4 strains of *Lactobacillus reuteri* to produce metabolites beneficial to the human gut (Saulnier et al. 2011). As researchers gain better understanding of interactions between microbiota and the human host, the utility of both human and microbial metabolic reconstructions in identifying how to manipulate the host environment will continue to rise.

Interactions within the microbiome

Few prior microbial community studies have integrated metabolic modeling into their approach; studies of competition over substrate utilization via differential equation modeling of uptake and growth rates have been conducted in chemostats (Hesseler et al. 2006; Behrends et al. 2009; Wintermute and Silver 2010), but the integration of genome-scale modeling approaches to study of interspecies interaction analysis has only just begun to gain momentum in the last few years. Metabolic models allow for broad assessment of substrate sharing and niche specialization by particular species within the model by enabling compound sharing through careful formulations of exchange reactions. Multiple computational approaches to analyzing a model community have been developed, and range from a community model created by linking individual species models using a simple compartmentalization approach to the sophisticated OptCom and d-OptCom modeling frameworks. OptCom enables multi-level, multi-objective optimization of a collection of models as well as application of an overarching 'community objective'; this approach allows that growth of the community as a whole may have different demands than growth requirements for a single species. d-OptCom is a dynamic extension of this approach providing for temporal analysis (Zomorrodi and Maranas 2012; Zomorrodi et al. 2013).

An early study in this area investigated the potential for interactions in large collections of minimally-curated models to assess synergy versus competition within unique species pairings (Freilich et al. 2011). A more recent publication details the

prediction of emergent synergy solely possible through community interactions (Chiu et al. 2014). With respect to clinical applications, several community models have been built to assess human gut microbiota. Bucci et al evaluated how doses of antibiotics might affect the species distributions within the human gut using paired model evaluations, while Shoaie et al compared predicted interactions among three representative species with transcriptomics data from model mouse microbiomes (Bucci et al. 2012; Shoaie et al. 2013). As detailed in a proposed road map for community systems biology (Zengler and Palsson 2012), the applications of these approaches will continue to grow as the study of human microbiomes expands beyond the gut to oral, nasopharyngeal, vaginal, and lung microbiomes that all impact human health.

Applications in infections by opportunistic pathogens

There is a wealth of opportunity to use the previously described approaches in the study of infections by clinically frustrating microbial pathogens. The need for improved understanding of pathogen metabolism and adaptation is great in the face of rising levels of antibiotic resistance. One in 20 hospital patients will acquire an infection by a drug-resistant pathogen and there are nearly 2 million cases of antibiotic resistant infections in the US every year with direct treatment costs estimated to be as high as $20 billion (Centers for Disease Control and Prevention 2014). Bacteria associated with drug resistant infections include obligate pathogens such as *Mycobacterium tuberculosis* as well as a growing class of opportunistic pathogens that take advantage of weakened immune function to cause serious infections in a variety of health contexts. Gram-negative opportunistic pathogens are a serious clinical concern; their double cell membranes and array of efflux systems contribute to high levels of intrinsic antibiotic resistance as well as resistance to host immune responses (Curcio 2014). Some of the modeling approaches already mentioned have involved the study of these important pathogens. Here, we provide further background on the characteristics of a few problematic and related Gram-negative opportunists, *Pseudomonas aeruginosa* and species of the *Burkholderia cepacia complex* (Bcc).

Pseudomonas aeruginosa is the most prevalent Gram-negative pathogen in pneumonia, surgical site infections, septicemia, and lung infections of cystic fibrosis (CF) patients; multi-drug resistant *P. aeruginosa* was recently classed as a serious clinical threat by the US Centers for Disease Control and Prevention (Centers for Disease Control and Prevention 2014; Curcio 2014). Nosocomial infections by Bcc species are increasing due to industrial contamination of aqueous healthcare products (antiseptics, IV solutions, etc), and Bcc infection is considered a death sentence in CF patients due to the potential for developing cepacia syndrome, a severe necrotizing pneumonia (Mahenthiralingam et al. 2008; Moehring et al. 2014; Ko et al. 2015). These Gram-negatives have large genomes that encompass a broad array of catabolic and anabolic pathways as well as complex regulatory networks. Strains of the most common model species of a Gram-negative pathogen, *Escherichia coli*, have genomes ranging from 4 Mb to 5.9 Mb in size. In contrast, strains of *P. aeruginosa* have genomes that range from 6.3-6.9 Mb (Centers for Disease Control and Prevention 2014) while members of the *Burkholderia cepacia complex* (Bcc) have genomes ranging from 6.5 Mb to 8.9 Mb (Mahenthiralingam et al. 2005). The redundant pathways and isozymes present in *P. aeruginosa* and Bcc species make them resistant to targeting of metabolic activity via antibiotics. These features also likely contribute to the ability of these pathogens to successfully adapt and persist in the human host for decades in diseases such as cystic fibrosis and chronic obstructive pulmonary disease. These bacteria are also capable of growth phenotypes such as single and multi-species communities that can form biofilms to further avoid eradication; persister cells that appear during chronic infections also pose a serious problem for successful treatment with bactericidal antibiotics (Eberl and Tümmler 2004; Mulcahy et al. 2010; Bragonzi et al. 2012; Rodríguez-Rojas et al. 2012; Van Acker et al. 2013).

Studies of opportunistic pathogen adaptation to the host are enhanced by using constrained models of long-term infection, such as bacterial persistence in the lungs of patients with cystic fibrosis (CF). Recent work has identified patterns of genetic

mutations that some highly successful strains of *P. aeruginosa* acquire in chronic infections (Yang et al. 2011). In addition, other work has highlighted surprising levels of bacterial community diversity, where an initial infection of a CF patient by a broad array of pathogenic bacteria narrows with time to infections of predominantly *P. aeruginosa* and/or Bcc species (Cox et al. 2010; Zhao et al. 2012a). How these particular species persist to become such troublesome chronic pathogens in the lung in contrast to other initial infecting species is poorly understood. Many studies have focused on the evaluation of virulence factors produced by pathogens that enhance the ability of the microbe to infect and/or persist in host tissue (Loutet and Valvano 2010a). Other studies focus on the potential for niche adaptation of microbial growth using nutrients provided by the host (Palmer et al. 2005). Evaluating the complex interrelationships between metabolic genes linked to growth and virulence that contribute to successful adaptation requires a systems approach. This approach can be provided by the framework of genome-scale metabolic reconstructions.

Metabolic adaptation

There is tremendous potential for adaptive evolution to different host environments and therapeutics during a chronic infection. Antibiotic resistance within an organism can develop from broad genetic adaptations to environmental factors such as growth within a multi-species biofilm or alterations in substrate availability as well as specific mutations in targeted genes, necessitating the development of new drugs. Genome-scale metabolic models aid this process through the evaluation of metabolic rewiring that may compensate for the inhibition of a growth-related enzyme by a chemotherapeutic. Models also enable evaluation of different potential drivers of this rewiring that produces systemic adaptation to novel stresses within the host. Metabolic drivers of adaptation may be a key player in the complex network of interactions with the host environment and surrounding microbial community. It has been shown that (1) metabolic adaptation and niche specialization occur over the course of chronic infection, (2) antibiotic resistance is correlated with altered metabolic phenotypes, and (3) antibiotic resistance of microbial communities is greater than resistance of monocultures of each species

within the community (Lawrence et al. 2012; Derewacz et al. 2013). However, the basic patterns of bacterial metabolic adaptation and interaction are often poorly understood in many disease contexts.

In infections of cystic fibrosis patients, pathogens such as *P. aeruginosa* and Bcc species face unique environmental conditions to which they must adapt in order to persist within host tissue. Studies of *in vitro* growth in synthetic CF sputum have shown clear preferences by *P. aeruginosa* for specific substrates such as arginine, proline, and aspartate within cystic fibrosis sputum (Palmer et al. 2007). These substrates can fluctuate in sputum depending on host physiology, location within the airway, and utilization by competing microbes (Folkesson et al. 2012). Limited oxygen promotes adaptation towards microaerobic growth, which may also protect against oxidative stresses compounded by host immune cells (Hogardt and Heesemann 2010). Studies have shown that long term adaptation within the CF lung results in repeated nonsynonymous mutation of a selection of pathoadaptive genes; late stage strains show an overall downregulation of pathways involved in central metabolism as well as broad regulatory remodeling (Marvig et al. 2014). While some of these genes can be directly linked to antibiotic resistance or phenotypic changes associated with biofilm growth, the functional impacts of other mutations are unknown. Further experimental study coupled with computational modeling can lend insight into how these adaptations functionally alter metabolism on a systems level, probe which optimized objectives best match experimental phenotyping, and provide new therapeutic targets for treating both early and late stage infections.

Virulence factors

In addition to adaptation of metabolic pathways critical to pathogen growth, the production of compounds that specifically enhance a pathogen's ability to infect a host can adapt over the course of colonization and establishment of chronic infection. This array of secondary metabolites, toxins, and enzymes are classed as microbial virulence factors. The definition of 'virulence factor' also usually includes

the caveat that loss of a virulence factor does not affect viability in rich media (Brown et al. 2012), which has complicated whether certain byproducts and enzymes should be classed as official virulence factors. Recent studies are identifying new roles for virulence factors that may indicate greater importance in general growth processes in certain environments (Price-Whelan et al. 2007; Brown et al. 2012). These studies particularly highlight the need for increased nuance in evaluating the role of virulence factors with respect to the metabolic processes in opportunistic pathogens.

Here, we provide a description of an array of compounds that are implemented in metabolic reconstructions of *P. aeruginosa* and Bcc species as metabolically synthesized virulence factors. Membrane lipopolysaccharides incite host immune responses that ultimately injure host tissue, enabling invasion and providing nutrients for the microbe to scavenge (King et al. 2009). Communication between microbes (occasionally across species) is enabled by the production of quorum sensing molecules that allow communities of bacteria to coordinate adaptive processes, response to stress, and nutrient scavenging by shared regulatory changes (Lewenza et al. 2002; Fazli et al. 2014). For example, quorum sensing molecules can signal for the production of other virulence factors including secondary metabolites such as siderophores (compounds that sequester essential iron) and phenazines (pigments that have antioxidant properties to fend off immune attacks) (Dietrich et al. 2006; Williams and Cámara 2009; Wilder et al. 2011). Quorum sensing signals are particularly relevant in the context of biofilm growth, where they can control the production of shared public goods such as these secondary metabolites as well as the production of rhamnolipid, a surfactant that helps regulate the biofilm structure (Pamp and Tolker-Nielsen 2007). Extracellular polysaccharides such as alginate that help form biofilm environments are also sometimes grouped with virulence factors (Franklin et al. 2011).

The virulence factors implemented in *P. aeruginosa* and Bcc models by no means represent the full array of virulence factors produced by these pathogens. However, certain virulence factors are straightforward to implement in a metabolic reconstruction. The synthesis pathways of lipopolysaccharides, siderophores, and phenazines have been studied in detail, and can be incorporated with the other metabolic pathways to provide the desired virulence factor as a compound to be optimized for production using constraint-based analyses. Other virulence factors pose challenges to model implementation. Enzymes and toxins that are considered virulence factors and operate as proteases or lipases could be incorporated as a metabolic reaction, but pose difficulties for optimization even if substrates or products of the enabled reaction are targeted given that these compounds may be extracellular. The many enzymes encoding the synthesis of complex structures such as flagella have not been mapped in enough detail for confident inclusion as a virulence factor synthesis pathway. In the context of metabolic reconstructions, we use the term virulence factor to describe compounds which we can successfully implement as metabolic byproducts, and intend for more compounds to be added in the future as other synthesis pathways are mapped in more detail. We also attempt to account for virulence-linked genes identified through animal infection assays and other experimental means; these genes do not always link to an explicit compound that could be termed a virulence factor.

Comparing growth and virulence-linked adaptation

Much of the work on metabolic pathway mapping and curation has been focused on understanding how pathways contribute to growth in different environmental conditions. The study of virulence factor production has been more limited and less systematic. However, both of these factors play important and interrelated roles in successful colonization and persistence in host tissue. Recent studies have been published evaluating the inhibition of quorum sensing molecules and other virulence factors as a therapeutic approach for treating *P. aeruginosa* and other pathogens (Bjarnsholt et al. 2010). Other studies have looked at connections between changes in pathogen metabolism and antibiotic susceptibility (Martínez and Rojo 2011).

There is a substantial need for development of methods to systematically evaluate critical metabolic activity and network rewiring contributing to colonization of host tissue, persistence in chronic infection by particular species, and interactions between these metabolic states and virulence-linked activity.

In this dissertation, I use metabolic reconstructions as a framework for comparing the metabolic capacity of opportunistic pathogens *P. aeruginosa* and the Bcc as they colonize and persist within human hosts. The models allow me to evaluate growth-linked activity, virulence-linked activity, and tradeoffs between these two metabolic objectives. In Chapter 3, I provide a comparative computational analysis of reconciled Bcc species models, providing novel functional assessments of genetic redundancy and the contribution of different clinically relevant substrates to the production of an array of virulence factors. In Chapter 4, I use updated models of *P. aeruginosa* strains to evaluate tradeoffs between metabolic activity and virulence factor production in an attempt to classify new therapeutic targets that inhibit virulence factor production alone or in addition to growth, identifying a novel set of central growth-linked pathways that are also critical for virulence factor production. In Chapter 5, I extend the comparative analyses implemented in Chapters 3 and 4 to the study of metabolic adaptation of *P. aeruginosa* during decades of chronic infection in CF patients. Guided by experimental characterization of representative isolates, I identify a novel mechanism of metabolic network rewiring that may contribute to improved oxidative stress resistance by persistent *P. aeruginosa* strains in the CF lung. These aims are linked in their use of comparative computational analyses to improve model predictions and evaluate differing pathogen objectives in the context of infection. In each aim, I identify critical pathways that can be targeted therapeutically, offering novel drug targets and optimized treatment regimens that may help combat antibiotic resistance.

Ultimately, my work has resulted in four new genome-scale metabolic reconstructions of CF pathogen strains. I assess these models using several different approaches described in this text; I use constraint-based analysis approaches to identify novel drug targets, compare genomic content, evaluate

adaptive evolution, and investigate the catabolic potential of nutrients provided by the host in the context of infections by CF pathogens. By comparing their metabolic capacity for virulence, adaptive rewiring, and alternate objectives in chronic infection conditions using a metabolic modeling approach, I provide novel insights into Gram-negative pathogen metabolism that improve our understanding of adaptation and virulence during infection by drug resistant pathogens.

Chapter 2: Whole-genome metabolic network reconstruction and constraint-based modeling

Acknowledgements: Charles Haggart[1], Jeff Saucerman[1], Jason Papin[1]

[1]Biomedical Engineering, University of Virginia

Reprinted from *Methods in Enzymology*, 500, Haggart, C. R.*, Bartell, J. A.*, Saucerman, J. J., & Papin, J. A., Whole-genome metabolic network reconstruction and constraint-based modeling, 411-33, Copyright (2011), Elsevier. * *co-first authorship*
doi:10.1016/B978-0-12-385118-5.00021-9

SYNOPSIS

With the advent of modern high throughput genomics, there is a significant need for genome-scale analysis techniques that can assist in complex systems analysis. Metabolic genome-scale network reconstructions (GENREs) paired with constraint-based modeling are an efficient method to integrate genomics, transcriptomics, and proteomics to conduct organism-specific analysis. This text explains key steps in the GENRE construction process and several methods of constraint-based modeling that can help elucidate basic life processes and development of disease treatment, bioenergy solutions, and industrial bioproduction applications.

Introduction

The rapid expansion of methods to utilize organism-specific whole genome sequences provided by high throughput sequencing technology has provided clarity in areas scientists have long puzzled over and improved our ability to probe for answers on a system-wide level in addition to our classic reductionist investigative strategy. The substantial amounts of data now being harvested on phenotypic, genetic, protein, and molecular scales is driving the development of computational systems analysis ever faster in a search for ways to organize and contextualize raw data into a coherent picture. One promising area within the field of systems biology is reconstruction of organism-specific genome-scale metabolic networks that has been accompanied by the development of a wide range of constraint-based modeling approaches. The metabolic genome-scale network reconstruction (GENRE) provides a framework to organize all available information about an organism's metabolism through careful construction and curation of a computational network that links the cell's genome and gene expression to metabolic reaction fluxes, biomass and energy production and consumption. Once completed, the power of each GENRE can be realized as a model for probing a cell's genotype-phenotype relationship via constraint-based modeling.

Building an organism-specific GENRE is a promising opportunity to improve metabolic understanding. If the reconstructed network model can be validated through experimental investigations of growth rates, phenotypes, reaction fluxes, and gene expression, it may be immediately useful as a method for drug discovery or the development of strategies for optimizing byproduct metabolite yields as examples. If the model predictions vary from experimental validation, the discrepancies offer the researcher a roadmap for iterative *in vivo* experimentation, *in silico* modeling, and annotation refinement that further the understanding of the organism's metabolism. Also, the GENRE itself, if well-constructed and curated, is a directed method of collecting and organizing all currently available knowledge of the organism's metabolism in a functional manner.

A cell's phenotype is subject to the constraints imposed by its genome and extracellular environment. By quantifying these constraints in an appropriate manner, one can predict the range of possible cellular phenotypes *in silico* without acquiring large quantities of kinetic metabolic data. While constraint-based modeling of metabolic networks is far more quantitative than interaction-based (i.e., graph-based) network models (Price et al. 2003), it does not allow for characterization of the dynamics of metabolism with the same level of confidence as a mechanistic, kinetic model (Raman and Chandra 2009). However, given the lack of experimentally measured metabolic reaction parameters for most organisms, kinetic modeling is not a widely applicable approach to modeling cellular biochemical networks. Constraint-based modeling presents an alternative approach to better understanding and predicting the behavior of a biochemical network under various environmental and genetic perturbations.

While excellent reviews of the reconstruction process are available (Durot et al. 2009; Feist et al. 2009), including highly detailed protocols for reconstructing genome-scale metabolic networks (Thiele and Palsson 2010), the present work focuses on providing an introduction to the reconstruction process, including important methods, tools, and validation techniques, as well as the constraint-based modeling approaches that utilize the curated information within each GENRE.

Metabolic Network Reconstruction

Metabolic network reconstructions are comprised of the stoichiometry of reactions necessary for nutrient usage in synthesis and degradation of basic metabolites and more complex compounds, specific genes whose protein products are associated with these biochemical reactions, and supporting annotation and literature references. This list can then be converted to mathematical form and combined with constraint-based modeling approaches detailed in the second half of this manuscript to predict metabolic phenotype. The reconstruction of a metabolic network of an organism with a completed genome sequence consists of the following steps: 1) genome annotation, 2) automated network reconstruction, 3) network refinement, 4) *in vitro* experimentation, and 5) gap analysis. These steps

are often completed concurrently or iteratively to improve the accuracy of the model as shown in Figure 2.1.

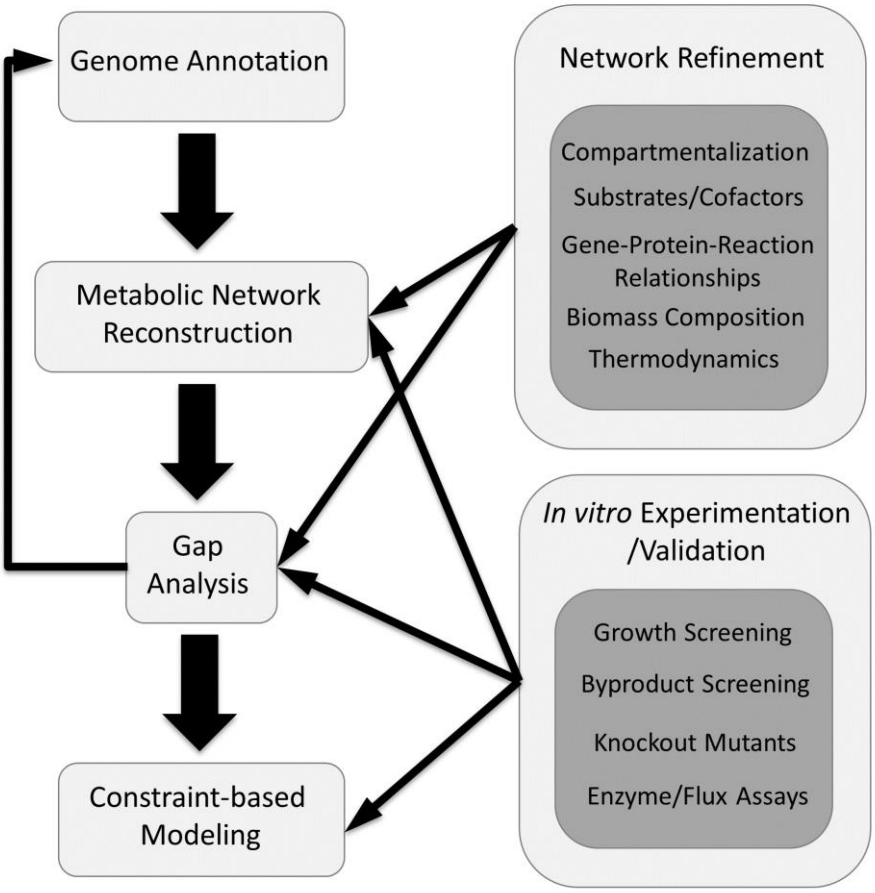


FIG 2.1. The reconstruction process. Flow diagram illustrates the iterative multi-step methodology for reconstructing and simulating the behavior of an organism-specific genome-scale metabolic network model.

Genome annotation

Since the first genome was sequenced via high-throughput technology in 1998, technological advancement has decreased the required cost and time to obtain a full genome sequence (Kircher and Kelso 2010). However, for completed genome sequences to be useful, they need to be annotated using standardized gene ontologies that provide consistency while identifying genes and cataloging their function and regulation in relation to biological processes (Giglio et al. 2009).

Automated annotation programs (ERGO, J. Craig Venter Institute, Integrated Microbial Genomes) provide annotations that usually require manual curation to

Table 2.1. Valuable resources for GENRE curation and analysis efforts.

Genome Annotation Resources		
ERGO	automated and manual annotation with pathway mapping and bioinformatics tools	http://ergo.integratedgenomics.com
Comprehensive Microbial Resource	database of all prokaryotic genomes sequenced to date with automated and/or manual annotation	http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi
Integrated Microbial Genomes	automated genome annotation and comparative genomics analysis	http://img.jgi.doe.gov
ENSEMBL	sequence access portal with annotation, visualization, and analysis tools	http://www.ensemblgenomes.org/
NCBI Entrez Genome	currently over 12,000 sequenced and/or annotated records covering all major organism groups	http://www.ncbi.nlm.nih.gov/sites/genome
KEGG	databases on sequenced genomes, orthology, hierarchy, reactions, proteins, and pathway maps	http://www.genome.jp/kegg/
UCSC Genome Bioinformatics	browser and database useful for comparing genomes and annotations	http://genome.ucsc.edu/
GO	database of standardized annotation terms	http://www.geneontology.org/
Gene Sequence Alignment Tools		
BLAST	local alignment search of protein/nucleotide sequences - does not allow gaps when matching sequences	http://blast.ncbi.nlm.nih.gov/
FASTA	local alignment search of protein/nucleotide sequences - allows gaps when matching sequences	http://www.ebi.ac.uk/Tools/fasta/
BLAT	local alignment search matching sequence greater than 25 bp long	http://genome.ucsc.edu/cgi-bin/hgBlat?org=human
Enzyme/Protein Databases		
UniProt/Swiss-Prot	manual annotation of protein sequence function	http://www.uniprot.org
UniProt/TrEMBL	automatic annotation of protein sequence function	http://www.uniprot.org
NCBI-RefSeq	curated DNA, RNA, and protein sequence annotations	http://www.ncbi.nlm.nih.gov/RefSeq/
BRENDA	enzymes with associated reactions, pathways, substrates, and cofactors	http://www.brenda-enzymes.org/

Table 2.1 Cont'd. Valuable resources for GENRE curation and analysis efforts.

Automated Reconstruction		
SEED	automated annotation and draft reconstruction of user-input genomes	http://www.theSEED.org/
PathwayTools	draft reconstruction with visualization tools and pathway gap prediction	http://bioinformatics.ai.sri.com/ptools/
metaSHARK	draft reconstruction with ability to overlay information on generic metabolic networks	http://bmbpcu36.leeds.ac.uk/shark/
KOBAS	automated annotation and reconstruction using KEGG orthology	http://kobas.cbi.pku.edu.cn/
ASGARD	automated annotation and reconstruction	http://sourceforge.net/projects/asgard-bio/
Analysis		
CellNetAnalyzer	Matlab-based environment for topological network analysis of metabolism and cell signaling	http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html
Metatool	analysis suite that finds a set of elementary modes (nondecomposable constrained steady state networks) for a reconstruction	http://pinguin.biologie.uni-jena.de/bioinformatik/networks/

supplement organism-specific depth and detail as well as verify the automated annotation assignments (Y. Yang et al. 2009). Some effort has been made to design methods for evaluating annotation quality dependent on the degree of sequence match and phylogenetic relatedness to other annotated genomes (Y. Yang et al. 2009). Alignment tools such as BLAST are commonly used to annotate gene function based on orthology with previously annotated genomes provided in online databases, followed by other more intensive methods relying on phylogenetic grouping (Kuzniar et al. 2008). Both general and organism-specific databases of genome annotations are available as listed in the appropriate section of Table 2.1, which also details other useful tools in the reconstruction process.

Automated network reconstruction

Most metabolic reconstruction efforts begin with an automated draft reconstruction from a publicly annotated genome available online. Much effort is being directed to development of comprehensive automation suites. One such effort, the Model SEED project, moves from an unannotated, sequenced genome to a draft metabolic

network with gap filling and verification features. Programs such as Pathway Tools and others listed in Table 2.1 require a previously annotated genome to build a draft reconstruction and offer various visualization and analysis tools. These programs produce at minimum lists of genes, their associated reactions, and corresponding Enzyme Commission (EC) numbers. It is also possible to use appropriate ontology keywords to retrieve core metabolic genes and reactions from an annotation. However, regardless of the automated tool used, manual reconstruction efforts are always necessary to ensure organism-specific metabolic characteristics are included (DeJongh et al. 2007; Karp et al. 2009).

Network refinement

Manual reconstruction begins with an initial evaluation of the completeness of the automated draft reconstruction. Each reaction must be evaluated for its necessity to the model, accurate stoichiometry, reaction direction and reversibility, and role in metabolite production, usage and recycling. Other concerns include thermodynamic feasibility and energy constraints. Integral to a successful reconstruction, this manual evaluation can take a significant amount of dedicated work to complete.

Key organizational tools. Research groups commonly use the few techniques listed below to speed the reconstruction process and provide consistency between different reconstruction projects.

(i) Spreadsheet organization. A thorough and well organized annotation record, usually stored in a spreadsheet, should include the gene name, all pertinent gene abbreviations, the reaction equation with consistent reactant, substrate, and product symbols, balanced stoichiometry, an indicator of reversibility, the metabolic subsystem that utilizes the reaction, the associated protein and its EC number, any literature references used to adjust or add the reaction, a rating of confidence in the annotation entry and comments about any adjustments or questions about the entry.

(ii) Reaction confidence level. In curating the reconstruction, it is helpful to keep track of the quality of the various annotations and changes. A standard method is to apply a classification system to each entry, ranging from class 1 (high confidence in the data) to class 4 (low confidence). This judgment is still qualitative, so consistency in class assignment within the reconstruction is important. When adding a reaction based solely on sequence matches with another organism, the confidence level of the added reaction should be based on the phylogenetic relatedness of the organism. It is also vital to record the addition of enzymes that have not been associated with a particular gene in the organism's sequence data but are known to be necessary in a certain pathway. Examples of metabolic reconstructions that have utilized reaction confidence ratings include iMO1056 (Oberhardt et al. 2008) and the human metabolic network reconstruction (Duarte et al. 2007).

Organism-specific curation. After canonical metabolic reactions have been collected via the automated draft reconstruction, a review of literature pertaining to the metabolism and function of the organism is necessary to identify organism-specific characteristics that should be integrated. Reconstructions of related organisms can be searched for specific functionality such as substrate usage preference and unique metabolic products that may be replicated in the organism of interest. For example, the reconstruction of *Aspergillus niger* included pathways related to steroid synthesis based on the biology of other aspergilli fungi, while a gap in lipopolysaccharide synthesis in the reconstruction of *Pseudomonas aeruginosa* was filled using gene similarity in the *Pseudomonas fluorescens* annotation (Andersen et al. 2008; Oberhardt et al. 2008).

(i) Compartmentalization and exchange reactions. When building the reconstruction, care must be taken to check the transport and exchange potential of a particular metabolite so that appropriate limits can implemented as constraints on model behavior. Localization and transport of a metabolite between the cytosol and organelles or within the cytosol may be complicated, particularly in eukaryotic organisms, and the order of transport may be critical to metabolic processes (Hao

et al. 2010). Exchange reactions can be implemented that account for the movement of metabolites across intracellular and intercellular membranes, creating compartments within the *in silico* network to represent these physical boundaries.

(ii) Gene-protein-reaction relationships. While data for the reconstruction is based on gene annotations, constraint-based modeling can proceed using solely the enzymes and associated reaction stoichiometries. However, to simulate the phenotypic effects of gene knockout (KO) experiments and enable integration of gene expression microarray data, genes can be connected to their associated enzymes. Boolean logic is often used to define this relationship. When multiple genes encode enzyme subunits that are each required for catalysis of a reaction, an "AND" statement is used to convey necessity of each gene for that reaction. Alternately, an "OR" statement is used to reflect the requirement of genes that encode isozymes that catalyze the same reaction independent of one another. Some enzymes may perform similar functions with unclear substrate specificity where it is difficult to correlate a particular gene with a particular pathway (Andersen et al. 2008) and may require approximation during modeling. The incorporation of gene-protein-reaction relationships (GPRs) have been incredibly valuable to researchers interested in studying the impact of different growth states, environment, or degree of infection measured with microarrays and connecting the gene expression results to a modeled reconstruction to investigate changes in metabolism (Oberhardt et al. 2010).

(iii) Biomass composition. Biomass, the key components from which an organism is built and sustains itself, may be determined from experimental measures, described here, or through more computational methods described in following sections. DNA, RNA, lipids, fatty acids, cell wall components, cytosolic solutes and ions, and cofactors must be present in sufficient amounts to enable cell operation (Senger 2010). Experimentally, biomass composition can be determined through fitting carbon content and amino acid content measured at different population levels and growth states (Gonzalez et al. 2010). The number of factors that should be included varies with the organism's metabolic characteristics, but even biomass composition

of the minimal cell model *Mycoplasma genitalium* includes 61 different factors (Suthers et al. 2009).

(iv) Thermodynamics. Assessing reaction thermodynamics can alleviate excessive constraining of the model's directionality (Henry et al. 2009). Directionality of reactions (reversible vs. irreversible) in the model and feasible rates of reaction are based on thermodynamic favorability determined from Gibbs free energy changes. Experimentally determined values can be found in literature, but available data is rarely sufficient to address a genome-scale reconstruction and more often has been obtained for smaller sections of metabolism or major canonical pathways. Methods for computationally estimating the free energy values of reactions include group contribution methods that can be performed online at Web GCM using reaction information from KEGG and BRENDA (Jankowski et al. 2008).

In vitro experimentation and validation

In conjunction with *in silico* reconstruction efforts that rely on annotations, databases, and literature review, valuable information can also be collected through *in vivo* and *in vitro* experimentation. Growing an organism on specific carbon sources via high throughput Biolog microplates or more traditional techniques is an easy method for initial model validation because carbon source can be limited *in silico*. The organism's carbon source usage and production of byproducts during growth can be investigated by measuring media metabolite concentration using high performance chromatography. Many organisms have also been examined to determine survival genes through mutagenesis studies that create single gene KO mutants (Jacobs et al. 2003; Gallagher et al. 2007; Cameron et al. 2008). Phenotypic characteristics such as substrate usage and byproduct synthesis of a particular mutant can be predicted *in silico* and then compared to *in vitro* experiments to investigate and improve GENRE accuracy. Investigating reaction fluxes and static and dynamic localization of specific enzymes in a living cell for comparison and incorporation with a GENRE is possible via metabolic mapping of fluorescent substrates and cofactors within live cells (Van Noorden 2010) and pulse

labeling, using 13C as a radiotracer to investigate distribution of labeled molecules (Niittylae et al. 2009).

Constraint-based modeling methods

Thus far, we have introduced the necessary steps to assemble an *in silico* metabolic GENRE whose components include stoichiometric coefficients for the reactants and products of each metabolic reaction, a Boolean GPR rule-set which defines the genes associated with each reaction, and notation for reversible and irreversible reactions. Each of these constrains the behavior of the metabolic model they comprise and limits the attainable metabolic phenotype of the cell. An organism's metabolic phenotype is often quantified by growth rate, the amount of biomass added per unit time. Because the reactions that produce individual biomass components are integrated into the GENRE (e.g., DNA, RNA, amino acids, lipids, carbohydrates), we can also characterize metabolic phenotype by the amount of flux (metabolite mass per dry organism mass per unit time) that is carried through these reactions in the GENRE. Constraint-based modeling aims to reduce the number of possible flux profiles and identify one that best predicts the metabolic phenotype of the organism under specified genetic and environmental conditions.

Constraints limit the number of possible flux profiles for a given organism to some finite number and additional constraints further reduce the flux space and allow us to make even more accurate *in silico* predictions of metabolic phenotype. These and other constraints are generally assigned to one of four categories: physico-chemical constraints (e.g., mass and energy balance), environmental constraints (e.g., temperature, pH, substrate availability), spatial/topological constraints (e.g., organelle compartmentalization), and self-regulatory constraints (Price et al. 2004). The rapid growth of complete genome sequencing and curation of metabolic GENREs and the utility of constraint-based modeling have driven the improvement and innovation of methods for constraint-based modeling. Here, we will introduce a range of these methods developed for constraint-based modeling of genome-scale metabolism using GENREs constructed with the methods detailed in following

sections. In particular, we will focus on the quantitative constraints and motivation for method development.

GENRE-to-model implementation

To test model predictions against experiments, the constraining elements of the GENRE are given mathematical structure, primarily within the *S* matrix, which quantifies the stoichiometric relationship between reactant and product metabolites for each reaction in the GENRE (Figure 2.2). The individual elements of *S* are the stoichiometric coefficients collected in the reconstruction process that correspond to the i^{th} metabolite within the j^{th} metabolic reaction, such that each of the *m* rows represents a metabolite ($M_{1,2,3}$) and each of *n* columns represents a reaction, including intracellular biochemical conversions ($R_{1,2}$), metabolite exchange (uptake, $X_{1,2}$; secretion, X_3), metabolite demand (cell maintenance, D_1), and inter-compartment metabolite transport (extracellular to cytosolic space, $T_{1,2}$). To account for a metabolite's presence in separate compartments, one row should be added to the *S* matrix for each compartment within which the metabolite exists. By convention reactants have a negative coefficient and products have a positive one, while flux, v_i, is positive for forward biochemical conversions, transport reactions into the cell, and exchange reactions out of the cell (i.e., secretion), and negative flux indicates a reverse intracellular reaction, an uptake exchange reaction and transport from the cytosol to another compartment. The GPR relationships between genes and reactions should be coded as a Boolean vector with an element for each reaction, while flux bounds and an objective function should be implemented in vector form, as described in the next section. One can carry out most of the constraint-based methods that follow with these few mathematical and Boolean structures.

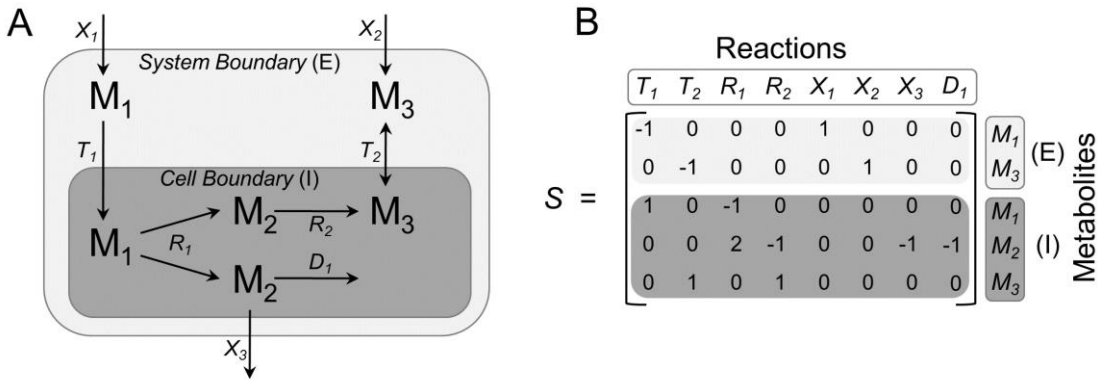


FIG 2.2. Conversion of GENRE reaction stoichiometries to *S* matrix. (A) Schematic of a toy cell/organism with defined boundaries (*E*/*I*), metabolites (*M*), and reactions (*X*,*T*,*R*,*D*). (B) – Five metabolites and seven reactions comprise the *S* matrix converted from the cell/organism of panel A.

Flux Balance Analysis

The most widely used constraint-based method in the field of metabolic engineering and systems analysis of metabolism is flux balance analysis (FBA), and most of the methods described herein are either explicit alternatives or add-on improvements to FBA. The number of review papers dedicated solely to FBA in recent years (Kauffman et al. 2003; J.M. Lee et al. 2006; Raman and Chandra 2009) gives an idea of the wide applicability of this approach. Two recent articles provide greater detail to support carrying out FBA (Oberhardt et al. 2009) as well as a concise, visually aided description of what is achieved by performing FBA (Orth et al. 2010). FBA allows for the *in silico* prediction of a flux profile that optimizes some pre-defined cellular objective without making any experimental measurements. It has become a standard tool for simulating the effects of genetic perturbations or environmental conditions on an organism's growth rate or rate of byproduct synthesis (Raman and Chandra 2009). FBA has also provided a means to simulate the metabolic effects of global gene expression and other large "omics" datasets for an organism or even a particular human tissue type (Becker and Palsson 2008; Shlomi et al. 2008; Jerby et al. 2010).

The underlying theory for FBA assumes that cellular metabolism is at a steady-state (i.e., constant growth rate) (Savinell and Palsson 1992b; Varma and Palsson 1994),

and all reaction fluxes (vi) and metabolite concentrations (Ci) have reached a steady state, such that the net amount of each metabolite added to and removed from the cell must equal zero (Eq. 1). Given this, FBA uses linear programming (LP) to find the steady-state flux profile (v) that optimizes the organism's objective function (e.g., maximal growth rate) while satisfying a set of physico-chemical and environmental constraints. The primary FBA constraint is mass balance, requiring that the amount of each metabolite that is (1) transported into or (2) produced within the cell be balanced by the amount of that metabolite that is (3) consumed within or (4) secreted from the cell. The mass balance constraint (Eq. 1) provides m metabolite balance equations to solve for n unknown reaction fluxes, and because there are usually more reactions than metabolites, the system of equations is under-determined. As such, the collection of flux vectors that satisfy the mass balance constraints span a multidimensional space, often referred to as the null space of the S matrix.

$$\frac{dC_i}{dt} = \sum_{j=1}^{n} S_{ij} \cdot v_j = 0 \qquad \forall i \in M \tag{1}$$

$$\alpha_j \le v_j \le \beta_j \qquad \forall j \in N \tag{2}$$

It is important to note that many of the flux profile vectors, v, in the null space are physically impossible for a number of reasons (Orth et al. 2010). In order to refine and reduce the size of the predicted flux space and exclude irrelevant flux profiles, the standard FBA implementation constrains individual reaction fluxes, v_j, by defining lower and upper bounds (α_j and β_j, respectively; Eq. 2). Some of these bounds are based on literature and experimentation and collected in the GENRE process. These bounds can be applied to enforce environmental constraints, such as the availability of a substrate in growth media, by setting the upper and lower uptake exchange bounds to the same negative value. Similarly setting the lower and upper bounds on an exchange reaction to zero and some arbitrarily high value (~1000), respectively, will prevent metabolite uptake, while leaving the rate of secretion of a desired output metabolite unconstrained. Bounds can also be set on internal biochemical reactions by constraining the lower bound to zero to enforce

thermodynamic irreversibility of a reaction, by setting the bounds to reflect the kinetic limitations of a particular enzyme or membrane transporter, or by constraining the flux to zero to simulate a gene KO. This approach is often used to identify genes essential for growth, both for identification of drug targets and maximizing the production of targeted byproducts (Feist and Palsson 2008; Oberhardt et al. 2009). Several of the methodologies described below impose more sophisticated or problem-specific constraints on the flux bounds to minimize the possible flux space and provide more accurate predictions of metabolic network behavior, though a standard FBA implementation typically enforces constraints on substrate uptake, reaction irreversibility, and flux capacity.

Once the mass balance constraint and flux bounds are set, an objective function (v_{obj}, Eq. 3) must be defined for the LP optimization problem. The optimization problem will find a flux profile, v, which maximizes or minimizes the objective function and satisfies the mass balance constraints and flux bounds described above. While the objective function can be unique for each FBA problem, a common choice for microorganisms is maximization of biomass (e.g., proteins, nucleic acids, carbohydrates, and lipids necessary for growth), based on the idea that evolutionary pressures select for higher growth rates of these organisms. In some organisms, however, alternative objective functions such as maximization (minimization) of ATP production (utilization) are more appropriate, as will be further discussed below. In those GENREs for which maximization of biomass production is the organism's objective, a "biomass reaction" is added as a column to the S matrix with stoichiometric coefficients in proportion to each metabolite's *in vivo* contribution to one unit of biomass determined as explained previously (Oberhardt et al. 2009). To implement the chosen objective function, a "cost" vector (c) is defined, whose elements correspond to each reaction (i.e., columns of S, elements of v). To maximize the desired reaction, all elements are set equal to zero except that corresponding to the chosen objective reaction (Eq. 4). FBA identifies an optimal flux vector whose elements corresponding to the non-zero elements of the cost vector are maximized, subject to the defined constraints.

$$\max \quad v_{obj} = c_j \cdot v_j \tag{3}$$

$$c = [0\,0\,0\,..\,c_{obj}\,..\,0\,0\,0] \qquad c_{obj} > 0 \tag{4}$$

Dynamic Flux Balance Analysis

The steady-state assumption required of FBA is not appropriate for *in silico* modeling of organisms in all situations, most notably during a "diauxic shift" from one carbon source to another. To address this, Dynamic Flux Balance Analysis (DFBA) was developed, based on an earlier approach (Varma and Palsson 1994), implementing both dynamic (non-linear programming, NLP) and static (LP) optimization of an objective function and the addition of constraints to the rates of change of flux in addition to the typical FBA constraints (Mahadevan et al. 2002). The static optimization – a more tractable computational approach for large biochemical networks – performs a series of FBA problems at the beginning of each discrete time interval. Using both approaches, Mahadevan and colleagues correctly predicted the timing of acetate production and the sequence of substrate utilization (glucose before acetate) in *Escherichia coli*, validating the use of this constraint-based method to simulate the dynamic behavior of genome-scale metabolism.

Flux Variability Analysis

The redundancy inherent in most biochemical networks corresponds to the possibility for numerous alternative flux distributions that yield the same maximal growth rate or ATP synthesis rate (Papin et al. 2002). Put in graphical terms, the multidimensional flux space of possible flux profiles often includes multiple points with an identical optimal objective function value. In metabolic pathway terms, different combinations of intermediate reaction fluxes give rise to the same flux through the objective reaction (e.g., biomass), and each of these alternative pathways may be biologically meaningful. FBA is often used to identify a reference *in silico* flux profile with which to compare constraint-based behavior of a genetically or environmentally perturbed counterpart (see Minimization of Metabolic Adjustment below). As such, it is critical to understand flux variability, as two

alternate FBA reference flux profiles could lead to a significantly different assessment of the perturbed state.

A report by Smallbone and Simeonidis introduced a computational method to identify a unique flux profile by analyzing the geometry of the multidimensional flux space defined by the standard FBA constraints (Smallbone and Simeonidis 2009). This method identifies a minimal flux solution without any thermodynamically infeasible fluxes. An earlier approach by Mahadevan and co-workers called Flux Variability Analysis (FVA), uses constraints to assess the variability of each reaction flux toward an optimized objective flux value (Mahadevan and Schilling 2003). FVA first uses FBA to determine the optimal objective flux (Eqs. 3-5), which is used with mass balance and flux bounds (Eqs. 1-2) as constraints for the subsequent one-by-one maximization and minimization of each reaction flux within the GENRE (Eqs. 6-7). This method quantifies the maximum and minimum fluxes for each reaction that are consistent with the optimal objective function value.

$$v_{obj} = \max\left(c_j \cdot v_j\right) \tag{5}$$

$$\max\ v_j, \quad \forall j \in N,\ j \neq obj \tag{6}$$

$$\min\ v_j, \quad \forall j \in N,\ j \neq obj \tag{7}$$

Minimization of Metabolic Adjustment

Flux balance analysis assumes that each organism's metabolic network has been tuned through evolution for some objective function, be it maximal growth rate or energy efficiency (e.g., minimal ATP utilization). While this assumption may be valid for wild-type (WT) organisms that have evolved over many hundreds or thousands of generations, it may be less appropriate for engineered mutants, genetically modified in a controlled laboratory environment and unexposed to the same evolutionary forces or number of reproduction cycles. As such, Segre and colleagues developed Minimization of Metabolic Adjustment (MOMA), hypothesizing that mutant organisms are unable to immediately adapt their metabolic network to achieve the WT objective function, but instead display some suboptimal flux profile

intermediate to the FBA-determined optima of both the WT (WT-FBA) and genetically perturbed (KO-FBA) organisms (Segrè et al. 2002). This approach finds a suboptimal flux profile (v) that is a minimal Euclidean distance from the WT-FBA flux profile (vFBA, Eq. 8), and is mathematically formalized as a quadratic programming (QP) problem (Eq. 9), subject to the standard FBA mass balance and flux capacity constraints (Eqs. 1-2), including constraining all KO gene-associated reaction fluxes to zero (Eq. 10). Ultimately, FBA combined with MOMA provides a more accurate prediction of the immediate metabolic response to KO than FBA does on its own (Segrè et al. 2002).

$$D\left(v^{FBA},v\right)=\sqrt{\sum_{j=1}^{N}\left(v_j^{FBA}-v_j\right)^2} \qquad \forall j \in N \tag{8}$$

$$\min \left(v^{FBA}-v\right)^T\left(v^{FBA}-v\right) \tag{9}$$

$$v_k = 0, \qquad \forall k \in A \tag{10}$$

Regulatory On/Off Minimization

The Euclidean distance metric introduced in MOMA finds a sub-optimal KO-associated flux distribution with many component fluxes slightly altered from their WT-FBA state. Shlomi and colleagues developed an alternative method, Regulatory On/Off Minimization (ROOM), which seeks a flux profile for the mutant organism with a minimal number of significant flux changes from the WT-FBA flux profile (Shlomi et al. 2005). This method is based on the assumptions that a KO organism will minimize the costs associated with adapting its gene regulatory network, and that these costs are quantifiable – using a Boolean on/off framework – and independent of the magnitude of change in gene expression and associated reaction flux.

ROOM does not explicitly incorporate any gene regulatory constraints into the model, but accounts for them via a binary cost variable for each reaction (y_j set to one for fluxes that are significantly different from the WT-FBA; Eq. 12). This quantification of an altered flux is assumed to reflect a concordant change in

expression of the necessary enzyme-encoding gene, and the associated cost to the organism of regulating that gene's expression. The objective of this mixed-integer linear programming (MILP) problem is to minimize the number of fluxes in the mutant GENRE that are different from their WT-FBA counterparts (Eq. 11), while satisfying the FBA mass balance, irreversibility, and flux capacity constraints (Eqs. 1-2), and constraining the set of fluxes (A) associated with the given knockouts to zero (Eq. 10). The upper and lower thresholds by which flux changes are deemed significant (*v* FBA+ and *v* FBA-, respectively) are user-defined via relative (δ) and absolute (ε) flux tolerance terms (Eqs. 13-14). Accordingly, if y_j is set to zero, v_j is constrained by these bounds, while setting y_j to one adds to the cost sum (Eq. 11) and leaves v_j unbounded (Eqs. 15-16). This method was shown to improve predictions of flux and steady-state growth rate over MOMA, though it performed very similarly to FBA (Shlomi et al. 2005).

$$\min \sum_{j=1}^{n} y_j \tag{11}$$

$$y_j \in \{0, 1\} \tag{12}$$

$$v_j^{FBA+} = v_j^{FBA} + \delta\left|v_j^{FBA}\right| + \varepsilon \tag{13}$$

$$v_j^{FBA-} = v_j^{FBA} - \delta\left|v_j^{FBA}\right| - \varepsilon \tag{14}$$

$$v_j - y_j(v_{\max,j} - v_j^{FBA+}) \le v_j^{FBA+} \tag{15}$$

$$v_j - y_j(v_{\max,j} - v_j^{FBA-}) \le v_j^{FBA-} \tag{16}$$

Objective function search methods

Both MOMA and ROOM proposed an alternative to the FBA assumption that organisms are evolved for a specific objective function. Though MOMA and ROOM were developed to address this assumption in the context of a gene knockout, the question of a universal WT objective function for an organism remains. To address this, Schuetz and colleagues tested the ability of 11 unique FBA objective functions (or combinations thereof) to accurately predict fluxes in an *E. coli* model of central metabolism (Schuetz et al. 2007), and reported that no single *in silico* objective

function was the most accurate predictor of 13C-determined *in vivo* fluxes across many different growth conditions. Several constraint-based approaches have been developed to identify the best objective function for an organism in a specified environment, including ObjFind, BOSS, and a Bayesian probability-based selection method.

ObjFind. ObjFind (Burgard and Maranas 2003), uses bi-level programming to find a set of positive weights, c_j, that maximize the sum of optimal fluxes (Eqs. 18-19), while minimizing the sum-squared difference between the optimal flux profile, v_j, and the experimentally measured fluxes, v_j* (Eq. 17, where *E* is the set of all experimentally measured fluxes and *P* is the set of all reactions that could potentially be cellular objectives). By comparing these optimal weights to those given by a hypothesized objective function, such as biomass, a researcher is able to assess the accuracy of the hypothesized objective function. However, this approach requires that all components of the true objective function be included within the *S* matrix, a priori, and will return the wrong objective if a component reaction is not included in the GENRE.

$$\min_{c_j} \quad \sum_{j \in E} (v_j - v_j^*)^2 \tag{17}$$

$$\max_{v_j} \quad \sum_{j \in P} c_j v_j \tag{18}$$

$$\sum_{j \in P} c_j = 1, \qquad c_j \geq 0, \qquad \forall j \in P \tag{19}$$

Biological Objective Solution Search. Gianchandani and co-workers addressed the a priori requirement of ObjFind by adding a generic "objective reaction," v_{obj}, to the stoichiometric *S* matrix (Gianchandani et al. 2008). This Biological Objective Solution Search (BOSS) requires only the *S* matrix and experimental isotopomer flux data. BOSS adds the generic objective reaction with unknown stoichiometry as an additional column in the *S* matrix (Eq. 20, where m is the total number of metabolites) and performs an FBA on this updated set of mass balance constraints

and flux bounds (Eqs. 1-2), defining v_{obj} as the objective function to maximize (Eq. 22).

$$S_{i,obj} = [S_{1,obj}\ S_{2,obj}\ \ldots S_{m,obj}] \tag{20}$$

$$\min \quad \sum_{j \in N} \left(v_j^{BOSS} - v_j^* \right)^2 \tag{21}$$

$$\max \quad v_{obj} \tag{22}$$

Similar to ObjFind, this inner optimization serves as a constraint on the outer optimization, which minimizes the sum-squared difference between the *in silico* flux profile (v_j^{BOSS}), and the measured *in vivo* fluxes (v_j*). This approach will identify the optimal objective reaction if it already exists in the network (i.e., with identical stoichiometry as an existing column in *S*), but also if it is a combination of existing reactions or a reaction that was omitted from the GENRE altogether. To validate this method, BOSS was used to identify the objective function in the central metabolic network of *S. cerevisiae*, and found the best two reactions to be nearly identical to the precursor biomass synthesis reaction and the ATP maintenance reaction, both commonly used FBA objective functions.

Bayesian discrimination. Knorr and colleagues introduced a method which uses FBA and *in vivo* flux measurements to compare any number of candidate objective functions (Knorr et al. 2007). This method determines the most probable objective function by calculating posterior probabilities. Specifically it calculates the probability of each objective function (F_x), given the product of differences between predicted and measured data (*Y*), normalized by the sum of all posterior probabilities (Eq. 23). While this approach again requires the definition of candidate objective functions a priori, the validation used only a few *in vivo* measurements (growth rate, oxygen uptake rate, succinate uptake rate and acetate production rate) to compare to *in silico* predictions.

$$\pi(F_x|Y) = \frac{p(F_x|Y)}{\sum_z p(F_z|Y)} \quad (23)$$

Multiple metabolic objectives

Selecting the most biologically relevant objective function is critical to accurately predicting cellular metabolism by FBA or other constraint-based methods, though more than one objective function is occasionally desired. For example, metabolic engineering often aims to identify the genetic manipulations for an organism that will optimize its synthesis rate of a desired byproduct (e.g., ethanol in *E. coli*). An appropriate approach must optimize the organism's intrinsic objective function (e.g., max growth) in parallel with this secondary engineered objective. Burgard and colleagues developed OptKnock, a constraint-based method, to tackle this problem of multiple objective functions (Burgard and Maranas 2003). Similar to the bi-level optimizations of ObjFind and BOSS, OptKnock uses an FBA framework to maximize the cellular objective, v_{obj}, subject to mass balance and flux bounds constraints (Eqs. 1-2), as well as gene knockout constraints (Eqs. 25-28, where K is the maximum number of allowable knock-outs across the set of all reactions, N). In parallel, this approach varies the number and identity of knockout genes to maximize byproduct secretion, $v_{byproduct}$ (Eq. 24). More recently, numerous constraint-based metabolic engineering methods have been developed to identify optimal gene deletions with more computational efficiency than OptKnock (OptGene and GDLS), and to identify optimal gene knock-ins (OptStrain) or manipulations to increase or decrease gene expression (OptReg) (Pharkya et al. 2004; Patil et al. 2005; Pharkya and Maranas 2006; Lun et al. 2009).

$$\max_{y_j} \quad v_{byproduct} \tag{24}$$

$$\max_{y_j} \quad v_{obj} \tag{25}$$

$$v_j^{\min} \cdot y_j \le v_j \le v_j^{\max} \cdot y_j, \qquad \forall j \in N \tag{26}$$

$$y_j = \{0, 1\} \tag{27}$$

$$\sum_{j \in N} \left(1 - y_j\right) \le K \tag{28}$$

Constraint-based modeling software

The bulk of the aforementioned constraint-based modeling approaches are LP problems, which can be solved using a standard solver of which there are many open source (LP_Solve, glpk) and proprietary versions (Gurobi, CPLEX, LINDO). The constraint-based reconstruction and analysis (COBRA) toolbox has become a popular means to apply many of these constraint-based methods (e.g., FBA, MOMA, and FVA) to GENRE-derived models (Becker et al. 2007). This toolbox is open source and has been written to operate in the MATLAB programming environment using the systems biology markup language (SBML). A recently developed open source software platform, OptFlux, combines strain optimization methods (e.g., OptKnock) as well as classic constraint-based approaches FBA, MOMA and ROOM (Rocha et al. 2010).

Summary

In this chapter we have attempted to detail the key steps in the metabolic reconstruction process as well as introduce several constraint-based modeling techniques useful for various applications and desired outcomes. Again, it is critical to use the highest quality annotation available and dedicate significant time to refinement of the network reconstructions with well-curated databases and quality literature references in addition to the results of automated reconstruction techniques. Some common methods of collecting experimental data for both model improvement and validation are detailed, but many more are available that may be more appropriate for specific applications. The constraint-based modeling

techniques presented are standards in the FBA field and accompanied by additional information helpful during difficult portions of the modeling process such as objective function definition and variability analysis. The field is a fast growing one, with an exponential rise in new reconstructions and further investigations into comparative model building techniques and reconstruction consistency, and we hope this text will be a helpful reference in these efforts.

Chapter 3: Differential metabolism in emerging opportunistic pathogens

Acknowledgements: Phillip Yen[1], John Varga[2], Joanna Goldberg[2], Jason Papin[1*]

[1]Biomedical Engineering, University of Virginia

[2]Pediatrics, Emory University School of Medicine

The text of this chapter has been published here:
Bartell, J. A.*, Yen, P.*, Varga, J. J., Goldberg, J. B., & Papin, J. a. (2014). Comparative metabolic systems analysis of pathogenic *Burkholderia*. Copyright © American Society for Microbiology, *Journal of Bacteriology*, 196(2), 210–226. doi:10.1128/JB.00997-13
* Co-first authorship

SYNOPSIS

Burkholderia cenocepacia and *Burkholderia multivorans* are opportunistic drug-resistant pathogens that account for the majority of *Burkholderia cepacia* complex infections in cystic fibrosis patients and also infect other immunocompromised individuals. While they share similar genetic compositions, *B. cenocepacia* and *B. multivorans* exhibit important differences in pathogenesis. We have developed reconciled genome-scale metabolic network reconstructions of *B. cenocepacia* J2315 and *B. multivorans* ATCC 17616 in parallel, designated iPY1537 and iJB1411, respectively, to compare metabolic ability and contextualize genetic differences between species. The reconstructions capture the metabolic function of the two species and give insight into similarities and differences of their virulence and growth capabilities. The two reconstructions have 1,437 reactions in common, and iPY1537 and iJB1411 have 67 and 36 metabolic reactions unique to each, respectively. After curating the extensive reservoir of metabolic genes in *Burkholderia*, we identified 6 genes essential to growth that are unique to iPY1513 and 13 genes uniquely essential to iJB1411. The reconstructions were refined and validated by comparing *in silico* growth predictions to *in vitro* growth capabilities of *B. cenocepacia* J2315, *B. cenocepacia* K56-2, and *B. multivorans* ATCC 17616 on 104 carbon sources. Overall, we identified functional pathways that indicate *B. cenocepacia* can produce a wider array of virulence factors compared to *B. multivorans*, which supports the clinical observation that *B. cenocepacia* is more virulent than *B. multivorans*. The reconciled reconstructions provide a framework for generating and testing hypotheses on the metabolic and virulence capabilities of these two related emerging pathogens.

INTRODUCTION

Multi-drug resistant pathogens are a severe health concern and can cause chronic infections in a variety of patient populations with limited recourse for treatment. Here, we investigate two multi-drug resistant species, *B. cenocepacia* and *B. multivorans*, of the *Burkholderia cepacia* complex (Bcc) which are considered dangerous and difficult to treat in patients with cystic fibrosis (CF), chronic granulomatous disease, or compromised immune systems (Mahenthiralingam et al. 2002). With larger genomes (8.06 Mbp and 7.01 Mbp, respectively) than many other multi-drug resistant pathogens, they also contain an expanded reservoir of genes that may assist their ability to avoid clinical eradication (Rice 2008; Slama 2008; Boucher et al. 2009). Nosocomial, transmissible between patients, and also routinely acquired (and reacquired) from the environment, *B. cenocepacia* and *B. multivorans* are the two Bcc species most commonly isolated from the sputum of CF patients (De Boeck et al. 2004; Jones et al. 2004; Vandamme and Coenye 2004; Horsley et al. 2011). In patients with CF, pulmonary infection with Bcc can contribute to the rapid deterioration of lung function known as cepacia syndrome, a necrotizing pneumonia that can lead to bacteremia, septicemia, and increased mortality (Isles et al. 1984; Glass and Govan 1986). A combination of high microbial antibiotic resistance and decreased immune function in patients makes *B. cenocepacia* and *B. multivorans* extremely difficult to eradicate once lung colonization is established (Hancock 1998; Aaron et al. 2000). The potential to cause chronic infection as well as rapid decline in health make Bcc clinically important emerging human pathogens.

Further complicating treatment is the variability in infection course and severity in patients infected with *B. cenocepacia* or *B. multivorans* (or occasionally both species). Compared to patients infected with *Pseudomonas aeruginosa*, the most common CF lung pathogen in adults, patients infected with Bcc can have a substantially worse prognosis (Snell et al. 1993; Marolda et al. 1999). Cepacia syndrome has historically been associated with *B. cenocepacia*, but infections with *B. multivorans* have on occasion also induced this syndrome (Horsley et al. 2011). *B. cenocepacia* causes greater mortality than *B. multivorans*, and has also been

seen to replace *B. multivorans* infections (Mahenthiralingam et al. 2001). However, many centers in multiple countries are showing a shift within Bcc infection incidence from predominantly *B. cenocepacia* to *B. multivorans* infections that is not fully understood (Horsley et al. 2011). These differences between *B. cenocepacia* and *B. multivorans* provide motivation for a comparative systems analysis of these two Bcc species. Comparing and contrasting the metabolic functions of *B. cenocepacia* and *B. multivorans* can help elucidate their metabolic adaptability, mechanisms of pathogenicity, and other underlying contributors to differing clinical outcomes such as cepacia syndrome.

Genome-scale metabolic reconstructions of multiple organisms can be used to study and predict phenotypic differences between related species or strains based on their genetic content (Oberhardt et al. 2009). For example, a recent study used the framework of two previously published reconstructions of *Staphylococcus aureus* N315, the ERGO bioinformatics suite, and the KEGG pathway suite to assemble a consensus species-level reconstruction of *S. aureus.* Strain-specific enzyme annotations were then incorporated to develop 13 reconciled models of *S. aureus* strains that identified differences in common essential enzymes (Lee et al. 2009). Meaningful biological comparison becomes more complex when using species-level reconstructions. These reconstructions must be intensively curated in such a way that all artifacts from the model building process are reconciled, as performed in the reconstruction-based comparison of two species of *Pseudomonas* (Oberhardt et al. 2011). Artifacts that must be reconciled include differences in naming conventions, reaction stoichiometries, and, importantly, annotations of gene functions and their implementation in gene-protein-reaction associations. Proper reconciliation of genome-scale metabolic reconstructions is crucial for studying the genetic and phenotypic differences between the organisms.

Here, we present reconciled genome-scale metabolic reconstructions of *B. cenocepacia* and *B. multivorans*, two of the largest bacterial reconstructions built to date in terms of the number of genes and reactions incorporated; this correlates with the large genomes of the pathogens. A previously published reconstruction of

B. cenocepacia J2315, iKF1028, using the ToBiN reconstruction platform accounted for 859 reactions, 834 metabolites, and 1,028 genes and was used as a reference for our curation efforts (Fang et al. 2011). However, we chose to begin with draft reconstructions both built using the Model SEED tool (Henry et al. 2010), currently the most widely used publicly available reconstruction platform, to aid our comparative analyses. Our *B. cenocepacia* reconstruction is substantially larger than iKF1028 while including much of the knowledge gained from this prior study. The reconstructions for *B. cenocepacia* J2315 and *B. multivorans* ATCC 17616 were developed and manually curated in parallel to ensure consistency in all aspects of the model-building process. In particular, the two reconstructions were reconciled in the annotation and assignment of orthologous gene functions which span 1,437 common metabolic reactions.

The reconstructions of *B. cenocepacia* J2315 and *B. multivorans* ATCC 17616 are denoted formally as iPY1537 and iJB1411, respectively, following established naming conventions (Reed et al. 2003). However, for the sake of clarity, iPY1537 and iJB1411 will be referred to as iBC and iBM throughout this manuscript for *in silico B. cenocepacia* and *in silico B. multivorans*, respectively. The reconstructions were validated by comparing growth predictions with substrate utilization experiments using *B. cenocepacia* J2315, *B. cenocepacia* K56-2, and *B. multivorans* ATCC 17616. Through the reconstruction and reconciliation process, we reannotated the functions of a collection of genes. Predictions of genes essential for growth were made by simulating *in silico* growth of iBC and iBM, which has relevance in drug target identification. Pathogenic characteristics of the two species were compared by predicting the production capacity of an array of metabolites involved in virulence. This study provides a framework for investigating the metabolic architectures of two clinically important Bcc species emphasizing metabolic connections to virulence and pathogenicity, as well as a guide for parallel reconstruction and comparison of genome-scale metabolic networks of related organisms.

MATERIALS AND METHODS

Metabolic reconstructions of B. cenocepacia *and* B. multivorans

The complete sequenced genomes of *B. cenocepacia* J2315 and *B. multivorans* ATCC 17616 were used as the starting points for the metabolic network reconstruction process shown in Figure 3.1. The J2315 strain of *B. cenocepacia* was isolated from a cystic fibrosis patient (Holden et al. 2009) and the ATCC 17616 strain of *B. multivorans* was an environmental isolate obtained from the soil (Stanier et al. 1966). The DNA genomic sequences of *B. cenocepacia* J2315 (Genbank accession numbers NC_011000 to NC_011003) and *B. multivorans* ATCC 17616 (Genbank accession numbers NC_010084, NC_010086, NC_010087, and NC_010070) were downloaded from the NCBI Genbank database.

The Model SEED was used to generate draft reconstructions of *B. cenocepacia* and *B. multivorans* (Henry et al. 2010). The Model SEED requires annotated genomes as input in order to produce draft reconstructions. While the finished annotation for *B. cenocepacia* had already been uploaded from NCBI and converted to the SEED format, the genome of *B. multivorans* was not included in the SEED database. The genome of *B. multivorans* was downloaded from the NCBI Genome database and was submitted to the RAST server for annotation (Aziz et al. 2008), and then the Model SEED was used to produce draft reconstructions of both species. Differences in annotation quality between these submission methods were evident in the resulting draft reconstructions and addressed during the manual curation process. We allowed SEED to gap-fill needed reactions to enable growth on SEED's complex media formulation (all exchange reactions open) and default biomass formulation in the draft construction process.

We then used biological databases such as KEGG (Ogata et al. 1999), Metacyc (Caspi et al. 2014) and the *Burkholderia* Genome Database (BGD) (Winsor et al. 2008) to evaluate the draft reconstructions. Specifically, we used the gene and reaction annotations in KEGG and compared them with those of the draft reconstructions. Protein BLAST was used to determine genes that likely code for isozymes, both within each genome as well as between the two genomes. This

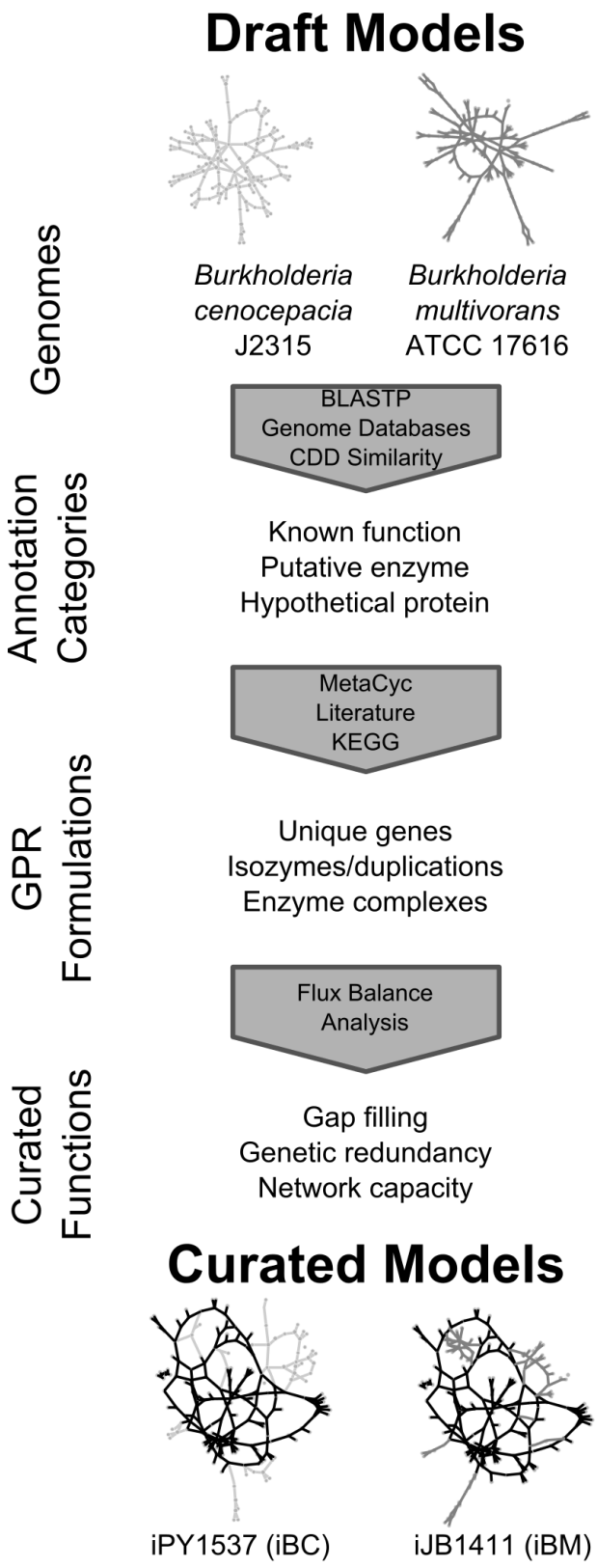


FIG 3.1. Overview of the reconstruction and reconciliation process. The annotated genomes of *B. cenocepacia* J2315 and *B. multivorans* ATCC17616 were used to generate automated draft metabolic reconstructions using Model SEED (20,60). The gene-protein-reaction associations of the draft reconstructions were then further manually curated and annotated using information from online databases and literature references. Curation of the two reconstructions was done in parallel to ensure the proper assignment of homologous and orthologous genes. The reconstructions were then functionally curated through iterative validations of *in silico* and experimental data. The final reconciled models for *B. cenocepacia* and *B. multivorans* are designated as iPY1539 and iJB1413, respectively, but are referred to throughout as iBC and iBM for clarity.

Abbreviations: GPR, Gene-Protein-Reaction; BLASTP, Protein Basic Local Alignment Search Tool; CDD, Conserved Domain Database; KEGG, Kyoto Encyclopedia of Genes and Genomes

process expanded the size of the models, as well as reduced the number of non-gene-associated reactions (gap-filled by SEED) needed in rich media to 15 reactions (11 unique enzymes). Ten more reactions were used in gap-filling virulence-related pathways and growth in particular minimal media conditions. Thirty-seven and 39 reactions were removed from the original SEED drafts of iBC and iBM, respectively, due to low annotation evidence, low likelihood of enzyme activity or infeasible loops formed in the model. The final curated models iPY1537 and iJB1411 are available as spreadsheets and SBML files on the Papin Lab website (http://bme.virginia.edu/csbl).

Flux Balance Analysis of iBC and iBM

Flux balance analysis (FBA) was used to assess *in silico* growth of iBC and iBM (Orth et al. 2010). Mathematically, the FBA problem is formulated as:

$$\begin{aligned} &\text{maximize } v_{biomass} \\ &\text{such that} \\ &S \bullet v = 0 \\ &v_{min} \leq v \leq v_{max} \end{aligned}$$

In the FBA framework, the reactions of a metabolic network are represented using a stoichiometric matrix, **S**, such that the rows of **S** correspond to the metabolites and the columns correspond to the reactions. The elements of the matrix are the stoichiometric coefficients so that S_{ij} is the coefficient for metabolite *i* in reaction *j*. At steady-state, we have the relation $\mathbf{S} \bullet v = 0$ where the vector *v* represents the fluxes of each of the reactions. The fluxes are conventionally in units of mmol/gDW/h (where gDW is gram dry cell weight), with the exception of $v_{biomass}$, which has units of h^{-1} (representative of the growth rate). Because $\mathbf{S} \bullet v = 0$ is underdetermined, FBA uses linear programming to optimize the flux through an objective reaction (typically, the biomass reaction, denoted as $v_{biomass}$). The biomass reaction includes the relative weights of the metabolites required to form biomass and details of its formulation are described below. v_{min} and v_{max} are vectors of the upper and lower bounds of the reaction fluxes, respectively. Consequently, media

conditions are imposed by constraining the bounds of the exchange reactions, which represent the uptake and secretion of metabolites in the model (e.g. $v_{min}<0$ for a metabolite that is imported and $v_{max}>0$ for a metabolite that is secreted). The formulations of the different media conditions are described on the Papin lab website. All FBA simulations were conducted using the COBRA Toolbox in MATLAB (Schellenberger et al. 2011).

Biomass reaction formulation

Burkholderia-specific biomass reactions were formulated for iBC and iBM. To evaluate capability of growth in a variety of environments and genetic backgrounds, the biomass reactions are often set as the objective functions in FBA and the numerical value of the flux through the biomass reactions corresponds to the theoretical maximal growth rates under a given media condition (Feist and Palsson 2010). The constituents of the biomass reactions represent the molecular components that are required for growth. These molecular components are grouped according to the main macromolecular components, which include protein, DNA, RNA, lipopolysaccharide, peptidoglycan, glycogen, lipids, and polyamine pools. *Burkholderia* species exhibit unique fatty acid and lipid compositions such as the presence of fatty acids containing cyclopropane rings, and these were accounted for in the formulation of the composition of the biomass reaction (Taylor et al. 1998). The biochemical reactions needed to synthesize many of these unique lipids and fatty acids have been included in the reconstructions based on work done previously (Fang et al. 2011). The stoichiometric coefficients for constituents of the biomass reactions of iBC and iBM were based on experimental evidence for *Burkholderia* and related organisms where possible. Otherwise, values from the well-annotated biomass reaction of *Escherichia coli* were used (Feist et al. 2007). Values specific to the two species were used when possible (e.g. RNA, DNA, protein). Otherwise, for constituents not specific to either species, the same value was used for both reconstructions (e.g. lipids, fatty acids, cell wall, ATP maintenance).

Predicting essential genes

Essential genes are defined as genes for which there is no *in silico* growth in a given media condition (maximum flux through $v_{biomass}$ is zero) when the reaction(s) associated with that gene is (are) removed from the network. To simulate the loss of a gene in the model, gene-protein-reaction (GPR) associations are evaluated to determine which reactions are not allowed to carry flux when the gene is removed. GPRs link genetic information such as isozymes, gene duplications, and enzyme complexes to the reactions in the reconstructions using Boolean logic. Perturbations of the GPRs enable the prediction of phenotypic changes as a function of genetic changes. For example, the fluxes for the reactions affected by a gene deletion are constrained by setting the corresponding v_{max} and v_{min} to zero, and the model is then optimized for growth to evaluate the effect of the missing gene.

in silico media conditions

Media conditions were formulated for lysogeny broth (LB), M9 minimal media, and synthetic cystic fibrosis media (SCFM) as in previous publications (Palmer et al. 2007; Oberhardt et al. 2010). The media conditions were defined by setting the lower and upper bounds of the exchange reactions to specify metabolites that are available or unavailable. Full descriptions of the media conditions are presented on the Papin lab website. All media conditions allow for the exchange of typical salts and water. LB contains amino acids, glucose and salts. SCFM contains lactate in addition to similar carbon sources to LB; in addition, the uptake rates of all nutrients present in SCFM are constrained to 10 mmol/gDW/h to represent the reduced nutrient availability in the cystic fibrosis lung environment as done elsewhere (Oberhardt et al. 2010). In M9 minimal media, a single carbon source is enabled for uptake at 10 mmol/gDW/h to represent tested compounds in *in vitro* Biolog growth screening (Oberhardt et al. 2008).

Prediction of virulence factor production capacity

Virulence factor production capacity was assessed by adding a demand reaction to the model for each virulence-related compound. Maximizing the flux through this reaction allows prediction of the maximum theoretical production level under a

given media condition such as with each of the individual carbon substrates in SCFM. Tradeoff between growth and virulence factor production was assessed by maximizing the flux through a demand reaction for each factor while constraining the biomass flux to various percentages of its maximum value when simulated under SCFM conditions.

Network visualization

MetDraw was used to generate visual representations of the metabolic networks (Jensen and Papin 2014). SBML versions of the models were used as inputs to MetDraw and the resulting network visualizations were output as SVG files.

Growth phenotype screening

We conducted growth screening using Biolog PM1 and PM2A microplates (Biolog, Inc., Hayward, CA). Instead of adding redox dye and measuring oxidation of each carbon source included in the screen, we evaluated growth in each well using optical density measurements at 700 nm to avoid error due to pigment production. Bacteria were scraped from an LB plate and resuspended in Biolog inoculating fluid PM IF-0a GN/GP to reach an optical density of 0.07. Biolog microplate wells were inoculated with 100 µl of this suspension, and plates were then incubated at 37°C for 48 h. Optical density at 700 nm was measured at time zero and every 12 h thereafter. Experiments with *B. cenocepacia* J2315 were carried out using two biological replicates each for PM1 and PM2A, while experiments with *B. cenocepacia* K56-2 and *B. multivorans* ATCC 17616 were performed using three biological replicates. Growth was evaluated based on the resulting growth curves and maximum change in OD for each carbon source compared to maximum change in OD of inoculated control wells with no carbon source. Additionally, we evaluated the ability of each organism to catabolize cysteine and tryptophan, which are amino acids present in cystic fibrosis lung sputum but not included in the Biolog plates (Palmer et al. 2007). These amino acids were each added at concentrations of 20 mM to M9 salts to create cysteine and tryptophan minimal media. Colonies of each *Burkholderia* strain were scraped from LB plates and resuspended to reach 0.07 OD in each respective minimal media. Three replicates of 200 µl of each inoculated

strain-media combination were plated in a 96-well microplate and grown in conditions replicating the Biolog screen.

An alternate strain-specific network reconstruction of *B. cenocepacia* K56-2 was created by conducting a 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). Protein sequences were derived from annotated ORFs in *B. cenocepacia* J2315 and *B. cenocepacia* K56-2 (Varga et al. 2013). Nearly all genes incorporated in iBC matched genes present in *B. cenocepacia* K56-2, and the 7 genes that did not have significant matches (E-value > 10^{-30}) did not affect predictions in this study.

RESULTS

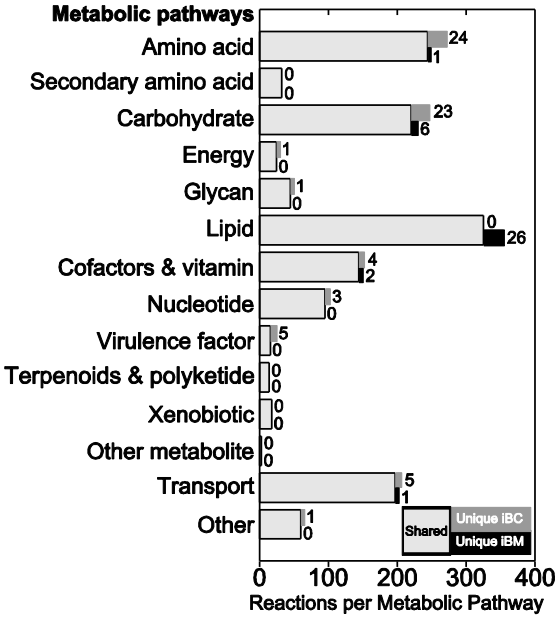
Metabolic network reconstructions

Here, we present the metabolic networks of *B. cenocepacia* and *B. multivorans* as valuable tools that we use as a framework to perform our comparative analyses (Figure 3.1). Figure 3.2A provides a quantitative comparison of the reconstructions in contrast to the previously published *B. cenocepacia* J2315 reconstruction (iKF1028) and the reconstruction of the closely related pathogen *Pseudomonas aeruginosa* PAO1 (iMO1086). It also highlights the substantial increase in scope of the reconciled network reconstructions. iBC accounts for the function of 1,537 genes, 1,506 reactions and 1,280 metabolites. iBM accounts for the function of 1,411 genes, 1,473 reactions and 1,245 metabolites. Most of the reactions for each reconstruction are broadly concentrated in lipid, amino acid, and carbohydrate metabolism (Figure 3.2B), but reactions in over 60 canonical metabolic pathways (as classified by KEGG) are included in the reconstructions. Figure 3.2C shows a map of the reconstructions that contrasts the shared and unique metabolic reactions of each species. iBC has nearly double the number of unique reactions compared to iBM as well as a substantially higher number of incorporated genes (Figure 3.2A). There are more unique reactions in iBC in the amino acid and carbohydrate subsystems; however, iBM has more unique reactions in the lipid subsystem (Figure 3.2B). The reconstructions enable functional comparative

A

Organism	*P. aeruginosa* PAO1	*B. cenocepacia* J2315		*B. multivorans* ATCC17616
Genome Size	6.3 Mbp	8.06 Mbp		7.01 Mbp
GC Content	66.6%	66.9%		66.7%
Coding DNA Sequences	5716	7116		6259
Model	iMO1086	iKF1028	**iBC**	**iBM**
Total Reactions	1028	862	**1504**	**1473**
Total Genes	1086	1028	**1537**	**1411**
Total Metabolites	1021	834	**1280**	**1245**
Unique Reactions	-	-	**67**	**36**
Non-gene-associated reactions	51	25	**23**	**24**

B



C

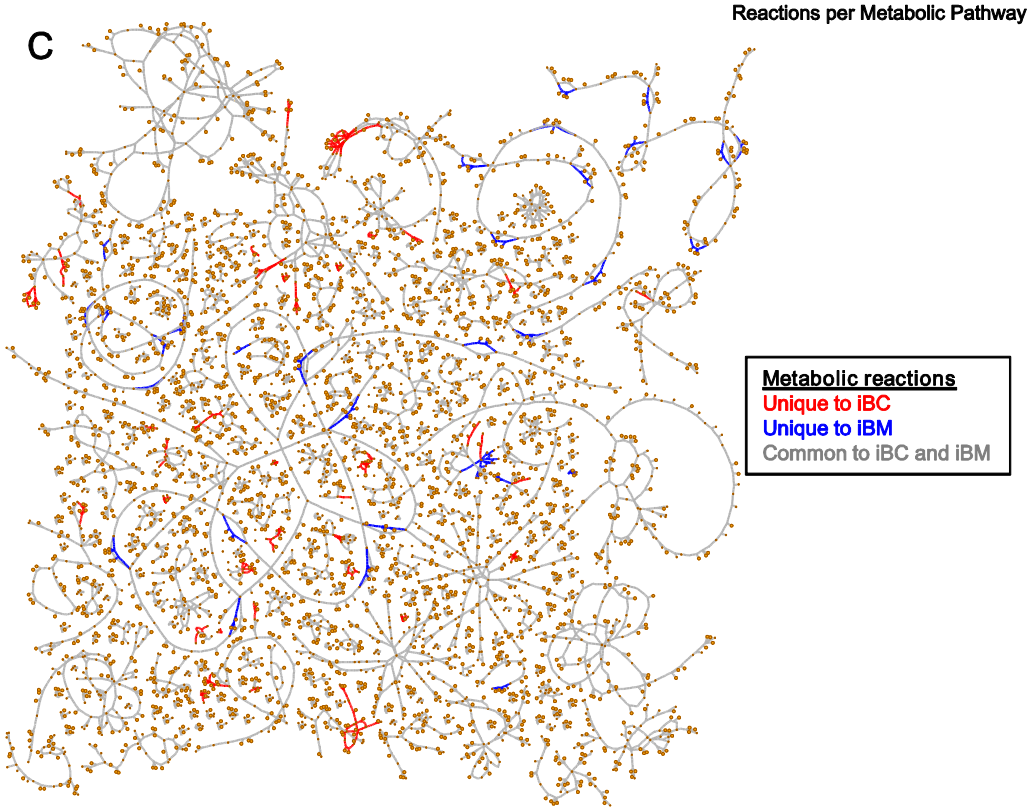


FIG 3.2. Statistics of metabolic network reconstructions. (A) Statistics of iBC and iBM are shown. For reference, statistics for the previously reconstructed models iMO1086 (*P. aeruginosa* PAO1) and iKF1028 (*B. cenocepacia* J2315) are listed. (B) Distribution of the reactions in iBC and iBM by metabolic pathways are presented. The dotted, gray, and black stacked bars show the reactions that are shared, unique to iBC, and unique to iBM, respectively. Numbers next to the gray and black bars indicate their numerical values. (C) Visualization of iBC and iBM shows the reactions unique to iBC, reactions unique to iBM, and reactions common to both reconstructions as red, blue, and gray lines, respectively.

analysis of these differences through examination of gene essentiality, growth capacity, and virulence factor production capabilities.

Model curation and validation

Substantial manual curation of the draft Model SEED reconstructions was performed in order to build an up-to-date knowledgebase of genetic and metabolic information for *B. cenocepacia* and *B. multivorans* and resulted in the addition of a substantial number of reactions. For example, we added 13 reactions to iBC involved in tryptophan catabolism and 17 reactions to iBM involved in phenylalanine catabolism that were missing from the original Model SEED draft reconstructions; species-specific functionality of the associated pathways had been investigated in *Burkholderia cenocepacia* J2315 in the literature (Colabroy and Begley 2005; Yudistira et al. 2011). After databases and literature were canvassed as explained in the Methods, *B. cenocepacia* J2315 and *B. multivorans* ATCC 17616 were experimentally tested for growth on Biolog phenotype microplates to generate *in vitro* data to guide further curation decisions and enable validation of substrate-utilization predictions.

The Biolog screen of *B. cenocepacia* J2315 showed markedly limited growth in many of the *in vitro* minimal media conditions, ranging from no apparent growth to growth at a substantially lower rate than *B. multivorans* ATCC 17616 on many substrates. To further investigate this outcome, we also performed identical Biolog screens on *B. cenocepacia* K56-2 after noting nearly identical genomic content with respect to the metabolic genes incorporated in iBC. Given the high sequence similarity between the *B. cenocepacia* strains, we used iBC to predict *B. cenocepacia* growth which we compared with equivalent *B. cenocepacia* K56-2 *in vitro* data (Figure 3.3) and *B. cenocepacia* J2315 *in vitro* data. The same analysis was performed with iBM and *B. multivorans* ATCC 17616. Complete *in vitro* data and *in silico* growth comparisons for all assessed strains are available.

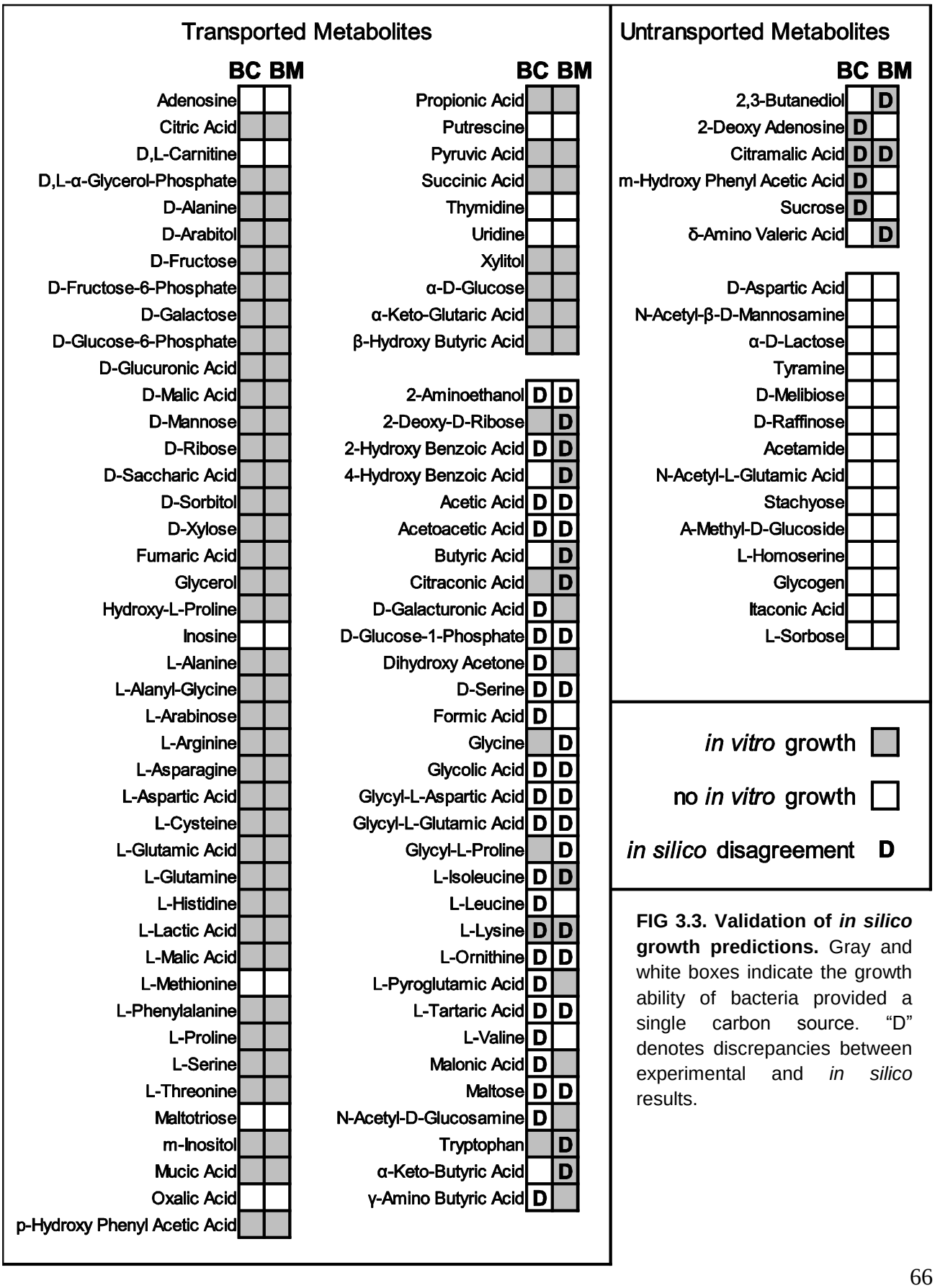


FIG 3.3. Validation of *in silico* growth predictions. Gray and white boxes indicate the growth ability of bacteria provided a single carbon source. "D" denotes discrepancies between experimental and *in silico* results.

When comparing growth between the two *B. cenocepacia* strains on the 192 carbon sources, 53 substrates enabled quantifiable growth of both J2315 and K56-2. Twenty eight substrates did not enable quantifiable growth of J2315 in contrast to successful utilization by K56-2, while only 4 substrates enabled quantifiable growth of J2315 but no quantifiable growth of K56-2. Biolog plates introduce temporal limitations to assessing growth because of potential effects of evaporation after 48 h of incubation that complicate the evaluation of a slow-growing strain like J2315. The limited growth timeframe as well as potential unaccounted for regulatory differences between strains may in combination explain the difference between *in vitro* results and *in silico* predictions that reflect the collected genetic evidence for the catabolic pathways in iBC. At minimum, this is an opportunity for comparative study in the future, as we found no metabolic evidence for the growth defects of J2315 compared to K56-2 within the context of our expansive models. Perhaps certain enzymes integral to basic metabolic processes are operating at low efficiency due to point mutations, or there are alterations in transcriptional control that could be addressed via a future integration of regulatory and metabolic reconstructions. In light of this, we chose to use *B. cenocepacia* K56-2 *in vitro* data when assessing the accuracy of our growth predictions as growth versus no growth on a given substrate was possible to assess confidently with the Biolog plates.

Of the 190 Biolog carbon source substrates experimentally tested, 102 substrates were accounted for in iBC and iBM and could be tested for *in silico* growth using FBA. In addition, growth on the two amino acids cysteine and tryptophan was separately evaluated given their presence in the CF lung and discrepancies identified during the curation process. After further curation efforts, iBC and iBM correctly matched the experimental growth phenotypes of 53 substrates for which transporters were incorporated in the reconstructions (44 correct growth and 9 correct no-growth, each). For the other 31 substrates for which transporters were incorporated, iBC, iBM, or both failed to match experimental growth phenotypes. *B. cenocepacia* and *B. multivorans* did not exhibit growth *in vitro* on 14 substrates, and because iBC and iBM lack transporters for these substrates, they predicted correct no-growth phenotypes. Lastly, 6 substrates did not have transporters in iBC

and iBM; however, *B. cenocepacia*, *B. multivorans*, or both grew on the substrates *in vitro*. Overall, iBC and iBM predicted the correct growth phenotype with 75% and 76% accuracy (78 and 79 out of 104 substrates), respectively. Sixty-seven substrates enabled the same accurate *in silico* and *in vitro* phenotypes in both models, 13 substrates showed the same type of discrepancy between *in silico* and *in vitro* phenotypes in both models, and the remaining 24 substrates showed discrepancies in phenotype between *B. cenocepacia* and *B. multivorans*. Though the remaining 88 carbon sources were not incorporated in the reconstructions and thus not analyzed *in silico*, *B. cenocepacia* and *B. multivorans* were capable of growing on 24 and 32 additional substrates *in vitro*, respectively.

With curation efforts guided by the *in vitro* data, iBC and iBM both underpredict growth on lysine, and iBM underpredicts growth on tryptophan (Figure 3.3). Genomic evidence to support the addition of reactions that would fix these inconsistencies could not be found. For example, no evidence for several key reactions in lysine degradation could be found in iBC (glutarate-CoA ligase E.C. 6.2.1.6, 3-hydroxyacyl-CoA dehydrogenase E.C. 1.1.1.35). The difference in branched chain amino acid catabolism between reconstructions is due to the identification of three *B. cenocepacia* genes that appear to have high sequence similarity with the *bkd* operon of other bacteria, which enables catabolism of isoleucine, leucine, and valine (Madhusudhan et al. 1999). These genes were not present in *B. multivorans*. However, neither species is capable of *in vitro* growth on valine or leucine, and *B. multivorans* is unexpectedly capable of *in vitro* growth on isoleucine (Figure 3.3).

These differences in experimental and *in silico* growth are potentially due to incomplete and/or incorrect annotations of genes in the databases. For example, while we explored homology of current hypothetical proteins with known enzymes in lysine catabolism, the homology may be below thresholds we used or novel lysine degradation reactions may exist. Another possible reason for the discrepancy is control by regulatory systems not accounted for in our current

metabolic network reconstruction. These differences present opportunities for further study of metabolism of the Bcc.

Through our curation of the models from the SEED drafts to the presented final versions, iBC gained 269 reactions and lost 116 reactions and GPRs were substantively altered in a further 324 reactions. iBM gained 297 reactions, lost 106 reactions, and had a further 352 reactions subjected to GPR alterations. In summary, more than a third of each model's individual content is different from the original draft reconstructions due to major additions and improvements to both reactions and GPR formulations during the curation and reconciliation process, resulting in the identification of 1,437 reactions common to both iBC and iBM. The scope of these changes emphasizes the value of manual curation after automated network reconstruction and the functional implications of these differences are presented below.

Refinement of genome annotation

Manual curation of iBC and iBM aided in the identification of many proposed genome annotation refinements of *B. cenocepacia* and *B. multivorans*. A subset of our changes and additions to reaction GPRs were based on the assignment of novel function to genes previously annotated as hypothetical or putative proteins. Many of the annotation refinements were the identification of potential isozymes and duplicated genes for a given reaction, but refinements also included genes that fill in gaps in model pathways and putative functional assignment of 10 hypothetical proteins. Table 3.1 summarizes these annotation refinements and a selection of other high confidence refinements of interest. A table of moderate confidence annotations and list of reactions that are missing gene associations is also available. Examples of genome annotation refinements are presented below.

Ornibactin synthesis. During our refinement process, we added pathways enabling the synthesis of ornibactin, a key *Burkholderia*-specific siderophore which literature indicates is associated with increased virulence of some clinical isolates (Darling et al. 1998). A model of ornibactin synthesis has been proposed based on a study of

TABLE 3.1. Proposed reannotations of gene functions resulting from the curation and reconciliation process.(Ogawa and Shimizu 1994; Hosie et al. 2002; Müller et al. 2005; Tarighi et al. 2008; Kohler et al. 2010; Nogales et al. 2011; Schwibbert et al. 2011; Bi et al. 2012;

B. cenocepacia		*B. multivorans*		Proposed Annotation	Gene Name	Protein ID	Evidence
Locus Tag	Current Annotation	Locus Tag	Current Annotation				
BCAS0031	Hypothetical protein	Bmul_6140	Hypothetical protein	L-2,4-diaminobutyric acid transaminase	*doeD*	EC 2.6.1.-	E-values of 1E-154 and 1E-155, respectively to Helo_3661 in *H. elongata* (44)
BCAM1003	Putative epimerase	Bmul_4613	NAD-dependent epimerase/dehydratase	GDP-D-mannose reductase	*bceM*	EC 1.1.1.187	E-values of 1E-141 and 1E-129, respectively to Bcep1808_4471 in *B. vietnamiensis* (45)
BCAS0105	Hypothetical protein			dTDP-4-dehydrorhamnose reductase	*rfbD*	EC 1.1.1.133	E-value of 1E-139 to Bmul_6045, which is annotated as a dTDP-4-dehydrorhamnose reductase
BCAM2148	Hypothetical protein	Bmul_3704	Hypothetical protein	Malate dehydrogenase	*mdh*	EC 1.1.1.37	E-values of 7E-30 and 3E-27 to BCAM1263 and Bmul_4451, respecitvely, which are annotated as malate/L-lactate dehydrogenases
BCAS0285	Hypothetical protein			Argininosuccinate lyase	*argH*	EC 4.3.2.1	CDD E-value of 3E-27 to argininosuccinate lyase; E-value of 9E-17 to Bcep1808_5167, which is annotated as an argininosuccinate lyase in *B. vietnamiensis*
		Bmul_6038	Hypothetical protein	4-carboxy-4-hydroxy-2-oxoadipate (CHA) aldolase	*galC*	EC 4.1.3.17	CDD E-value of 1E-95 to CHA aldolase; E-value of 2E-80 and 95% coverage to PP_2514 in *P. putida* (43); No ortholog in *B. cenocepacia* .
BCAM0581	enoyl-CoA hydratase	Bmul_5121	enoyl-CoA hydratase	3-hydroxydodecanoyl-ACP hydratase/cis-2-dodecenoyl-ACP thioesterase	-	-	in vitro characterization of BCAM0581 showed bifunctional product (44); Bmul_5121 has E-value of 1E-162 to BCAM0581
		Bmul_3969	autoinducer synthesis protein	N-acylhomoserine lactone synthase	*cepI*	EC 2.3.1.184	E-value of 8E-81 to BCAM1870, which is annotated as *cepI*

TABLE 3.1 Cont'd. Proposed reannotations of gene functions resulting from the curation and reconciliation process.

BCAL2418	Hypothetical protein	Bmul_0955	6-phosphogluconolactonase	Gluconolactonase	*ppgL*	EC 3.1.1.17	PA4204 in *P. aeruginosa*, which encodes for a characterized periplasmic gluconolactonase (48)
BCAL1668	putative amino acid transport system	Bmul_1621	extracellular solute-binding protein	Cysteine ABC Transporter	*cjaA*	CjaA	E-values of 1E-25 and 1E-25 to Cj0982 in *C. jejuni*, which was shown to be a high-affinity cysteine transporter (49)
BCAS0145; BCAL1224; BCAS0735; BCAM1221; BCAM0361	metallo peptidase; allantoate amidohydrolase	Bmul_5349; Bmul_4490; Bmul_2063; Bmul_5235	allantoate amidohydrolase	beta-Ureidopropionase	-	EC 3.5.1.6	All genes have E-values of < 1E-100 to PP_0614 in *P. putida*, which was shown to be a beta-ureidopropionase (47)
BCAL1428	putative amine catabolism-related protein	Bmul_1894	xylose isomerase domain-containing protein	myo-inosose-2 dehydratase	*iolE*	EC 4.2.1.44	E-values of 2E-56 and 4E-58 to SMc00433 in *S. meliloti* (52)
BCAL1429	putative TPP-binding acetolactate synthase	Bmul_1893	thiamine pyrophosphate protein	3D-(3,5/4)-trihydroxycyclohexane-1,2-dione hydrolase	*iolD*	-	E-values of 1E-168 and 1E-166 to SMc01166 in *S. meliloti* (52)
BCAL1427	myo-inositol catabolism protein	Bmul_1895	myo-inositol catabolism IolB domain-containing protein	5-deoxy-D-glucuronate isomerase	*iolB*	EC 5.3.1.-	E-values of 3E-71 and 2E-71 to SMc00432 in *S. meliloti* (52)
BCAL1430	putative carbohydrate kinase	Bmul_1892	ribokinase-like domain-containing protein	5-dehydro-2-deoxygluconokinase	*iolC*	EC 2.7.1.92	E-values of 1E-147 and 1E-143 to SMc01165 in *S. meliloti* (52)
BCAL2435	Na+/solute symporter family protein	Bmul_0940	Na+/solute symporter	monocarboxylate permease	*mctP*	MctP	E-values of 1E-161 and 1E-155 to *mctP* in *R. leguminosarum* which was shown to code for a pyruvate transporter (51)

B. cenocepacia 715j transposon mutants identifying nonribosomal peptide synthases and associated enzymes grouped into an operon (the *orb* operon) that enables ornibactin production (Agnoli et al. 2012). This operon of 14 genes involved in ornibactin synthesis, export, and uptake was annotated in the BGD for *B. cenocepacia* J2315. However, despite identification of a likely *orb* operon induced by iron-responsive regulator Fur in *B. multivorans ATCC 17616* (Yuhara et al. 2008), most of the *B. multivorans* genes are annotated as putative enzymes or hypothetical proteins in the BGD. We have incorporated the proposed synthesis pathways in both reconstructions.

Ectoine degradation. As another example of a refinement that involves novel assignment of gene function, Figure 3.4A shows our proposed annotation of orthologous hypothetical proteins encoded by BCAS0031 and Bmul_6140 as an essential enzyme in ectoine degradation. Ectoine, produced by many prokaryotes as a solute that assists in maintaining an osmotic equilibrium, can be used as a growth substrate by environmental bacteria including some halophile and soil bacteria (Schwibbert et al. 2011). Ectoine production has been shown to be upregulated under stress conditions and, interestingly, in mutant strains of soil bacteria that are resistant to antibiotics targeting the cell envelope (e.g. *Streptomyces coelicolor*) and protein synthesis (e.g. *Nocardiopsis sp.* FU40) (Hesketh et al. 2011; Derewacz et al. 2012). These studies theorize that ectoine upregulation may be part of a broad secondary metabolism response to changes in regulation due to resistance-related mutations. Through literature mining and similarity searches via BLASTP, we determined that BCAS0031 and Bmul_6140 have high similarities to Helo_3661 in *H. elongata* DSM 2581 (E-values of 10^{-154} and 10^{-155}, respectively) (Schwibbert et al. 2011). Through mutational studies, Helo_3661 was found to be an L-2,4-diaminobutyric acid transaminase (DoeD) and BCAS0031 and Bmul_6140 and the associated reaction in iBC and iBM were subsequently included. These genes are also part of apparent operons that also contain *doeABC*, lending credence to this reannotation (Schwibbert et al. 2011). A search for sequences similar to BCAS0031 and Bmul_6140 across all bacteria via BLASTP showed that many of the orthologous genes were annotated as encoding hypothetical proteins or generic

class III aminotransferases. This refinement enabled a fully functional ectoine degradation pathway in the models that had not been previously identified in *B. cenocepacia* and *B. multivorans* genome databases or literature.

Cepacian synthesis. Another type of model refinement is the gap-filling of metabolic pathways that are known to be present in *Burkholderia*. For example, as part of our effort to include pathways for the biosynthesis of virulence-related compounds, reactions required for the production of the *Burkholderia* exopolysaccharide cepacian were added. Literature evidence suggests that the production of GDP-D-rhamnose, one of the sugar residues of cepacian, is catalyzed by a GDP-D-mannose reductase (RMD) (Sousa et al. 2013). In *B. vietnamiensis* G4, the RMD protein is encoded by Bcep1808_4471, also known as *bceM (55).* A similarity search of Bcep1808_4471 via BLASTP shows that *bceM* in *B. cenocepacia* and *B. multivorans* is encoded by BCAM1003 and Bmul_4613 (E-values 10^{-141} and 10^{-129}, respectively). Though BCAM1003 and Bmul_4613 are currently listed as a putative epimerase and an NAD-dependent epimerase/dehydratase in the BGD, respectively, they have been assigned as RMDs in iBC and iBM and we suggest a consistent update to the database annotations.

Tryptophan catabolism. The *in vitro* carbon-source growth screens further aided in identifying gaps in iBC and iBM. In our growth screens, *B. cenocepacia* and *B. multivorans* were observed to grow on M9 minimal media supplemented with tryptophan as the sole carbon source. However, the draft reconstructions initially could not grow *in silico* in these conditions. Literature mining revealed that kynurenine-3-monooxygenase (Kmo) catalyzes an essential reaction in the tryptophan degradation pathway (Figure 3.4B). A study identified orthologs of a cluster of tryptophan catabolic genes similar to those in *Bacillus cereus* 10897 in *B. cenocepacia* J2315 with the exception of the gene encoding Kmo, which suggests the existence of a nonorthologous form of Kmo in *Burkholderia* (Colabroy and Begley 2005). The reaction catalyzed by Kmo was then added to iBC and iBM without a gene assignment. This model refinement identifies a current gap in our genomic knowledge of *Burkholderia* even though there is evidence for the

enzymatic reaction (Figure 3.4B) and the experimental screen provides evidence of its assumed function.

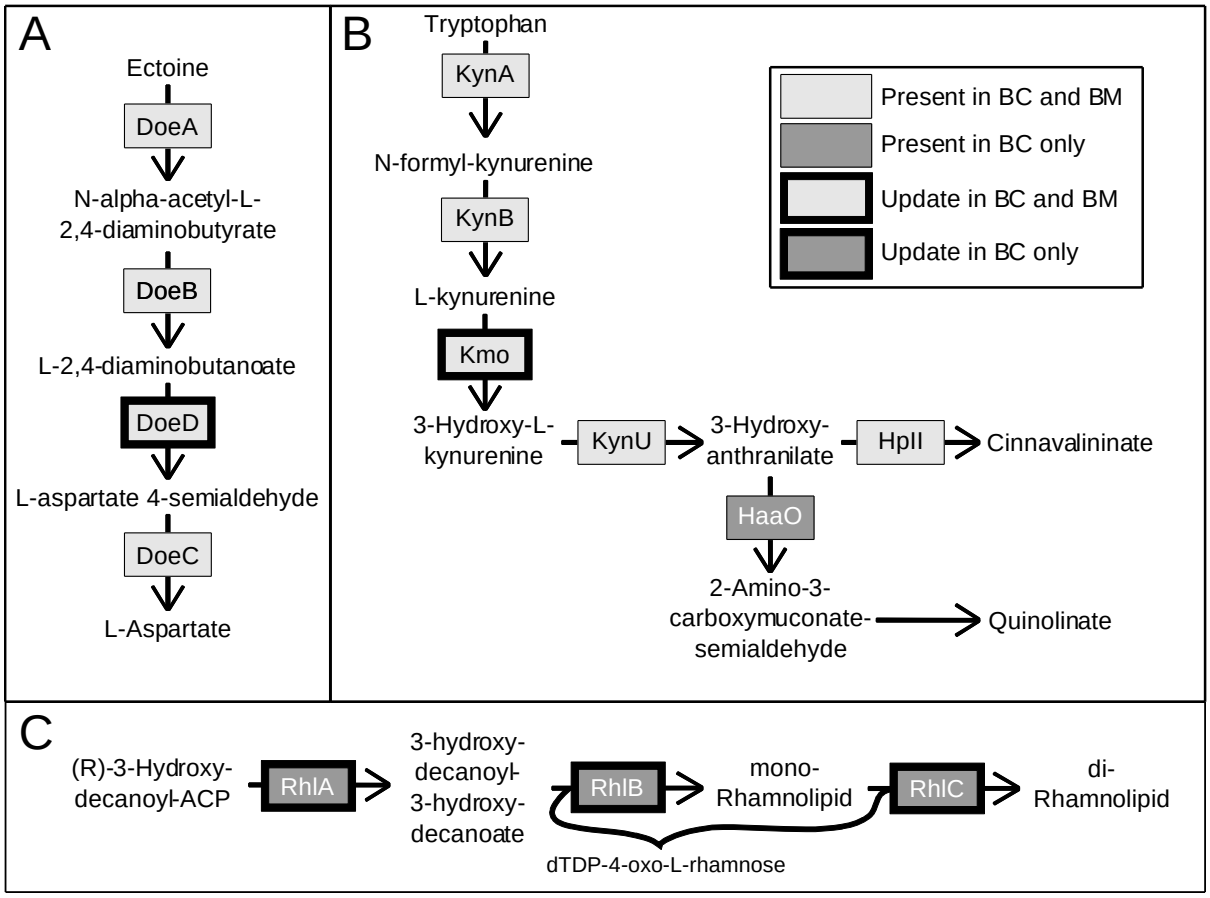


FIG 3.4. Examples of network refinements and reannotations. (A) Both *B. cenocepacia* (BC) and *B. multivorans* (BM) genomes include genes that allow for the degradation of ectoine. BCAS0031 and Bmul_6140, which are currently annotated as hypothetical proteins, were identified as orthologs of Helo_3661 in *H. elongata*, which encodes DoeD. (B) Curation of the tryptophan degradation pathway identified a gap for the essential reaction catalyzed by Kmo which led to the addition of Kmo to both iBC and iBM without an assigned gene. The gene for HaaO was present only in BC and allows for the synthesis of quinolinate. (C) BC has the *rhlABC* genes necessary for rhamnolipid synthesis while BM does not. The intermediate reaction that converts beta-hydroxydecanoyl-beta-hydroxydecanoyl-S-CoA to 3-hydroxydecanoyl-3-hydroxydecanoate is catalyzed by an unknown enzyme but because there is experimental and modeling evidence that this reaction takes place in *P. aeruginosa* PAO1, it was included in iBC and iBM.

Abbreviations: DoeA, ectoine hydrolase; DoeB, N-alpha-acetyl-L-2,4-diaminobutyric acid deacetylase; DoeC, aspartate-semialdehyde dehydrogenase; DoeD, diaminobutyric acid transaminase; KynA, tryptophan 2,3-dioxygenase (EC 1.13.11.11); KynB, kynurenine formamidase (EC 3.5.1.9); KynU, kynureninase (EC 3.7.1.3); Kmo, kynurenine 3-monooxygenase (EC 1.14.13.9); HaaO, 3-hydroxyanthranilate-3,4-dioxygenase (EC 1.13.11.6); RhlA, rhamnosyltransferase chain A; RhlB, rhamnosyltransferase chain B; RhlC, rhamnosyltransferase chain 2; PhaC, poly(3-hydroxyalkanoic acid) synthase

Further curation of the tryptophan degradation pathway led to refinements that showed species-specific differences. Literature evidence has shown that hydroxyanthranilate-3,4-dioxygenase (HaaO) also catalyzes an essential step required for the degradation of tryptophan in *B. cenocepacia* (Colabroy and Begley 2005). BCAM2130 encodes HaaO in *B. cenocepacia*; however, an ortholog could not be found via a BLASTP search in *B. multivorans* (Figure 3.4B). In this case, the reaction was included in iBC but not in iBM. Lack of HaaO and its reaction contributes to the inability of iBM to grow *in silico* with tryptophan as a sole carbon source. However, artificial inclusion of the reaction in iBM still did not allow for *in silico* growth, suggesting that other reactions needed for the catabolism of tryptophan are missing in iBM. This was one of the inconsistencies between the *in vitro* and *in silico* growth screens that we were unable to rectify through model refinement. This inconsistency demonstrates how the capability of the entire metabolic network is assessed when testing for growth; network gaps are not always direct or trivial and provide hypotheses for further experimentation.

Rhamnolipid synthesis. Species-specific differences are also exemplified within the rhamnolipid synthesis pathway. Rhamnolipid production has been linked to enhanced *P. aeruginosa* virulence in the cystic fibrosis lung by enabling invasion of host epithelial cells and affecting biofilm assembly, structural maintenance and dispersion (Zulianello et al. 2006; Pamp and Tolker-Nielsen 2007). We located genes in *B. cenocepacia* that were previously unidentified as the rhamnolipid synthesis operon *rhlABC* but could not find genes encoding enzymes in this operon in the rhamnolipid synthesis pathway in *B. multivorans* (Figure 3.4C). The *rhlABC* genes in *B. cenocepacia* are BCAM2340, BCAM2338, and BCAM2336, which have BLASTP E-values of 7×10^{-63}, 5×10^{-86} and 4×10^{-63} to the orthologous genes in *P. aeruginosa* PAO1 (PA3479, PA3478 and PA1130, respectively). Our additions resulted in a functional rhamnolipid synthesis pathway in iBC, while iBM was unable to produce rhamnolipids. Currently we are unaware of literature supporting rhamnolipid production in *B. cenocepacia* or *B. multivorans*, but nonpathogenic *Burkholderia thailandensis* and *Burkholderia plantarii* as well as pathogenic *Burkholderia pseudomallei* can produce rhamnolipids in specific conditions (Häußler

et al. 1998; Andrä et al. 2006; Dubeau et al. 2009). However, a transcriptomics screening study by Sass *et al*. showed significant upregulation of the genes we identified as the *rhl* operon in *B. cenocepacia* during stationary phase growth in minimal medium (Sass et al. 2013). These studies support the likelihood of the ability of *B. cenocepacia* to produce rhamnolipids, and may contribute to the enhanced virulence of *B. cenocepacia* over *B. multivorans* during chronic lung infections.

Comparison of gene content

Our model curation efforts resulted in a high number of genes associated with each reaction in the reconstructions compared to other large, well-curated reconstructions. This increase in genes per reaction was quantified by counting the number of unique genes in the GPR of each reaction in each model, resulting in an average of 3.24 genes per reaction in iBC, 3.01 genes per reaction in iBM, and 1.9 genes per reaction in the *P. aeruginosa* reconstruction iMO1086. This difference in genes per reaction in the Bcc reconstructions appears to be due to higher numbers of isozymes and gene duplications as implemented in GPRs and may be evidenced by the larger genomes of the Bcc species compared to *P. aeruginosa.* By dividing the total number of genes associated with the reactions in each subsystem by the number of reactions in each subsystem, the average number of genes per reaction per subsystem was calculated (Figure 3.5). For context, the average number of genes associated with each reaction grouped by subsystems in iBC, iBM, and iMO1086 is compared.

While iMO1086 shows a higher number of average genes per reaction in a subset of amino acid metabolism pathways and a few other subsystems, overall, iBC has the most genes per reaction across all subsystems, followed by iBM and iMO1086. The most notable increases in the average number of genes associated with reactions in iBC over the other reconstructions are incorporated in lipid metabolism. However, iBM may have higher metabolic capacity in certain energy and carbohydrate metabolic pathways as it has a higher gene per reaction average in these pathways. Pathways where iBC and iBM have similar numbers of genes associated with a

reaction include metabolism of amino acids as well as terpenoid and polyketide metabolism. The reconstructions offer a method of probing specific subsystems to identify gene duplications or isozymes that may indicate concentrated genetic redundancy.

Gene essentiality

An important consequence of increased genes per reaction in iBC and iBM compared to reconstructions of other bacteria was the reduction in the number of genes predicted to be essential for growth. During *in silico* growth in LB medium, our models predicted 66 essential genes in iBC and 73 essential genes in iBM (Table 3.2). Sixty of these genes were orthologs between iBC and iBM, 6 genes were uniquely essential in iBC, and 13 genes were uniquely essential in iBM. In comparison, iMO1086 required 150 genes to grow *in silico* on LB (with an accuracy of 83.9%) (Oberhardt et al. 2011). Because iBM and iBC are not currently reconciled with iMO1086, we compared our predicted essential genes with a list of potentially essential *P. aeruginosa* PAO1 genes identified experimentally; these PAO1 genes had no recorded transposon mutants as identified in the Pseudomonas Genome Database based on genome scale transposon mutagenesis libraries (Winsor et al. 2011). Thirty five out of 78 predicted essential *Burkholderia* genes matched *P. aeruginosa* PAO1 probable essential genes using a BLAST comparison. Another 12 predicted essential *Burkholderia* genes matched PAO1 genes with likely isozymes or duplications. Two predicted essential *Burkholderia* genes had no match to any PAO1 gene locus. The low number of potential essential *Burkholderia* genes compared to other species corresponds with a recent study that created promoter-based conditional mutants to identify essential *Burkholderia cenocepacia* K56-2 genes (Bloodworth et al. 2013). However, the authors saw unexpectedly low essential operon hit rates during their mutant library screening.

The difference between the overall number of essential genes in iBC and iBM compared to *P. aeruginosa* is likely due to the high number of isozymes and duplicate genes included in our GPRs as well as the comprehensive GPR formulation, which added robustness to the performance of particular functions.

Many of the predicted essential genes were located in expected pathways such as nucleotide metabolism, energy metabolism, and lipid metabolism. The common essential genes represent opportunities to target both species with the same treatment. The high number of unique essential genes predicted for each species offers new hypotheses regarding species-specific targets in developing novel treatments for *B. cenocepacia* versus *B. multivorans* infections.

When comparing the unique essential genes predicted for each model shown in Table 3.2, a gene associated with dihydrofolate reductase (DHFR) was only present in iBM. DHFR, encoded by Bmul_2221, is the target of trimethoprim, an antibiotic commonly effective against pathogens such as *E. coli* and *Haemophilus influenzae*. However, trimethoprim has shown variable effectiveness against different strains of the Bcc (Nzula 2002) and our essentiality analysis indicates no apparent matching essential gene i n iBC. Instead, two genes were associated with this reaction in iBC by the Model SEED tool. One gene, BCAL2915, is annotated as *dfrA*, a trimethoprim-resistant variant of DHFR (Kehrenberg and Schwarz 2005), and has high similarity to Bmul_2221. The other gene, BCAL1859, is annotated on the BGD as a hypothetical protein. When conducting a BLASTP search with BCAL1859, the most similar functionally annotated matches were to DHFR genes in *Deinococcus* and *Aeromonas* species (E-values of $\sim 10^{-50}$ and $\sim 10^{-40}$, respectively) with no other highly similar matches. Interestingly, a study of *in vitro* antimicrobial susceptibility determined that *B. multivorans* ATCC 17616 is notably more susceptible to trimethoprim than *B. cenocepacia* J2315, with a minimum inhibitory concentration of 2 mg/L compared to 64 mg/L (Nzula 2002). Determining whether this apparently rare variant of DHFR (BCAL1859) is connected with the elevated trimethoprim resistance of *B. cenocepacia* J2315 would require further experimental study, but the identification of a previously unannotated DHFR isozyme in the more virulent of our two Bcc species of interest is an important consequence of our comparative study of essential genes.

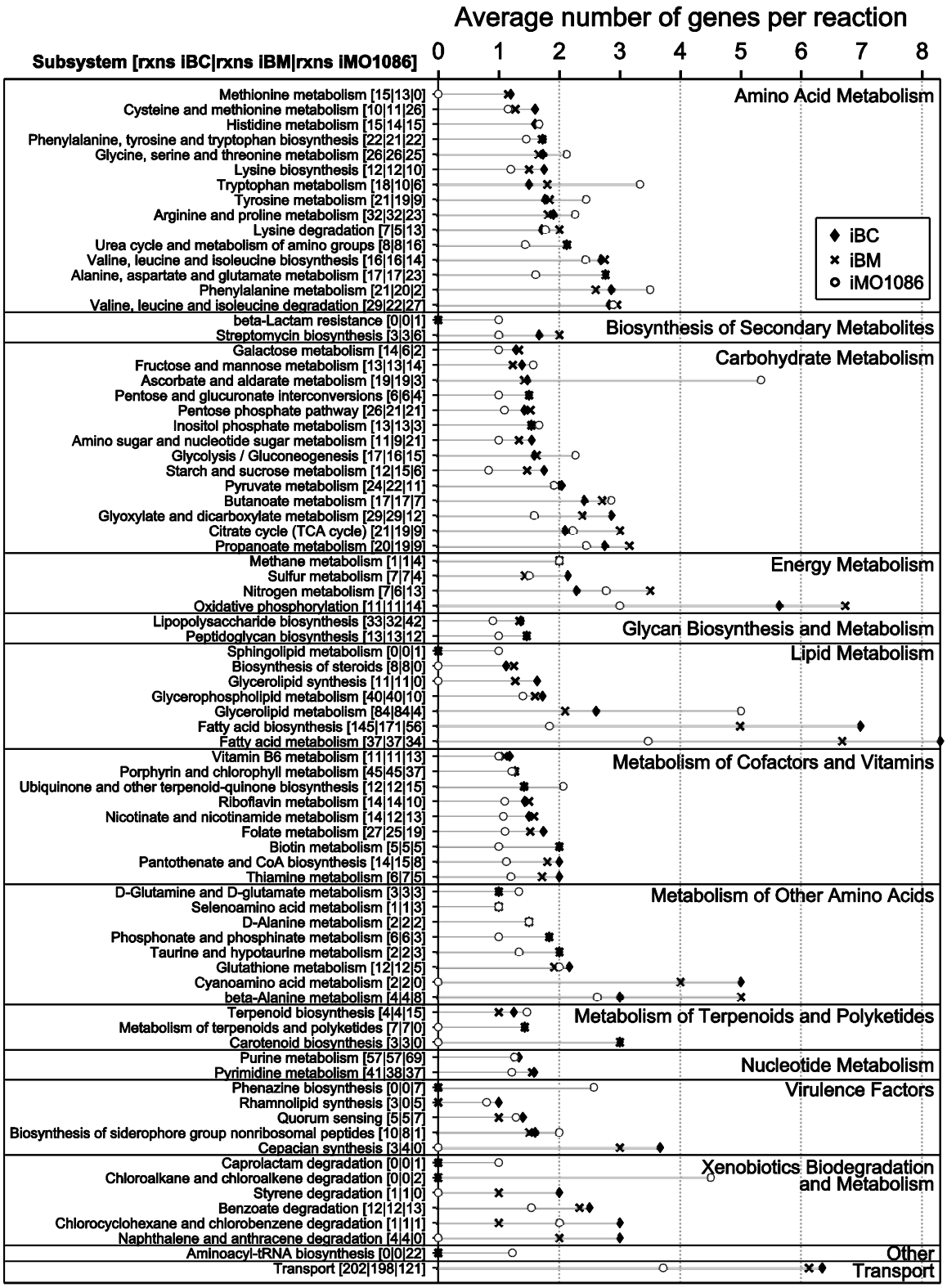


FIG 3.5. Distribution of the number of genes per reaction in iBC and iBM by subsystem. The average number of genes per reaction per subsystem was calculated by dividing the total number of genes in the GPRs of the reactions in each subsystem by the number of reactions in each subsystem.

FIG 3.5 Cont'd. Diamonds and crosses show the average number of genes per reaction for iBC and iBM, respectively. For reference, the average number of genes per reaction for iMO1086 (*P. aeruginosa* PAO1) is shown as open circles. For each subsystem, the three values in brackets denote the number of reactions for iBC, iBM, and iMO1086.

Virulence factor production capacity

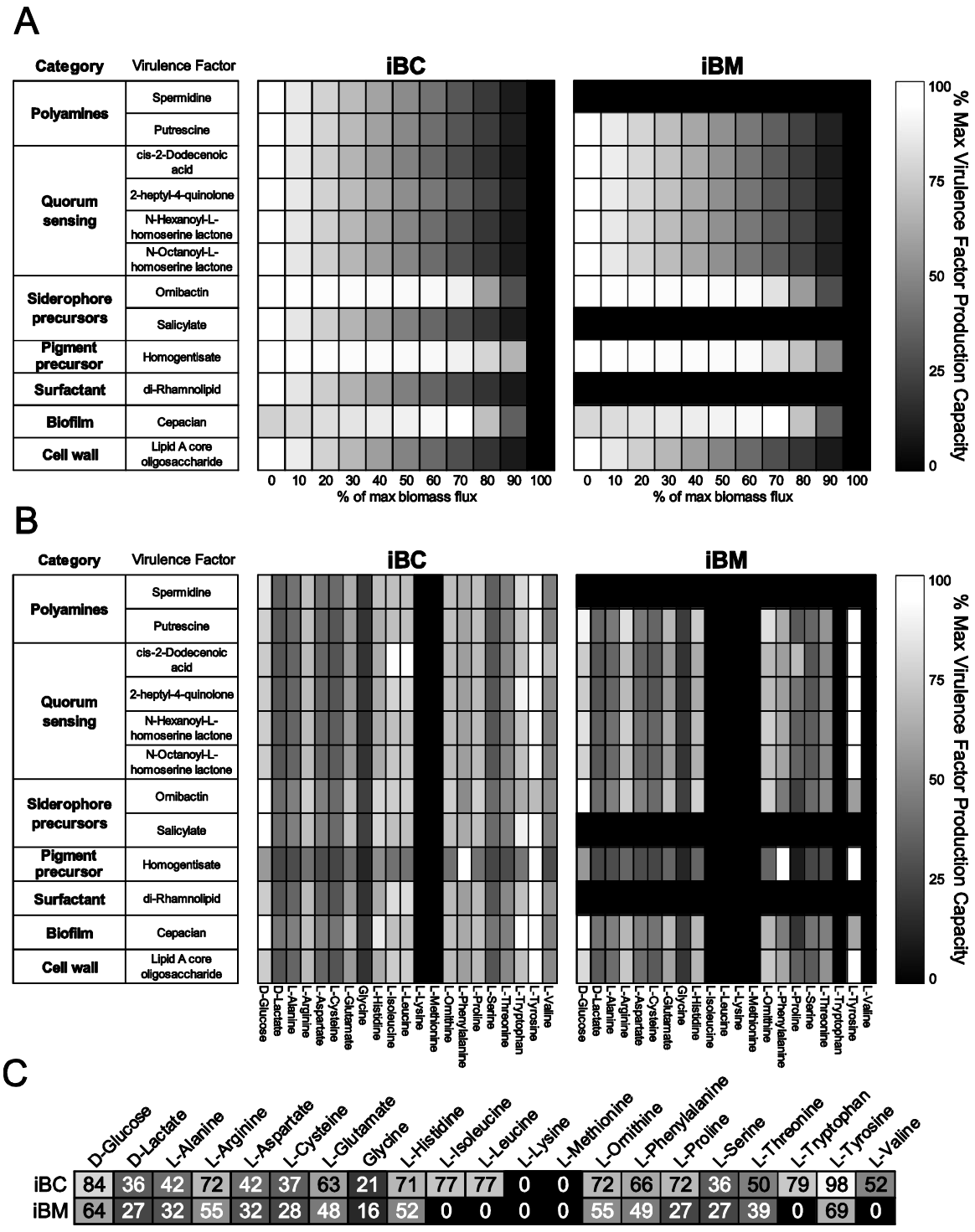
Our models enable *in silico* investigation of how the nutritional environment of the cystic fibrosis lung contributes to production of important factors in initial colonization and chronic infection by *Burkholderia*. iBC and iBM were used to predict the ability of each species to produce an array of virulence-related molecules while under varied growth demands or when presented with limited nutritional resources. We include prototypical virulence factors such as biofilm-related exopolysaccharide cepacian, immune response-triggering, phagocytosis-resistant LPS, and quorum sensing signals that enable communication with other bacteria (Loutet and Valvano 2010b). We also include the production of rhamnolipids due to their role in biofilm formation and regulation as previously mentioned. Intracellular signaling molecules such as polyamines putrescine and spermidine are not essential for growth in many environments, but their loss induces significant phenotypic changes in various bacteria and they are considered important in the regulation of biofilm production and other virulence pathways (Williams et al. 2010). Ornibactin enables iron acquisition, and salicylate potentially acts as a siderophore and is also required for production of other siderophores; both compounds have been connected with enhanced virulence both clinically and in animal models (Darling et al. 1998). Homogentisate is a precursor of the melanin-like pigment produced by some *Burkholderia* species that provides protection from reactive oxygen and nitrogen species (Keith et al. 2007; Liu and Nizet 2009).

TABLE 3.2. Comparison of essential gene predictions.

	B. cenocepacia Locus Tag	*B. multivorans* Locus Tag	Gene Abbreviation	Enzyme Name	EC #
iBC Unique Essential Genes	BCAL0800		*prs*	ribose-phosphate pyrophosphokinase	EC 2.7.6.1
	BCAL0902			D,D-heptose 1,7-bisphosphate phosphatase	
	BCAL2356			phosphatidylserine decarboxylase	EC 4.1.1.65
	BCAL3389		*tktA*	transketolase	EC 2.2.1.1
	BCAL3428		*nrdB*	ribonucleotide-diphosphate reductase subunit beta	EC 1.17.4.1
	BCAL3429			ribonucleotide-diphosphate reductase subunit alpha	EC 1.17.4.1
Common Essential Genes	BCAL0162	Bmul_0198	*gmhA*	phosphoheptose isomerase	EC 5.3.1.28
	BCAL0508	Bmul_3079		lipid A biosynthesis lauroyl acyltransferase	EC 2.3.1.-
	BCAL0509	Bmul_3078	*metK*	S-adenosylmethionine synthetase	EC 2.5.1.6
	BCAL0612	Bmul_2976	*glmU*	UDP-N-acetylglucosamine pyrophosphorylase	EC 2.7.7.23
	BCAL0743	Bmul_0450	*gpsA*	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	EC 1.1.1.94
	BCAL0817	Bmul_0528	*kdsC*	putative 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	EC 3.1.3.45
	BCAL0818	Bmul_0529		putative arabinose 5-phosphate isomerase	EC 5.3.1.13
	BCAL1269	Bmul_2021	*glmM*	phosphoglucosamine mutase (EC:5.4.2.10)	
	BCAL1281	Bmul_1997		hypothetical protein (ornithine-acyl-ACP N-acyltransferase in Burkholderia sp. 383)	
	BCAL1556	Bmul_1683	*rpiA*	ribose-5-phosphate isomerase A	EC 5.3.1.6
	BCAL1887	Bmul_1460	*ndk*	nucleoside diphosphate kinase	EC 2.7.4.6
	BCAL1987	Bmul_1360	*purL*	phosphoribosylformylglycinamidine synthase	EC 6.3.5.3
	BCAL2061	Bmul_1287	*guaA*	GMP synthase	EC 6.3.5.2
	BCAL2078	Bmul_1270	*lpxB*	lipid-A-disaccharide synthase	EC 2.4.1.182
	BCAL2079	Bmul_1269	*lpxA*	UDP-N-acetylglucosamine acyltransferase	EC 2.3.1.129
	BCAL2080	Bmul_1268	*fabZ*	(3R)-hydroxymyristoyl-ACP dehydratase	EC 4.2.1.-
	BCAL2081	Bmul_1267	*lpxD*	UDP-3-O-(3-hydroxymyristoyl) glucosamine N-acyltransferase	EC 2.3.1.191
	BCAL2089	Bmul_1259	*pyrH*	uridylate kinase	EC 2.7.4.-
	BCAL2101	Bmul_1247	*dapD*	2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase	EC 2.3.1.117
	BCAL2103	Bmul_1245	*dapE*	succinyl-diaminopimelate desuccinylase	EC 3.5.1.18
	BCAL2146	Bmul_1207	*ask*	aspartate kinase	EC 2.7.2.4
	BCAL2154	Bmul_1199		UDP-2,3-diacylglucosamine hydrolase	EC 3.6.1.54
	BCAL2180	Bmul_1162		2-dehydro-3-deoxyphosphooctonate aldolase	EC 2.5.1.55
	BCAL2181	Bmul_1161	*pyrG*	CTP synthetase	EC 6.3.4.2
	BCAL2289	Bmul_1080		glutamate racemase	
	BCAL2355	Bmul_1017		putative phosphatidyltransferase	EC 2.7.8.8
	BCAL2388	Bmul_0984		hypothetical protein (Glucose-1-phosphate adenylyltransferase in iKF1028)	
	BCAL2389	Bmul_0983	*purD*	phosphoribosylamine--glycine ligase	EC 6.3.4.13
	BCAL2403	Bmul_0965		putative LPS core biosynthesis protein	EC 2.4.-.-
	BCAL2759	Bmul_0751		tetraacyldisaccharide 4'-kinase	EC 2.7.1.130
	BCAL2761	Bmul_0749	*kdsB*	3-deoxy-manno-octulosonate cytidylyltransferase	EC 2.7.7.38
	BCAL2770	Bmul_0740		putative glycerol-3-phosphate acyltransferase PlsY	EC 2.3.1.15
	BCAL2836	Bmul_0675	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	EC 4.1.1.21
	BCAL2837	Bmul_0674	*purE*	phosphoribosylaminoimidazole carboxylase catalytic subunit	EC 4.1.1.21
	BCAL2838	Bmul_0673	*purC*	phosphoribosylaminoimidazole-succinocarboxamide synthase	EC 6.3.2.6
	BCAL2912	Bmul_2218	*thyA*	thymidylate synthase	EC 2.1.1.45
	BCAL2944	Bmul_2251	*hldD*	ADP-l-glycero-D-manno-heptose-6-epimerase	EC 5.1.3.20
	BCAL2951	Bmul_2258	*cmk*	cytidylate kinase	EC 2.7.4.14
	BCAL3012	Bmul_2401	*gmk*	guanylate kinase	EC 2.7.4.8
	BCAL3110	Bmul_2494	*waaA*	3-deoxy-D-manno-octulosonic-acid transferase	EC 2.4.99.12
	BCAL3113	Bmul_2497	*manB*	phosphomannomutase	EC 5.4.2.2
	BCAL3133	Bmul_2598	*rmlC*	dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase	EC 5.1.3.13
	BCAL3134	Bmul_2597	*rmlA*	glucose-1-phosphate thymidylyltransferase	EC 2.7.7.24
	BCAL3239	Bmul_4484		glucosyltransferase	
	BCAL3261	Bmul_2624	*purM*	phosphoribosylaminoimidazole synthetase	EC 6.3.3.1
	BCAL3336	Bmul_2693	*purH*	phosphoribosylaminoimidazolecarboxamide formyltransferase	EC 2.1.2.3
	BCAL3361	Bmul_2717	*purB*	adenylosuccinate lyase	EC 4.3.2.2
	BCAL3397	Bmul_2756		putative phosphatidylglycerophosphatase	EC 3.1.3.27
	BCAL3460	Bmul_2834	*ddl*	D-alanine--D-alanine ligase	EC 6.3.2.4
	BCAL3461	Bmul_2835	*murC*	UDP-N-acetylmuramate--L-alanine ligase	EC 6.3.2.8
	BCAL3462	Bmul_2836	*murG*	undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase	EC 2.4.1.227
	BCAL3464	Bmul_2838	*murD*	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	EC 6.3.2.9
	BCAL3465	Bmul_2839	*mraY*	phospho-N-acetylmuramoyl-pentapeptide-transferase	EC 2.7.8.13
	BCAL3466	Bmul_2840	*murF*	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diami nopimelate--D-alanyl-D-alanyl ligase	EC 6.3.2.10
	BCAL3467	Bmul_2841	*murE*	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	EC 6.3.2.13
	BCAM0683	Bmul_3371	*ceoR*	LysR family regulatory protein	
	BCAM0986	Bmul_4626	*asd*	aspartate-semialdehyde dehydrogenase	EC 1.2.1.11
	BCAM0998	Bmul_4615	*purF*	amidophosphoribosyltransferase	EC 2.4.2.14
	BCAM1337	Bmul_4402		glycosyltransferase	
	BCAM2044	Bmul_3751		putative asparagine synthetase	EC 6.3.5.4
iBM Unique Essential Genes		Bmul_0742	*murB*	UDP-N-acetylenolpyruvoylglucosamine reductase	EC 1.1.1.158
		Bmul_1262		phosphatidate cytidylyltransferase	EC 2.7.7.41
		Bmul_2179		3-oxoacyl-(acyl carrier protein) synthase II	EC 2.3.1.179
		Bmul_2182	*fabD1, fabD2*	UDP-3-O-(3-hydroxymyristoyl) N-acetylglucosamine deacetylase	EC 3.5.1.108
		Bmul_2183	*fabH1, fabH2*	3-oxoacyl-ACP synthase	EC 2.3.1.41
		Bmul_2184	*plsX*	putative glycerol-3-phosphate acyltransferase	EC 2.3.1.15
		Bmul_2221		dihydrofolate reductase	EC 1.5.1.3
		Bmul_2735	*dapB*	dihydrodipicolinate reductase	EC 1.3.1.26
		Bmul_2829	*lpxC*	UDP-3-O-(3-hydroxymyristoyl) N-acetylglucosamine deacetylase	EC 3.5.1.108
		Bmul_3080	*dapF*	diaminopimelate epimerase	EC 5.1.1.7
		Bmul_3287	*prpB*	2-methylisocitrate lyase	EC 4.1.3.30
		Bmul_3288		methylcitrate synthase	EC 2.3.3.5
		Bmul_4654	*prpD*	2-methylcitrate dehydratase	EC 4.2.1.79

We evaluated the *in silico* production capacities of each of these compounds under different growth and media constraints as explained in Methods (Figure 3.6). We analyzed the tradeoff between virulence factors and biomass by predicting the maximum production of each virulence factor under a given requirement for biomass flux in SCFM (Figure 3.6A). For the majority of the virulence factors that iBC and iBM can produce, production capacity decreases approximately linearly as the percent of maximum biomass flux constraint is increased (Figure 3.6A). However, both models predict that ornibactin and homogentisate can be produced at a sustained level over a wide range of growth rates until constrained by demand for resources necessary for near-optimal levels of biomass production. Also, both models predict that maximum cepacian production occurs at non-zero biomass production requirements. This dependency between cepacian production and growth is a result of complex interrelationships between cofactors and byproducts upstream of the cepacian biosynthesis pathway. Additionally, while iBC is capable of *de novo* synthesis of all of the tested virulence factors, iBM contains incomplete biosynthetic pathways for the production of spermidine, salicylate, and di-rhamnolipid, resulting in the inability to produce these molecules. The models enable the identification of both the missing genes that entirely prevent production and altered upstream pathways that limit final production levels of a given virulence factor in iBM. The relationship among changing *in vivo* nutrient resources, colonization and infection has been highlighted as an understudied potential target of novel therapies in an array of host-pathogen relationships (Brown et al. 2008).

To better understand the role of the individual components of SCFM in the production of the various virulence factors, we assessed *in silico* virulence factor production capacities on each individual carbon source present in SCFM without any biomass flux constraints (Figure 3.6B). The production capacities for each virulence factor were normalized by the maximum possible value across the different individual carbon sources for that molecule. For both models, on average, tyrosine and glucose were the carbon sources that contributed the most towards the production of the virulence factors. An example of an exception is that tyrosine



C

	D-Glucose	D-Lactate	L-Alanine	L-Arginine	L-Aspartate	L-Cysteine	L-Glutamate	Glycine	L-Histidine	L-Isoleucine	L-Leucine	L-Lysine	L-Methionine	L-Ornithine	L-Phenylalanine	L-Proline	L-Serine	L-Threonine	L-Tryptophan	L-Tyrosine	L-Valine
iBC	84	36	42	72	42	37	63	21	71	77	77	0	0	72	66	72	36	50	79	98	52
iBM	64	27	32	55	32	28	48	16	52	0	0	0	0	55	49	27	27	39	0	69	0

FIG 3.6. iBC and iBM virulence factor production capacities.

FIG 3.6 (Cont'd). (A) Tradeoff between virulence factor production and growth (biomass production) was predicted by optimizing the production of an array of virulence factors while constraining the flux of the biomass reaction to various percentages of its maximum value under SCFM conditions. Presented values were normalized from 0-100% for each virulence factor across all biomass flux constraint conditions. While iBC is capable of producing all tested virulence factors, iBM contains incomplete biosynthetic pathways for the production of spermidine, cis-2-dodecenoic acid, salicylate, and di-rhamnolipid. (B) The contribution of the carbon source components of SCFM to the production of the virulence factors was predicted. For each virulence factor, production was maximized under minimal media conditions supplemented with each SCFM carbon source individually. Presented values were normalized from 0-100% for each virulence factor across all single substrate conditions. (C) Each column of production capacities was averaged along all virulence factors to calculate an average production capacity on each substrate for each model as shown in the heatmap via text and shading. Capacity averages below 50 are in white text for contrast. Colors are consistent with the color bars of panels A-B. For both iBC and iBM, on average, L-tyrosine supports the highest average production capacity for both models.

contributes greatly to cepacian production in iBC but not in iBM (Figure 3.6B). Homogentisate is produced at high levels on L-phenylalanine for both iBC and iBM, but other factors are only produced at average levels on the same substrate. The models also predict that each virulence factor can be produced at some level on any single carbon source that also enables growth in that model as shown in Figure 3.3. These results suggest variabilities in virulence factor production given changing substrate availability and predict potential pathways to target for therapeutic virulence inhibition.

To evaluate the potential production advantage imparted by each analyzed substrate to the full set of virulence factors, we averaged the virulence factor production capacity for all compounds by each substrate in Figure 3.6C. The results suggest that general virulence factor production is well supported by catabolism of L-tyrosine, D-glucose, L-arginine and L-ornithine, and poorly supported by glycine, D-lactate, and L-serine. *B. cenocepacia* has higher capacity averages on all substrates capable of supporting virulence compound production. The difference in averages of substrates supporting any virulence factor production in both models highlight a potential species-specific substrate preference for L-proline by *B. cenocepacia*. Ultimately, these analyses offers an interesting set of hypotheses regarding substrates that may be important to maximal virulence activity during

infection of cystic fibrosis patients, identify potential variations in substrate preferences between each *Burkholderia* species, and emphasize the enhanced pathogenic abilities of *B. cenocepacia* in comparison to *B. multivorans*.

DISCUSSION

This study involved the parallel generation and comparison of two new genome-scale metabolic reconstructions for the multi-drug resistant pathogens *B. cenocepacia* and *B. multivorans*. The reconstructions share 1,437 reactions, with 68 additional reactions unique to iBC and 36 additional reactions unique to iBM. iBC accounts for the function of 1,537 genes, while iBM accounts for the function of 1,411 genes. The reconstructions incorporated pathways necessary for growth on an array of substrates, species-specific biomass formulations, and virulence factor synthesis pathways common to related bacteria as well as specific to Bcc pathogens. Models were validated using experimental growth screens on over 100 carbon sources including substrates abundant in the cystic fibrosis lung environment. This process enabled us to refine the genome annotations of each species by reannotating an array of hypothetical and putative proteins, evaluate the consequences of the large genomes of *B. cenocepacia* and *B. multivorans* via GPR comparisons, make predictions regarding unique and shared essential genes for each pathogen, and evaluate the enhanced virulence factor production capacity of *B. cenocepacia* in comparison to *B. multivorans*.

The reconstructions provide a framework for contextualizing the genes of *B. cenocepacia* and *B. multivorans* in relation to growth and virulence factor production capacities. iBC and iBM are new tools that can enable the interrogation of interdependent functions of the large Bcc genomes. The reconstruction process guided our evaluation of the current genome annotations, and in several cases, led us to propose annotation refinements. Additionally, our comparison of *in silico* predictions with *in vitro* growth screening identified pathways which were clearly functional *in vitro*, such as lysine degradation and tryptophan degradation, that we could not make function *in silico* by addition of any known enzymes/genes. Thus, this analysis highlighted potentially uncharacterized enzymes that enable the

function of certain pathways in *Burkholderia* and may also be active in other organisms. Our genome annotation refinements also generated model-driven hypotheses that can be followed up experimentally such as the species-specific production of rhamnolipids and spermidine and the degradation of ectoine.

Tools for semi-automatic generation of genome-scale metabolic models have proved useful for expediting the reconstruction process (Agren et al. 2013; Devoid et al. 2013). In this study, using the Model SEED pipeline to concurrently generate the two reconstructions aided the manual curation and reconciliation phases since the differences in content and function between the two reconstructions could more easily be tracked and evaluated in parallel. The parallel curation process also ensured that the two reconstructions were of higher detail and quality than if the two reconstructions were developed in isolation since genetic evidence for reactions were exhaustively cross-checked between the two species. It is unlikely that our reconstructions have incorporated every metabolic function of each species and further curation will be required as the Bcc knowledgebase grows. Growth prediction discrepancies on lysine, tryptophan, and branched chain amino acids have highlighted areas to be investigated further experimentally.

The size and scope of the reconstructions in addition to high average gene count per reaction likely contributed to the reduced number of predicted essential genes in comparison to other reconstructions. To date, while there are emerging computational efforts to predict essential genes (Juhas et al. 2012), no complete transposon mutagenesis library of *B. cenocepacia* or *B. multivorans* is available, and our network-driven *in silico* gene essentiality predictions serve as hypotheses for future work. The essential genes predicted to be unique to each species also offer interesting opportunities for targeted therapeutics. Given the general similarity of the network architectures in terms of the common incorporated reactions, the unique essential gene sets are a novel outcome of our comprehensive formulation and comparison of the GPR associations.

Our comparative analysis enabled system-wide evaluations as well as the evaluation of single reaction or gene functionality that could be tied to a particular model prediction, which is a strength of genome-scale metabolic modeling. Modeling reaction activity using the reconstructions and constraint-based analysis allowed us to evaluate altered production capacity between models due to specific reaction differences in upstream pathways that limited final production levels. These analyses provide specific, testable hypotheses regarding important metabolic functions that are less intuitive than hypotheses generated through gene essentiality analysis alone.

Ultimately, each comparative analysis performed in this study supports the conclusion that *B. cenocepacia* has enhanced metabolic capacity over *B. multivorans*. The predicted growth rate of iBC on most *in silico* media was equal to or higher than that of iBM, and notably higher on SCFM, a medium designed to replicate the environment of the cystic fibrosis lung. iBC also showed higher capacity for virulence factor production over iBM on a range of substrates found in CF lung sputum. Our reconstruction process enabled the organization and functional evaluation of the increased genetic redundancy (e.g. isozymes and gene duplications) across a range of pathways that were available to potentially supplement any disrupted gene function in iBC compared to iBM. This added genetic redundancy resulted in both metabolic robustness and increased opportunity for the evolutionary divergence of duplicated genes to perform new functions (Innan and Kondrashov 2010). We have shown that these reconciled metabolic reconstructions offer a way to develop and investigate intriguing hypotheses for the enhanced virulence of *B. cenocepacia* – a higher capacity for genetic adaptation in the CF lung in addition to the flexibility provided by a large number of distinct catabolic pathways.

In summary, our reconciled metabolic models of the emerging multi-drug resistant pathogens *B. cenocepacia* and *B. multivorans* are valuable tools that can aid in comparative analysis of metabolic capacity, identification of novel therapeutic targets and strategies, and prediction of key phenotypes of pathogenesis. Here, we

have thoroughly characterized the metabolism of the type strains of *B. cenocepacia* and *B. multivorans*. Our network reconstructions can be used to contextualize high throughput transcriptomic and proteomic data (Yoder-Himes et al. 2010; Zlosnik and Speert 2010; Sass et al. 2011; Blazier and Papin 2012; Sass et al. 2013) to provide further insight into gene regulation and downstream phenotypes. Additionally, our reconstructions can be used as established starting points to analyze pathogenesis and physiology of the growing list of sequenced *Burkholderia* strains, including both clinical and environmental isolates (Mukhopadhyay et al. 2010). The other members of the Bcc are also human pathogens, while related species *B. mallei* and *B. pseudomallei* are dangerous bioterror agents that would be ideal candidates for *in silico* study. Other non-pathogenic *Burkholderia* species are of great interest to metabolic engineering projects in bioremediation and agricultural biotechnology as soil pathogens that excel in catabolism of certain pollutants as well as the production of antimicrobials that protect plant health (Mahenthiralingam et al. 2005; Suárez-Moreno et al. 2012). Given their notably large genomes and capacity for metabolic adaptation as a key factor in pathogenesis, the reconstructions of *B. cenocepacia* and *B. multivorans* offer tremendous potential for future development of treatment strategies to combat chronic infections and a model for the metabolic analysis and comparison of similar dangerous pathogens.

Chapter 4: Targeting virulence-related metabolism in a prominent opportunistic pathogen

Acknowledgments: Anna Blazier, Phillip Yen, Juliane Thoegersen[2], Jason Papin

[1]Biomedical Engineering, University of Virginia

[2]Department of Systems Biology, Technical University of Denmark

SYNOPSIS

Alternative approaches to treating opportunistic pathogens are desperately needed to stem the rising incidence of antibiotic resistant infections. Traditional antibiotics often target gene products essential to growth of a microbe; we investigate virulence-related targets in combatting infection that can be used as an alternative or supplement to current chemotherapy. We evaluate complex interrelationships between growth and virulence-linked pathways using a new genome-scale reconstruction of *P. aeruginosa* PA14, a model organism known for its intrinsically high virulence. Our new model is a substantial expansion of previous models of *P. aeruginosa* PAO1; additional genes and pathways account for the potential function of 73 hypothetical proteins and putative gene functions as well as the activity of 77 virulence-linked genes, including expansion of virulence factor synthesis pathways to 19 unique compounds. Computational screening on clinically relevant media identified gene deletions resulting in production inhibition using a quantitative metric of tradeoff between synthesis of virulence factor versus biomass. This systems approach allowed us to class gene deletions that are inhibitory to virulence factor synthesis, growth, or both as unique therapeutic targets; successful experimental validation of selected gene targets using PA14 mutants justifies our approach and quality of predictions as well as provides potential new targets for therapy that sidestep traditional selectors for antibiotic resistance.

INTRODUCTION

Antibiotic resistance is a major health concern. Each year in the United States alone, two million illnesses are caused by antibiotic resistant bacteria, resulting in more than 23,000 deaths and an economic burden of over $55 billion (CDC, 2013). These numbers continue to grow as fewer antibiotics are brought to the market each year. Thus, there is a need to develop new approaches to study resistance that contribute to the design of novel treatment strategies.

Many traditional antibiotics kill bacteria by targeting pathways essential for growth and, consequently, promote the rapid development of resistance (Allen et al. 2014). Here, we explore an alternative therapeutic approach that prioritizes inhibition of mechanisms of infection rather than mechanisms of growth. Virulence factors are pathogen-produced molecules that are involved in a variety of activities that promote the maintenance of an infection, such as signaling or sequestration of essential compounds such as iron, and enhance the organism's potential to cause disease, such as the synthesis of toxins. Since virulence factors are non-essential for growth, it has been argued that targeting virulence factor production may result in a reduced selection pressure for the development of drug-resistant mutations within a bacterial population compared to traditional antibiotics (Clatworthy et al. 2007; Rasko and Sperandio 2010). Additionally, targeting virulence factor production might have a less detrimental effect on the host microbiota (Cegelski et al. 2008). Thus, therapeutics that inhibit these mechanisms of infection may prove to be a promising alternative to traditional antibiotics to treat resistant infections. At minimum, pairing current clinical therapeutics with new treatments that reduce virulence will handicap a pathogen's ability to thrive and adapt to the human host, reducing bacterial load, inflammation, and community coordination that contributes to the development of resistance to traditional antibiotics.

We approach this search for novel virulence-related targets from a systems perspective. Genome-scale metabolic network reconstructions can be used to study an organism's metabolic capability to produce virulence factors. Assembled from

annotated genomes, metabolic reconstructions incorporate biochemical, genetic, and cell phenotype data and account for hundreds to thousands of gene-protein-reaction relationships as well as physicochemical constraints, such as reaction stoichiometry and directionality. Importantly, these reconstructions can be represented as a mathematical framework that can be analyzed using a variety of constraint-based modeling techniques, such as flux balance analysis (FBA), to probe the organism's capability of catabolizing different substrates and synthesizing various metabolites. This allows us to optimize activity for different metabolic objectives, whether that is maximization of virulence factor production in a particular growth environment or a quantitative evaluation of tradeoffs between the production of biomass versus a virulence factor.

We use *Pseudomonas aeruginosa* as our model organism of choice; *P. aeruginosa* is a Gram-negative opportunistic pathogen known to infect immunocompromised patients and develop substantial antibiotic resistance during chronic infections of the lungs of cystic fibrosis (CF) patients. Two genome-scale metabolic network reconstructions of *P. aeruginosa* PAO1 have been previously published (Oberhardt et al. 2008; Oberhardt et al. 2011). These models have been used to study *P. aeruginosa* metabolism in CF infections (Oberhardt et al. 2010), metabolic activity within bacterial biofilms (Sigurdsson et al. 2012), and, importantly, serve as a knowledgebase of *P. aeruginosa* metabolism. We updated this model to improve predictions and account for new discoveries in literature, enhance usability by updating model syntax, and expand the model's ability to predict virulence-related metabolic activity. We then used this model as a base for creating a strain-specific model of *P. aeruginosa* PA14. While these strains are closely related, PA14 is a primary clinical isolate that has increased in use as a model strain due to its significantly higher virulence in a variety of hosts (D.G. Lee et al. 2006). This study specifically describes analysis of this new *P. aeruginosa* PA14 model though statistics, curation, and accuracy of the updated PAO1 model are also included in certain sections.

Aiming to enhance the functionality of the model, incorporate new understanding of *P. aeruginosa* metabolism, and specifically probe the complex network of virulence-related metabolism in Pseudomonad pathogens, we present an updated version of the *P. aeruginosa PAO1* metabolic network reconstruction, iPae1148, and a new reconstruction for *P. aeruginosa* PA14, iPau1131. This new model accounts for 1,131 genes, 1,499 reactions, and 1,286 metabolites. We validated our models using data from carbon source utilization experiments conducted for PAO1 and PA14 as well as published genome-wide transposon mutant data from PAO1 and PA14 libraries. We then used iPAu1131 to probe *P. aeruginosa* virulence factor production capabilities on different carbon sources and mutant backgrounds, predicting genes essential for virulence factor production as well as biomass production. Finally, we tested these predictions for the virulence factor pyoverdine, a siderophore involved in iron acquisition, and validated 7 genes predicted to be critical for pyoverdine production on different carbon sources *in vitro*.

MATERIALS AND METHODS

Metabolic network reconstruction

The iMO1056 model (Oberhardt et al. 2008) and iMO1086 model (Oberhardt et al. 2011) were both used as resources during reconstruction efforts. iMO1056 was created using field-standard syntax based off the software Simpheny and consistent with many models in the BiGG database, while iMO1086 was built using the ToBiN platform which is not publically available. Since these original models were published, the modelSEED has become a favored draft reconstruction resource, and offers a comprehensive library of balanced reactions and metabolites from which hundreds of draft models have been created for use within the modeling community (Henry et al. 2010). In light of this, we used a draft conversion of iMO1056 to model SEED format as the starting point for our reconstruction update to enable consistency with our past *P. aeruginosa* models and easy comparison with a large collection of models created and curated by other groups. 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 as well

as correct conversion errors in reaction stoichiometry, directionality, and 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.

Genome comparisons. The genomic contents of *P. aeruginosa* PAO1 and *P. aeruginosa* PA14 and two related pathogens from the *Burkholderia cepacia* complex were compared to enable development of new, reconciled reconstructions for each strain from previously built models. *P. aeruginosa* PA14, *P. aeruginosa* PAO1, *B. cenocepacia* J2315, and *B. 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 between iPae1148 and iPau1131. Genes with hits that received a higher E-value score were manually evaluated based on predicted function, PseudoCAP category, and gene descriptions on the Pseudomonas Genome Database (PGD) for inclusion in the models (Winsor et al. 2011).

Biomass. Updated, strain-specific biomass formulas were created using methods detailed in Bartell et al (Bartell et al. 2014). This effort expanded the number of components considered necessary for growth according to improved formulation development and an updated search of literature pertaining to *Pseudomonas* species. More specific lipids were implemented using recent studies from literature and as enabled by the expanded lipid reactions used in modelSEED draft reconstructions. Additional changes to *Pseudomonas*-specific requirements such as preference for ubiquinone-9 versus ubiquinone-8 as a key cofactor in respiration were also included.

Model curation. To fill gaps and improve predictions, additional model components were added from iMO1086 and the recently published reconstructions of *Burkholderia* species after identifying potentially orthologous genes between strains. Reactions implemented in curated *B. subtilis* SEED model iBsu1103 as well

as reactions included in the MetaCyc and MetRxn databases were also used as a resource (Henry et al. 2009; Kumar et al. 2012; Caspi et al. 2014). PAO1 and PA14 genes categorized as linked to virulence via data from experimental studies incorporated into the PGD were specifically evaluated for inclusion in the models to expand clinically-relevant functional prediction ability.

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 datasets motivated careful re-assessment of data sets and experimental confirmation of results.

Gene essentiality. 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 (Becker et al. 2007). Predicted essential genes were compared with a list of genes that did not show effective transposon insertions in both genome-scale transposon insertion libraries of *P. aeruginosa* PAO1 (Jacobs et al. 2003) and *P. aeruginosa* PA14 (Liberati et al. 2006). By using genes missing 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 to 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 addition of new components to the biomass formula.

Substrate utilization. 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. Carbon utilization data was compiled from literature for both PAO1 and PA14, but discrepancies motivated us to perform our own growth screens for both strains using Biolog phenotype arrays PM1 and PM2. Growth curve screens were performed in triplicate

using a microplate reader with shaking at 37C for 48 hours. Curves were evaluated to identify substrates enabling growth versus no growth as performed in Bartell et al (Bartell et al. 2014). Results guided specific curation of catabolic pathways and expansion of transport systems included in the model to improve prediction accuracy.

Prediction of virulence-related production capacities and growth tradeoffs.

Virulence factor production capacity was first evaluated by optimizing the flux through an artificial 'demand' reaction for each virulence-related metabolite. Genes essential for the production of each virulence factor were predicted by evaluation of production flux after *in silico* deletion of a single gene at a time. Tradeoffs between virulence factor production and growth under WT and mutant strain conditions were evaluated by construction of Pareto optimum fronts. Pareto fronts were calculated for WT and all single gene knockouts by constraining biomass production flux at 20 different levels ranging from 0 to 100% of biomass production capacity in the given strain background and calculating the optimal virulence factor production level at each step. The area under this curve was used as an index of production that incorporated both virulence factor synthesis and biomass synthesis. This analysis was repeated for 'enzyme' knockouts that were represented by simultaneous knockout of all of the genes involved in each unique GPR within the model. From these analyses, the Pareto area for WT, single gene knockout strains, and enzyme knockout strains were predicted for each virulence factor under a range of growth conditions including SCFM and clinically relevant single substrate minimal media. The resulting Pareto areas for each knockout strain condition were normalized by WT Pareto area in that condition to create a Pareto Tradeoff Area (PTA) index to enable uniform comparisons of change in tradeoff relationships across all combinations.

Network visualization

iPau1131 was visualized using an in-house command line implementation of MetDraw that enables color overlay. A public version that builds the maps is available at www.Metdraw.com (Jensen and Papin 2014).

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 (Liberati et al. 2006) were grown in Luria-Bertani media supplemented with 15 ug/ml gentamycin as necessary at 37C with aeration for liquid cultures. Single substrate growth conditions were evaluated using 20mM concentrations of the carbon source in M9 minimal media with 15 ug/ml gentamycin as necessary at 37C.

Pyoverdine assay

To measure pyoverdine production, strains were grown in minimal media to stationary phase in 50 ml flasks and the absorbances of culture supernatants were measured at 405 nm according to a previously published protocol (Wurtzel et al. 2012). All measurements were normalized to culture density as determined by the absorbance of the bacterial culture at 600 nm.

RESULTS

Metabolic network reconstruction of P. aeruginosa

Here, we present an updated genome-scale metabolic network reconstruction of *P.aeruginosa* PAO1, iPae1148, and a new genome-scale metabolic reconstruction of *P. aeruginosa* PA14, iPau1131, that accounts for 1131 genes associated with 1499 reactions across a variety of functional categories as defined by KEGG. Briefly, the reconstruction process began with a Model SEED version of a previously published *P.aeruginosa* metabolic reconstruction, iMO1056. After verifying proper conversion of the SEED model from iMO156, we implemented an expanded biomass equation and ensured that all biomass components could be synthesized under rich media conditions. Using an iterative process, we subsequently curated and validated the draft reconstruction according to carbon source utilization, gene essentiality, and drug activity data. During this curation, we relied on several literature and database resources, namely MetaCyc and KEGG. We then used this model to develop a *P. aeruginosa* PA14 reconstruction using bidirectional BLAST hits for automated

conversion of high confidence orthologs paired with manual curation of low confidence orthologs, hypothetical proteins, and virulence factor pathways.

Of the 1499 reactions in iPau1131, 1085 are metabolic reactions, while 242 are transport reactions, and 172 are exchange reactions (Figure 4.1A). Furthermore, 797 of the total number of reactions in the model are associated with a single gene, while 470 reactions are associated with multiple genes. Conversely, 59 reactions have no gene association. In order to assess the coverage of iPau1131, we grouped the model reactions, genes, and metabolites according to functional categories as defined by KEGG (Figure 4.1B). Because reactions have their own unique identifiers, they were each assigned to a single functional category within the model despite their ability to potentially function in multiple pathways. Metabolites and genes can occur multiple times in a functional category as well as across multiple functional categories. To account for this redundancy within categories, we only counted metabolites and genes once per the functional categories in which they appeared.

A

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

Model	**iMO1056**	**iMO1086**	**iPae1148**	**iPau1131**
Genes	1056	1086	1148	1131
Unique EC identifiers	502	NA	675	675
Metabolites	760	1021	1284	1286
Cytoplasmic	760	916	1083	1085
Extracellular	NA	105	201	201
Reactions	883	1031	1497	1499
Metabolic	750	898	1083	1085
Transport	133	133	242	242
Exchange	NA	105	172	172
Gene-protein-reaction associations				
Single gene	839	839	800	797
Multi-gene			468	470
None	44	47	56	59

B

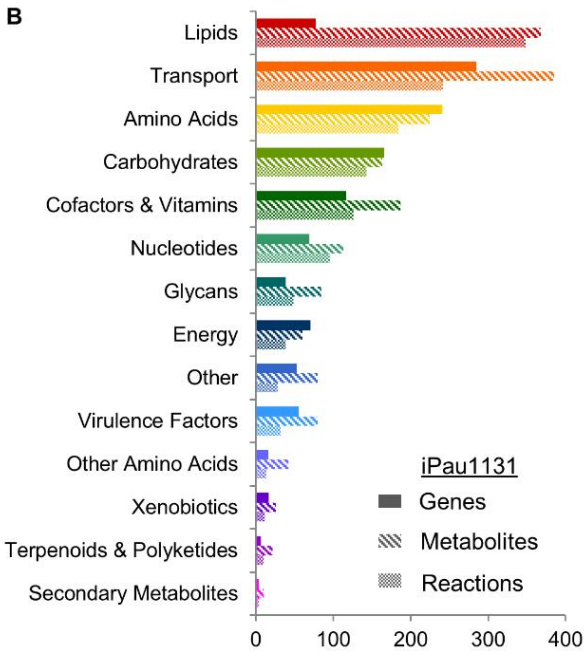


FIG 4.1. **iPau1131 reconstruction characteristics.** (**A**) Properties of iPau1131 (for comparison with iMO1056, see Supplementary Table S1). (**B**) The number of genes, metabolites, and reactions in iPau1131 grouped into functional categories as defined by KEGG (Kanehisa *et al*, 2006).

This classification revealed that reactions involved in lipid metabolism are most abundant, followed by transport reactions and reactions involved in amino acid metabolism. This high prevalence of lipid metabolism reactions is due to the implementation of a more detailed biomass equation, which includes an updated accounting of lipid constituents based on an expanded implementation of lipid reactions in SEED format as well as updated analysis of lipid composition from literature. Thus, a major manual curation effort of this *P. aeruginosa* model version focused on adding and curating lipid metabolism reactions to ensure that the lipid components in the biomass equation could be synthesized. In total, the number of reactions in lipid metabolism increased by 240 from iMO1086 to iPau1131.

In contrast to reactions, genes involved in transport, amino acid metabolism, and carbohydrate metabolism are most abundant. In an effort to expand carbon source utilization capabilities of the model according to BIOLOG phenotyping data, 109 transport reactions were added to the model compared to iMO1086 in a manual curation step that improved GPR formulations and metabolite specificity. Interestingly, the number of genes in the lipid metabolism category is relatively low compared to the number of reactions in this functional category; lipid metabolism has the lowest gene-to-reaction ratio out of all of the functional categories (approximately .2 genes per reaction) due to the same gene being able to catabolize several reactions. Virulence factors, on the other hand, have a gene-to-reaction ratio of 2, the highest of all the functional categories.

Finally, unique metabolites are most abundant in transport, followed by lipid metabolism, and amino acid metabolism.

Prediction of growth on single carbon sources

In order to guide curation efforts and validate our updated model, we used Biolog phenotyping microarrays to predict substrate utilization of 190 carbon sources for *P. aeruginosa* PAO1 and PA14. While these screens had been previously performed and published in literature, we discovered discrepancies in the results from different screens and decided to replicate them for both strains using the same experimental

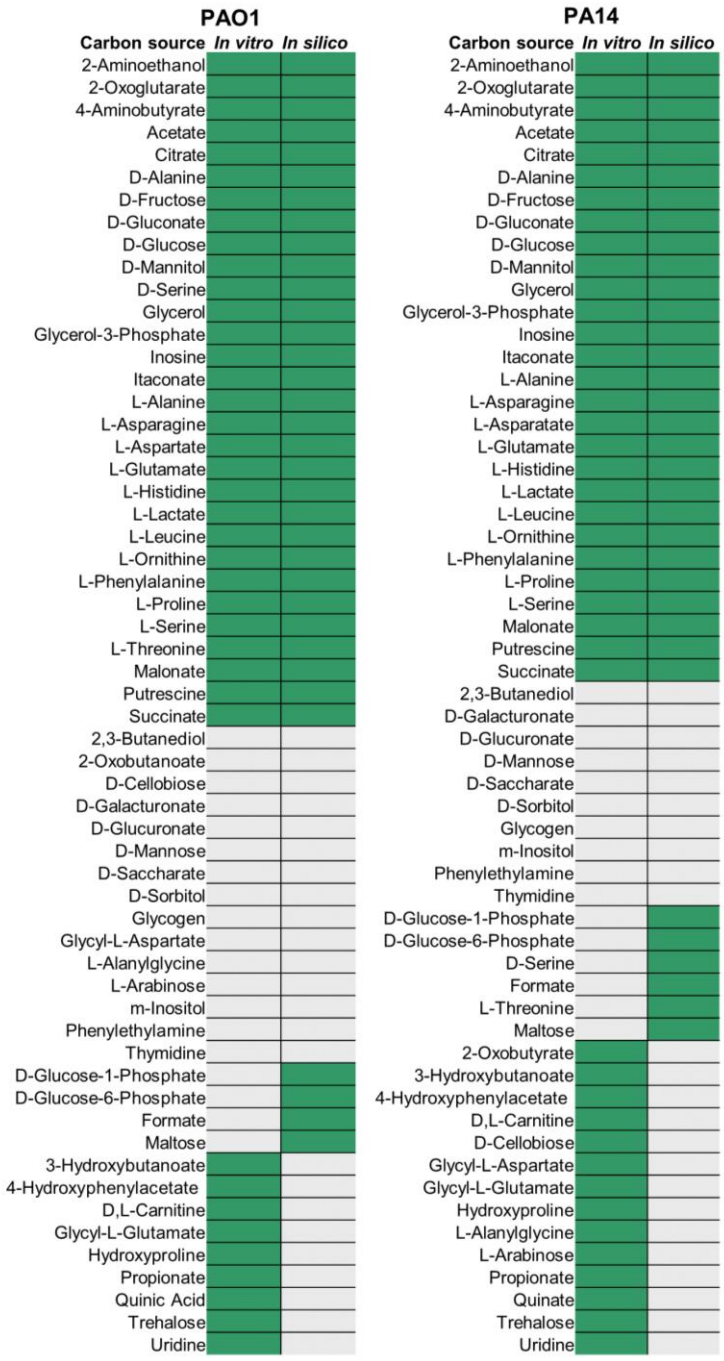


FIG 4.2. **Substrate utilization by iPae1148 and iPau1131.** Green indicates growth, grey indicates no growth *in vitro* versus *in silico*, respectively. Overall, iPae1148 and iPau1131 predict *P. aeruginosa* carbon source utilization with an accuracy of 75% and 63.3%, respectively.

conditions and analysis to ensure accuracy. Instead of evaluating optical density at 0 versus 48 hours as is often performed with these arrays, we collected full growth curves from absorbance measurements over 48 hours in a shaking, incubated

microplate reader. A previously published *in vitro* dataset had been used to evaluate the ability of *P. aeruginosa* PAO1 to grow on 95 substrates using BIOLOG phenotyping plate PM1 (Oberhardt et al. 2008). *P. aeruginosa* PAO1 model iMO1086 enabled growth predictions for 51 of these substrates. In iPae1148 and iPau1131, transport reactions and catabolic pathways for substrates were curated to account for a total of 60 Biolog carbon sources.

Using these updated models, we performed growth simulations to assess their ability to synthesize biomass on 60 single-carbon source minimal media. We then compared these results to the *in vitro* dataset to determine the predictive accuracy of the model (Figure 4.2). The results show a surprising number of differences between strains given the high degree of similarity between their metabolic genes. Ultimately, iPau1131 achieved an accuracy of 63.3% compared to the 75% accuracy of iPae1148 which was comparable to the 78% accuracy of iMO1086 and iMO1056. The lower accuracy of iPau1131 is primarily due to an increase in *in vitro* catabolic abilities compared to *in silico* predictions, though two false extra false positives were also predicted in the utilization of two amino acids, D-serine and L-threonine. These false positives highlight superfluous functionality of the model; possibly PA14 has lost this catabolic ability as an adaptation towards more efficient use of key substrates during infection. The false negatives may possibly be enabled by genes of currently unknown function within the PA14 genome or expanded capacity of transporters. The associated hypothetical proteins could also be present in PAO1, but simply aren't expressed based on differences in gene regulation between the strains. These discrepancies offer an opportunity for more targeted experimental assays to fill in missing knowledge regarding PA14-specific catabolic pathways.

Gene essentiality predictions

In addition to carbon source utilization data, we also validated our model with previously published gene essentiality data derived from two transposon mutagenesis screens for PAO1 and PA14 (Jacobs et al. 2003; Liberati et al. 2006). Though initially we planned to use each library screen as a validation data set for

each respective model, comparison of the genes without transposon insertions in both libraries revealed greater variability than expected as shown in Figure 4.3. Liberati et al. included a detailed explanation of potential causes of this discrepancy between library-derived essential genes in their 2006 publication describing the PA14 library which included biases in transposon insertion sites, difficulty in hitting genes with short sequences, failed insertion in operons upstream of essential genes, and insertions at the 3' end of genes that do not actually inhibit function. In that publication, the authors identified 335 candidate orthologous essential genes between the two libraries. We compared the up-to-date information on transposon insertions for both libraries from the Pseudomonas Genome Database and identified 404 overlapping orthologous essential genes; 101 of these genes were classed in a metabolism-related PseudoCAP category. Strains that had a transposon inserted in only one library but were predicted to be essential by the models were deemed to be probable essentials if the location of the transposon insertion was at the 3' end of the sequence.

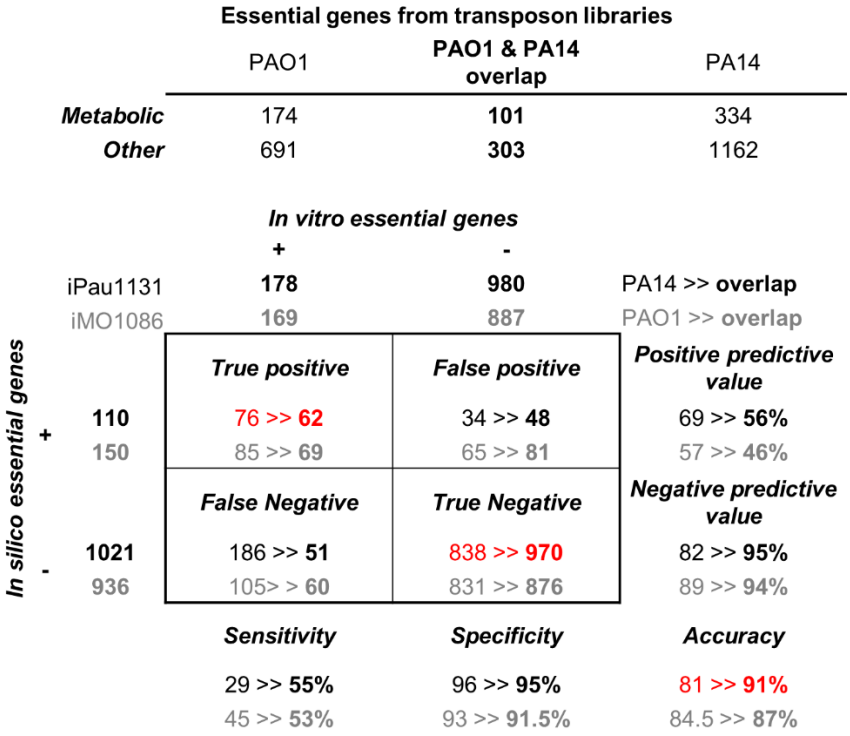


FIG 4.3. **Gene essentiality improvements in iPau1131.** Green indicates growth, grey indicates no growth *in vitro* versus *in silico*, respectively. Overall, iPae1148 and iPau1131 predict *P. aeruginosa* carbon source utilization with an accuracy of 75% and 63.3%, respectively.

We then evaluated *in silico* gene essentiality using single gene knockout analysis while optimizing for biomass production. We compared predictions from our new models and iMO1086 to both single strain *in vitro* essential gene sets and the overlapped essentials set. We identified 110 genes to be essential for rich media growth by iPau1131 compared to 150 in iMO1086 using the overlapped essentials set as shown in Figure 4.3. Encouragingly, the overall accuracy of iMO1086 at 87% improved with the overlapped essentials set compared to an accuracy of 84.5% using just PAO1 essentials. iPau1131 increased to 91% accurately predicted essential genes from the overlapped set from an original accuracy of 81% using the lower transposon insertion coverage PA14 essentials set, showing that both the new overlapped essentials and our updated reconstructions contribute to a better picture of gene essentiality in rich media. Interestingly, iPau1131 decreases in positive predictive value when using the overlapped essentials set versus the PA14 essentials set, but all other metrics improve in the new model or match those of iMO1086. The most substantial improvement in using the overlapped essentials set was the reduction of false negative predictions, while the most detrimental change is in the ability to predict true positive essential genes; however, the reduction in this number is not surprising given that the number of *in vitro* essential genes has been halved by the use of the overlapped essentials set. In summary, all but one metric of predictive quality (both counts and values) of iPau1131 equaled or improved from the prior predictions made by iMO1086 using only the PAO1 transposon library essentials. A final list of genes proposed to be essential to rich media growth based on *in silico* predictions and the combined transposon library analysis is also available.

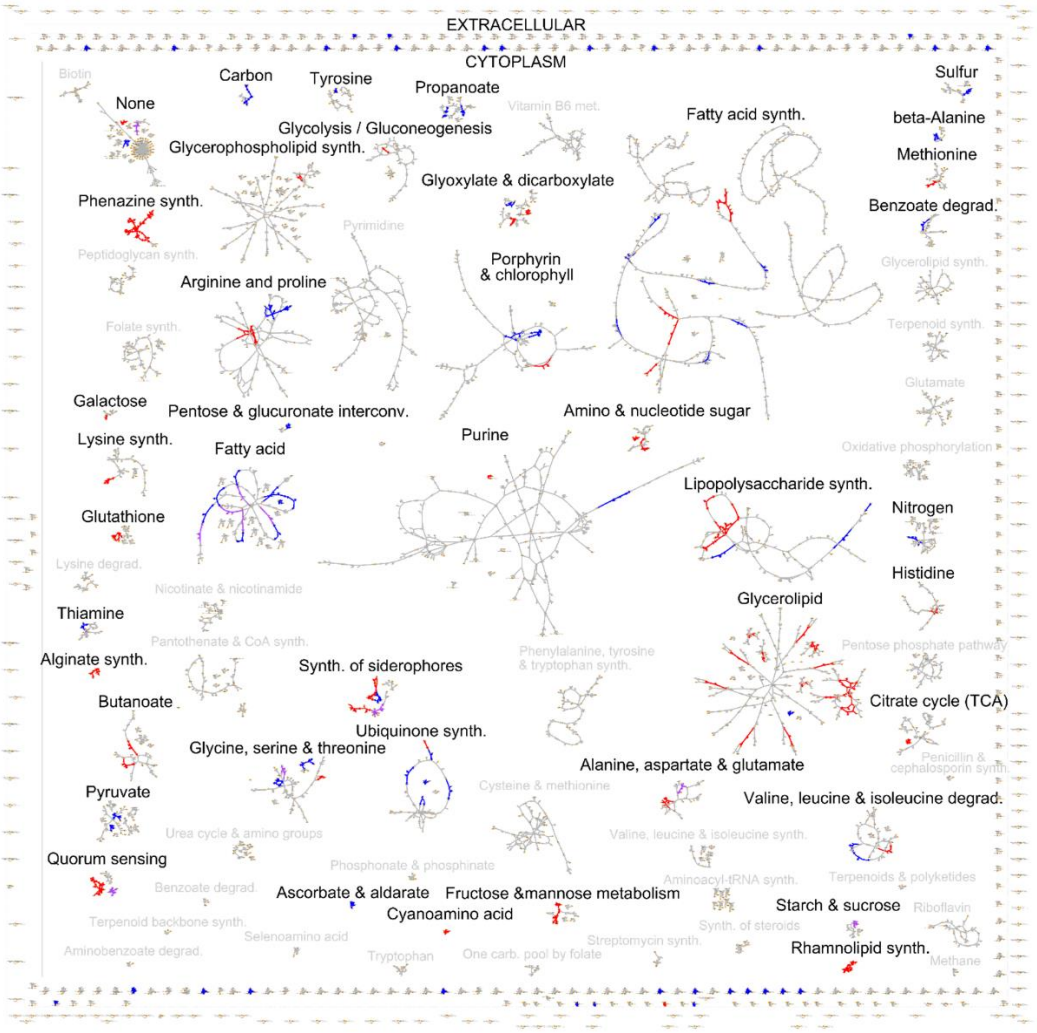
Pathway refinement

Despite the extensive study of *P. aeruginosa* as a model pathogen, genomic analyses of P. aeruginosa strains have noted that about 40% of sequenced proteins are uncharacterized, consistent with the same estimate for all currently sequenced organisms (Blaby-Haas and De Crécy-Lagard 2011; Pohl et al. 2014). Hypothetical proteins are considered a likely reservoir for novel metabolic and virulence-related

functionality (Hernández et al. 2009). Virulence-related genes and secondary metabolites are still being identified in various strains, and attempts to connect associated pathways to the broader metabolic network have been limited. Many high-throughput screening and bioinformatics-based approaches are being developed to improve these poor annotations; within this effort, genome-scale modeling provides an excellent quantitative framework for systematic curation and interrogation of less studied pathways and functionality (Blais et al. 2013). Thus, our model curation process was guided by multiple objectives. We wanted to improve our prediction accuracy when accounting for essential gene function and substrate catabolism as described above, and we simultaneously wanted to expand our ability to make functional predictions regarding putative protein activity and virulence-related genes. Figure 4.4A shows the distribution of pathway curation efforts using a map of all model reactions categorized by KEGG subsystem. Reactions associated with hypothetical protein refinements are highlighted in blue, reactions associated with virulence-linked genes are highlighted in red, and refinements linked to genes with putative functional assignments and virulence are labeled purple in Figure 4.4A. These refinements are described in more detail in the following sections.

Hypothetical proteins. Currently the Pseudomonas Genome Database (PGD) includes annotations of more than 2200 hypothetical proteins in each annotation of PAO1 and PA14, respectively. More than a thousand additional genes in PAO1 are annotated with putative or probable functions; the function of these genes are likely also putative/probable in the less-curated annotation for PA14. In iPau1131, we refined 50 hypothetical proteins or putative function gene annotations with a high degree of confidence, and included another 23 refinements at a lower degree of confidence.

A



B

Category	Virulence Factor	Metabolite ID
Exopolysaccharide	Alginate	cpd17074[c]
Lipopolysaccharide	A-band-O-antigen	cpd17056[c]
	B-band-O-antigen	cpd17057[c]
	Core Oligosaccharide	cpd17065[c]
	LipidA	cpd17066[c]
Phenazines	Chorismate	cpd00216[c]
	Phenazine-1-carboxylic acid	*cpd17083[c]*
	Pyocyanin	cpd01206[c]
Quroum Sensing	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]

New to iPae and iPau* ***iPau specific*

FIG 4.4. Curated pathways of iPau1131.

FIG 4.4 Cont'd. **Curated pathways of iPau1131.** (A) Refinement distribution of hypothetical proteins and virulence-related genes visualized across all iPau1131 reactions using MetDraw. (B) Completed virulence factor synthesis pathways for iPae1148 and iPau1131. Italicized virulence factors are additions to the new reconstructions, and bolded virulence factors are unique to PA14.

Further annotation refinements are accounted for in the model if low specificity functional assignments (example: ABC transporter) are included. Twenty seven of the genes are dehydrogenases of varying specificity, and many of these are classed in the low confidence refinement group. Other genes have been linked to substrate transport systems as guided by necessary gapfilling for experimental Biolog results; for example, an operon was identified with likely association with itaconate transport and catabolism that resulted in the refinement of hypothetical proteins and non-specific transport genes in the GPRs for five reactions with the assistance of a publication from 1964 (Cooper and Kornberg 1964). Further subsets are associated with other metabolic systems as shown in Fig. 4.3A. In summary, these genes span 28 KEGG subsystems, with the highest number of genes involved in transport reactions (29 genes), fatty acid metabolism (21 genes), and glycine, serine, and threonine metabolism (17 genes), and their implementation in the models offers a method of functional assessment in the context of broader system interactions that can guide experimental confirmation of function.

Virulence-linked genes. The PGD has incorporated two different resources describing virulence-linked genes in PAO1 and PA14 that are accessible via the 'Browse Virulence Factor' feature in their Database Search tool. The virulence-linked genes identified in PAO1 are sourced from the Virulence Factors Database (VFDB), a reference database gathered from literature on virulence of an array of pathogens including *P. aeruginosa* (Yang et al. 2008). PA14-specific virulence-linked genes are assembled from a transposon mutagenesis library screen in *C. elegans* (Feinbaum et al. 2012). Comparison of these virulence-linked gene sets showed that virulence was associated with 223 PAO1 genes and 56 PA14 genes. Evaluation of these genes showed that 177 out of the total 279 genes had metabolism-associated PseudoCAP assignments. While 261 of 279 genes had orthologs in each strain, only 12 orthologous genes were linked to virulence in both strains. Twenty-two virulence-linked genes were unique to PAO1 while 7 were unique to PA14. In our curation

process, we evaluated addition of the 177 metabolism-associated genes to both iPae1148 and iPau1131 because the unique approaches for identifying the virulence-linked genes for each strain likely contributed substantially to the low overlap between gene sets. We also focused on adding as many of the strain-specific virulence-linked genes as possible. Finally, a subset of hypothetical proteins linked to virulence were also evaluated in detail in efforts to propose reasonable functional assignments.

Our curation of associated pathways enabled iPAu1131 to account for the function of 77 virulence-linked genes. They are distributed through 32 KEGG subsystems as shown via the red reactions of Figure 4.4A that are associated with these genes. Subsystems most impacted by virulence linked genes include glycerolipid metabolism, lipopolysaccharide biosynthesis, and phenazine synthesis. While many genes are associated with the synthesis of specific virulence factors (discussed below), other more central metabolic genes were also linked to virulence such as hahahthose encoding the activity of imidazoleglycerol-phosphate synthase and arginine succinyltransferase. Accounting for the functions of these more distant virulence-related genes such that their activity can be specifically monitored in future computational analyses is an important update in these new reconstructions.

Virulence factor synthesis. Virulence factors are compounds or enzymes that enhance the ability of a pathogen to infect host tissue but are not essential to growth (Allen et al. 2014). The nature of our metabolic models leads us to focus specifically on virulence factor compounds that can be incorporated via metabolic synthesis pathways in our reaction networks. *P. aeruginosa* produces an array of such compounds which can be grouped into several categories: exopolysaccharides, lipopolysaccharides, phenazines, quorum sensing signal molecules, siderophores, and surfactant (Williams and Cámara 2009; Ballok and O'Toole 2013). Within these virulence categories, we expanded the number of complete virulence factor synthesis pathways that are accounted for in the models. Figure 4.4B lists the compounds that can be synthesized by iPau1131 and are considered virulence factors that contribute to successful pathogenesis. The six italicized factors are new

additions to iPae1148 and iPau1131 compared to iMO1086, and bolded dihydroaeruginoic acid is a recently identified PA14-specific virulence factor included only in iPau1131 (Maspoli et al. 2014).

A subset of the virulence genes included in these pathways are also hypothetical proteins, and their associated reactions are noted in purple in Figure 4.4A. A non-specific acyl-CoA dehydrogenase, PA14_52900, is associated with several fatty acid metabolism reactions as well as a reaction in valine, leucine, and isoleucine degradation. Additionally, PA14_54640 is annotated as an enoyl-CoA hydratase on the PGD, but literature identifies it as *dspI*, a gene required for synthesis of diffusible signal factor cis-2-decenoic acid (Amari et al. 2013). Other virulence factor genes are annotated as more typical hypothetical proteins; for example, several genes of the highly divergent B-band O-antigen operon have not been characterized for PA14, though gene functions have been studied in detail for PAO1 (Raymond et al. 2002; Lam et al. 2011). In iPau1131, we chose to incorporate the associated reactions as formulated for iPA1148 (and previously iMO1086) and fill in genes with high and moderate confidence annotations as possible. These pathways will need further curation once the characterization of this locus has been performed experimentally, but for now they represent a best estimate of how these genes may function.

Modeling virulence factor production capabilities

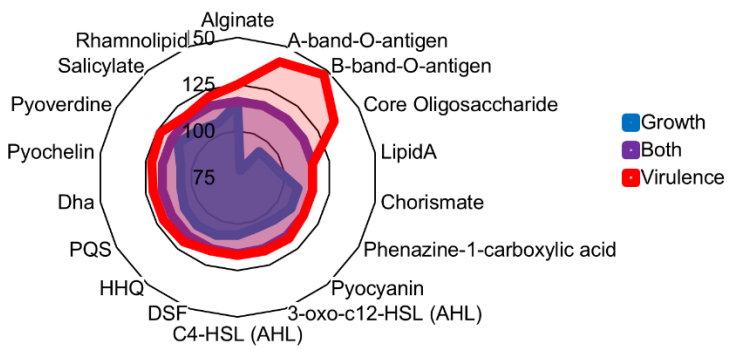
A strength of using genome-scale metabolic modeling as an interrogation approach is the ability to systematically compare the effects of a broad range of small changes such as single gene knockouts or altered media conditions. This comparative approach is valuable for predictions made in the context of complex environmental conditions such as growth in CF sputum. Previous modeling studies have used an *in silico* synthetic cystic fibrosis medium from a defined media developed from a chromatographic analysis of CF patient sputum samples (Palmer et al. 2007; Oberhardt et al. 2010). We use this *in silico* medium to represent the growth conditions that strains of *P. aeruginosa* encounter during CF lung

colonization while evaluating the roles of genes and enzymes in both the production of biomass and the 19 virulence factors accounted for in iPau1131.

Gene essentiality on SCFM. We probed the effect of single gene knockouts on the production of each virulence factor alone and compared essential gene sets for each factor to the genes predicted to be essential for production of biomass on SCFM. Analysis showed that unique subsets of genes were required for production of growth and virulence factors, respectively, though this set changed depending on the virulence factor being evaluated. However, there was a third subset of genes that was required for production of both biomass and a given virulence factor as shown in Figure 4.5A. The blue inner boundary can be interpreted as essential genes solely contributing to production of biomass. The gene count represented by the difference between the blue boundary and the purple boundary are the genes that are required for both biomass production and production of a given virulence factor. The difference between the red boundary and the purple boundary is the number of genes that are essential solely for the production of that virulence factor; knockout of these genes does not affect biomass production. As shown in Figure 4.5A, the O-antigens require the most genes for production; they have the highest number of virulence-only essential genes as well as the highest number of shared genes required for production of both biomass and virulence factors. In contrast, the analysis indicates that the synthesis pathways for quorum signaling molecules PQS, HHQ, and DSF are much simpler in comparison. Far fewer genes are essential for both virulence factor production and biomass production, and even fewer are required solely for synthesis of the signals.

What this figure also highlights is an unexpected consistency in the pool of genes required for both virulence factor production and growth. One hundred sixteen genes are essential for growth on SCFM, and 46 of these genes are also essential for the production of at least one virulence factor. These 46 genes critical to both biomass production and virulence are listed in Figure 4.5B along with their PseudoCAP category and function, and the number of virulence factors that cannot be synthesized if they are knocked out is listed in the far right column. The

A



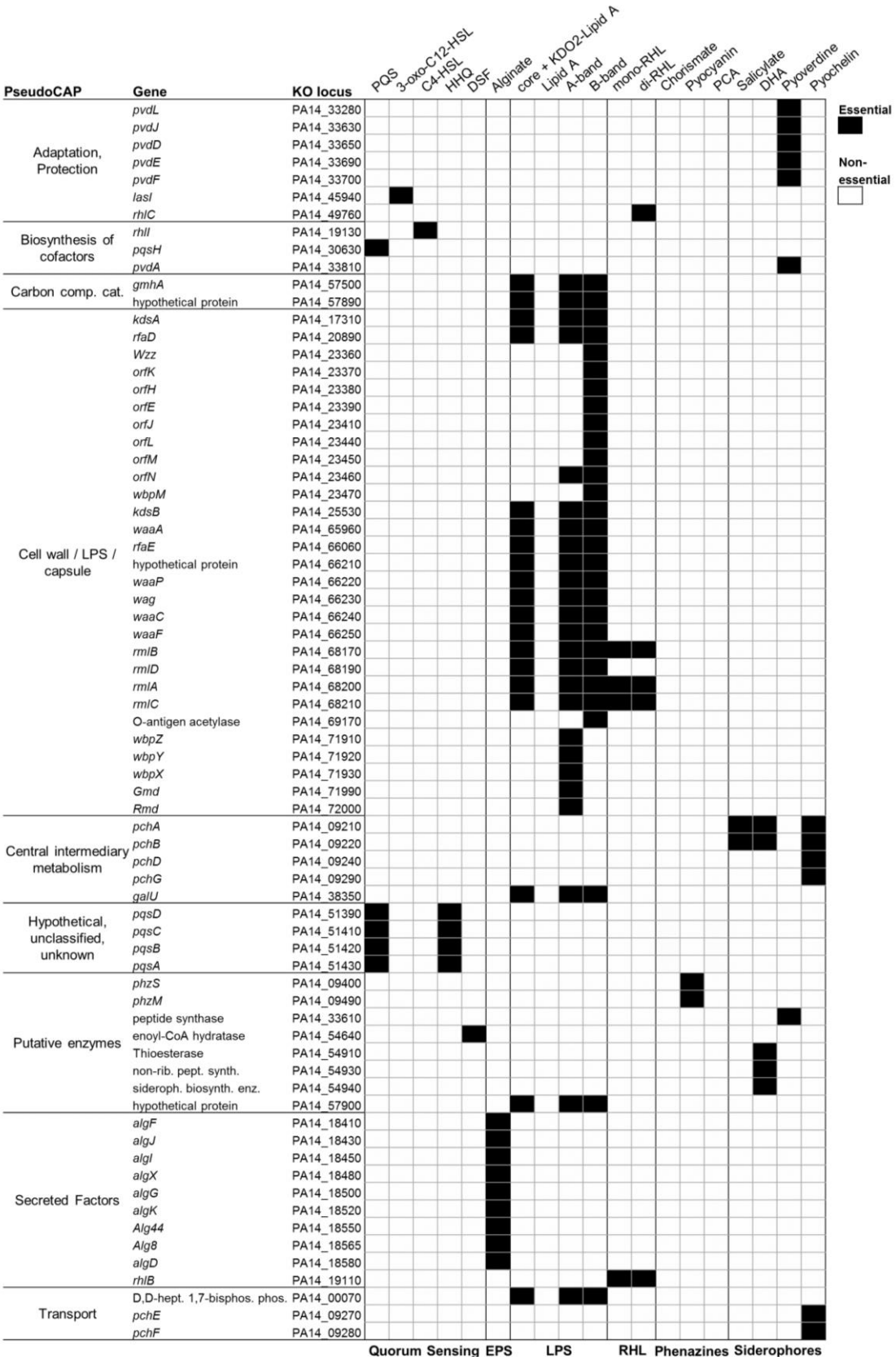
B

PseudoCAP	Locus Tag	Gene	Function	VF Count
Amino acid biosynthesis & metabolism	PA14_00290	*aroE*	shikimate 5-dehydrogenase	6
	PA14_23310		3-phosphoshikimate 1-carboxyvinyltransferase	6
	PA14_42760	*aroC*	chorismate synthase	6
	PA14_66600	*aroB*	3-dehydroquinate synthase	6
	PA14_66610	*aroK*	shikimate kinase	6
	PA14_23800	*asd*	aspartate-semialdehyde dehydrogenase	1
	PA14_23920	*purF*	amidophosphoribosyltransferase	1
	PA14_52580	*lysC*	aspartate kinase	1
Antibiotic resistance & susceptibility	PA14_07090	*metK*	S-adenosylmethionine synthetase	5
	PA14_39190	*bacA*	UDP pyrophosphate phosphatase	2
Biosynth. of cofactors, etc.	PA14_41350	*folD*	5,10-methylene-tetrahydrofolate dehydrogenase/cyclohydrolase	2
Carbon compound catabolism	PA14_07130	*tktA*	transketolase	11
Cell wall / LPS / capsule	PA14_17180	*lpxD*	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	4
	PA14_17210	*lpxA*	UDP-N-acetylglucosamine acyltransferase	4
	PA14_17220	*lpxB*	lipid-A-disaccharide synthase	4
	PA14_25510	*lpxK*	tetraacyldisaccharide 4'-kinase	4
	PA14_57260	*lpxC*	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	4
	PA14_62840	*glmM*	phosphoglucosamine mutase	4
	PA14_73220	*glmU*	glucosamine-1-phosphate acetyltransferase	4
Central intermediary metabolism	PA14_07910	*rpe*	ribulose-phosphate 3-epimerase	11
	PA14_70270	*algC*	phosphomannomutase	6
	PA14_73170	*glmS*	glucosamine--fructose-6-phosphate aminotransferase	4
	PA14_05620	*sahH*	S-adenosyl-L-homocysteine hydrolase	3
Energy metabolism	PA14_04310	*rpiA*	ribose-5-phosphate isomerase A	9
Fatty acid & phospholipid metabolism	PA14_17270	*accA*	acetyl-CoA carboxylase carboxyltransferase subunit alpha	11
	PA14_23860	*accD*	acetyl-CoA carboxylase subunit beta	11
	PA14_25650	*fabD*	malonyl-CoA-ACP transacylase	11
	PA14_25660	*fabG*	3-ketoacyl-ACP reductase	11
	PA14_64100	*accB*	acetyl-CoA carboxylase biotin carboxyl carrier protein	11
	PA14_64110	*accC*	acetyl-CoA carboxylase biotin carboxylase subunit	11
	PA14_43690	*fabB*	3-oxoacyl-ACP synthase	10
	PA14_43680	*fabA*	3-hydroxydecanoyl-ACP dehydratase	8
Nucleotide biosynthesis & metabolism	PA14_14820	*ndk*	nucleoside diphosphate kinase	6
	PA14_17080	*pyrH*	uridylate kinase	4
	PA14_15310	*guaB*	inosine 5'-monophosphate dehydrogenase	1
	PA14_15340	*guaA*	GMP synthase	1
	PA14_15740	*purL*	phosphoribosylformylglycinamidine synthase	1
	PA14_30110	*purB*	adenylosuccinate lyase	1
	PA14_51240	*purC*	phosphoribosylaminoimidazole-succinocarboxamide synthase	1
	PA14_52040	*purM*	phosphoribosylaminoimidazole synthetase	1
	PA14_61770	*prs*	ribose-phosphate pyrophosphokinase	1
	PA14_64200	*purH*	phosphoribosylaminoimidazolecarboxamide formyltransferase	1
	PA14_64220	*purD*	phosphoribosylamine--glycine ligase	1
	PA14_70440	*gmk*	guanylate kinase	1
	PA14_71600	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	1
Putative enzymes	PA14_41400		UDP-2,3-diacylglucosamine hydrolase	4

FIG 4.5. **Genes essential for virulence factor synthesis versus growth in SCFM.** (A) Overlap of gene sets involved in virulence versus growth. (B) Genes that are essential for production of biomass and at least one virulence factor.

PseudoCAP category that appears to be critical for the largest number of virulence factors is fatty acid and phospholipid metabolism, where six genes are predicted to be essential for the production of 11 virulence factors in addition to biomass production. Several genes in the *aro* operon that contributes to aromatic amino acid synthesis are essential for the production of 6 virulence factors in the phenazine and siderophore families, while an array of genes involved in purine metabolism only impact 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 makes them attractive new targets for therapeutic intervention.

Mutant production tradeoffs on SCFM. The models enable us to compare quantitative levels of production inhibition due to gene knockouts as well as complete prevention of byproduct and biomass synthesis. We began our analysis by comparing sets of genes inhibitory to production of each virulence factor versus inhibitory to growth. Figure 4.6 shows all 71 genes that are essential for production of at least one virulence factor, but their deletion has no effect on biomass production levels. Interestingly, these genes are all located within synthesis pathways directly connected to production of each virulence factor, as shown by the effect of deletions within the *pvd* operon on pyoverdine alone. The number of essential genes per virulence factor is also an indicator of the complexity of these synthesis pathways. While this analysis does not identify any novel central metabolic targets that solely prevent virulence factor production, it does show the connections between the hypothetical genes and is also an indicator of the complexity of these synthesis pathways. While this analysis does not identify any novel central metabolic targets that solely prevent virulence factor production, it does show the connections between the hypothetical genes and putative enzymes reannotated in our curation of virulence factor synthesis pathways to affected virulence factors.



PseudoCAP	Gene	KO locus
Adaptation, Protection	*pvdL*	PA14_33280
	pvdJ	PA14_33630
	pvdD	PA14_33650
	pvdE	PA14_33690
	pvdF	PA14_33700
	lasI	PA14_45940
	rhlC	PA14_49760
Biosynthesis of cofactors	*rhlI*	PA14_19130
	pqsH	PA14_30630
	pvdA	PA14_33810
Carbon comp. cat.	*gmhA*	PA14_57500
	hypothetical protein	PA14_57890
Cell wall / LPS / capsule	*kdsA*	PA14_17310
	rfaD	PA14_20890
	Wzz	PA14_23360
	orfK	PA14_23370
	orfH	PA14_23380
	orfE	PA14_23390
	orfJ	PA14_23410
	orfL	PA14_23440
	orfM	PA14_23450
	orfN	PA14_23460
	wbpM	PA14_23470
	kdsB	PA14_25530
	waaA	PA14_65960
	rfaE	PA14_66060
	hypothetical protein	PA14_66210
	waaP	PA14_66220
	wag	PA14_66230
	waaC	PA14_66240
	waaF	PA14_66250
	rmlB	PA14_68170
	rmlD	PA14_68190
	rmlA	PA14_68200
	rmlC	PA14_68210
	O-antigen acetylase	PA14_69170
	wbpZ	PA14_71910
	wbpY	PA14_71920
	wbpX	PA14_71930
	Gmd	PA14_71990
	Rmd	PA14_72000
Central intermediary metabolism	*pchA*	PA14_09210
	pchB	PA14_09220
	pchD	PA14_09240
	pchG	PA14_09290
	galU	PA14_38350
Hypothetical, unclassified, unknown	*pqsD*	PA14_51390
	pqsC	PA14_51410
	pqsB	PA14_51420
	pqsA	PA14_51430
Putative enzymes	*phzS*	PA14_09400
	phzM	PA14_09490
	peptide synthase	PA14_33610
	enoyl-CoA hydratase	PA14_54640
	Thioesterase	PA14_54910
	non-rib. pept. synth.	PA14_54930
	sideroph. biosynth. enz.	PA14_54940
	hypothetical protein	PA14_57900
Secreted Factors	*algF*	PA14_18410
	algJ	PA14_18430
	algI	PA14_18450
	algX	PA14_18480
	algG	PA14_18500
	algK	PA14_18520
	Alg44	PA14_18550
	Alg8	PA14_18565
	algD	PA14_18580
	rhlB	PA14_19110
Transport	D,D-hept. 1,7-bisphos. phos.	PA14_00070
	pchE	PA14_09270
	pchF	PA14_09280

FIG 4.6. **Genes only essential for synthesis of each virulence factor.**

FIG 4.6 Cont'd. **Genes only essential for synthesis of each virulence factor.** Single gene knockout was performed while optimizing for production of each virulence factor separately on SCFM. All genes that when deleted also resulted in growth inhibition were removed from results.

To evaluate more complex effects of gene deletions in addition to essentiality, we calculated Pareto curves by maximizing for the production of a given virulence factor at intervals of possible biomass production levels for that particular genetic phenotype (WT or single gene knockout) as explained in more detail in the methods. After calculating the area under this curve for single gene knockouts of all the genes in the model, we normalized the value by the area calculated for wild-type production levels for the given virulence factor; this normalized area was then used as a representative production tradeoff (PTA) presented as 0 to 100% inhibition (white to black) compared to WT. Inhibition could be linked to lowered production of of a virulence factor, biomass, or both. We replicated this analysis for each virulence factor versus growth for wild-type and single gene knockouts of all genes in the model under SCFM growth constraints.

Two hundred forty seven gene deletions were identified that reduced PTA for at least one virulence factor. This number includes genes essential solely for biomass production or virulence factor production as presented in the preceding section. We then specifically examined inhibition of virulence factor production within this dataset, which was associated with 134 genes (71 of which are essential to VF production as shown previously); this analysis left us with 63 inhibitory gene deletions for further analysis. Figure 4.7A shows the 17 gene knockouts within this dataset that resulted in inhibited production of a virulence factor but did not affect biomass production, while the 24 gene knockouts of Figure 4.7B reduced the PTA by at least 20% from WT when knocked out and were also associated with reduction of biomass production as indicated by the 'Growth Alone' column on the right. Knockout production phenotypes range from minimal inhibition for a few virulence factors to complete inhibition of most byproducts. At times, all virulence factor within a category are affected at a similar inhibition level by what is clearly inhibition of the same pathway.

The gene deletions presented in Figure 4.7A highlight the importance of tryptophan catabolic pathways to the synthesis of quorum sensing molecules (tryptophan and kynurenine-related genes), sulfate assimilation to certain siderophores (*cys* genes), and impact on both phenazines and select siderophores by succinate dehydrogenase (*sdh* genes). However, the most prominent inhibition-causing deletion presented in Figure 4.7A is the knockout of PA14_07230, fructose-1,6-bisphosphate aldolase, which plays a key role in glycolysis. The PGD indicates that this gene is essential for growth in rich media; there are no transposon mutants available from the PA14 or PAO1 transposon mutagenesis library. So, while this is theoretically a false negative result in our gene essentiality analysis, the gene still shows clear impacts on production of metabolic byproducts such as virulence factors. Recent studies have identified novel alternative rewiring of glycolysis pathways in response to gene deletions in *Escherichia coli* that involve novel functionality of *fbaA* (Nakahigashi et al. 2009). Evolving knowledge of the role of this enzyme in glycolysis as well as its apparent importance to virulence factor production merits further experimental study in *P. aeruginosa* and potential consideration of the gene product as a therapeutic target that may be more effective against virulence factor production than growth *in vivo*.

In Figure 4.7B, the effects on PTA of growth inhibition by gene deletion are often amplified by inhibited virulence factor production. For example, the deletion of phosphoglycerate kinase (PA14_07190) broadly reduces the PTA for most virulence factors as it significantly affects biomass production; the most interesting artifacts of this gene knockout may actually be the virulence factors that appear to be much less inhibited than the average. Other virulence factors are substantially reduced in production while growth remains relatively unaffected, such as the effects of folate metabolism genes *metH* and *metF* on phenazine and siderophore production.

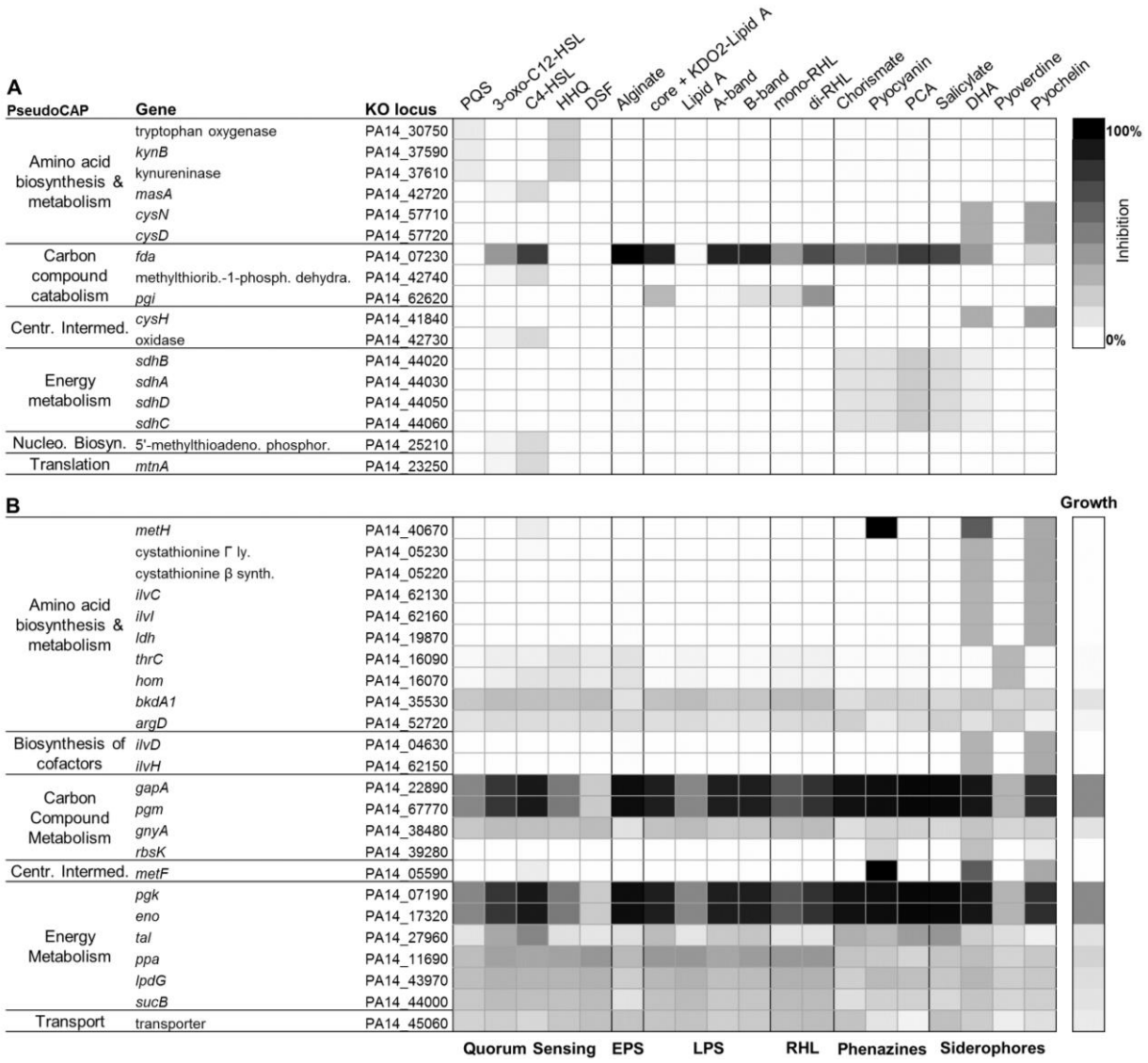


FIG 4.7. **Gene deletions that reduce production of virulence factors.** All values presented are of production tradeoff area assessing a mutant's ability to produce a given virulence factor at varying levels of biomass production. This value is normalized by the same area calculated for the wildtype under that condition, resulting in the PTA index. (A) Gene deletions that result in a reduction of PTA but deletions have zero effect on biomass production in isolation. (B) Gene deletions that reduce PTA and separately reduce biomass production.

On average, PTA inhibition is low for many of the mutants in Figure 4.7, but there are certain noteworthy patterns in the results. Highlighted relationships include the close connection between the synthesis of dihydroaeruginoic acid and pyochelin that results in a similar pattern of inhibition across the gene knockouts, though Dha production appears more susceptible to the knockout of a handful of genes.

Interestingly, four virulence factors from each of the large functional categories stood out in particular due to the variance of their inhibition pattern from the other members of their groups: C4-homoserine-lactone, lipid A, pyocyanin, and pyoverdine.

Pyoverdine production on SCFM substrates. Studies have shown that particular substrates within a complex medium may be preferred by bacteria in different growth states or under adapting regulatory networks due to exposure to stress or competition with other microbes (Shrout et al. 2006; Brown et al. 2008; Xavier et al. 2009; Frimmersdorf et al. 2010). These subtle changes are more difficult to capture using typical flux balance analyses of metabolic models without overlaying the effects of regulatory changes via reaction activity constraints based on experimental 'omics' expression data, but potential contributions of each individual media subcomponent to production of the desired product can be easily parsed. As an *in silico* case study of the potential contributions of different substrates to byproduct synthesis, we evaluated production tradeoffs between biomass and the virulence factor pyoverdine through calculation of Pareto fronts while individually providing 16 different carbon sources from SCFM. We calculated these fronts across the 16 substrates in minimal media using a single gene knockout screen as performed previously on SCFM.

Pyoverdine is a siderophore, a molecule that solubilizes iron for use by essential processes in bacterial metabolism. In fluorescent Pseudomonads, pyoverdine is the main siderophore and primary iron uptake mechanism in use, has been implicated in bacterial interactions in biofilms, and is essential for successful colonization of burn wounds; it is upregulated in initial colonization of the CF lung (Meyer et al. 1996; L. Yang et al. 2009; Yeterian et al. 2010). It was also noted for having an unusual production index inhibition pattern on SCFM when compared to other siderophores accounted for in the model. In examining our results from our screen of production tradeoffs under single gene knockouts replicated across SCFM and 16 individual SCFM substrates, we identified 189 genes that were essential for pyoverdine production during growth maintenance in at least one media condition

and another 13 genes that resulted in reduction of the production index from wildtype in at least one media condition when deleted. The number of essential genes varied across the different substrate conditions; aromatic amino acids isoleucine and leucine required the most genes for successful growth and virulence factor production at 171 genes, while glucose required the fewest at 159 genes.

In Figure 4.8, we show PTA results for 74 gene deletions that result in inhibition of pyoverdine production on at least one substrate but are not essential for growth on SCFM. The deletions resulting in 100% inhibition could therefore be tied to growth on that particular substrate, virulence factor production, or both. We find 20 genes essential across all substrate conditions; these are likely tied to biomass production inhibition in most cases. We also see genes that are clearly tied to production activity on a single substrate. Gene deletions of particular interest are the 15 those that show a distribution of inhibition phenotypes across the range of media; these hits indicate broad impacts across relevant substrates and offer a novel multi-targeted therapeutic approach in a growth environment where successful competition for the same resources results in preferential substrate switching.

Experimental evaluation of pyoverdine mutants

Our *in silico* study of production tradeoffs of pyoverdine on SCFM and 16 individual SCFM carbon sources mapped the effects of gene targeting across a range of conditions. We tested a subset of our predictions in an effort to validate novel therapeutic targets that were designed with respect to virulence factor production rather than growth. Evaluation of our computational results guided us to choose four carbon sources for inhibition comparisons. Glucose is used as a standard in minimal media studies, lactate is a known preferred substrate in growth of *P. aeruginosa* on CF sputum, arginine is a preferred substrate particularly in the low oxygen conditions of the CF lung, and isoleucine is predicted to have the highest number of essential genes for pyoverdine production (Palmer et al. 2007; Palmer et al. 2010).

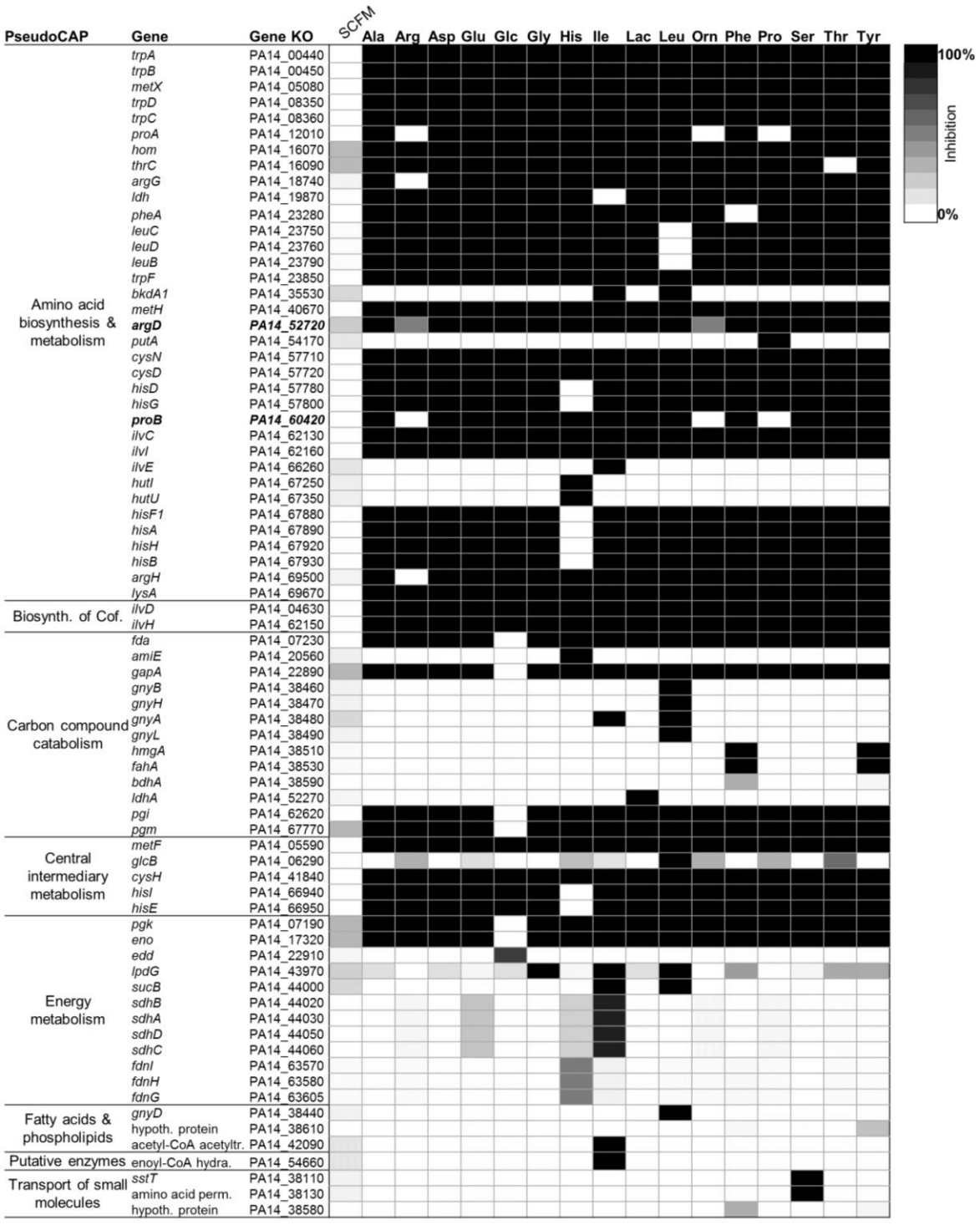


FIG 4.8. **Genes important to pyoverdine synthesis on SCFM and SCFM components.** Gene deletions that reduce PTA on at least one of a range of SCFM substrates but are not essential for pyoverdine or biomass production on SCFM.

We chose gene targets that were connected to multiple different inhibition phenotypes. Metabolic genes directly connected to the pyoverdine synthesis operon were used as controls; *pvdA*, *pvdD*, and *pvdF* were predicted to be essential to virulence factor production but had no effect on biomass production in all media conditions for pyoverdine production. Aspartate kinase (*lysC)* was chosen as an upstream metabolic gene predicted to be essential for growth on all media. Deletion of gamma-glutamyl kinase (*proB*) resulted in inhibition of pyoverdine and biomass production on arginine, ornithine, and proline, while it was predicted to be essential on all other substrates. Bifunctional N-succinyldiaminopimelate-aminotransferase/acetylornithine transaminase (*argD*) was also essential for pyoverdine and biomass production on all substrates but arginine and ornithine, on which pyoverdine production was inhibited.

Absorbance-based assays of pyoverdine production and growth of strains with transposon insertions in these genes from the PA14 library were conducted on glucose, lactate, arginine, and isoleucine M9 minimal media as described in Methods. Ratios of pyoverdine to growth for wild-type PA14 and each mutant strain are shown in Figure 4.9. The mutants show production inhibition from WT levels in most conditions, with the *pvd* gene deletions having the greatest impact on production in all cases. All mutants successfully inhibited pyoverdine production to varying levels on glucose and arginine as predicted by iPau1131. We also validate *proB* as a potential target for pyoverdine production on all substrates in accordance with model predictions. However, we also identified substantial resilience to pyoverdine inhibition when PA14 mutants were grown on lactate, contrary to model predictions. As mentioned previously, lactate is a preferred growth substrate of *P. aeruginosa* in the CF lung environment (Palmer et al. 2007.). The discrepancies between our model predictions and experimental results for lactate highlight potential knowledge gaps in our understanding of metabolic capacity and regulation during growth on this key substrate that may contribute to enhanced infections by *P. aeruginosa* in CF patients.

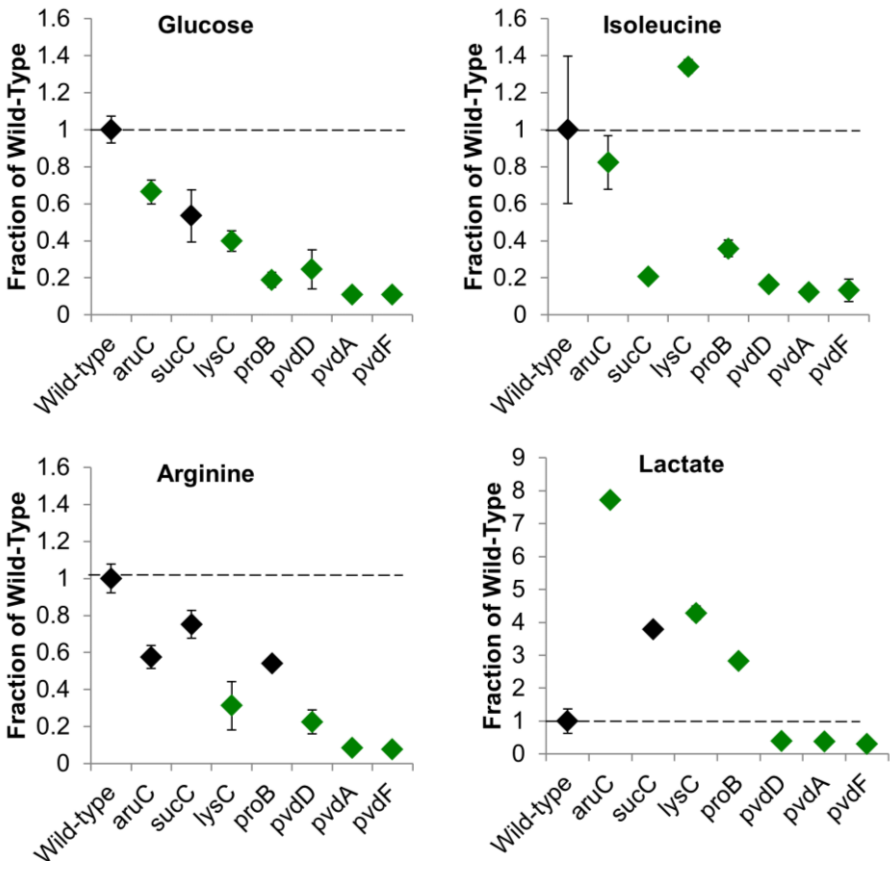


FIG 4.9. **Pyoverdine synthesis capabilities on 4 CF-relevant substrates *in vitro*.** Values presented are the log of the ratio of optical density at 405nm to evaluate pyoverdine production and optical density at 600 nm to evaluate growth normalized by the same value for PA14 WT (i.e. log(OD405/OD600) for mutant / log(OD405/OD600) for WT).

We also present experimental data for pyoverdine production on a mutant of succinyl-CoA synthetase (*sucC*) which was initially identified as inducing variable inhibition of pyoverdine across substrates using a previous iteration of iPau1131. Our experimental results showed limited production inhibition on all substrates but isoleucine; further examination of the model resulted in identification of missing *aru* operon genes linked to a pathway that utilized succinyl-CoA which we had not incorporated into the model. When this pathway was completed, our predictions of inhibition disappeared. In this way, an iterative approach of prediction and experimentation can be used to better capture the true metabolic capabilities of *P. aeruginosa* through our reconstructions.

DISCUSSION

In this study, we present a substantial update to the genome-scale metabolic reconstruction of *P. aeruginosa* PAO1 and contribute a new reconstruction of *P. aeruginosa* PA14, a strain increasing in use as a highly virulent model organism. We expand both reconstructions to account for more than 60 additional genes in iPae1148, and more than 40 additional genes in iPau1131 in comparison to previous *P. aeruginosa* reconstructions. The new models are reconciled with SEED database nomenclature to expand their comparative applications and improve consistency within the modeling field. Validating data sets regarding gene essentiality and substrate utilization were improved through both bioinformatics and experimental assays, and contributed to advances in prediction accuracy of the new reconstructions. Substantial expansion of pathway specificity and coverage resulted in an increase in the number of reactions included in the new models. Focused curation of hypothetical proteins and virulence-related genes enabled novel predictions of genes essential to virulence factor production and quantification of tradeoffs between synthesis of virulence factors and biomass. Genes that were predicted to be essential or important to synthesis of these products were evaluated experimentally on a range of CF-relevant substrates. Ultimately we identified novel therapeutic targets that could be used to prevent synthesis of virulence factors as an alternative or supplement to traditional approaches targeting growth of *P. aeruginosa*.

Targeting virulence-related metabolic pathways as a new avenue of therapeutic treatment is an approach that has been attracting more attention and investment from a field struggling to find effective ways to combat drug-resistant pathogens (D.G. Lee et al. 2006). Quorum sensing inhibitors have been investigated through high-throughput small molecule screening for a range of pathogens including *P. aeruginosa* (Fletcher et al. 2007; Bjarnsholt et al. 2010; Valentine et al. 2013 Aug 2) because of their importance to growth and cooperation of pathogens growing in biofilms. Other virulence factors such as rhamnolipids, phenazines, and siderophores have also been connected to phenotypes of biofilm growth, but have not been targeted at the same level of comprehensive evaluation (L. Yang et al.

2009; Martin et al. 2011; Mavrodi et al. 2013; Fazli et al. 2014). While these virulence factors are regulated by quorum sensing molecules, the involved signaling networks are complex; more direct routes of inhibition provide a valuable and efficient avenue for high confidence, controlled treatment. Our analysis in this study provides curated sets of hypothetical new targets for diminishing or preventing the production of a large array of virulence factors, which we also evaluate experimentally. Reducing experimental costs and time investment to identify targets while simultaneously providing a route to understanding of the underlying mechanisms by which our proposed targets enable inhibition are major contributions of our models to effective development of new therapies.

Our evaluation of virulence-linked genes from disparate studies in PAO1 and PA14 highlights the complexity of pathogenesis in the context of unique strains. The genes identified as linked to virulence were remarkably different between PAO1 and PA14 even in consideration of the different study approaches despite the majority of the genes being orthologous in both strains. This may indicate complex regulation of virulence-related pathways that provide adaptive flexibility and resilience to a strain adapting to varied stressful environments and cooperating or competing with other microbes in the same environment (D.G. Lee et al. 2006). The reconciled models enable overlay of 'omics' data that captures this varied regulation; future studies paired with modeling constraints based on such experimental data may elucidate the combinatorial nature of these genes in different strain backgrounds that enable novel routes of virulence.

The ability to compare predictions of *in silico* and *in vitro* essential genes using the models as a framework highlights the importance of careful, contextualized analysis of transposon insertion libraries. The results of the *P. aeruginosa* transposon library screens are currently the best experimental estimate of gene essentiality from high-throughput screening, but clear differences in final essential gene sets result from varying experimental approaches and analysis techniques. We have attempted to capture an accurate representation of gene essentiality within our models and have improved many metrics of prediction accuracy, but discrepancies in predictions

remain. However, these discrepancies can be used to guide future targeted *in vitro* evaluation of probable essential genes both in rich media and other more clinically relevant conditions.

Our investigation of potential inhibitors of virulence factor synthesis used a novel combinatorial analysis of gene deletions and tradeoffs between virulence factor and biomass production to reduce and rank potential inhibitor targets. The resulting tradeoff indices enable quantitative comparison of multifactorial inhibition that identify essential and important genes for product synthesis. PTA inhibition can be interpreted as an indicator of potential areas of metabolic prioritization in addition to identification of blocked synthesis pathways. Genes resulting in near 100% PTA inhibition while growth is only moderately impacted may strain the capacity of metabolic processes to the point that all resources are utilized for growth, preventing secondary pathways like virulence factor synthesis from functioning. These subtleties can be probed further for gene targets that show promising results when evaluated experimentally to identify mechanistic explanations of successful inhibition.

Treatment with virulence factor inhibitors may ultimately select for "nonvirulent" pathogens that stop wasting resources on attempted synthesis of these factors only to be thwarted in receiving any benefit from their spent metabolic costs while neighboring bacteria expend all of their energy on biomass production. Thus, we may be selecting for 'social cheaters' that do less damage to the host during infection and are less successful at surviving in the stressful host environment without the production of virulence-related shared goods (Dandekar et al. 2012). Interestingly, recent studies have indicated that traditional virulence factors produced by opportunistic pathogens such as *P. aeruginosa* may also serve important roles in bacterial survival other than promotion of virulence (Brown et al. 2012) This indicates that we could also be inhibiting unknown contributions to basic metabolic processes that may in fact also result in resistance to virulence inhibitors. Ultimately, these broader concerns require further examination of the intricacies of virulence versus growth inhibition; our models are an excellent platform for

pursuing fundamental understanding of these relationships from a systems perspective in the future.

Our analysis of genes essential or important to the synthesis of virulence factors is the most comprehensive genome-scale computational screen to date of virulence-related metabolism. We account for the interrelated nature of virulence factor synthesis and growth in our predictions, providing different classes of targets that may prove beneficial in unique therapeutic contexts. Concerns regarding treatment side-effects or resistance to growth-targeting antibiotics in the context of multi-drug treatments may benefit from incorporation of new therapeutics that only target virulence factor synthesis (Chait et al. 2007; Torella et al. 2010; Ejim et al. 2011). However, new proposals regarding careful design of sequences of single drug treatments in an attempt to avoid drug resistance may favor drugs that inhibit virulence factor production and growth simultaneously (Imamovic and Sommer 2013; Kim et al. 2014). Ultimately, our updated models are valuable tools for quantitatively assessing these relationships that would be challenging to interrogate experimentally at a genome-scale level. Our experimental validation of a subset of model-based predictions indicates that our approach provides testable hypotheses of gene function from a systems perspective that can be used to develop novel therapeutic virulence inhibitors to treat drug-resistant pathogens.

Chapter 5: Metabolic rewiring in an adapting pathogen during chronic infection in the human host

Acknowledgments: Juliane Thøgersen[2*], Jette Thykaer[2], Kristian F Nielsen[2], Helle K. Johansen[3], Søren Molin[2,4], Lars Jelsbak[2], Jason Papin[1]

[1]Biomedical Engineering, University of Virginia

[2]Department of Systems Biology, Technical University of Denmark

[3] Department of Clinical Microbiology, Copenhagen University Hospital

[4]Center for Biosustainability, Technical University of Denmark

The text of this chapter is currently under review at *mBio*:
Juliane C. Thøgersen[a*], Jennifer A. Bartell*, Jette Thykaer[a], Kristian F Nielsen[a], Helle K. Johansen[1], Jason A. Papin, Søren Molin[ac], and Lars Jelsbak[a]. Systems-based analysis of metabolic evolution during pathogen adaptation to the human host.
* Co-first authorship

SYNOPSIS

Successful bacterial pathogens must satisfy specific metabolic requirements to avoid eradication by the human host during chronic infections. Identification of metabolic pathways that change during the course of an infection provides novel targets for potential therapeutic intervention. We use long-term *Pseudomonas aeruginosa* infections in cystic fibrosis patients as a model system to study this evolutionary process. After experimentally determining activity in central metabolism of longitudinal isolates, we apply genome-scale metabolic models to contextualize our experimental findings from a systems perspective and elucidate systemic metabolic adaptations during chronic infection. We find strong evidence for a shift in metabolism towards fixation of carbon dioxide through reversal of the glycine cleavage system, which may operate as an alternative redox recycling reaction. Redox-related metabolic adaptation merits greater consideration as an important enabler of pathogen persistence and a potential therapeutic target in *Pseudomonas aeruginosa* and other emerging pathogens.

INTRODUCTION

Opportunistic pathogens change their metabolism in response to the conditions they encounter when they colonize their host. This metabolic reprogramming is facilitated through complex regulatory and metabolic networks encoded in the genomes of bacterial pathogens. Metabolic adaptation is necessary to capitalize on available nutrients for growth and survival and such shifts are essential for successful pathogenesis (Brown et al. 2008; Mitchell et al. 2009; Eisenreich et al. 2010). However, the underlying metabolic mechanisms that contribute to colonization and persistence are unclear in many bacteria and how these metabolic systems develop during pathogen adaptation remains unknown. The opportunistic pathogen *Pseudomonas aeruginosa* is an ideal model system for understanding these processes. It has principally evolved in its natural habitat outside the human host, where its specific regulatory and metabolic repertoire enables growth in soil and water environments. However, during chronic infections with *P. aeruginosa* in cystic fibrosis (CF) patients, some clinical *P. aeruginosa* strains have developed into host-specific organisms by adaptive mutations that enhance survival in the human lung environment (Yang et al. 2011; Folkesson et al. 2012). How these particular *P. aeruginosa* strains persist to become the dominant, chronic pathogens in the lung in contrast to other initial infecting species is poorly understood. Improving our understanding of microbial adaptation to the human host will have important implications in treatment of infectious disease, development of probiotic therapies, and other applications.

Compared to the natural environment of *P. aeruginosa*, the host environment is characterized by a complex and novel combination of stressors that could be mitigated by various adaptive strategies. In the CF lung, most of the bacterial population grows within CF sputum, which is rich in nutrient sources and has a low oxygen tension (Ohman and Chakrabarty 1982; Worlitzsch et al. 2002; Palmer et al. 2005). Patient airways have elevated numbers of polymorphonuclear neutrophilic leukocytes (PMNs), alveolar macrophages and antibodies; phagocytosis of bacteria by PMNs promotes the generation of reactive oxygen species (ROS) by host cells (Hoiby 2006). In addition to the host immune response, *P. aeruginosa* is

exposed to a range of antibiotics during the course of infection in CF patients and resistance towards antibiotics is a common feature observed for adapted strains (Govan and Nelson 1993; Burns et al. 1998; Döring et al. 2000).

We know from previous studies that adaptation of *P. aeruginosa* to the CF lung environment involves many gene regulatory mutations that affect metabolism as well as mutations in metabolic enzymes (Yang et al. 2011), but connecting these underlying mechanisms to their global impacts on pathogen behaviour is difficult. The aim of this study is to identify novel metabolic systems that contribute to successful adaptation of *P. aeruginosa* in the host. Specific metabolic pathways that are undergoing changes in activity during the course of adaptation may be essential for the bacteria in order to persist in the lungs of the patients and could serve as targets for future antibiotics.

We identify these pathways of interest using a systems-level computational and experimental approach to characterize and compare the metabolic activity of clinical bacterial isolates. By integrating and contextualizing our multi-scale experimental data using a genome-scale computational metabolic model, we can streamline the prediction and comparison of early and late stage isolate phenotypes to connect shifts in the activity of a single enzyme to systems-level changes in metabolism. The results have broad implications in understanding mechanisms underlying pathogen adaptation in chronic infections and microbial evolution under selective pressures.

MATERIALS AND METHODS

We have chosen three monoclonal longitudinal isolates of *P. aeruginosa* to represent adaptation during cystic fibrosis infection. Full genome sequences are available for all strains (Marvig et al. 2013). Our experimental approach includes isotope-labelling experiments paired with growth profiling and characterization of metabolite excretion. In addition, we perform transcriptome profiling to support genome-wide characterization of metabolism under our specific experimental growth conditions. We apply genome-scale metabolic models as an analytical tool

instead of analysing the omics data sets individually. We integrate SNP data and gene expression data into the most recent validated genome scale metabolic model of *P. aeruginosa*, iPA1139. The changes between the derived models of the three strains are therefore based on information stored in the genomes or transcriptomes of the strains. A comparative study of these individual models can identify potential metabolic pathways that have changed during the course of adaptation. We want to emphasize that our derived models are based on experimental data and confidence in these models and their derived predictions can partly be gained by comparison to the original data sets. We are able to trace specific constraints that are responsible for the predicted changes, by inverting the SNP or gene expression constraints applied to the models. The advantage of using this systemic approach of integrating multiple data sets is that we may discover patterns in metabolism that we are not able to extract from the individual data sets - e. g. it is possible that we do not see any genetic variation or differential gene expression in an otherwise important metabolic subsystem, which is affected by genetic or regulatory changes in adjacent metabolic pathways and only identified through our systemic characterization.

***Pseudomonas aeruginosa* strains used in this study**

We selected three isolates of the DK2 clone type for our analysis. Two of them, DK2-91 and DK2-07, are late-stage clinical isolates isolated from the same patient in 1991 and 2007, respectively (DK2-91 and DK2-07 are referred to as CF333-1991 and CF333-2007 respectively in (Jelsbak et al. 2007)). The third isolate, DK2-WT (referred to as CF510-2006 in (Rau et al. 2012 Jun 5)) also shares the DK2 clone type, but has a phenotype similar to strains isolated from outside the CF lung (*P. aeruginosa* PAO1 (Holloway et al. 1979) and *P. aeruginosa* PA14 (Rahme et al. 1995)) and its genotype is similar to the predicted most common recent ancestor for DK2 dated back to 1970 (Rau et al. 2012 Jun 5). DK2-WT therefore resembles a non-adapted isolate of DK2 and this isolate serves as our point of reference for the DK2 lineage. Other early isolates of the DK2-lineage collected in the early 1970s exist. However, we chose DK2-WT as our reference for the DK2 lineage since its phylogenetic branching from the most common recent ancestor is distinct from the

adaptation path of DK2-91 and DK2-07 in contrast to the other early isolates. We therefore expect to capture most adaptive events in DK2-91 and DK2-07 by comparing to DK2-WT. *P. aeruginosa* PAO1 (PAO1) is also included as reference throughout most of our experimental and *in silico* analyses. PAO1 was originally isolated from a burn wound (Holloway et al. 1979) and has been widely used as a reference strain for studies of *P. aeruginosa*.

Cell storage and cultivation

Cells were stored at -80C in a 20% glycerol solution. DK2-91 and DK2-07 were streaked on a Luria-Bertani (LB) agar plate and incubated at 37C for 48 hours. Individual colonies were inoculated in 10 mL of morpholinepropanesulfonic acid (MOPS)-buffered medium supplemented with glucose and grown aerobically at 37C for 24-36 hours (depending on growth rate). The total composition of the MOPS minimal medium was 40 mM MOPS, 9.5 mM NH_4Cl, 0.28 mM K_2SO_4, 1.3 mM KH_2PO_4, 10 mM glucose and vitamins (0.4 µM biotin, 10 µM pyroxidal-HCl, 2.3 µM folic acid, 2.6 µM riboflavin, 8 µM niacinamide, 3 µM thiamine-HCl and 2 µM pantotheneate) (Jensen and Hammer 1993).

DK2-WT and PA01 were streaked on LB agar plates and incubated at 37C for 24 hours. Individual colonies were inoculated in 10 mL of MOPS minimal medium supplemented with glucose and grown aerobically at 37C for 16 hours. After initial incubation cells were transferred to a 250 mL baffled flask with 50 mL MOPS minimal medium supplemented with 10 mM of defined carbon source to an optical density (OD_{600}) of 0.01 measured at 600 nm.

Cell growth was determined by measuring OD_{600} during growth. Cells were harvested for GC-MS and DNA microarray analyses at OD_{600} = 0.4 during the exponential growth phase. Supernatant was collected for a glucose determination assay during growth in order to make biomass yield calculations.

Biomass yield calculations

Glucose concentrations were determined enzymatically using a glucose reagent (catalogue no. 7200-017A, from Thermo Electron, Australia). The dry weight biomass concentration was estimated using a correlation factor of 0.360 g cellular dry weight per OD unit. This correlation factor was determined for an *Escherichia coli* strain (Kiviharju et al. 2007) and is assumed to be valid for *P. aeruginosa*. The biomass yield on glucose was determined using the concentration data for biomass and glucose, respectively.

Labelling experiments

The experimental protocol for labelling determination was modified from (Christensen and Nielsen 1999). Cells were grown in MOPS minimal medium to an OD_{600} of 0.4. 10 mM [1-^{13}C]-glucose (D-glucose-^{13}C, 99% ^{13}C, from Isotec, Miamisburg, Ohio, USA, CAS no. 297046) was used as a carbon source. For some experiments, a mixture of 44 mol-% [1-^{13}C]-glucose and 56 mol-% $^{13}C_6$ glucose (D-glucose-$^{13}C_6$, 99 % ^{13}C, from Isotec, Miamisburg, Ohio, USA, CAS no. 110187-42-3) were used to give a final glucose concentration of 10 mM.

30 ml culture was harvested and the samples were spun down for 10 minutes at 5,000 rpm at 4C. The pellet was resuspended in 2 mL 0.9 % NaCl and the volume was divided into two Eppendorf tubes. The Eppendorf tubes were further spun down for two minutes at 10,000 rpm at 4C and the pellets were finally stored at -80C until hydrolysis and subsequent derivatization and amino acid analysis by GC-MS. The supernatant was stored in 4 individual Eppendorf tubes of 1 mL at -80C for later GC-MS analysis of extracellular metabolites.

GC-MS analysis for extracellular metabolites. The experimental procedure of labelling determination was modified from Kind et al. 2009 (Kind et al. 2009). Supernatants were centrifuged at 15000 g and 100µL supernatant lyophilized in 2-ml silanized glass vials, and then derivatised by 20 µL O-methylhydroxylamine in pyridine for 2 hrs, before adding 180 µL of N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (Thermo Fischer) heating in an oven at 40°C for 30 min. The samples were analysed by GC-MS on a

Thermo Electron DSQII GCMS systems using the same parameters as described in 46 for their Agilent GC-MS, and peaks matched in the Fiehn Lib (Agilent technologies) using the AMDIS 2.71 (http://chemdata.nist.gov/mass-spc/amdis/downloads/). Reference standards of glucose, gluconate, 5-ketogluconate and 2-ketogluconate were co-analysed in the sequence with real samples for verification.

Proteinogenic amino acid analysis from 13C-labeled biomass. Hydrolysis: The pellet was resuspended in 600 µL of 6 M hydrochloric acid and the volume was transferred to a 2 mL glass vial. The vial was capped with an aluminium cap (able to withstand high temperatures) and kept at 105°C for hydrolysis overnight. After overnight hydrolysis the content of the vial was transferred to an Eppendorf tube and centrifuged at 15000 rpm for two minutes. Supernatant was transferred to two clean glass vials (280 µL each). The vials were dried for three hours at 105°C without caps. After drying one of the vials was capped and stored at -80°C for backup. The other vial was added 200 µL of milliQ water and vortexed for 30 seconds. Another 800 µL of milliQ water was added followed by vortexing. A control sample containing bovine serum albumin (BSA) was included to test if the hydrolysis step was completed successfully.

Purification: The biomass hydrolysate was loaded on a cat-ion exchange solid phase extraction column packed in a 1–ml syringe (200 mg Dowex 50W X8, 200- 400 mesh, H+-form, Sigma-Aldrich, St. Louis, MO), that had been conditioned by 1 ml methanol and 1 ml water, and the sample was passed through by gravity. Waste was discarded. The sample was washed with 1 mL of ethanol in water (1:1). 0.2 mL of 1 M NaOH was added to increase the pH of the column and waste was discarded. A 2 mL glass vial was placed under the column to collect the purified amino acids. 1 mL of a mixture of 1% (wt/v) NaOH in saline, ethanol and pyridine in a 9:5:1 proportion was added and the eluate was collected. The content was divided into two parts, 500 µL in an Eppendorf tube for ethylchloroformate (ECF) derivatization and 500 µL in a glass vial for N-dimethyl-amino-methylene-methyl-esters (DMFDMA) derivatization respectively. The samples were kept at -20°C until

derivatization. A control sample containing a mixture of amino acids was included to test if the purification step was completed successfully.

ECF Derivatization: 50 µL of ethylchloroformate was added to the 500 µL SPE column eluate. Pipetting in and out using a 1 mL pipette followed by a gentle vortexing gently mixed the content. The Eppendorf tube was uncapped to release the pressure. This step was repeated until no CO_2 was observed. 5 additional µL of ECF was added followed by vortexing and release of pressure. 200 µL of propyl acetate was added, the tube was vortexed for 30 seconds and pressure was released. 50 µL of 1 M HCl was added followed by vortexing and release of pressure. The fluid was allowed to separate for 1 minute. Thereafter 175 µL of the upper organic layer was transferred to a new Eppendorf tube. A small amount of anhydrous $NaSO_4$ or $MgSO_4$ was added followed by vortexing. The supernatant was transferred to a 2 mL glass vial and kept at -20°C until GC-MS analysis.

DMFDMA Derivatization: 200 µL 1 M HCl was added to the 500 µL SPE column eluate and mixed well. The vial was kept for drying for 2 to 4 hours at 105°C without cap. The vial was allowed to cool down for ten minutes. 200 µL DMFDMA and 200 µL acetonitrile was added to the vial. The vial was capped with a screw cap and kept for derivatization at 100°C for 20 minutes. After derivatization the vial was placed at -20°C for 10 min. The supernatant was transferred to an Eppendorf tube and centrifuged at 15.000 rpm for 2 min. The supernatant was transferred to a new glass vial with a screw cap and kept at -20°C until GC-MS.

GC-MS analysis of proteinogenic amino acids. Samples were analysed by GC-MS on an Agilent 6890 gas chromatograph (Agilent Technologies, Waldbronn, Germany) coupled to an Agilent 5973 quadruple MS run in electron impact ionization (EI+) mode using an electron energy of 70 eV. The GC was equipped with a 4.0 mm i.d. Siltek gooseneck splitless deactivated liner (Restek, Bellefonte, PA, USA), and a Supelco (Bellefonte, PA, .US) Equity®-1701 (15 m, 0.25 mm i.d., 0.25 µm film) column. Helium was used as carrier gas at a constant linear gas velocity of 38 cm/s. Transfer line temperature was 280ºC, quadruple temperature 150 ºC and MS

source 230 ºC. The GC-MS system was controlled from Agilent MSD Chemstation v. D.01.02.16, and auto tuned for prior to every sequence. Samples of 1 µL was injected using a Combi PAL autosampler (CTC Analytics AG, Zwingen, Switzerland). Analysis of amino acid-ECF derivatives was done at an injection temperature of 220ºC, and oven temperature was initially held at 75 °C for one min. Hereafter the temperature was raised 40 °C min^{-1} until 165 °C, then 4 °C min^{-1} until 190 °C and then 40 °C min^{-1} to 240 °C. At the end, temperature was increased to 260 °C at 4 °C min^{-1} and held constant for 4 minutes. Analysis of the amino acid-DMFDMA derivatives was done at an injection temperature of 230ºC, and oven temperature was initially held at 60 °C for one min. Hereafter the temperature was raised at 20 °C min^{-1} until 130 °C, then 4 °C min^{-1} until 150 °C and 40 °C min^{-1} to 260 °C and held constant for 4.25 minutes.

Testing for CO2 incorporation into glycine. PAO1 and DK2-91 were grown in MOPS minimal medium supplemented with 10 mM unlabelled glucose. At OD_{600}=0.01 20 mM of $NaH^{13}CO_3$ was added. Cells were harvested at OD_{600}=0.2 and OD_{600}=0.4 and labelling patterns of amino acids were determined as described above. We chose PAO1 instead of DK2-WT to find out if the potential of carbon fixation into glycine is general for *P. aeruginosa* or just a feature of the DK2 lineage. We chose DK2-91 to represent the late-stage clinical isolates, since the growth rate of DK2-91 was higher than that for DK2-07 (slow growth of the cells would allow more bicarbonate to vaporize before cell harvest). We used a high concentration of bicarbonate (20 mM) to make sure that some bicarbonate would remain in the medium at the time of harvesting despite dilution with unlabelled bicarbonate/carbon dioxide and vaporization. Ideally, the experiment would be carried out with concentrations of bicarbonate and carbon dioxide corresponding to the initial experiments. However, the labelled bicarbonate and carbon dioxide would be diluted out from unlabelled carbon dioxide produced under glycolysis in the growing culture. Therefore this experimental setup only addresses the question whether carbon dioxide is fixated in glycine synthesis and the results cannot be used quantitatively.

DNA microarray analysis

Cells were grown in MOPS minimal medium supplemented with 10 mM glucose to an OD600 of 0.4 prior to Affymetrix *P. aeruginosa* GeneChip microarray analysis. Microarray data were generated using Affymetrix protocols as previously described (Yang et al. 2011). Data processing was carried out according to Thøgersen et al, 2013 (Thøgersen et al. 2013). The raw cel-files were extracted in R by use of the package *affy* (Gautier et al. 2004) followed by qspline normalization (Workman et al. 2002) and calculation of gene expression index values using *robust multiarray average* expression measure (Irizarry et al. 2003). Differentially expressed genes for DK2-91 and DK2-07 compared to DK2-WT were determined with Bonferroni adjusted p-values (significance level p=0.05) using the R package "*limma*" (Smyth 2005). Enriched gene ontology classes among differentially expressed genes were identified by the Hypergeometric distribution test with significance level p = 0.01.

Sorting Intolerant from Tolerant (SIFT) Analysis of SNPs

The SNP data were obtained from previous studies of the DK2-WT, DK2-91 and DK2-07 strains (Yang et al. 2011; Rau et al. 2012 Jun 5). A list of strain specific SNPs is available and includes our SIFT (Sorting Intolerant From Tolerant (Kumar et al. 2009)) analysis of missense mutations in metabolic genes. The SIFT analysis is used to predict if a missense mutation would affect protein function of the given gene product, providing numeric scores that indicate the degree to which a missense SNP is tolerated or affects protein function.

Isolate-specific genome-scale metabolic models

An earlier iteration of the new genome scale metabolic reconstruction for *P. aeruginosa* PAO1 described in the previous chapter, iPA1139, was used as the base for all computational modelling in this study. This model accounts for the function of 1139 genes, 1491 reactions, and 1280 metabolites involved in the metabolism of *P. aeruginosa*. Isolate-specific genome-scale metabolic models were created by a semi-automated approach in order to incorporate both single nucleotide polymorphisms (SNPs) and gene expression-based constraints using the TIGER Toolbox 1.2.0 (Zur et al. 2010; Jensen et al. 2011).

Raw expression levels were used to develop proposed 'off' and 'on' gene activity levels using 25th and 75th percentile cutoffs of the expression data similar to methods described by Machado and Herrgard (2014) (Machado and Herrgård 2014) These gene levels were converted into tri-valued logic levels ('off' – 0, 'unconstrained' – 1, and 'on' – 2) as the input for the TIGER implementation of iMAT. Different levels of SNP constraints were also used, ranging from minor impact (silent and SIFT-predicted tolerated missense SNPs), moderate impact (missense SNPs with SIFT-predicted functional impact), and maximum impact (nonsense SNPs). In order to integrate these datasets before iMAT was used to create strain-specific models, any Boolean gene-to-protein-to-reaction (GPR) relationship that incorporated a gene associated with a SNP was manually evaluated in the context of the gene expression levels. If only the SNP-affected gene was associated with the reaction, the activity of the connected reaction was limited by modifying the reaction bounds. If the GPR was a more complex Boolean statement involving multiple genes (gene duplications, isozymes, or multiple subunits of an enzyme), the GPR was evaluated to see whether any genes were present that could compensate for the affected function of the SNP-associated gene. If these compensatory genes also had 'off' expression levels, the SNP-based constraint was applied. If the compensatory genes were unconstrained or 'on', the SNP-based constraint was not applied. Instead of reducing the potential activity of SNP-targeted reactions by their base model bounds (usually -1000 to 1000 for reversible reactions and 0 to 1000 for irreversible reactions), we conducted flux variability analysis of the base model at 100% biomass production to calculate the normal range of activity of each reaction in glucose minimal medium conditions. Any minor impact SNP being implemented resulted in a 10% reduction of the FVA-predicted base activity range enforced via reaction bounds while a moderate impact SNP resulted in a 50% reduction applied in the same manner. SNPs implemented with maximum impact resulted in associated reactions being turned entirely off via modification of reaction bounds.

The above SNP integration method was applied to each strain-specific model prior to the use of the TIGER implementation of iMAT. Using an objective function threshold of 10% of the maximum and Gurobi 5.6.2 as the solver, iMAT predicted new sets of genes that could feasibly be turned 'off' or 'on' while maintaining production of biomass at 10% in each SNP-constrained isolate-specific model. The 'off' genes were inactivated in the model. The predicted 'on' genes were implemented by applying a lower bound constraint of 0.001 or -0.001 to ensure a minimum level of activity in the appropriate direction of reaction activity. Reaction direction was evaluated via FVA, and if there was not a clear preference for direction of activity (for example, the FVA max and min indicated the reaction was fully reversible (a range of -0.001 to 0.001 or larger)), then the 'on' minimum constraint was not applied to avoid inappropriate/unsupported bias in reaction directionality. This evaluation meant that it was not feasible to apply all constraints predicted by iMAT, and a summary of the gene constraints and the difference between predicted and applied isolate-specific model constraints is presented in Figure 5.1. The resulting isolate specific models, base model iPA1139, and SBML versions of the isolate-specific models are available at http://bme.virginia.edu/csbl/downloads-pseudomonas-v3.php.

		DK2-WT	DK2-91	DK2-07
		constrained genes		
Gene Exp.	*off*	323	287	239
	on	324	271	260
SNPs	*off*	0	2	2
	50%	3	5	9
	90%	11	8	14
		TIGER-predicted constraints		
iMAT	*off*	299	273	221
	on	174	153	147
		constrained model reactions		
combined integration:		364	296	271

FIG 5.1. Converting 'omics' data to isolate-specific model constraints. Counts of genes and SNPs binned into their respective functional categories (shown in top two tables) are manually evaluated for combined expression-SNP functional impact and then provided to the TIGER implementation of iMAT for constraint development. Resulting iMAT predictions of 'off' genes that should be inactivated and 'on' genes that should result in associated reactions carrying at least a minimum level of flux during growth are shown. After enforcing the requirement that all models must be able to produce biomass (grow), the resulting number of reactions with constrained flux activity in each isolate-specific model is presented in the last table.

The isolate specific models were then used to evaluate metabolic activity using several methods of constraint-based modelling. Flux balance analysis (FBA) was used to predict the ability of each isolate model to grow (produce biomass) in 'wild type' conditions as well as with single genes deleted to identify *in silico* genes essential for growth. Flux variability analysis (FVA) was used to predict changes in potential reaction activity by calculating the minimum and maximum flux of a given reaction when the model was required to produce maximum biomass. We calculated the flux range from the maximum and minimum flux values for each reaction, and then determined whether the range increased or decreased compared to unconstrained iPA1139, sorting results into 5 categories: decreased range in mDK2-07 compared to mDK2-WT, increased range in mDK2-07 compared to mDK2-WT, or comparable changes in the ranges in both strains that are increased, the same, or decreased compared to iPA1139. To identify subsystems with high concentrations of changes in activity that we interpret here as potential adaptive processes, we counted the number of reactions in these altered activity categories within each subsystem and then normalized by the number of active subsystem reactions in iPA1139.

The redox cofactor production analysis presented in Figure 5.3 was performed by optimizing for the maximum production capacity of NAD+ and NADH separately while constraining the maximum uptake rates of O2 and CO2 to 0.2, 2, 5, and 20 mmol/gDW/hr (low to high uptake rates) and fixing the production rate of biomass at 0 to 100% of optimum production when only growth is maximized. The command-line implementation of Metdraw (also available at www.metdraw.com) was used to build a full-sized map of the base model on which FVA results were overlaid automatically (Jensen and Papin 2014).

Statistical Analysis

All experimental data presented are based on biological triplicates. Significant differences in data are evaluated using two-sided Student's t-test with significance level p = 0.05 unless otherwise stated. For DNA microarray data the statistical

analysis includes corrections for multiple testing (see above under DNA microarray analysis).

RESULTS

Previously, we studied genomic evolution in an epidemic *P. aeruginosa* clone type (DK2) during its dissemination across multiple patients over a 40 year time period (Yang et al. 2011; Rau et al. 2012 Jun 5). The DK2 clone has been successfully transmitted between patients and replaced previously colonizing *P. aeruginosa* clone types (Jelsbak et al. 2007). Thus, DK2 is highly adapted to the CF airway environment, which likely includes optimization of its metabolic activity for growth within the CF lung. Here, we use DK2 patient isolates to study metabolic adaptation, focusing on DK2-WT (which represents the ancestral genotype at the time of first colonization of the CF niche), and two isolates collected at later stages of clone evolution (DK2-91 and DK2-07) representing host-adapted isolates. We also included the well-studied reference strain *P. aeruginosa* PAO1 (PAO1) (see Materials and Methods for detailed description of the strains).

Major changes in central metabolism occur during adaptation

Growth experiments with DK2-WT, DK2-91 and DK2-07 in glucose minimal medium showed a significant reduction in growth rate for DK2-91 (μmax = 0.46 h-1) and DK2-07 (μmax = 0.23 h-1) compared to DK2-WT (μmax = 0.87 h-1). The growth rate of DK2-WT was higher but similar to the growth rate of PAO1 (μmax = 0.63 h-1). Closer inspection of the growth curves revealed diauxic growth curves for DK2-91 and DK2-07, which were not observed for PAO1 and DK2-WT. This observation led us to hypothesize that DK2-91 and DK2-07 excreted one or more metabolites that were later degraded and metabolized after glucose depletion. We then measured extracellular metabolites via GC-MS analysis, revealing that the oxidized glucose derivatives gluconate and 2-ketogluconate were accumulating in the medium for DK2-91 and DK2-07. Gluconate and 2-ketogluconate were not detected for the reference strain PAO1 and only a small amount of gluconate was detected for DK2-WT. These results indicate activity in the oxidative route of glucose degradation via gluconate and 2-ketogluconate for DK2-91 and DK2-07 as an

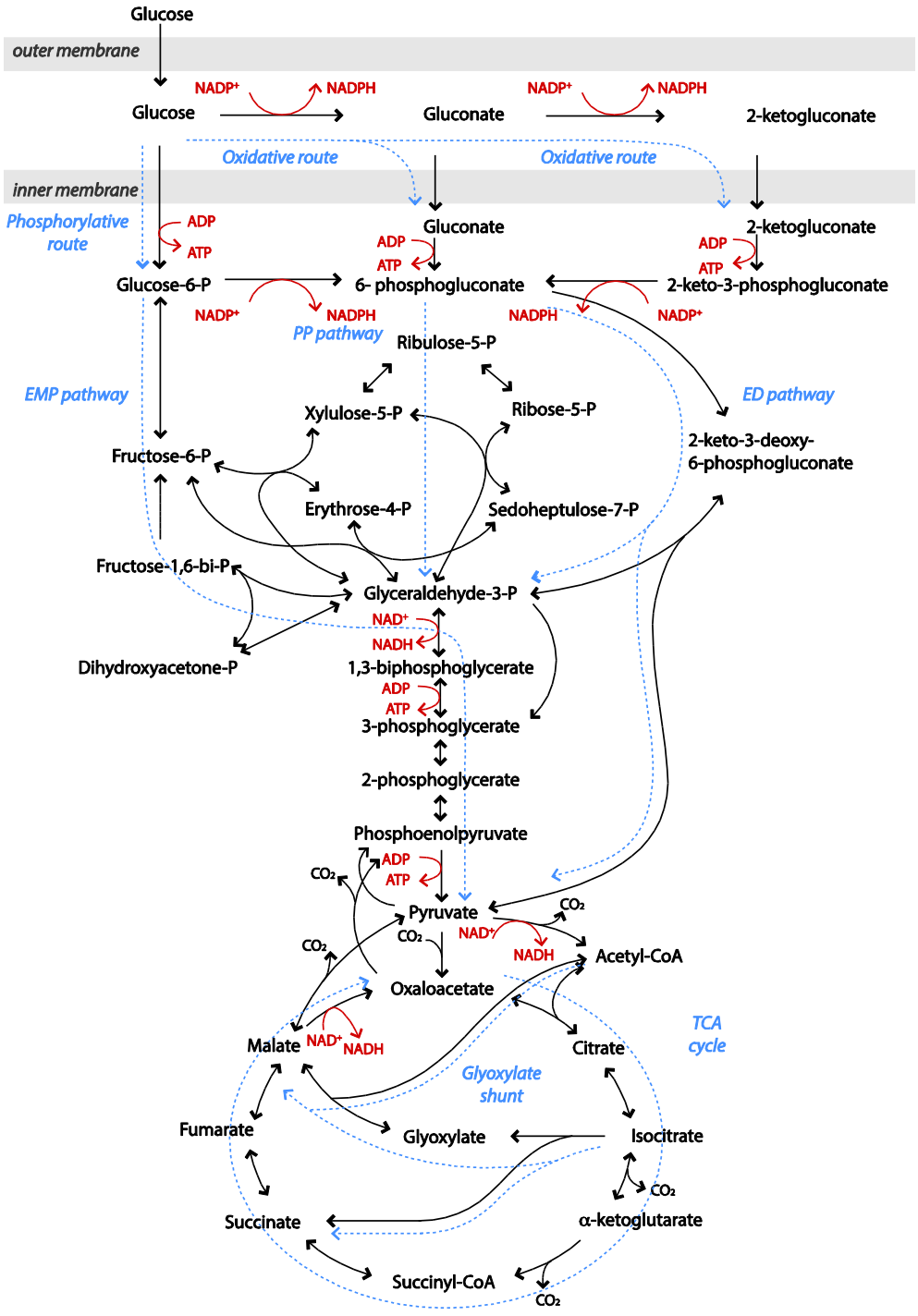


FIG 5.2. Glucose metabolism in *Pseudomonas aeruginosa*. Glucose can enter the cell through the phosphorylative or the oxidative route. The oxidative route involves conversion of glucose into gluconate or 2-ketogluconate. In the cytoplasm, further degradation of pyruvate may occur through three alternate pathways. The blue arrows indicate alternative convergent pathways and their respective names17–19,57–60. Abbreviations: ED (Entner-Doudoroff), PP (pentose phosphate), EMP (Embden-Meyerhof-Parnas), TCA (Tricarboxylic acid), P (phosphate).

alternative to the phosphorylative route where glucose is phosphorylated to glucose-6-phosphate (Fig. 5.2).

Pseudomonas shifts its metabolism towards fixation of carbon dioxide

To further investigate the central metabolism of *P. aeruginosa*, we performed substrate-labelling experiments. We used a mixture of [1-^{13}C]-labelled glucose and [$^{13}C_6$]-labelled glucose. Using uniformly [$^{13}C_6$]-labelled glucose has been referred to as reciprocal labelling and it is particularly useful for investigating catabolism of co-substrates (Christensen and Nielsen 2002). By increasing the background labelling of position 2-5 of glucose, incorporation of an unlabelled carbon source (e.g. carbon dioxide) could be detected. The [1-^{13}C]-labelled glucose can be used to track activities of different convergent pathways since the labelled C-atom will end up at different positions in the carbon skeleton of metabolic intermediates depending on which pathway is used to degrade glucose. This method cannot be used to differentiate between the phosphorylative and oxidative route of glucose degradation to 6-phosphogluconate (Fig. 5.2) since the resulting carbon skeleton of 6-phosphogluconate is the same regardless of the two alternative routes. However, the method makes it possible to distinguish between the three glycolytic pathways: the Embden-Meyerhof Parnas (EMP) pathway, the pentose phosphate (PP) pathway and the Entner-Doudoroff (ED) pathway (Gunnarsson et al. 2004). Different labelling patterns of pyruvate occur depending on which pathway is used to catabolize glucose. By inspecting the labelling patterns of amino acids derived from pyruvate, we found that the labelling degree of the carbon atom at position 1 in pyruvate was around 50% for all strains grown in 100% [1-^{13}C]-glucose indicating that most if not all glucose was degraded through the ED pathway. It is well known that the EMP pathway is inactive in *Pseudomonas* species due to a missing enzyme and the PP pathway has previously been found only to serve biosynthetic purposes for other *Pseudomonas* species including *P. putida* and *P. fluorescens* (Fuhrer et al. 2005; del Castillo et al. 2007; Chavarría et al. 2013).

When we further examined the labelling patterns of amino acids derived from central metabolites from the combined [1-^{13}C]-glucose and $^{13}C_6$-glucose experiment, we found that glycine had a significantly lower labelling degree than

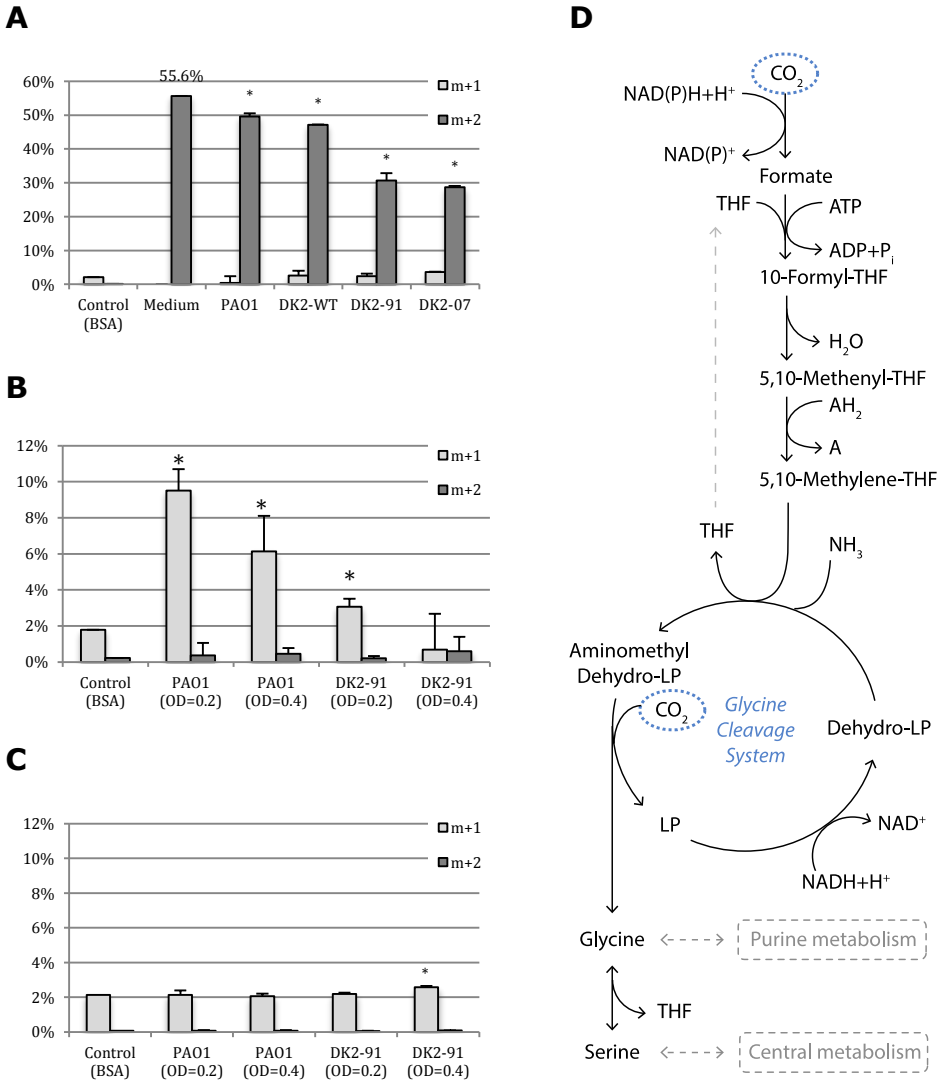


FIG 5.3. Carbon dioxide fixation into glycine through the glycine cleavage system. (A) Labelling of glycine derived from cultivation in glucose minimal medium (56% $^{13}C_6$-glucose and 44% [1-^{13}C]-glucose). The (m+1)-columns indicate the percentages of compounds with one labelled C-atom, whereas the (m+2)-columns indicate the percentages of compound with both carbon atoms labelled. The control is a measure of the naturally occurring ^{13}C-isotope in bovine serum albumin (BSA). The amount of background labelling from $^{13}C_6$-glucose (56%) is indicated as a separate column. In (A), (*) denotes where the labelling percentages of (m+2)-labelling are significantly lower (Student's *t*-test, two-sided, significance level, $p = 0.05$) than the background level from labelled glucose in the medium. **(B)** Labelling of glycine derived from cultivation in unlabelled glucose and labelled bicarbonate ($H^{13}CO_3^-$). In (B-C), (*) denotes where the percentages of (m+1)-labelling of strains are significantly higher (Student's *t*-test, significance level, two-sided, $p = 0.05$) than the level of the naturally occurring isotope (control). Note that panels (A-C) have different scales.

FIG 5.3 Cont'd (C) Labelling of serine derived from cultivation in unlabelled glucose and labelled bicarbonate ($H^{13}CO_3$). **(D)** The glycine cleavage system in reverse. Two molecules of carbon dioxide are fixated into glycine - one of them via formate formation. Figure adapted from (Bar-Even et al. 2012). Abbreviations: Reduced electron donor (AH_2), oxidized electron donor (A), tetrahydrofolate (THF), lipoyl protein (LP).

the labelled carbon substrate the cells were growing on for all strains (Fig. 5.3A). This observation was most noteworthy for DK2-91 and DK2-07. The minimal medium contained 56% $^{13}C_6$-glucose and 44% [1-^{13}C]-glucose and we would therefore expect the average labelling degree for each carbon atom to be 56% at minimum. Surprisingly, the data showed that the carbon atoms in glycine had an average labelling degree of approximately 30% for DK2-91 and DK2-07 compared to approximately 50% for PAO1 and DK2-WT; all were significantly lower than 56%. Since the labelled substrate was the only carbon source available for the cells in the experiment, we hypothesized that the cells have the capacity to fix carbon through glycine metabolism.

A literature search identified instances of non-canonical reversal of the glycine cleavage system (GCS) in *Clostridia* species (Bar-Even et al. 2012), using CO2 as a carbon source for the synthesis of glycine. To test this, we added ^{13}C-labeled bicarbonate into a growing culture in minimal medium with unlabelled glucose. Since the bicarbonate was the only source of the ^{13}C isotope (except for 1.1% natural prevalence), any excess labelling on glycine would indicate CO2 fixation. Bicarbonate was added during exponential growth and DK2-91 and PAO1 cells were harvested after one and two generation times. The results were a qualitative measure of the ability of the cells to fix CO2 (see Materials and Methods). Fig. 5.3B shows the labelling patterns of glycine for PAO1 and DK2-91 from the ^{13}C-bicarbonate experiment. We find a significant enrichment of ^{13}C in glycine for both PAO1 and DK2-91 one generation time after ^{13}C-bicarbonate addition and for PAO1 the same observation was made after two generation times. Based on these results we confirmed that CO2 could be fixed into glycine when *P. aeruginosa* is growing in glucose minimal medium. No significant enrichment of ^{13}C-isotope was measured after two generation times in DK2-91, but since the generation time of DK2-91 is 1.4 fold longer than PAO1, dilution and vaporization of ^{13}C-bicarbonate during the

course of the experiment can account for this difference. We included the labelling patterns of serine in Fig. 5.2C since serine can be produced from glycine. We find a significant enrichment of the ^{13}C-isotope for DK2-91 harvested two generation times after ^{13}C-bicarbonate addition. The lack of ^{13}C-labeling in serine for the other samples confirms that carbon dioxide is fixed directly into glycine and not into upstream metabolic intermediates in glycolysis, since otherwise we would expect at least the same degree of labelling in serine as for glycine. In conclusion, we find that the *Pseudomonas* strains are incorporating carbon dioxide into glycine and under normal laboratory conditions with normal carbon dioxide pressure in glucose minimal medium (Fig. 5.3A), this observation is more pronounced in DK2-91 and DK2-07 compared to DK2-WT and PAO1. Our experimental analysis of central metabolism therefore resulted in two specific findings related to metabolic adaptation in the late stage isolates: (1) metabolism is shifted towards excretion of gluconate and 2-ketogluconate and (2) the activity of the glycine cleavage system is altered to enable fixation of carbon dioxide in a non-canonical reversal of associated pathways.

The glycine cleavage system may operate as an alternative redox recycling reaction

The ability of *P. aeruginosa* to fix carbon dioxide into glycine has not been reported previously, and alterations in glycine synthesis indicated by our labelling experiments support our hypothesis that carbon fixation is occurring through the glycine cleavage system (Fig. 5.3D) (Bar-Even et al. 2012). A potential selective advantage for the activity of this altered glycine synthesis route may be linked to the regeneration of NAD+ from NADH coupled to this reaction. We propose that this unconventional pathway phenotype operates as an electron sink for the recycling of reduced electron carriers, alleviating redox stress as also suggested for some anaerobic bacteria (Bar-Even et al. 2012). Different factors in the lung environment may contribute to redox stress in *P. aeruginosa* including oxidative stress from immune system defenses and low availability of electron acceptors. The relative contribution of antibiotic exposure to increased oxidative stress in bacteria is currently under debate (Kohanski et al. 2010; Liu and Imlay 2013), but most

recently, Dwyer et al. (Dwyer et al. 2014) provided evidence for antibiotic-induced redox alterations in E. coli. We cannot specify whether the source of redox stress is limited oxygen, antibiotic exposure or the immune defense within the CF lung; it is possibly a combination of all three factors. However, we hypothesize that the impact of these stressors is substantial, driving the enhanced carbon fixation into glycine for the late-stage clinical isolates and therefore improving the balance of redox equivalents.

Metabolic models evaluate the feasibility of adapted redox metabolism

In light of our experimental observations, we focused our studies on the global effects of our proposed isolate-specific phenotypes of glycine metabolism via a well-curated and recently updated genome-scale metabolic model of *P. aeruginosa*, iPA1139 (http://bme.virginia.edu/csbl/downloads-pseudomonas-v3.php). This approach allowed us to systematically evaluate our observed phenotypes in context with model-integrated transcriptomics and sequencing data. Our experimental examination of glycine metabolism supports non-canonical activity for two connected routes of carbon fixation: glycine dehydrogenase and formate dehydrogenase (as shown in Fig. 5.3D). To create isolate-specific models using iPA1139, we first altered the possible activity of the glycine cleavage system and formate dehydrogenase (both canonically modelled in the forward direction); we allowed these reactions to run only in the reverse direction in our *in silico* models of DK2-91 and DK2-07 (mDK2-91 and mDK2-07) while they were modelled as reversible in our *in silico* model of DK2-WT (mDK2-WT) and the base model (iPA1139) during our data integration process. This enabled us to evaluate the feasibility and systemic impacts of the novel carbon fixation phenotypes indicated by our labelling experiments.

We additionally constrained the models to replicate isolate-specific phenotypes in our experimental conditions by integrating isolate-specific single nucleotide polymorphisms (SNP) and transcriptome data collected during growth on glucose minimal medium; this effort substantially expands our data integration approach from our earlier study of metabolic activity within CF isolates using a previous

genome-scale model of *P. aeruginosa* (Oberhardt et al. 2010). In brief, a SNP introducing a nonsense mutation in a given gene resulted in inactivation of that gene in the model; the Sorting Intolerant From Tolerant (SIFT) algorithm (Kumar et al. 2009) was used to predict the functional impact of other SNPs resulting in missense mutations. These data were interpreted as "levels" of gene activity reduction (minimal, moderate, or maximal) implemented in context with transcriptome expression levels (off, potentially active, on) consistent with the gene-protein-reaction relationships to develop activity constraints for associated reactions (further details of our constraint-based integration of SNP and transcriptome data are described in Materials and Methods). These methods resulted in isolate-specific models that are consistent with the substantial activity changes in pathways suggested by our experimental observations, and also enabled prediction of activity changes that were not highlighted by analysis of the *in vitro* data.

Constrained purine metabolism activity contributes to improved redox balance during adaptation

Results from the constraint-based flux modelling support the feasibility of alterations in glycine metabolism that result in novel carbon fixation; our isolate-specific models predict comparable levels of optimal biomass production regardless of GCS and formate dehydrogenase directionality. We hypothesized that the experimental phenotypes shown by the late stage isolates might indicate a shift from aerobic growth with high biomass production to microaerobic conditions where redox cofactor recycling was prioritized in addition to biomass production. Given our additional experimental evidence for a novel route of CO2 fixation, we evaluated the effects of limitations of O2 and CO2 uptake and biomass production levels on the ability for each strain model to produce redox cofactors. We specifically compared the ratio of maximized NADH vs. NAD+ production fluxes under varied uptake constraints and growth requirements for mDK2-WT and mDK2-07, as shown in Fig. 5.4A. mDK2-07 predicts a stable redox cofactor production ratio across varied O2 and CO2 uptake conditions while the redox ratio of mDK2-WT varies with

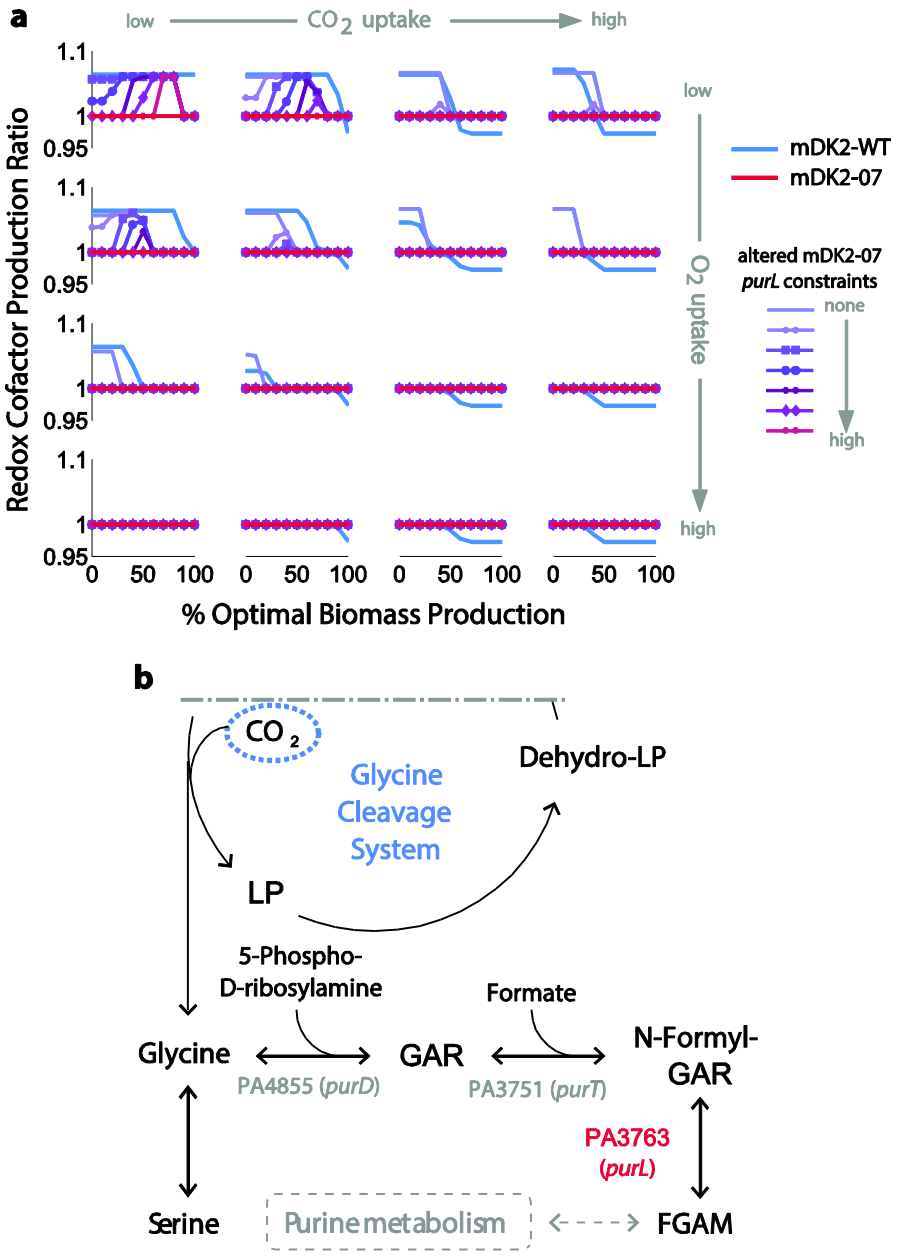


FIG 5.4. Redox cofactor production differences between mDK2-WT and mDK2-07 due to SNP in purine metabolism. (a) An evaluation of the effects of altered O_2 and CO_2 uptake on the ratio of NADH production to NAD^+ production under a range of biomass production constraints for mDK2-WT (blue), mDK2-07 (red), and mDK2-07 with reduced *purL* activity constraints (shades of purple). **(b)** Pathway illustration of the connection between glycine metabolism and purine metabolism, specifically highlighting *purL*, a gene that contains a SNP in DK2-07 that the model predicts is connected to differential redox metabolism activity between strains. Abbreviations: Glycinamide ribonucleotide (GAR), 5'-phosphoribosylformylglycinamidine (FGAM), lipoyl protein (LP).

O2 uptake, CO2 fixation, and biomass production.

To identify contributors to these differential predictions between mDK2-WT and mDK2-07, we modulated the gene and SNP-based constraints applied to each model. We identified the restriction of the purine metabolism enzyme phosphoribosylformylglycinamidine synthase (*purL*) due to an applied SNP constraint as the main contributor to the stability of the redox ratio in mDK2-07. While mDK2-WT has several SNPs resulting in model reaction constraints including a SNP affecting GMP synthase (*guaA*) that also plays a role in purine metabolism, a SNP in *purL* is not present and thus the associated activity of this reaction is unconstrained. We were surprised to find that a single SNP contributed so substantially to this phenotype of a stable redox ratio under varying uptake and growth constraints; further investigation of our model identified the functional relationship between glycine metabolism and *purL* as shown in Fig. 5.4B, which is a non-canonical mapping of pathways that usually would not be obviously linked together. By incrementally increasing the constraints applied to the phosphoribosylformylglycinamidine synthase reaction due to the SNP in *purL* from unconstrained (mDK2-WT phenotype) to the moderate constraints applied in mDK2-07, we showed that the redox ratio of mDK2-07 transitions to a balanced state as *purL* activity is constrained. The graded impact on redox metabolism due to the *purL* constraint is clear in microaerobic conditions at low levels of CO2 uptake; high biomass production requirements magnify the impact of *purL* constraints on the transition to a balanced redox state. We propose that the close connections between the altered glycine metabolism reactions and *purL* as shown in Fig. 5.4B support the potential role of *purL* as a modulator of redox recycling via reversal of the glycine cleavage pathway. Our original constraints based on the SIFT predictions of SNP impact in *purL* were a broad estimate of how function might be altered; further fine-tuning may reflect the actual degree of impact of the SNP in connection to the experimentally-observed phenotype. Ultimately, our models predict that the *purL* SNP is tightly tied to improved redox balance via novel CO2 fixation in the late stage isolates.

Many metabolic systems may contribute to the redox balance of a cell in addition to the contributions we have shown from glycine and purine metabolism. Using flux variability analysis (FVA) (Mahadevan and Schilling 2003) we evaluated potential changes in redox metabolism by comparing changes in reaction activity within reactions where redox cofactors (here defined as NAD+, NADH, NADP+, NADPH, FAD+, and/or FADH) participate versus changes in reaction activity across all reactions. We then used a global metric of total flux activity (the sum of the ranges between minimum and maximum potential flux predicted for all reactions using FVA in a given model divided by the same calculation performed for iPA1139) for each isolate model normalized by the same measure in iPA1139. Albeit a coarse representation of "metabolic capability" of the network, this metric provides a single snapshot of changes in metabolism as a function of changes in the underlying network characteristics. The late stage models predict 73.2% and 74.7% of the iPA1139 global activity metric compared to 69% by mDK2-WT, indicating a total flux activity increase in mDK2-91 and mDK2-07 compared to mDK2-WT with this global metabolic metric. However, the late stage models predict 85% and 84.4% of the iPA1139 redox activity metric compared to 90.1% by mDK2-WT within the subset of reactions that utilize redox cofactors, showing a reduction in the redox-related flux activity of the late stage strain models compared to mDK2-WT. We interpret these opposing changes between global and redox metabolism potential activity as an indication of systemic shifts in redox-related reaction activity between the wild-type and late stage isolates.

Genome-scale metabolic modelling contextualizes global metabolic changes

The isolate-specific metabolic models allow us to evaluate altered activity across a far greater expanse of metabolic systems than just glucose and glycine metabolism. They account for the effects of other SNPs in addition to the *purL* SNP that we previously highlighted as well as the reprogramming of the transcriptome due to adapted regulation and/or mutation. We can readily perturb specific genes and reactions computationally to investigate both the underlying drivers and potential consequences of genetic and transcriptomic adaptations at the genome scale. Here,

we performed routine predictions of essential genes and flux variability that are often used to identify novel treatment targets by prioritizing genes and reactions important for growth (Chavali et al. 2012). Our results indicate broad systemic rewiring in the late stage isolates that both complement our conclusions about glycine and redox metabolism as well as highlight other potential therapeutic targets important to adaptation during adaptation to a host environment.

Essential metabolic activity alters during adaptation

The isolate-specific models enable us to evaluate genes essential to strain growth phenotypes in glucose minimal medium by inactivating a given gene in the models and then predicting maximum possible growth *in silico*. Figure 5.5 shows a Venn diagram categorizing all essential genes across our base model iPA1139, mDK2-WT, mDK2-91, and mDK2-07 together with a stacked histogram of reactions associated with the DK2-specific essential genes. Isolate-specific SNPs were located in six genes predicted to be essential for growth in all models. Of these, constraints applied due to the SNP in PA3769, encoding GMP synthase (*guaA*), were the main driver of reduced *in silico* growth in mDK2-WT compared to the base model; constraints applied due to a SNP in PA1609, encoding beta-ketoacyl-ACP synthase I (*fabB*), affected growth to a lesser degree in the same strain. In contrast, constraints based on the *purL* SNP located in PA3763 were the main driver of reduced *in silico* growth in mDK2-91 and mDK2-07. The presence of SNPs in these genes predicted to be critical in metabolic activity according to our computational models adds emphasis to their potential importance to adaptive selection during infection.

While an array of interesting pathways have altered gene essentiality between strains, we found the changes in glucose metabolism, glycine metabolism, and oxidative phosphorylation as indicated in Fig. 5.5B to be of particular interest when compared with the results of our previous experiments. These changes are the result of integrating SNP and expression data into our models; we therefore can identify the experimental data underlying the specific constraints that contribute to these gains in gene essentiality. Upregulated pentose phosphate pathway genes in

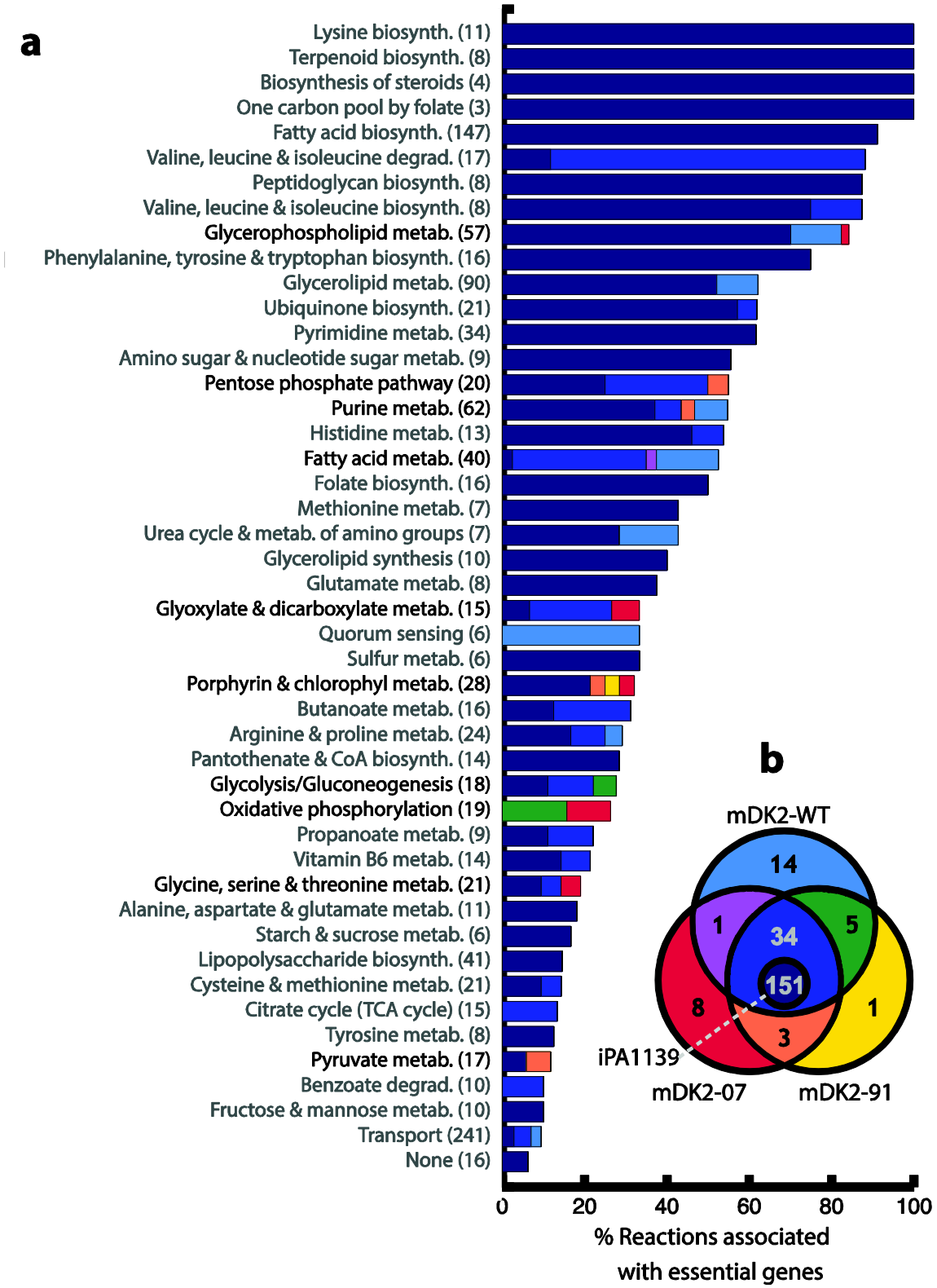


FIG 5.5. Isolate-specific gene essentiality and associated functions.

FIG 5.5 Cont'd. (a) Stacked histogram of reactions associated with DK2-specific essential genes, as shown by % associated reactions within a particular KEGG subsystem. . Total reactions assigned in the KEGG subsystem are included in parentheses in each subsystem label. Results for the essential reaction distribution across the base model and three isolate-specific models are shown in each subsystem category as indicated by colours corresponding to the categories of the Venn diagram in (b). Bolded histogram labels highlight subsystems that show variation in reaction distributions between isolate models. **(b)** Venn diagram of the distribution of *in silico* essential gene predictions, highlighting the differences in unique versus shared essential genes between mDK2-WT, mDK2-91, mDK2-07 and the base model (iPA1139).

DK2-91 and DK2-07 contribute to differences in essentiality predicted by the late stage isolate models, highlighting adaptation in glucose metabolism. Select glycine cleavage system genes are essential in mDK2-07 due to expression-based constraints; glycine dehydrogenase is identified as an essential reaction in mDK2-91 and mDK2-07 in contrast to mDK2-WT for similar reasons. In oxidative phosphorylation, there is a switch in preferred cytochrome complexes in oxidative phosphorylation between model mDK2-WT and mDK2-91, which rely on cytochrome bc1 complex genes (PA4429-4431), while model mDK2-07 relies on cytochrome c oxidase genes (PA1317-1321). This phenotype results from transcriptomic changes in DK2-91 and DK2-07 compared to DK2-WT in glucose minimal medium that shows significant downregulation of the *nuo* operon encoding NADH dehydrogenase (complex I of the electron chain) in the late stage isolates. The lack of active oxidative phosphorylation could explain the need for alternative redox recycling reactions such as glycine synthesis through the glycine cleavage system. These hypotheses regarding mechanistic drivers of altered essentiality between strains are a key contribution enabled by our integrated systems approach. Identifying the strain-specific genes important to the adaptations occurring in the DK2 lineage allows us to highlight functionally impactful SNPs and offer specific, novel treatment targets within key pathways reprogrammed during evolution within the host.

Changes in pathway activity highlight adapting systems

We evaluated the results of flux variability analysis that predicts the minimum and maximum levels of a reaction's flux while maintaining maximum biomass production; this enables calculation of the range of potential activity for a given

reaction. Fig. 5.6 shows a full-scale map of the metabolic network where directional differences in adapting reaction activity in mDK2-WT and mDK2-07 are identified by reaction colour and dashed lines identify SNPs in associated reactions. Decreases in the range of reaction activity likely indicate a SNP- or gene expression-associated constraint, while increases in range could be interpreted as increased flexibility of this pathway that is required to enable the expression-associated constraints or produce necessary biomass components by an alternate pathway; broadly, altered range in either direction may indicate areas of potentially important metabolic adaptation.

Notable trends visualized on the map include increased constraint of "Purine metabolism" flexibility in mDK2-07 and changes in range of reaction activity in "Glycine, serine & threonine metabolism". These specific metabolic pathways were also identified through our study of central metabolism. However, the network map includes a list of additional metabolic pathways with differential activity including pathways related to "Lysine degradation", "Folate metabolism", "Valine, leucine, and isoleucine degradation", "Pyrimidine metabolism", and "Histidine metabolism". The mentioned pathways showed the highest degree of altered system activity in comparisons between early and late stage isolates (see Materials and Methods). Our systems analysis highlights areas of potential adaptation due to SNPs and altered transcriptomics in a broad array of pathways, suggesting new avenues of future experimental investigation that could elucidate other important mechanisms of adaptation in addition to our novel relationship between altered carbon fixation and redox metabolism.

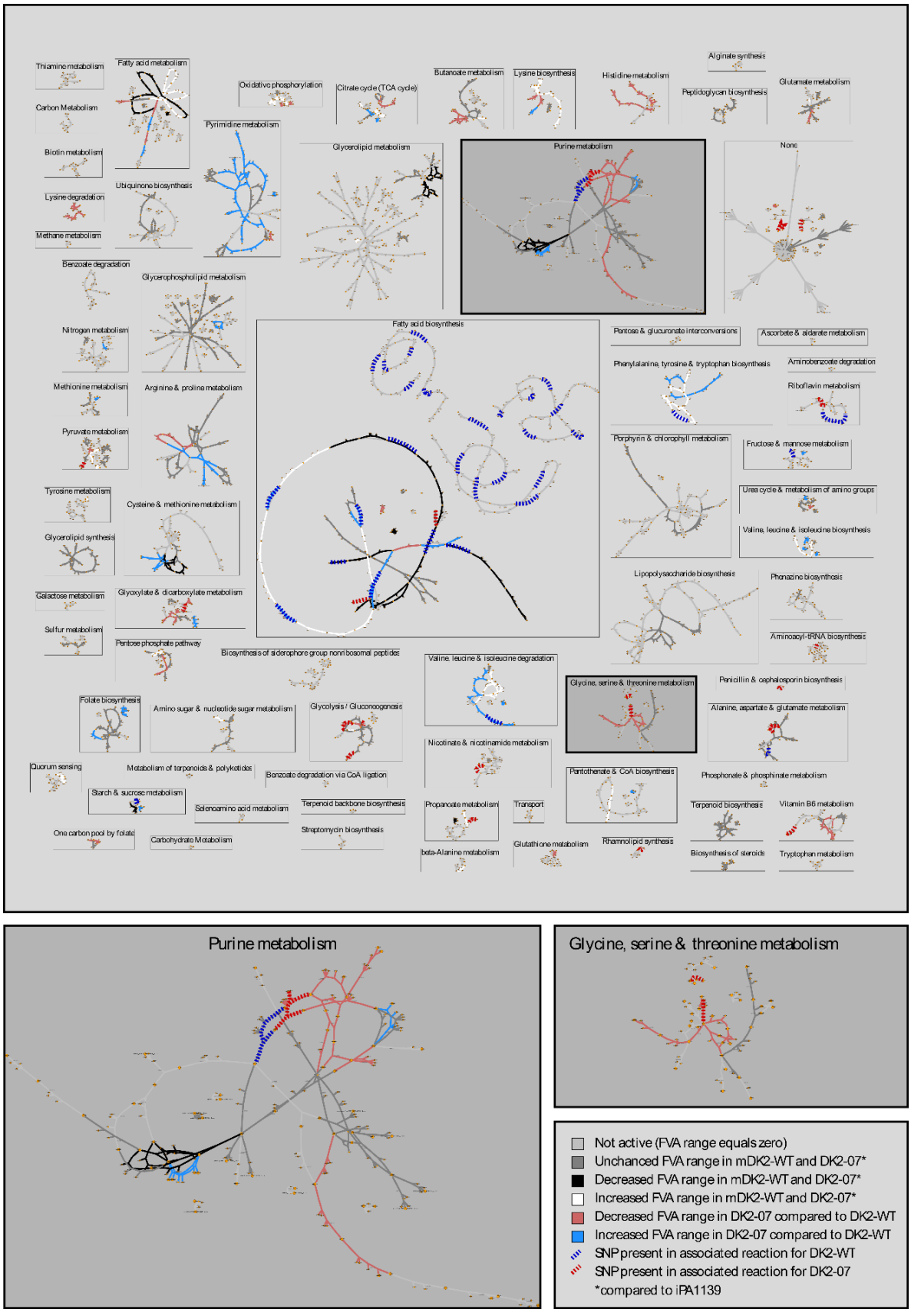


FIG 5.6. Flux variability analysis displayed on global metabolic map.

FIG 5.6. Flux variability analysis displayed on global metabolic map. Differential reaction activity ranges between mDK2-WT and mDK2-07 predicted by flux variability analysis under 100% biomass demands. Increase/decrease in flexibility was identified through comparison of mDK2-WT and mDK2-07 reaction predictions with base model iPA1139 reaction predictions. Dashed lines indicate SNPs present in DK2-WT and DK2-07. The map provides an overview of metabolic changes between DK2-WT and DK2-07, with enlarged panels of purine metabolism and glycine, serine and threonine metabolism presented to highlight the important changes identified in these subsystems. Users can zoom in to identify specific compounds and reactions connected to highlighted areas of differential activity.

DISCUSSION

In this study we have used a systems biology approach to investigate metabolic behaviour during adaptation of a pathogen to the human host. We used genome-scale metabolic models integrated with high throughput data to evaluate the feasibility and potential impacts of novel metabolic adaptations suggested by experimental characterization of glucose metabolism in *P. aeruginosa* clinical isolates. There is value in evaluating both broad changes in high level systems and specific, detailed molecular mechanisms using systems biology approaches; the former enables the prediction of systemic network production and quantification of network elements while the latter offers specific hypotheses regarding functional roles of the smallest network components (Heinemann and Sauer 2010). Here, we provide a systems level perspective of key pathways connected to metabolic adaptation, but focus our analysis on specific systems suggested by targeted experiments that indicated major changes in metabolism between initial infecting strains of *P. aeruginosa* and late-stage clinical isolates. We confirmed that the ED pathway is the only active glycolytic pathway in *P. aeruginosa*, consistent with other *Pseudomonas* species. Experimental profiling identified a transition towards accumulation of gluconate and 2-ketogluconate and enhanced fixation of carbon dioxide into glycine specifically in the late stage isolates. Computational modelling supported the feasibility of reversed utilization of the glycine cleavage system, enabling a novel route of carbon fixation that in combination with a previously inconspicuous mutation in purine metabolism contributed to improved redox balance in the late stage isolates. We identify genes and pathways key to the adaptive processes we see in the DK2 lineage using gene essentiality and flux variability analysis, which may contribute to the design of novel treatment

strategies. Our approach results in a metabolic map that provides mechanistic insight into how SNP and transcriptional changes affect metabolism at a genome scale, bridging the difficult gap between molecular mechanisms and broad, system-wide adaptation and prioritizing novel areas of metabolic reprogramming that can be targeted therapeutically.

The production of gluconate has previously been observed for clinical isolates of *P. aeruginosa*. Behrends et al. (Behrends et al. 2013) found that gluconate excretion is associated with higher tolerance towards antibiotics and another study by Galet et al. (Galet et al. 2014) found that gluconate produced by *P. aeruginosa* inhibits production of an antibiotic in *Streptomyces coelicolor*. In the context of the above analysis indicating the shift in redox balancing, it might also be possible that the accumulation of gluconate and 2-ketogluconate is driven by the production of two equivalents of NADPH coupled to the oxidation reactions of glucose to 2-ketogluconate via gluconate in the periplasmic space (Fig. 5.1). This suggestion is not necessarily in disagreement with the correlation between gluconate and antibiotic resistance since there may also be a link between NADPH generation and antibiotic resistance given the literature on antibiotics and oxidative stress (Kohanski et al. 2010; Derewacz et al. 2013). The identification of the ED pathway as the only active glycolytic route in *P. aeruginosa* can also be linked to generation of NADPH. The ED pathway is found to be essential for glucose metabolism in other *Pseudomonas* species; in *P. putida*, its activity has recently been associated with resistance towards oxidative stress (Chavarría et al. 2013). The activity of these pathways can thereby be a mechanism to accommodate the conditions within the lung environment including both antibiotics and ROS generated by PMNs, both of which are sources of oxidative stress.

The genome scale models support the potential for novel carbon fixation routes in the late stage isolates; they also enable us to connect the late stage isolate glycine metabolism phenotype and altered redox balance in microaerobic conditions to a specific SNP in purine metabolism through network analysis. A study by Ryssel et al. recently identified upregulated purine metabolism activity as a contributor to

poor stress response in *Lactococcus lactis*, citing the production of guanine nucleotides in inducing stress sensitivity (Ryssel et al. 2014 Aug 20) while a prior study had noted the essentiality of purine synthesis in *Escherichia coli* during blood infections (Samant et al. 2008). To our knowledge, adaptation in purine metabolism has not been identified as noteworthy in cystic fibrosis infections; we evaluated published genotyping studies of cystic fibrosis isolates and identified purine SNPs in other clinical isolates (Bezuidt et al. 2013). We suggest that altered purine metabolism may be tied to the reversal of the glycine cleavage system and contributes to resultant altered redox physiology. Whether the *purL* SNP also contributes to the need for glycine production via the GCS or is a simple way to modulate effects of the reversed GCS phenotype is currently uncertain. However, it is likely that the downregulated oxidative phosphorylation highlighted by our late stage isolate gene essentiality predictions is a way to avoid generation of oxygen radicals through the electron chain. The bacteria therefore need to redirect the metabolic flux through the glycine cleavage system to ensure regeneration of NAD+ that is used in glycolysis.

Our hypothesis regarding the role of the glycine cleavage system as an important mediator of successful adaptation in *P. aeruginosa* led to our investigation of other cases where the glycine cleavage system is important. The glycine cleavage system is not only present in bacteria but is present across all domains of life (Bar-Even et al. 2012). In cancer cells, elevated activity of the glycine cleavage system has been associated with tumorigenesis; glycine decarboxylase activity was correlated with reduced survival of patients with lung cancer (Zhang et al. 2012). Further investigation of the GCS in other bacterial pathogens and disordered human cells such as cancer cells may merit evaluation of a potential reversal of the pathway that enables beneficial adaptation in redox metabolism during cell proliferation in stressful environments.

Changes in metabolism in *P. aeruginosa* during adaptation have previously been considered as pleiotropic effects of regulatory actions on other targets, such as virulence factor production (Nguyen and Singh 2006; Smith et al. 2006). Here, we

suggest that changes in metabolism are a direct target of adaptation and a driving force is selection for improved redox balance. Our systems-based analysis highlights important genes and metabolic activities involved in these adaptive processes, proposing specific pathways for novel therapeutic measures that could be used to pre-emptively combat an organism's evolutionary goals such as rewired redox metabolism. We suggest a concrete example of redox balancing through the glycine cleavage system, identifying a future target of interest for unwanted cell growth in the human body.

Chapter 6: Reflections and future directions

A Summary

My dissertation work has been a collection of projects revolving around systems analyses of the metabolic activity of Gram-negative opportunistic pathogens. Focusing specifically on pathogens that induce chronic infections in cystic fibrosis patients, I investigate the inherent differences in metabolic capacity and virulence encoded into their genomes by converting this information into quantitative models of metabolic activity. The models have allowed me to predict the functional impacts of differences in genetic content between species and extend these comparisons to differences within species and strains under specific growth states and adaptive stages using omics-based constraints. Parsing predictions of reaction fluxes from linear optimization techniques such as flux balance analysis and flux variability analysis has led to the identification of network components critical for pathogen growth and unique capacities of virulence-linked metabolic pathways. The models are also an excellent framework for contextualized analysis of information from genome annotations of bacteria with large and complex reservoirs of metabolic genes. Curation of model content and predictions guides annotation refinements and provides an efficient organizational structure for comparing genomic content between species and strains from a functional perspective.

Contributions

Reconstructions

Building and updating the four new reconstructions at a high level of detail and coverage was a considerable portion of my PhD work; these models are tools for the entire community to use and have hopefully been built in an accessible format that will make future improvements an easier process than what I underwent in creating them in standardized syntax with tools that were still under development. While model construction can be tedious even with the use of semi-automated tools, the experience provides invaluable insight into a given model's strengths and potential weaknesses that is critical to the design of future analyses. Additionally,

these models are then disseminated throughout the modeling community, multiplying the impact of any time devoted to model construction and refinement.

The models that I (and co-authors) have provided were curated using the field-standard approaches of single substrate utilization assays and proposed rich-media essential genes identified through transposon mutagenesis studies. I also devoted a substantial amount of time to inclusion of peripheral pathways and curation of GPR formulas because the bacteria I was modeling were known for their large, under-annotated genomes that encoded broad and flexible catabolic potential. These efforts resulted in very large models that predict high optimal growth rates and a limited number of essential genes in complex media. Future curation should be focused on improving the predictions of biomass and byproduct production rates as well as GPR assignments under a broader range of growth conditions.

Of these future areas of work, validating the predicted impacts of gene and pathway redundancy seems more immediately accessible. Growing interest in the idea of underground metabolism and enzyme promiscuity motivates an expanded analysis of model component essentiality. Using transposon mutant libraries and TN-seq screening to experimentally map essential genes and pathways in a broader array of growth conditions would provide validating data sets that could be used to substantially refine model predictions. Our models would be particularly useful to couple with these studies as they enable systemic assessments of the function of particular genes and enzymes.

Experimental measurement of production rates and yields would also be feasible using transposon library screening, but requires a transition from the binary (or at least binned) outcomes of gene essentiality assessment to a continuum of predicted rates. For example, achieving an accurate prediction of a gene deletion’s impact on growth (essential, inhibitory, or no impact) is more straightforward than accurate predictions of the specific production level of a siderophore.

Virulence factor modeling

My focus on virulence factor synthesis pathways and virulence-linked metabolic capacity was an extension of ideas originally developed in Matthew Oberhardt's work with the first iterations of the *P. aeruginosa* model that has been mentioned throughout this thesis. Methodical expansions to his approach have improved the accessibility of modeling predictions to non-modelers. More emphasis has been placed on the role of single genes and substrates in the context of virulence factor production to simply experimental validation. The idea of targeting virulence factors has been occasionally maligned as a distraction from the need to understand mechanisms of existing antibiotic resistance. However, the method of choice for rapidly getting new therapeutics out to patients is high throughput screening of small molecules and drugs, repurposing approved chemotherapies, and identifying drug synergy that increases effectiveness and/or prevents resistance from arising. Thus, targeting alternate pathways that contribute to successful infections is a compelling and under-explored area that aligns perfectly with the capabilities of genome-scale metabolic modeling.

This dissertation contains preliminary results of experimental validation of the provided targets for inhibition of virulence factor production. Ultimately, screening of gene deletions predicted to result in complete or partial inhibition based on quantitative production metrics should be performed using transposon mutant libraries or other engineered strains. Methods are available to quantitatively evaluate production levels of the virulence factors included in the model, though ease of performance and precision of the results can vary. Quantification of pyocyanin production requires a few extraction steps similar to those used for pyoverdine as described in Chapter 4 such that concentration can be determined via optical density measurements at a specified wavelength (Wurtzel et al. 2012). The production of alginate and rhamnolipids would be other easily accessible measurements to conduct given that they can be assayed with dyes which are then quantified using optical density measurements at specific wavelengths (Hay et al. 2009; Pinzon and Ju 2009). In contrast, experimental evaluation of LPS production

requires more specialized expertise using LPS Western blots (Davis, Jr. and Goldberg 2012).

The production tradeoff metric used to comprehensively account for both biomass and virulence factor production defects due to gene inhibition could also be further explored. Computational evaluation of the specific shape of the Pareto front (the curve of the Pareto area boundary) could enable finer categorization of different tradeoff 'phenotypes' that might offer insight into the underlying mechanics of the defects. For example, gene inhibition in similarly functioning pathways might result in the same Pareto front shape. Additionally, optimized growth versus virulence factor production could be more rigorously evaluated with these predictions by comparing *in vitro* versus *in silico* virulence factor production at a given growth rate. Confirmation of growth-linked versus virulence-linked defects would be important in choosing a therapeutic target that may or may not result in growth-linked drug resistance.

Evaluating adaptation via omics integration

My work analyzing adaptation within longitudinal clinical isolates from Danish CF patients was an interesting challenge in data integration. There are a tremendous number of ways to analyze omics data, much less integrate the dataset into a giant model (Machado and Herrgård 2014). I tried many methods based on expression level cutoffs and statistical significance before implementing a comparatively simple approach modeled after the paper by Machado et al.. My novel method for SNP reconciliation and integration in conjunction with the expression data was the most substantive contribution to altered predictions between the isolate models. This is sensible given that any permanent change that appears to be actively selected for within temporally consecutive strains is likely of higher import than impermanent fluctuations in expression data. Viewing the expression data as an outward extension and artifact of the changes induced by the SNPs was an important mental transition to make. While not a perfect representation of the complex layers of metabolic rewiring that appear to occur during long term adaptation, a considered understanding of the layered relationship between SNPs and expression data was

certainly helpful during the methods development process and analysis of results. The contextualization of functional impacts of adaptation with my co-author Julianne's experimental characterization of flux was also a challenging aspect of this project. However, our integrated approach helped us develop our hypotheses regarding redox stress and the connection of the glycine cleavage pathway to purine metabolism is a novel contribution that would not have been feasible without the metabolic modeling.

Design decisions and caveats

Tradeoffs in annotation refinement and model expansion

In my methodical refinement of the models, the incorporation of genes with lower quality functional assignments has contributed to the substantial size of the models. I have attempted to balance between coverage and specificity; when gap-filling for a needed reaction, I considered lower confidence annotations (BLAST E-value scores up to 1E-10 and/or assessment of protein functional domain predictions depending on how badly the gap needs to be filled based on modeling results). However, I also added genes of higher similarity and more specific secondary function annotations on genome databases even if gapfilling was not required (BLAST E-value scores up to ~1E-35 and more specific domain predictions). Given the recent publications supporting unexpected levels of enzyme promiscuity and underground metabolism (Patrick et al. 2007; Notebaart et al. 2014), I feel these additions were justified.

In my view, the known metabolic flexibility of these pathogens, their large genomes, and their complex regulatory networks already motivate the incorporation of omics constraints to make specific, non-comparative predictions of function. I treat the models more as a reservoir of potential functionality that then can be reduced with further constraints to tailor them for specific questions. By including a more than strictly necessary number of genes, I attempt to increase the flexibility of the models in capturing behavior in an expanded range of scenarios. I would rather overpredict function in the context of a hunt for new therapeutic

targets; to me, it increases the confidence I have in the targets I do identify as being critical for a particular application or synthesis product.

There are obviously drawbacks and caveats to this approach that must be taken into account when designing a study using the models. If I argue that I need to include more hypothetical gene assignments in favor of increased coverage and flexibility, this may lower the accuracy of my predictions in certain conditions if I do not incorporate extra constraints. The models currently predict very high optimal metabolic capacities in rich media because of the many catabolic pathways present in *Pseudomonas* and *Burkholderia* genomes as well as the high level of pathway redundancy via either gene duplications or isozymes. Whether the bacteria are realistically capable of evolving to the high levels of growth predicted under these conditions is less certain due to regulatory constraints as well as alternate biological objectives that may compete or interfere with growth (i.e. quorum sensing, competition for nutrients as a community of pathogens grows, the benefits of social cheating versus production of public goods). Studies have also shown clear substrate utilization preferences by *P. aeruginosa* in the CF lung that cannot be justified by current model predictions based on optimality (Palmer et al. 2007), though this may also hinge on incomplete or unknown metabolic pathways not currently included in the model.

Further development of multi-objective modeling (Schuetz et al. 2012; Zomorrodi et al. 2013), omics integration techniques, and the construction of a complete regulatory network to integrate with our metabolic models as is being attempted for other organisms (Karr et al. 2012; Lerman et al. 2012) are all avenues to improve predictions in the context of high content reconstructions. It would also be worthwhile to conduct evolution experiments on relevant rich media such as SCFM and improve the accuracy of applied media conditions by improving the accuracy of substrate uptake constraints specifically.

Benefits of comparative analysis

This obviously draws questions regarding the predictions made in this thesis without incorporating extra constraints via omics data. While our updated *P. aeruginosa* reconstructions have still improved in prediction accuracy for gene essentiality and substrate utilization via curation and expansion of included reactions and their GPR formulations despite the moderate confidence addition of genes, there are likely limitations to the model's ability to accurately predict growth and production yields as mentioned. However, our comparative analysis approach allows us to identify gene deletions which induce differences from the base model state, mitigating concerns regarding specific rate and yield predictions in favor of evaluating relative changes.

Future iterations of the model can be curated to provide improved assessments of production rate via more specific constraints of flux bounds, but this would also induce more specific tailoring to specific media conditions and reduce the flexibility of the model. It would be worthwhile to compare the model with later versions of the many *E. coli* reconstructions that have been built using more tailored constraints to achieve accurate yield predictions when compared with *in vitro* evolution experiments. Here, the investigation of gene deletions that indicate novel targets for virulence factor inhibition supports our focus on improving GPR relationships and expanding genomic coverage of the models.

Iteration in integrated analyses

The specific mechanistic predictions made in my evaluation of adaption in chronic infections do use omics constraints to tailor the models to isolate-specific growth states. However, we still did not achieve perfect alignment with the experimental results obtained regarding the glycine cleavage system. Ultimately we used the model as an organized resource in our initial hunt for any reactions that might be contributing to the unexpected glycine accumulation in late stage isolates. We then had to specifically constrain the model to replicate the pathway reversal; creating the models using the SNP and expression constraints still did not create any specific demand for this pathway to be reversed. This illustrates how there still needs to be

active involvement in choosing how to best achieve your investigative goals when using reconstructions. Ultimately we used the model to seed a hypothesis, went back to the lab to test said hypothesis, and then evaluated what the repercussions of this hypothesis would be if it were true from a broad perspective by returning to the model and adjusting constraints.

Some of our specific mechanistic predictions definitely need further experimental validation. Other CF isolates should be examined to see if SNPs appear in purine metabolism pathways; ideally, a pool of isolates with and without this SNP could be evaluated for glycine accumulation. Redox stress could also be artificially induced in the two groups to see which pool adapted more quickly. An alternate approach would be to see if reversal of the purine SNP in our DK2 isolates through allelic replacement caused any change in glycine accumulation. While we are considering which avenue to pursue, we feel our combined experimental and computational approach is an impressive representation of an integrated systems analysis that offers mechanistic insight, validation, and expansion to broader, important hypotheses regarding novel biological functions.

Future (i.e. partially completed) work

CF community model

At the beginning of my time in Jason's lab, we planned to develop a multi-pathogen model of cystic fibrosis infection. Initially our interest revolved around the competition and differential virulence between *P. aeruginosa*, *B. multivorans* and *B. cenocepacia*. Growing interest in microbiome studies has resulted in some of the first studies of bacterial species composition within CF sputum as patients age (Zhao et al. 2012b), showing that *P. aeruginosa* outcompetes other infecting species such as *Staphylococcus aureus* in children with CF to become the dominant pathogen after decades of infection. *Burkholderia* species also manage to persist for decades, and *B. cenocepacia* is capable of outcompeting and replacing other strains such as *B. multivorans* to cause severe morbidity and mortality in patients; however, overall incidence of Bcc infections is low. We wanted to investigate both inherent advantages in metabolic capacity as performed in Chapter 3 of this

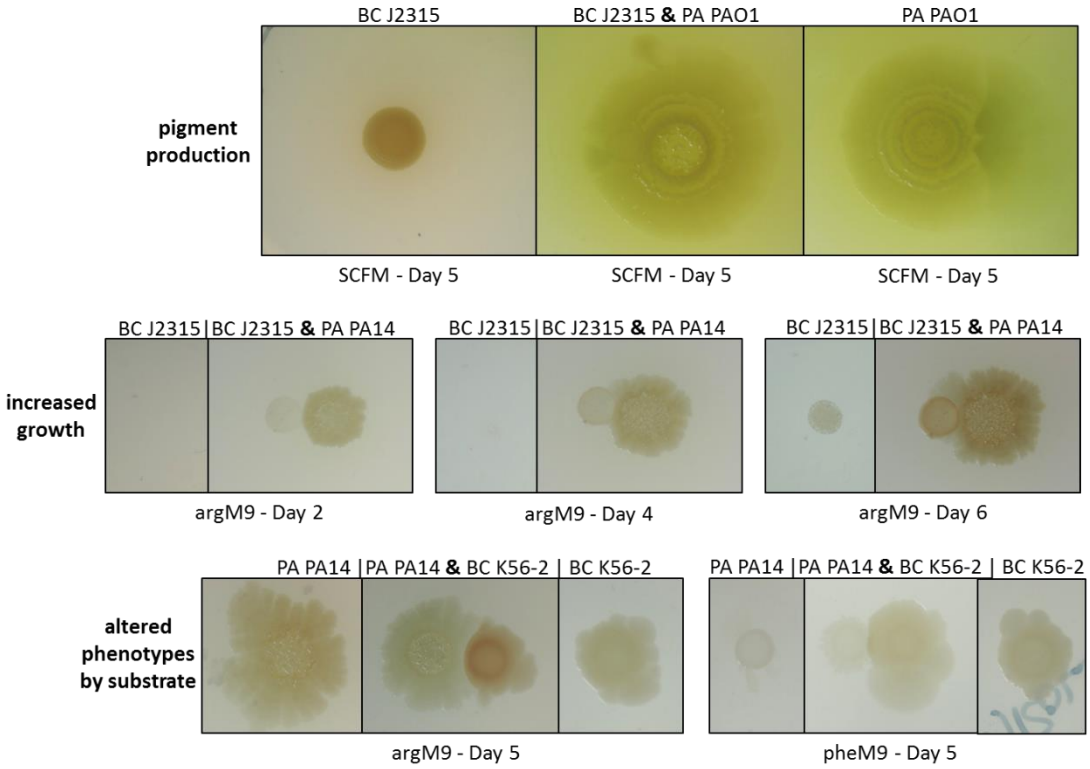
dissertation and potential metabolic interactions such as resource sharing or competition, adaptive niche specialization, and the deployment of virulence factors such as phenazines or quorum sensing crosstalk between species.

Inspired by a metabolomics study of secreted metabolites in *Streptomyces coelicolor* co-culture (Traxler et al. 2013), I have screened pairings of five strains of our three CF pathogen species on agar plates for four media conditions. 2 ul of culture at a standardized 0.25 OD_{600} in dilute liquid SCFM were plated on glucose-M9, arginine-M9, phenylalanine-M9, and SCFM solid media plates. Strain pairings were created by spotting two cultures of distinct strains 5 mm apart and incubated at 37C for 6 days, imaging each colony pair every 24 hours (2 replicates per pairing). Single colonies for each strain (2 replicates) from the same starting culture were plated in isolation as a control on each media.

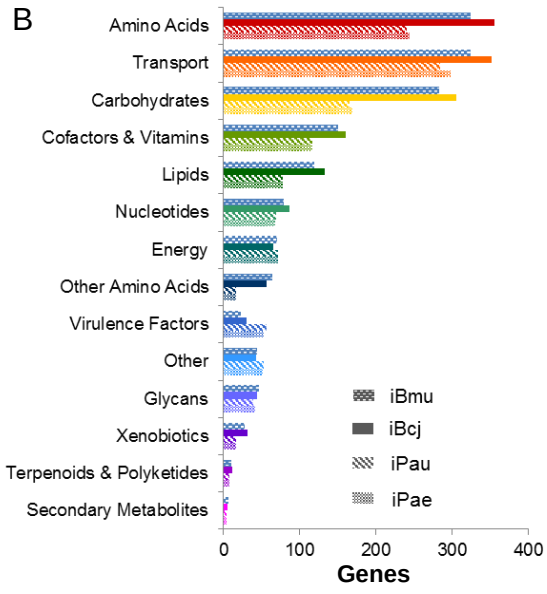
Unique interactions were apparent via visual inspection of pigment production, colony size, and colony morphology between multiple strain pairings as shown via representative pairings in Fig. 6.1A. These results, combined with prior studies of growth enhancement in biofilms and during antibiotic exposure showing unique interrelationships between *Pseudomonas* and *Burkholderia* species, supported our hypotheses regarding potential metabolic interactions of these species. The *in vitro* phenotyping shows increased Bcc strain growth rate when paired with PA14, changes in secondary metabolite production, and strain-specific killing that is also substrate-dependent. These experiments need to be replicated several more times and evaluated using image-based quantitative measurements such as colony diameter, pigment intensity, and surface area to enable statistical analysis.

I have performed some initial comparative analyses of the four CF pathogen reconstructions for *B. cenocepacia* J2315, *B. multivorans* ATCC17616, *P. aeruginosa* PAO1 and *P. aeruginosa* PA14. Figure 6.1B and 6.1C show the distribution of genes and reactions across KEGG subsystems within the 4 models. Bcc species have a larger repertoire of secondary metabolism pathways than the *P. aeruginosa* strains, which may provide additional avenues for substrate utilization that enable the

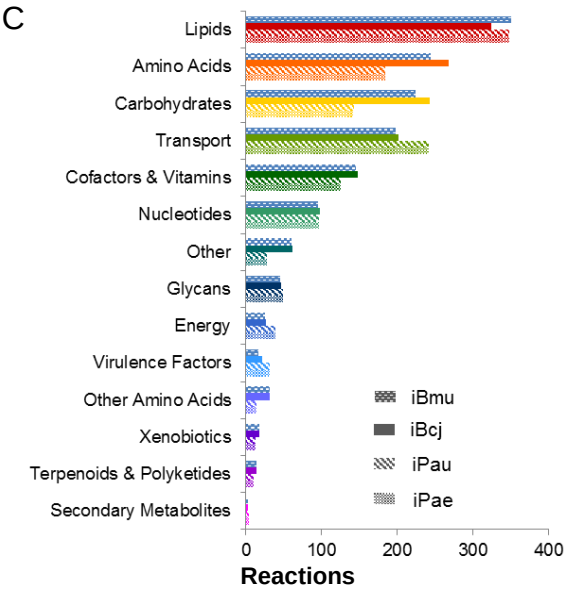
A



B



C



D

	iPY1537	iJB1411	iPae1148	iPau1131
Cystic fibrosis sputum (synthetic)	0.41	0.34	1.00	1.00
Arginine minimal media	0.08	0.08	0.08	0.08
Glucose minimal media	0.09	0.09	0.09	0.09
Phenylalanine minimal media	0.07	0.06	0.05	0.05

FIG 6.1. Preliminary CF pathogen community assessments.

FIG 6.1 Cont'd. (A) A subset of phenotypic differences in co-culture of 5 strains on 4 different media over the course of 6 days at 37C. Initial colonies were spotted in 2 µl .5X SCFM at 0.25 OD at a distance of 5 mm apart. (B) Distribution of genes within KEGG subsystems across 4 metabolic reconstructions. (C) Distribution of reactions within KEGG subsystems across 4 metabolic reconstructions. (D) Normalized differences in growth capacity across 4 metabolic reconstructions for 4 different media.

avoidance of metabolic competition and/or enhance growth of *P. aeruginosa*. Figure 6.1D shows normalized predictions of growth capacity (biomass production flux) for the four models on relevant media; *P. aeruginosa* strains are predicted to have a growth advantage on CF sputum, but Bcc strains have an equal to slightly higher growth capacity on the other substrates. Subsystem distributions of predicted essential reactions indicate that a broader set of unique pathways are vital to Bcc's growth in nutrient limited environments; this inefficiency may contribute Bcc's improved survival in certain co-culture conditions with *P. aeruginosa* strains *in vitro*.

I plan to perform a more rigorous analysis of potential metabolic interactions by actually connecting the models to create a community model, possibly using OptCom as a platform (Zomorrodi and Maranas 2012). I would like to improve some of my phenotyping measures using more quantitative methods, and potentially integrate omics characterization via RNAseq of colony pairs.

Broader goals

My dissertation research has provided the opportunity to build a depth of knowledge of systems analysis, metabolic modeling, and comparative genomics. I have also substantially developed my understanding of the biology of opportunistic Gram-negative pathogens; I find their position as both human pathogens and beneficial soil bacteria to be a fascinating convergence of metabolic potential. Their complexity presents a real challenge to biomedical researchers; I have therefore placed high value on truly investing in both understanding underlying biology and developing computational skills to improve my ability to contribute to a growing health crisis and communicate with collaborators with strengths on either side of the field.

I am interested in pursuing the study of interactions among the growing list of bacteria that co-habit in these chronic infections. Most microbiota studies have so far focused on the human gut. However, chronic infections are a serious and growing problem that affect more than just CF patients; the number of patients with pneumonia linked to COPD is expanding as rising wealth in developing countries has led to longer lives and a surge in later-life noncommunicable diseases that are compounded by increased tobacco use. Hospital-acquired infections are another key component of the drug-resistance health crisis. Comparisons between the adaptive rewiring and pathogenic communities that develop in these health conditions have not yet been performed using any kind of systems modeling analysis to my knowledge.

I am also interested in expanding my evaluation of tradeoffs between adaptive drivers during chronic infections. Few studies have used robust quantitative modeling to assess competing objectives during bacterial colonization of the host. An understanding of the complex factors that contribute to successful adaptation will offer novel therapeutic targets. Given the rapid development of high throughput phenotyping and increasing sophistication of modeling techniques, this goal is no longer far from reach and represents a unique opportunity for a metabolic modeling approach.

In summary, the field of metabolic modeling has been slowly building its reputation as an important tool in addressing human disease. Its utility will only increase as more modelers are trained and reconstruction and analysis techniques are improved and made more accessible. Metabolic reconstruction and analysis is quickly merging with high quality experimental approaches that promise to revolutionize our understanding of and ability to combat infections by opportunistic pathogens.

References

Aaron SD, Ferris W, Henry D a, Speert DP, Macdonald NE. 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with Burkholderia cepacia. Am. J. Respir. Crit. Care Med. 161:1206–12.

Van Acker H, Sass A, Bazzini S, De Roy K, Udine C, Messiaen T, Riccardi G, Boon N, Nelis HJ, Mahenthiralingam E, et al. 2013. Biofilm-Grown Burkholderia cepacia Complex Cells Survive Antibiotic Treatment by Avoiding Production of Reactive Oxygen Species. PLoS One 8:e58943. [accessed 2013 Mar 22]. http://www.ncbi.nlm.nih.gov/pubmed/23516582

Agnoli K, Schwager S, Uehlinger S, Vergunst a, Viteri DF, Nguyen DT, Sokol P a, Carlier a, Eberl L. 2012. Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid. Mol. Microbiol. 83:362–78. [accessed 2012 Mar 15]. http://www.ncbi.nlm.nih.gov/pubmed/22171913

Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput. Biol. 9:e1002980. [accessed 2014 Jul 11]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3605104&tool=pmcentrez&rendertype=abstract

Allen RC, Popat R, Diggle SP, Brown SP. 2014. Targeting virulence: can we make evolution-proof drugs? Nat. Rev. Microbiol. 12:300–8. [accessed 2014 May 26]. http://www.ncbi.nlm.nih.gov/pubmed/24625893

Amari DT, Marques CNH, Davies DG. 2013. The putative enoyl-coenzyme A hydratase DspI is required for production of the Pseudomonas aeruginosa biofilm dispersion autoinducer cis-2-decenoic acid. J. Bacteriol. 195:4600–10. [accessed 2014 Dec 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3807434&tool=pmcentrez&rendertype=abstract

Andersen MR, Nielsen ML, Nielsen J. 2008. Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. Mol. Syst. Biol. 4:178.

Andrä J, Rademann J, Koch MHJ, Heine H, Zähringer U, Brandenburg K. 2006. Endotoxin-like properties of a rhamnolipid exotoxin from Burkholderia (Pseudomonas) plantarii: Immune cell stimulation and biophysical characterization. Biol. Chem. 387:301–310.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Ballok AE, O'Toole G a. 2013. Pouring salt on a wound: Pseudomonas aeruginosa virulence factors alter Na+ and Cl- flux in the lung. J. Bacteriol. 195:4013–9. [accessed 2014 Jan 11]. http://www.ncbi.nlm.nih.gov/pubmed/23836869

Bar-Even A, Noor E, Milo R. 2012. A survey of carbon fixation pathways through a quantitative lens. J. Exp. Bot. 63:2325–42.

Bartell JA, Yen P, Varga JJ, Goldberg JB, Papin J a. 2014. Comparative metabolic systems analysis of pathogenic Burkholderia. J. Bacteriol. 196:210–226. [accessed 2013 Dec 20]. http://www.ncbi.nlm.nih.gov/pubmed/24163337

Becker S a, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2:727–38. [accessed 2012 Oct 30]. http://www.ncbi.nlm.nih.gov/pubmed/17406635

Becker S a, Palsson BO. 2008. Context-specific metabolic networks are consistent with experiments. PLoS Comput. Biol. 4:e1000082. [accessed 2012 Mar 8]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2366062&tool=pmcentrez&rendertype=abstract

Behrends V, Ebbels TMD, Williams HD, Bundy JG. 2009. Time-resolved metabolic footprinting for nonlinear modeling of bacterial substrate utilization. Appl. Environ. Microbiol. 75:2453–63. [accessed 2012 Jul 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2675220&tool=pmcentrez&rendertype=abstract

Behrends V, Ryall B, Zlosnik JEA, Speert DP, Bundy JG, Williams HD. 2013. Metabolic adaptations of Pseudomonas aeruginosa during cystic fibrosis chronic lung infections. Environ. Microbiol. 15:398–408.

Bezuidt OK, Klockgether J, Elsen S, Attree I, Davenport CF, Tümmler B. 2013. Intraclonal genome diversity of Pseudomonas aeruginosa clones CHA and TB. BMC Genomics 14:416. [accessed 2014 Aug 4]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3697988&tool=pmcentrez&rendertype=abstract

Bi H, Christensen QH, Feng Y, Wang H, Cronan JE. 2012. The Burkholderia cenocepacia BDSF quorum sensing fatty acid is synthesized by a bifunctional crotonase homologue having both dehydratase and thioesterase activities. Mol. Microbiol. 83:840–55. [accessed 2013 Feb 4]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3276249&tool=pmcentrez&rendertype=abstract

Bjarnsholt T, van Gennip M, Jakobsen TH, Christensen LD, Jensen PØ, Givskov M. 2010. In vitro screens for quorum sensing inhibitors and in vivo confirmation of their effect. Nat. Protoc. 5:282–93. [accessed 2013 Aug 22]. http://www.ncbi.nlm.nih.gov/pubmed/20134428

Blaby-Haas CE, De Crécy-Lagard V. 2011. Mining high-throughput experimental data to link gene and function. Trends Biotechnol. 29:174–182.

Blais EM, Chavali AK, Papin JA. 2013. Linking genome-scale metabolic modeling and genome annotation. METHODS Mol. Biol. 985:149 – 173.

Blazier AS, Papin J a. 2012. Integration of expression data in genome-scale metabolic network reconstructions. Front. Physiol. 3:299. [accessed 2012 Nov 1]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3429070&tool=pmcentrez&rendertype=abstract

Bloodworth R a M, Gislason AS, Cardona ST. 2013. Burkholderia cenocepacia conditional growth mutant library created by random promoter replacement of essential genes. Microbiologyopen 2:243–58. [accessed 2013 Sep 3]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3633349&tool=pmcentrez&rendertype=abstract

De Boeck K, Malfroot A, Van Schil L, Lebecque P, Knoop C, Govan JRW, Doherty C, Laevens S, Vandamme P, Van Schil L, et al. 2004. Epidemiology of Burkholderia cepacia complex colonisation in cystic fibrosis patients. Eur. Respir. J. 23:851–856.

Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J. 2009. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin. Infect. Dis. 48:1–12. [accessed 2013 May 25]. http://www.ncbi.nlm.nih.gov/pubmed/19035777

Bragonzi A, Farulla I, Paroni M, Twomey KB, Pirone L, Lorè NI, Bianconi I, Dalmastri C, Ryan RP, Bevivino A. 2012. Modelling co-infection of the cystic fibrosis lung by Pseudomonas aeruginosa and Burkholderia cenocepacia reveals influences on biofilm formation and host response. PLoS One 7:e52330. [accessed 2014 Feb 26]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3528780&tool=pmcentrez&rendertype=abstract

Brown S a, Palmer KL, Whiteley M. 2008. Revisiting the host as a growth medium. Nat. Rev. Microbiol. 6:657–66. [accessed 2014 Nov 24]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3115587&tool=pmcentrez&rendertype=abstract

Brown SP, Cornforth DM, Mideo N. 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. Trends Microbiol.:1–7. [accessed 2012 May 15]. http://www.ncbi.nlm.nih.gov/pubmed/22564248

Bucci V, Bradde S, Biroli G, Xavier JB. 2012. Social Interaction, Noise and Antibiotic-Mediated Switches in the Intestinal Microbiota. De Boer RJ, editor. PLoS Comput. Biol. 8:e1002497. [accessed 2012 May 4]. http://dx.plos.org/10.1371/journal.pcbi.1002497

Burgard AP, Maranas CD. 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol. Bioeng. 82:670–7. [accessed 2012 Nov 6]. http://www.ncbi.nlm.nih.gov/pubmed/12673766

Burns JL, Emerson J, Stapp JR, Yim DL, Krzewinski J, Louden L, Ramsey BW, Clausen CR. 1998. Microbiology of sputum from patients at cystic fibrosis centers in the United States. Clin. Infect. Dis. 27:158–163.

Cameron DE, Urbach JM, Mekalanos JJ. 2008. A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. Proc. Natl. Acad. Sci. U. S. A. 105:8736–8741.

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, et al. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 42.

Del Castillo T, Ramos JL, Rodríguez-Herva JJ, Fuhrer T, Sauer U, Duque E. 2007. Convergent peripheral pathways catalyze initial glucose catabolism in Pseudomonas putida: genomic and flux analysis. J. Bacteriol. 189:5142–52. [accessed 2014 Jun 30]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1951859&tool=pmcentrez&rendertype=abstract

Cegelski L, Marshall GR, Eldridge GR, Hultgren SJ. 2008. The biology and future prospects of antivirulence therapies. Nat. Rev. Microbiol. 6:17–27.

Centers for Disease Control and Prevention. 2014. Antibiotic resistance threats in the United States, 2013.

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

Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin J a. 2012. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol. 20:113–23. [accessed 2013 May 21]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3299924&tool=pmcentrez&rendertype=abstract

Chavarría M, Nikel PI, Pérez-Pantoja D, de Lorenzo V. 2013. The Entner-Doudoroff pathway empowers Pseudomonas putida KT2440 with a high tolerance to oxidative stress. Environ. Microbiol. 15:1772–85. [accessed 2013 Aug 13]. http://www.ncbi.nlm.nih.gov/pubmed/23301697

Chiu H-C, Levy R, Borenstein E. 2014. Emergent biosynthetic capacity in simple microbial communities. PLoS Comput. Biol. 10:e1003695. [accessed 2014 Jul 10]. http://www.ncbi.nlm.nih.gov/pubmed/24992662

Christensen B, Nielsen J. 1999. Isotopomer analysis using GC-MS. Metab. Eng. 1:282–290.

Christensen B, Nielsen J. 2002. Reciprocal 13C-labeling: A method for investigating the catabolism of cosubstrates. Biotechnol. Prog. 18:163–166.

Clatworthy AE, Pierson E, Hung DT. 2007. Targeting virulence: a new paradigm for antimicrobial therapy. Nat. Chem. Biol. 3:541–8. [accessed 2013 Aug 7]. http://www.ncbi.nlm.nih.gov/pubmed/17710100

Colabroy KL, Begley TP. 2005. Tryptophan Catabolism : Identification and Characterization of a New Degradative Pathway Tryptophan Catabolism : Identification and Characterization of a New Degradative Pathway. 187.

Cooper R a, Kornberg HL. 1964. The utilization of itaconate by Pseudomonas sp. Biochem. J. 91:82–91.

Cox MJ, Allgaier M, Taylor B, Baek MS, Huang YJ, Daly R a, Karaoz U, Andersen GL, Brown R, Fujimura KE, et al. 2010. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. PLoS One 5:e11044. [accessed 2012 Feb 29]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2890402&tool=pmcentrez&rendertype=abstract

Curcio D. 2014. Multidrug-Resistant Gram-Negative Bacterial Infections: Are you Ready for the Challenge? Curr. Clin. Pharmacol. 9:27–38.

Dandekar A a, Chugani S, Greenberg EP. 2012. Bacterial quorum sensing and metabolic incentives to cooperate. Science 338:264–6. [accessed 2013 Aug 8]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3587168&tool=pmcentrez&rendertype=abstract

Darling P, Chan M, Cox AD, Pamela A. 1998. Siderophore Production by Cystic Fibrosis Isolates of Burkholderia cepacia Siderophore Production by Cystic Fibrosis Isolates of Burkholderia cepacia. 66:2–6.

David L a, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V, Devlin a S, Varma Y, Fischbach M a, et al. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–63.

Davies K. 2013. The era of genomic medicine. Clin. Med. J. R. Coll. Physicians London 13:594–601.

Davis, Jr. MR, Goldberg JB. 2012. Purification and Visualization of Lipopolysaccharide from Gram-negative Bacteria by Hot Aqueous-phenol Extraction. J. Vis. Exp.

DeJongh M, Formsma K, Boillot P, Gould J, Rycenga M, Best A. 2007. Toward the automated generation of genome-scale metabolic networks in the SEED. BMC Bioinformatics 8:139.

Derewacz D, Goodwin C, McNees C, McLean J, Bachmann B. 2013. Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. PNAS 110:2336–2341. [accessed 2013 Sep 16]. http://www.pnas.org/content/110/6/2336.short

Derewacz DK, Goodwin CR, Mcnees CR, Mclean JA, Bachmann BO. 2012. Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism.

Devoid S, Overbeek R, Dejongh M, Vonstein V, Best AA, Henry C. 2013. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. Alper HS, editor. METHODS Mol. Biol. 985:17–45. [accessed 2013 Feb 28]. http://link.springer.com/10.1007/978-1-62703-299-5

Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, Newman DK. 2006. The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of Pseudomonas aeruginosa. Mol. Microbiol. 61:1308–21. [accessed 2013 Oct 28]. http://www.ncbi.nlm.nih.gov/pubmed/16879411

Döring G, Conway SP, Heijerman HG, Hodson ME, Høiby N, Smyth a, Touw DJ. 2000. Antibiotic therapy against Pseudomonas aeruginosa in cystic fibrosis: a European consensus. Eur. Respir. J. 16:749–67.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl. Acad. Sci. U. S. A. 104:1777–1782.

Dubeau D, Déziel E, Woods DE, Lépine F. 2009. Burkholderia thailandensis harbors two identical rhl gene clusters responsible for the biosynthesis of rhamnolipids. BMC Microbiol. 9:263. [accessed 2013 Jun 28]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2804600&tool=pmcentrez&rendertype=abstract

Dumas ME, Kinross J, Nicholson JK. 2014. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. Gastroenterology 146:46–62.

Durot M, Bourguignon PY, Schachter V. 2009. Genome-scale models of bacterial metabolism: Reconstruction and applications. FEMS Microbiol. Rev. 33:164–190.

Dwyer DJ, Belenky P a, Yang JH, MacDonald IC, Martell JD, Takahashi N, Chan CTY, Lobritz M a, Braff D, Schwarz EG, et al. 2014. Antibiotics induce redox-related physiological alterations as part of their lethality. Proc. Natl. Acad. Sci. U. S. A.

111:E2100–9. [accessed 2014 Jul 21]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4034191&tool=pmcentrez&rendertype=abstract

Eberl L, Tümmler B. 2004. Pseudomonas aeruginosa and Burkholderia cepacia in cystic fibrosis: genome evolution, interactions and adaptation. Int. J. Med. Microbiol. 294:123–31. [accessed 2013 Sep 9]. http://www.ncbi.nlm.nih.gov/pubmed/15493822

Ebrahim A, Lerman J a, Palsson BO, Hyduke DR. 2013. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. 7:74.

Edwards JS, Palsson BO. 1999. Systems properties of the *Haemophilus influenzae Rd* Metabolic Genotype. Cell Biol. Metab. 274:17410–17416.

Eisenreich W, Dandekar T, Heesemann J, Goebel W. 2010. Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. Nat. Rev. Microbiol. 8:401–412.

Ejim L, Farha M a, Falconer SB, Wildenhain J, Coombes BK, Tyers M, Brown ED, Wright GD. 2011. Combinations of antibiotics and nonantibiotic drugs enhance antimicrobial efficacy. Nat. Chem. Biol. 7:348–350. [accessed 2011 Apr 26]. http://www.ncbi.nlm.nih.gov/pubmed/21516114

Fang K, Zhao H, Sun C, Lam CMC, Chang S, Zhang K, Panda G, Godinho M, Martins dos Santos V a PV a P, Wang J. 2011. Exploring the metabolic network of the epidemic pathogen Burkholderia cenocepacia J2315 via genome-scale reconstruction. BMC Syst. Biol. 5:83. [accessed 2011 May 27]. http://www.ncbi.nlm.nih.gov/pubmed/21609491

Fazli M, Almblad H, Rybtke ML, Givskov M, Eberl L, Tolker-Nielsen T. 2014. Regulation of biofilm formation in Pseudomonas and Burkholderia species. Environ. Microbiol. 16:1961–81. [accessed 2014 Nov 26]. http://www.ncbi.nlm.nih.gov/pubmed/24592823

Feinbaum RL, Urbach JM, Liberati NT, Djonovic S, Adonizio A, Carvunis A-R, Ausubel FM. 2012. Genome-wide identification of Pseudomonas aeruginosa virulence-related genes using a Caenorhabditis elegans infection model. PLoS Pathog. 8:e1002813. [accessed 2015 Jan 16]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3406104&tool=pmcentrez&rendertype=abstract

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3:121. [accessed 2012 Nov 6]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1911197&tool=pmcentrez&rendertype=abstract

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. 2009. Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7:129–43. [accessed 2011 Jul 18]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3119670&tool=pmcentrez&rendertype=abstract

Feist AM, Palsson BO. 2010. The biomass objective function. Curr. Opin. Microbiol. 13:344–349.

Feist AM, Palsson BØ. 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat. Biotechnol. 26:659–667.

Fletcher MP, Diggle SP, Cámara M, Williams P. 2007. Biosensor-based assays for PQS, HHQ and related 2-alkyl-4-quinolone quorum sensing signal molecules. Nat. Protoc. 2:1254–62. [accessed 2013 Feb 4]. http://www.ncbi.nlm.nih.gov/pubmed/17546021

Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. 2011. Predicting selective drug targets in cancer through metabolic networks. Mol. Syst. Biol. 7:501.

Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Høiby N, Molin S. 2012. Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nat. Rev. Microbiol. 10:841–51. [accessed 2013 May 27]. http://www.ncbi.nlm.nih.gov/pubmed/23147702

Franklin MJ, Nivens DE, Weadge JT, Howell PL. 2011. Biosynthesis of the Pseudomonas aeruginosa Extracellular Polysaccharides, Alginate, Pel, and Psl. Front. Microbiol. 2:167. [accessed 2014 Nov 25]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3159412&tool=pmcentrez&rendertype=abstract

Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E. 2011. Competitive and cooperative metabolic interactions in bacterial communities. Nat. Commun. 2:589. [accessed 2012 Mar 12]. http://www.ncbi.nlm.nih.gov/pubmed/22158444

Frimmersdorf E, Horatzek S, Pelnikevich A, Wiehlmann L, Schomburg D. 2010. How Pseudomonas aeruginosa adapts to various environments: a metabolomic approach. Environ. Microbiol. 12:1734–47. [accessed 2011 Apr 26]. http://www.ncbi.nlm.nih.gov/pubmed/20553553

Fuhrer T, Fischer E, Sauer U. 2005. Experimental identification and quantification of glucose metabolism in seven bacterial species. J. Bacteriol. 187:1581–1590.

Galet J, Deveau A, Hôtel L, Leblond P, Frey-Klett P, Aigle B. 2014. Gluconic acid-producing Pseudomonas sp. prevent γ-actinorhodin biosynthesis by Streptomyces coelicolor A3(2). Arch. Microbiol.

Gallagher LA, Ramage E, Jacobs MA, Kaul R, Brittnacher M, Manoil C. 2007. A comprehensive transposon mutant library of Francisella novicida, a bioweapon surrogate. Proc. Natl. Acad. Sci. U. S. A. 104:1009–1014.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. Affy - Analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20:307–315.

Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA. 2008. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformatics 9:43.

Giglio MG, Collmer CW, Lomax J, Ireland A. 2009. Applying the Gene Ontology in microbial annotation. Trends Microbiol. 17:262–268.

Glass S, Govan JR. 1986. Pseudomonas cepacia--fatal pulmonary infection in a patient with cystic fibrosis.

Gonzalez O, Oberwinkler T, Mansueto L, Pfeiffer F, Mendoza E, Zimmer R, Oesterhelt D. 2010. Characterization of growth and metabolism of the haloalkaliphile Natronomonas pharaonis. PLoS Comput. Biol. 6:1–10.

Govan JR, Nelson JW. 1993. Microbiology of cystic fibrosis lung infections: themes and issues. J. R. Soc. Med. 86:11–18.

Gunnarsson N, Mortensen UH, Sosio M, Nielsen J. 2004. Identification of the Entner-Doudoroff pathway in an antibiotic-producing actinomycete species. Mol. Microbiol. 52:895–902.

Haggart CR, Bartell J a, Saucerman JJ, Papin J a. 2011. Whole-genome metabolic network reconstruction and constraint-based modeling. Methods Enzymol. 500:411–33. [accessed 2011 Nov 17]. http://www.ncbi.nlm.nih.gov/pubmed/21943909

Hancock RE. 1998. Resistance mechanisms in Pseudomonas aeruginosa and other nonfermentative gram-negative bacteria. Clin. Infect. Dis. 27 Suppl 1:S93–S99.

Hao T, Ma H-W, Zhao X-M, Goryanin I. 2010. Compartmentalization of the Edinburgh Human Metabolic Network. BMC Bioinformatics 11:393.

Häußler S, Nimtz M, Domke T, Wray V, Steinmetz I. 1998. Purification and Characterization of a Cytotoxic Exolipid of Burkholderia pseudomallei Purification and Characterization of a Cytotoxic Exolipid of Burkholderia pseudomallei. 66.

Hay ID, Gatland K, Campisano A, Jordens JZ, Rehm BH a. 2009. Impact of alginate overproduction on attachment and biofilm architecture of a supermucoid Pseudomonas aeruginosa strain. Appl. Environ. Microbiol. 75:6022–6025.

Heinemann M, Sauer U. 2010. Systems biology of microbial metabolism. Curr. Opin. Microbiol. 13:337–343.

Henry CS, DeJongh M, Best A a, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28:977–82. [accessed 2014 Mar 20]. http://www.ncbi.nlm.nih.gov/pubmed/20802497

Henry CS, Zinner JF, Cohoon MP, Stevens RL. 2009. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. Genome Biol. 10:R69. [accessed 2014 Dec 3]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2718503&tool=pmcentrez&rendertype=abstract

Hernández S, Gómez A, Cedano J, Querol E. 2009. Bioinformatics annotation of the hypothetical proteins found by omics techniques can help to disclose additional virulence factors. Curr. Microbiol. 59:451–456.

Hesketh A, Hill C, Mokhtar J, Novotna G, Tran N, Bibb M, Hong H-J. 2011. Genome-wide dynamics of a bacterial response to antibiotics that target the cell envelope. BMC Genomics 12:226. [accessed 2013 Sep 11]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3123327&tool=pmcentrez&rendertype=abstract

Hesseler J, Schmidt JK, Reichl U, Flockerzi D. 2006. Coexistence in the chemostat as a result of metabolic by-products. J. Math. Biol. 53:556–84. [accessed 2012 Apr 1]. http://www.ncbi.nlm.nih.gov/pubmed/16819650

Hogardt M, Heesemann J. 2010. Adaptation of Pseudomonas aeruginosa during persistence in the cystic fibrosis lung. Int. J. Med. Microbiol. 300:557–62. [accessed 2014 Nov 19]. http://www.ncbi.nlm.nih.gov/pubmed/20943439

Hoiby N. 2006. P. aeruginosa in cystic fibrosis patients resists host defenses antibiotics. Microbe.

Holden MTG, Seth-Smith HMB, Crossman LC, Sebaihia M, Bentley SD, Cerdeño-Tárraga AM, Thomson NR, Bason N, Quail M a, Sharp S, et al. 2009. The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients. J. Bacteriol. 191:261–77. [accessed 2011 Jul 22]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2612433&tool=pmcentrez&rendertype=abstract

Holloway BW, Krishnapillai V, Morgan AF. 1979. Chromosomal genetics of Pseudomonas. Microbiol. Rev. 43:73–102.

Hood L, Heath JR, Phelps ME, Lin B. 2004. Systems biology and new technologies enable predictive and preventative medicine. Science 306:640–643.

Horsley A, Webb K, Bright-Thomas R, Govan J, Jones A. 2011. Can early Burkholderia cepacia complex infection in cystic fibrosis be eradicated with antibiotic therapy? Front. Cell. Infect. Microbiol. 1:18. [accessed 2013 Jul 8]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3417365&tool=pmcentrez&rendertype=abstract

Hosie AHF, Allaway D, Poole PS. 2002. A Monocarboxylate Permease of Rhizobium leguminosarum Is the First Member of a New Subfamily of Transporters A Monocarboxylate Permease of Rhizobium leguminosarum Is the First Member of a New Subfamily of Transporters. 184.

Huthmacher C, Hoppe A, Bulik S, Holzhütter H-G. 2010. Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis. BMC Syst. Biol. 4:120.

Ibarra RU, Edwards JS, Palsson BO. 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420:20–23.

Imamovic L, Sommer MO a. 2013. Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. Sci. Transl. Med. 5:204ra132. [accessed 2013 Oct 17]. http://www.ncbi.nlm.nih.gov/pubmed/24068739

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11:97–108. [accessed 2013 Mar 2]. http://www.ncbi.nlm.nih.gov/pubmed/20051986

Irizarry R a, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–64.

Isles A, Maclusky I, Corey M, Gold R, Prober C, Fleming P, Levison H. 1984. Pseudomonas cepacia infection in cystic fibrosis: an emerging problem. J. Pediatr. 104:206–210.

Jacobs M a, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, et al. 2003. Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. U. S. A. 100:14339–44.

Jamshidi N, Palsson BØ. 2007. Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. BMC Syst. Biol. 1:26.

Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V. 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys. J. 95:1487–99. [accessed 2012 Mar 26].

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2479599&tool=pmcentrez&rendertype=abstract

Jelsbak L, Johansen HK, Frost A-L, Thøgersen R, Thomsen LE, Ciofu O, Yang L, Haagensen J a J, Høiby N, Molin S. 2007. Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. Infect. Immun. 75:2214–24. [accessed 2013 Jan 6]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1865789&tool=pmcentrez&rendertype=abstract

Jensen P a, Lutz K a, Papin J a. 2011. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst. Biol. 5:147. [accessed 2011 Sep 24]. http://www.biomedcentral.com/1752-0509/5/147

Jensen PA, Papin JA. 2014. MetDraw : automated visualization of genome-scale metabolic network reconstructions and associated high-throughput data. Bioinformatics 30:1–2.

Jensen PR, Hammer K. 1993. Minimal requirements for exponential growth of Lactococcus lactis. Appl. Environ. Microbiol. 59:4363–4366.

Jerby L, Shlomi T, Ruppin E. 2010. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol. Syst. Biol. 6:401.

Jones a M, Dodd ME, Govan JRW, Barcus V, Doherty CJ, Morris J, Webb a K. 2004. Burkholderia cenocepacia and Burkholderia multivorans: influence on survival in cystic fibrosis. Thorax 59:948–51. [accessed 2014 Dec 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1746874&tool=pmcentrez&rendertype=abstract

Juhas M, Stark M, von Mering C, Lumjiaktase P, Crook DW, Valvano M a, Eberl L. 2012. High confidence prediction of essential genes in burkholderia cenocepacia. PLoS One 7:e40064. [accessed 2012 Jul 23]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3386938&tool=pmcentrez&rendertype=abstract

Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L, et al. 2009. Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. Brief. Bioinform. 11:40–79.

Karr JR, Sanghvi JC, Macklin DN, Gutschow M V, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW. 2012. A whole-cell computational model predicts phenotype from genotype. Cell 150:389–401.

Kauffman KJ, Prakash P, Edwards JS. 2003. Advances in flux balance analysis. Curr. Opin. Biotechnol. 14:491–496.

Kehrenberg C, Schwarz S. 2005. dfrA20 , a Novel Trimethoprim Resistance Gene from Pasteurella multocida dfrA20 , a Novel Trimethoprim Resistance Gene from Pasteurella multocida. 49.

Keith KE, Killip L, He P, Moran GR, Valvano M a. 2007. Burkholderia cenocepacia C5424 produces a pigment with antioxidant properties using a homogentisate intermediate. J. Bacteriol. 189:9057–65. [accessed 2013 Sep 25]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2168628&tool=pmcentrez&rendertype=abstract

Kim S, Lieberman TD, Kishony R. 2014. Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. Proc. Natl. Acad. Sci. U. S. A. 111. [accessed 2014 Sep 24]. http://www.ncbi.nlm.nih.gov/pubmed/25246554

Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. 2009. FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal. Chem. 81:10038–10048.

King JD, Kocíncová D, Westman EL, Lam JS. 2009. Review: Lipopolysaccharide biosynthesis in Pseudomonas aeruginosa. Innate Immun. 15:261–312. [accessed 2013 Oct 17]. http://www.ncbi.nlm.nih.gov/pubmed/19710102

Kircher M, Kelso J. 2010. High-throughput DNA sequencing - Concepts and limitations. BioEssays 32:524–536.

Kiviharju K, Salonen K, Moilanen U, Meskanen E, Leisola M, Eerikäinen T. 2007. On-line biomass measurements in bioreactor cultivations: Comparison study of two on-line probes. J. Ind. Microbiol. Biotechnol. 34:561–566.

Klitgord N, Segre D, Segrè D, Papin JA. 2010. Environments that Induce Synthetic Microbial Ecosystems. Public Library of Science San Francisco, USA. [accessed 2011 Jan 6]. http://dx.plos.org/10.1371/journal.pcbi.1001002

Knorr AL, Jain R, Srivastava R. 2007. Bayesian-based selection of metabolic objective functions. Bioinformatics 23:351–7. [accessed 2013 Aug 7]. http://www.ncbi.nlm.nih.gov/pubmed/17150997

Ko S, An H, Bang JH, Park S-W. 2015. An outbreak of Burkholderia cepacia complex pseudobacteremia associated with intrinsically contaminated commercial 0.5% chlorhexidine solution. Am. J. Infect. Control 43:266–268.

Kohanski M, Dwyer D, Collins J, Manuscript A. 2010. How antibiotics kill bacteria: from targets to networks. Nat. Rev. Microbiol. 8:423–435. [accessed 2013 Sep 4]. http://www.nature.com/nrmicro/journal/vaop/ncurrent/full/nrmicro2333.html

Kohler PR a, Zheng JY, Schoffers E, Rossbach S. 2010. Inositol catabolism, a key pathway in sinorhizobium meliloti for competitive host nodulation. Appl. Environ. Microbiol. 76:7972–80. [accessed 2014 Dec 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3008233&tool=pmcentrez&rendertype=abstract

Kristensen VN, Lingjærde OC, Russnes HG, Vollan HKM, Frigessi A, Børresen-Dale A-L. 2014. Principles and methods of integrative genomic analyses in cancer. Nat. Rev. Cancer 14:299–313.

Kumar A, Suthers PF, Maranas CD. 2012. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics 13:6. [accessed 2012 Mar 19]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3277463&tool=pmcentrez&rendertype=abstract

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4:1073–81. [accessed 2014 Mar 20]. http://www.ncbi.nlm.nih.gov/pubmed/19561590

Kuzniar A, van Ham RCHJ, Pongor S, Leunissen JAM. 2008. The quest for orthologs: finding the corresponding gene across genomes. Trends Genet. 24:539–551.

Kyrtopoulos SA. 2013. Making sense of OMICS data in population-based environmental health studies. Environ. Mol. Mutagen. 54:468–479.

Lam JS, Taylor VL, Islam ST, Hao Y, Kocíncová D. 2011. Genetic and Functional Diversity of Pseudomonas aeruginosa Lipopolysaccharide. Front. Microbiol. 2:118. [accessed 2011 Jul 28]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3108286&tool=pmcentrez&rendertype=abstract

Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, Barraclough TG. 2012. Species interactions alter evolutionary responses to a novel environment. PLoS Biol. 10:e1001330. [accessed 2012 Jul 16]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3352820&tool=pmcentrez&rendertype=abstract

Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, Diggins LT, He J, Saucier M, Déziel E, et al. 2006. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. Genome Biol. 7:R90. [accessed 2012 Mar 9]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1794565&tool=pmcentrez&rendertype=abstract

Lee D-S, Burd H, Liu J, Almaas E, Wiest O, Barabási A-L, Oltvai ZN, Kapatral V. 2009. Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple Staphylococcus aureus genomes identify novel antimicrobial drug targets. J. Bacteriol. 191:4015–24. [accessed 2012 Mar 12]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2698402&tool=pmcentrez&rendertype=abstract

Lee JM, Gianchandani EP, Papin JA. 2006. Flux balance analysis in the era of metabolomics. Brief. Bioinform. 7:140–150.

Lerman J a, Hyduke DR, Latif H, Portnoy V a, Lewis NE, Orth JD, Schrimpe-Rutledge AC, Smith RD, Adkins JN, Zengler K, et al. 2012. In silico method for modelling metabolism and gene product expression at genome scale. Nat. Commun. 3:929. [accessed 2012 Oct 29]. http://www.ncbi.nlm.nih.gov/pubmed/22760628

Lewenza S, Visser MB, Sokol PA. 2002. Interspecies communication between Burkholderia cepacia and Pseudomonas aeruginosa.

Lewis NE, Nagarajan H, Palsson BO. 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10:291–305. [accessed 2012 Mar 2]. http://www.ncbi.nlm.nih.gov/pubmed/22367118

Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T, Ausubel FM. 2006. An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proc. Natl. Acad. Sci. U. S. A. 103:2833–8.

Liu GY, Nizet V. 2009. Color me bad: microbial pigments as virulence factors. Trends Microbiol. 17:406–13. [accessed 2013 Feb 16]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2743764&tool=pmcentrez&rendertype=abstract

Liu Y, Imlay J a. 2013. Cell death from antibiotics without the involvement of reactive oxygen species. Science 339:1210–3. [accessed 2014 Aug 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3731989&tool=pmcentrez&rendertype=abstract

Loutet S a, Valvano M a. 2010a. A decade of Burkholderia cenocepacia virulence determinant research. Infect. Immun. 78:4088–100. [accessed 2012 Jul 19]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2950345&tool=pmcentrez&rendertype=abstract

Loutet S a, Valvano M a. 2010b. A decade of Burkholderia cenocepacia virulence determinant research. Infect. Immun. 78:4088–100. [accessed 2013 Sep 9]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2950345&tool=pmcentrez&rendertype=abstract

Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM. 2009. Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol. 5:296.

Machado D, Herrgård M. 2014. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. Maranas CD, editor. PLoS Comput. Biol. 10:e1003580. [accessed 2014 Apr 24]. http://dx.plos.org/10.1371/journal.pcbi.1003580

Madhusudhan KT, Luo J, Sokatch JR. 1999. In vitro transcriptional studies of the bkd operon of Pseudomonas putida: L-branched-chain amino acids and D-leucine are the inducers. J. Bacteriol. 181:2889–2894.

Mahadevan R, Edwards JS, Doyle FJ. 2002. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys. J. 83:1331–40. [accessed 2011 Jun 30]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1302231&tool=pmcentrez&rendertype=abstract

Mahadevan R, Schilling CH. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5:264–276. [accessed 2014 Jul 17]. http://linkinghub.elsevier.com/retrieve/pii/S1096717603000582

Mahenthiralingam E, Baldwin a, Dowson CG. 2008. Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. J. Appl. Microbiol. 104:1539–51. [accessed 2014 Mar 24]. http://www.ncbi.nlm.nih.gov/pubmed/18217926

Mahenthiralingam E, Baldwin A, Vandamme P. 2002. Burkholderia cepacia complex infection in patients with cystic fibrosis. J. Med. Microbiol. 51:533–538.

Mahenthiralingam E, Urban T a, Goldberg JB. 2005. The multifarious, multireplicon Burkholderia cepacia complex. Nat. Rev. Microbiol. 3:144–56. [accessed 2011 Feb 3]. http://www.ncbi.nlm.nih.gov/pubmed/15643431

Mahenthiralingam E, Vandamme P, Campbell ME, Henry D a, Gravelle a M, Wong LT, Davidson a G, Wilcox PG, Nakielna B, Speert DP. 2001. Infection with Burkholderia cepacia complex genomovars in patients with cystic fibrosis: virulent transmissible strains of genomovar III can replace Burkholderia multivorans. Clin. Infect. Dis. 33:1469–75.

Marolda CL, Hauröder B, John MA, Michel R, Valvano MA. 1999. Intracellular survival and saprophytic growth of isolates from the Burkholderia cepacia complex in free-living amoebae. Microbiology 145:1509–1517.

Martin LW, Reid DW, Sharples KJ, Lamont IL. 2011. Pseudomonas siderophores in the sputum of patients with cystic fibrosis. BioMetals 24:1059–1067.

Martínez JL, Rojo F. 2011. Metabolic regulation of antibiotic resistance. FEMS Microbiol. Rev. 35:768–89. [accessed 2014 Jul 11]. http://www.ncbi.nlm.nih.gov/pubmed/21645016

Marvig RL, Johansen HK, Molin S, Jelsbak L. 2013. Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. Hughes D, editor. PLoS Genet. 9:e1003741. [accessed 2013 Sep 16]. http://dx.plos.org/10.1371/journal.pgen.1003741

Marvig RL, Sommer LM, Molin S, Johansen HK. 2014. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat. Genet. [accessed 2014 Nov 18]. http://www.ncbi.nlm.nih.gov/pubmed/25401299

Maspoli A, Wenner N, Mislin GL a, Reimmann C. 2014. Functional analysis of pyochelin-/enantiopyochelin-related genes from a pathogenicity island of Pseudomonas aeruginosa strain PA14. BioMetals 27:559–573.

Mavrodi D V, Parejko J a, Mavrodi O V, Kwak Y-S, Weller DM, Blankenfeldt W, Thomashow LS. 2013. Recent insights into the diversity, frequency and ecological roles of phenazines in fluorescent Pseudomonas spp. Environ. Microbiol. 15:675–86. [accessed 2014 Mar 20]. http://www.ncbi.nlm.nih.gov/pubmed/22882648

Meyer JM, Neely a, Stintzi a, Georges C, Holder I a, Meyer J, Neely A, Stintzi A, Georges C. 1996. Pyoverdin is essential for virulence of Pseudomonas aeruginosa . These include : Pyoverdin Is Essential for Virulence of Pseudomonas aeruginosa. Infect. Immun. 64:518–523.

Mitchell A, Romano GH, Groisman B, Yona A, Dekel E, Kupiec M, Dahan O, Pilpel Y. 2009. Adaptive prediction of environmental changes by microorganisms. Nature 460:220–4. [accessed 2012 Mar 2]. http://www.ncbi.nlm.nih.gov/pubmed/19536156

Moehring RW, Lewis SS, Isaacs PJ, Schell W a, Thomann WR, Althaus MM, Hazen KC, Dicks K V, Lipuma JJ, Chen LF, et al. 2014. Outbreak of bacteremia due to Burkholderia contaminans linked to intravenous fentanyl from an institutional compounding pharmacy. JAMA Intern Med 174:606–612.

Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD. 2013. Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-speci fi c adaptations to nutritional environments.

Mukhopadhyay S, Thomason MK, Lentz S, Nolan N, Willner K, Gee JE, Glass MB, Inglis TJJ, Merritt A, Levy A, et al. 2010. High-redundancy draft sequencing of 15 clinical and environmental Burkholderia strains. J. Bacteriol. 192:6313–6314.

Mulcahy LR, Burns JL, Lory S, Lewis K. 2010. Emergence of Pseudomonas aeruginosa strains producing high levels of persister cells in patients with cystic fibrosis. J. Bacteriol. 192:6191–9. [accessed 2012 Aug 22]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2981199&tool=pmcentrez&rendertype=abstract

Müller A, Thomas GH, Horler R, Brannigan J a, Blagova E, Levdikov VM, Fogg MJ, Wilson KS, Wilkinson AJ. 2005. An ATP-binding cassette-type cysteine transporter in Campylobacter jejuni inferred from the structure of an extracytoplasmic solute receptor protein. Mol. Microbiol. 57:143–55. [accessed 2012 Aug 9]. http://www.ncbi.nlm.nih.gov/pubmed/15948956

Nakahigashi K, Toya Y, Ishii N, Soga T, Hasegawa M, Watanabe H, Takai Y, Honma M, Mori H, Tomita M. 2009. Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol. Syst. Biol. 5:306.

Nguyen D, Singh PK. 2006. Evolving stealth: genetic adaptation of Pseudomonas aeruginosa during cystic fibrosis infections. Proc. Natl. Acad. Sci. U. S. A. 103:8305–8306.

Niittylae T, Chaudhuri B, Sauer U, Frommer WB. 2009. Comparison of quantitative metabolite imaging tools and carbon-13 techniques for fluxomics. Methods Mol. Biol. 553:355–372.

Nogales J, Canales A, Jiménez-Barbero J, Serra B, Pingarrón JM, García JL, Díaz E. 2011. Unravelling the gallic acid degradation pathway in bacteria: the gal cluster from Pseudomonas putida. Mol. Microbiol. 79:359–74. [accessed 2013 Oct 30]. http://www.ncbi.nlm.nih.gov/pubmed/21219457

Van Noorden CJF. 2010. Imaging enzymes at work: metabolic mapping by enzyme histochemistry. J. Histochem. Cytochem. 58:481–497.

Notebaart R a., Szappanos B, Kintses B, Pal F, Gyorkei a., Bogos B, Lazar V, Spohn R, Csorg B, Wagner a., et al. 2014. Network-level architecture and the evolutionary potential of underground metabolism. Proc. Natl. Acad. Sci. 111:11762–7.

Nzula S. 2002. Influence of taxonomic status on the in vitro antimicrobial susceptibility of the Burkholderia cepacia complex. J. Antimicrob. Chemother. 50:265–269. [accessed 2013 Aug 21]. http://www.jac.oupjournals.org/cgi/doi/10.1093/jac/dkf137

Oberhardt M a, Goldberg JB, Hogardt M, Papin J a. 2010. Metabolic Network Analysis of Pseudomonas aeruginosa during Chronic Cystic Fibrosis Lung Infection. J. Bacteriol. 192:5534–5548. [accessed 2010 Aug 17]. http://www.ncbi.nlm.nih.gov/pubmed/20709898

Oberhardt M a, Palsson BØ, Papin J a. 2009. Applications of genome-scale metabolic reconstructions. Mol. Syst. Biol. 5:320.

Oberhardt M a, Puchałka J, Fryer KE, Martins dos Santos V a P, Papin J a. 2008. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. J. Bacteriol. 190:2790–803.

Oberhardt M a, Puchałka J, Martins dos Santos V a P, Papin J a. 2011. Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. PLoS Comput. Biol. 7:e1001116. [accessed 2014 Dec 10]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3068926&tool=pmcentrez&rendertype=abstract

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 27:29–34.

Ogawa J, Shimizu S. 1994. Beta-ureidopropionase with N-carbamoyl-alpha-L-amino acid amidohydrolase activity from an aerobic bacterium, Pseudomonas putida IFO 12996. Eur. J. Biochem. 223:625–30.

Ohman DE, Chakrabarty AM. 1982. Utilization of human respiratory secretions by mucoid Pseudomonas aeruginosa of cystic fibrosis origin. Infect. Immun. 37:662–669.

Orth JD, Thiele I, Palsson BØ. 2010. What is flux balance analysis? Nat. Biotechnol. 28:245–8. [accessed 2013 Aug 7]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3108565&tool=pmcentrez&rendertype=abstract

Palmer GC, Palmer KL, Jorth P a, Whiteley M. 2010. Characterization of the Pseudomonas aeruginosa transcriptional response to phenylalanine and tyrosine. J. Bacteriol. 192:2722–8. [accessed 2012 Jul 16]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2876504&tool=pmcentrez&rendertype=abstract

Palmer KL, Aye LM, Whiteley M. 2007. Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum. J. Bacteriol. 189:8079–87. [accessed 2011 May 16]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2168676&tool=pmcentrez&rendertype=abstract

Palmer KL, Mashburn LM, Singh PK, Whiteley M. 2005. Cystic Fibrosis Sputum Supports Growth and Cues Key Aspects of Pseudomonas aeruginosa Physiology. Society 187:5267–5277.

Pamp SJ, Tolker-Nielsen T. 2007. Multiple roles of biosurfactants in structural biofilm development by Pseudomonas aeruginosa. J. Bacteriol. 189:2531–9.

Papin J a, Price ND, Edwards JS, Palsson B BØ. 2002. The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy. J. Theor. Biol. 215:67–82. [accessed 2012 Mar 8]. http://www.ncbi.nlm.nih.gov/pubmed/12051985

Patil KR, Rocha I, Förster J, Nielsen J. 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6:308.

Patrick WM, Quandt EM, Swartzlander DB, Matsumura I. 2007. Multicopy suppression underpins metabolic evolvability. Mol. Biol. Evol. 24:2716–2722.

Pharkya P, Burgard AP, Maranas CD. 2004. OptStrain: A computational framework for redesign of microbial production systems. Genome Res. 14:2367–2376.

Pharkya P, Maranas CD. 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8:1–13.

Pinzon NM, Ju LK. 2009. Analysis of rhamnolipid biosurfactants by methylene blue complexation. Appl. Microbiol. Biotechnol. 82:975–981.

Plata G, Hsiao T-L, Olszewski KL, Llinás M, Vitkup D. 2010. Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Mol. Syst. Biol. 6:408.

Poblete-Castro I, Binger D, Rodrigues A, Becker J, Martins Dos Santos V a P, Wittmann C. 2013. In-silico-driven metabolic engineering of Pseudomonas putida for enhanced production of poly-hydroxyalkanoates. Metab. Eng. 15:113–23. [accessed 2013 Aug 8]. http://www.ncbi.nlm.nih.gov/pubmed/23164576

Pohl S, Klockgether J, Eckweiler D, Khaledi A, Schniederjans M, Chouvarine P, Tümmler B, Häussler S. 2014. The extensive set of accessory Pseudomonas aeruginosa genomic components. FEMS Microbiol. Lett. 356:235–241.

Preidis G a., Versalovic J. 2009. Targeting the Human Microbiome With Antibiotics, Probiotics, and Prebiotics: Gastroenterology Enters the Metagenomics Era. Gastroenterology 136:2015–2031.

Price ND, Papin J a, Schilling CH, Palsson BO. 2003. Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol. 21:162–169. [accessed 2011 Jun 3]. http://linkinghub.elsevier.com/retrieve/pii/S0167779903000301

Price ND, Schellenberger J, Palsson BO. 2004. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. Biophys. J. 87:2172–86. [accessed 2014 Jul 22].

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1304643&tool=pmcentrez&rendertype=abstract

Price-Whelan A, Dietrich LEP, Newman DK. 2007. Pyocyanin alters redox homeostasis and carbon flux through central metabolic pathways in Pseudomonas aeruginosa PA14. J. Bacteriol. 189:6372–81. [accessed 2014 Mar 20]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1951912&tool=pmcentrez&rendertype=abstract

Proctor R a., Kriegeskorte A, Kahl BC, Becker K, LÃ¶ffler B, Peters G. 2014. Staphylococcus aureus Small Colony Variants (SCVs): a road map for the metabolic pathways involved in persistent infections. Front. Cell. Infect. Microbiol. 4:1–8.

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

Raman K, Chandra N. 2009. Flux balance analysis of biological systems: Applications and challenges. Brief. Bioinform. 10:435–449.

Rangel-vega A, Bernstein LR, Mandujano-Tinoco E-A, Garcia-Contreras S-J, Garcia-Contreras R. 2015. Drug repurposing as an alternative for the treatment of recalcitrant bacterial infections. Front. Microbiol. 6.

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

Rau MH, Marvig RL, Ehrlich GD, Molin S, Jelsbak L. 2012 Jun 5. Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. Environ. Microbiol. [accessed 2012 Jun 12]. http://www.ncbi.nlm.nih.gov/pubmed/22672046

Raymond CK, Sims EH, Kas a., Spencer DH, Kutyavin T V., Ivey RG, Zhou Y, Kaul R, Clendenning JB, Olson M V. 2002. Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas aeruginosa. J. Bacteriol. 184:3614–3622.

Reed JL, Vo TD, Schilling CH, Palsson BO. 2003. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol. 4:R54.

Rice LB. 2008. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. J. Infect. Dis. 197:1079–81. [accessed 2013 Jun 1]. http://www.ncbi.nlm.nih.gov/pubmed/18419525

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype – phenotype interactions. Nat. Rev. Genet. 16:85–97.

Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M. 2010. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst. Biol. 4:45.

Rodríguez-Rojas A, Oliver A, Blázquez J. 2012. Intrinsic and environmental mutagenesis drive diversification and persistence of Pseudomonas aeruginosa in chronic lung infections. J. Infect. Dis. 205:121–7. [accessed 2012 Mar 9]. http://www.ncbi.nlm.nih.gov/pubmed/22080096

Ryssel M, Hviid A-MM, Dawish MS, Haaber J, Hammer K, Martinussen J, Kilstrup M. 2014 Aug 20. Multi-stress resistance in Lactococcus lactis is actually escape from purine induced stress sensitivity. Microbiology. [accessed 2014 Aug 26]. http://www.ncbi.nlm.nih.gov/pubmed/25143058

Samant S, Lee H, Ghassemi M, Chen J, Cook JL, Mankin AS, Neyfakh A a. 2008. Nucleotide biosynthesis is critical for growth of bacteria in human blood. PLoS Pathog. 4:e37. [accessed 2014 Aug 8]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2242838&tool=pmcentrez&rendertype=abstract

Sass A, Marchbank A, Tullis E, Lipuma JJ, Mahenthiralingam E. 2011. Spontaneous and evolutionary changes in the antibiotic resistance of Burkholderia cenocepacia observed by global gene expression analysis. BMC Genomics 12:373. [accessed 2012 Jul 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3155924&tool=pmcentrez&rendertype=abstract

Sass AM, Schmerk C, Agnoli K, Norville PJ, Eberl L, Valvano MA, Mahenthiralingam E. 2013. The unexpected discovery of a novel low-oxygen-activated locus for the anoxic persistence of Burkholderia cenocepacia. ISME J.:1–14.

Saulnier DM, Santos F, Roos S, Mistretta T-A, Spinler JK, Molenaar D, Teusink B, Versalovic J. 2011. Exploring Metabolic Pathway Reconstruction and Genome-Wide Expression Profiling in Lactobacillus reuteri to Define Functional Probiotic Features. Gibas C, editor. PLoS One 6:e18783. [accessed 2011 Apr 29]. http://dx.plos.org/10.1371/journal.pone.0018783

Savinell JM, Palsson BO. 1992a. Optimal selection of metabolic fluxes for in vivo measurement. II. Application to Escherichia coli and hybridoma cell metabolism. J. Theor. Biol. 155:215–42.

Savinell JM, Palsson BO. 1992b. Optimal selection of metabolic fluxes for in vivo measurement. I. Development of mathematical methods. J. Theor. Biol. 155:201–14.

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, et al. 2011. Quantitative prediction of cellular

metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6:1290–1307.

Schuetz R, Kuepfer L, Sauer U. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol. Syst. Biol. 3:119. [accessed 2012 Nov 3]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1949037&tool=pmcentrez&rendertype=abstract

Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. 2012. Multidimensional Optimality of Microbial Metabolism. Science (80-.). 336:601–604. [accessed 2012 May 3]. http://www.sciencemag.org/cgi/doi/10.1126/science.1216882

Schwibbert K, Marin-Sanguino A, Bagyan I, Heidrich G, Lentzen G, Seitz H, Rampp M, Schuster SC, Klenk H-P, Pfeiffer F, et al. 2011. A blueprint of ectoine metabolism from the genome of the industrial producer Halomonas elongata DSM 2581 T. Environ. Microbiol. 13:1973–94. [accessed 2013 Mar 1]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3187862&tool=pmcentrez&rendertype=abstract

Segrè D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U. S. A. 99:15112–15117.

Senger RS. 2010. Biofuel production improvement with genome-scale models: The role of cell composition. Biotechnol. J. 5:671–685.

Shlomi T, Berkman O, Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. U. S. A. 102:7695–7700.

Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. 2008. Network-based prediction of human tissue-specific metabolism. Nat. Biotechnol. 26:1003–10. [accessed 2013 Aug 18]. http://www.ncbi.nlm.nih.gov/pubmed/18711341

Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. 2013. Understanding the interactions between bacteria in the human gut through metabolic modeling. Sci. Rep. 3:2532. [accessed 2014 Dec 4]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3755282&tool=pmcentrez&rendertype=abstract

Shrout JD, Chopp DL, Just CL, Hentzer M, Givskov M, Parsek MR. 2006. The impact of quorum sensing and swarming motility on Pseudomonas aeruginosa biofilm formation is nutritionally conditional. Mol. Microbiol. 62:1264–77. [accessed 2012 Mar 10]. http://www.ncbi.nlm.nih.gov/pubmed/17059568

Sigurdsson G, Fleming RMT, Heinken A, Thiele I. 2012. A systems biology approach to drug targets in Pseudomonas aeruginosa biofilm. PLoS One 7:e34337. [accessed

2013 Jun 9]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3327687&tool=pmcentrez&rendertype=abstract

Singh V, Brecik M, Mukherjee R, Evans JC, Svetlíková Z, Blaško J, Surade S, Blackburn J, Warner DF, Mikušová K, et al. 2015. The Complex Mechanism of Antimycobacterial Action of 5-Fluorouracil. Chem. Biol. 22:63–75.

Slama TG. 2008. Gram-negative antibiotic resistance : there is a price to pay. 7.

Smallbone K, Simeonidis E. 2009. Flux balance analysis: A geometric perspective. J. Theor. Biol. 258:311–315.

Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio D a, Miller SI, Ramsey BW, Speert DP, Moskowitz SM, et al. 2006. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. Proc. Natl. Acad. Sci. U. S. A. 103:8487–92.

Smyth G. 2005. limma: Linear Models for Microarray Data. In: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. p. 397–420.

Snell GI, De Hoyos A, Krajden M, Winton T, Maurer JR. 1993. Pseudomonas cepacia in lung transplant recipients with cystic fibrosis. Chest 103:466–471.

Sousa SA, Feliciano JR, Pinheiro PF, Leitão JH. 2013. Biochemical and Functional Studies on the Burkholderia cepacia Complex bceN Gene, Encoding a GDP-D-Mannose 4,6-Dehydratase. PLoS One 8.

Staib L, Fuchs TM. 2014. From food to cell: Nutrient exploitation strategies of enteropathogens. Microbiol. (United Kingdom) 160:1020–1039.

Stanier RY, Palleroni NJ, Doudoroff M. 1966. The aerobic pseudomonads: a taxonomic study. J. Gen. Microbiol. 43:159–271.

Suárez-Moreno ZR, Caballero-Mellado J, Coutinho BG, Mendonça-Previato L, James EK, Venturi V. 2012. Common features of environmental and potentially beneficial plant-associated Burkholderia. Microb. Ecol. 63:249–66. [accessed 2013 Jun 12]. http://www.ncbi.nlm.nih.gov/pubmed/21850446

Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, Maranas CD. 2009. Genome-scale metabolic reconstruction Of mycoplasma genitalium, iPS189. PLoS Comput. Biol. 5.

Tarighi S, Wei Q, Cámara M, Williams P, Fletcher MP, Kajander T, Cornelis P. 2008. The PA4204 gene encodes a periplasmic gluconolactonase (PpgL) which is important for fitness of Pseudomonas aeruginosa. Microbiology 154:2979–90. [accessed 2014 Dec 13]. http://www.ncbi.nlm.nih.gov/pubmed/18832304

Taylor CJ, Anderson a J, Wilkinson SG. 1998. Phenotypic variation of lipid composition in Burkholderia cepacia: a response to increased growth temperature is a greater content of 2-hydroxy acids in phosphatidylethanolamine and ornithine amide lipid. Microbiology 144 (Pt 7:1737–45.

Thiele I, Palsson BØ. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5:93–121. [accessed 2013 Sep 20]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3125167&tool=pmcentrez&rendertype=abstract

Thøgersen JC, Mørup M, Damkiær S, Molin S, Jelsbak L. 2013. Archetypal analysis of diverse Pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis airways. BMC Bioinformatics 14:279.

Torella JP, Chait R, Kishony R. 2010. Optimal drug synergy in antimicrobial treatments. PLoS Comput. Biol. 6:e1000796. [accessed 2011 Jan 14]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2880566&tool=pmcentrez&rendertype=abstract

Traxler M, Watrous J, Alexandrov T. 2013. Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome. MBio 4:e00459–13. [accessed 2014 Feb 6]. http://mbio.asm.org/content/4/4/e00459-13.short

Valentine CD, Zhang H, Phuan P-W, Nguyen J, Verkman a S, Haggie PM. 2013 Aug 2. Small Molecule Screen Yields Inhibitors of Pseudomonas Homoserine Lactone-Induced Host Responses. Cell. Microbiol. [accessed 2013 Aug 28]. http://www.ncbi.nlm.nih.gov/pubmed/23910799

Vandamme P, Coenye T. 2004. Taxonomy of the genus Cupriavidus: a tale of lost and found. Int. J. Syst. Evol. Microbiol. 54:2285–9. [accessed 2013 Jan 31]. http://www.ncbi.nlm.nih.gov/pubmed/15545472

Varga JJ, Losada L, Zelazny AM, Kim M, McCorrison J, Brinkac L, Sampaio EP, Greenberg DE, Singh I, Heiner C, et al. 2013. Draft Genome Sequences of Burkholderia cenocepacia ET12 Lineage Strains K56-2 and BC7. Genome Announc. 1:e00841–13.

Varma A, Palsson BO. 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Nat. Biotechnol. 12:994–998. [accessed 2011 Jul 2]. http://www.nature.com/nbt/journal/v12/n10/abs/nbt1094-994.html

Vrijheid M. 2014. The exposome: a new paradigm to study the impact of environment on health. Thorax 0:1–3.

Wessel AK, Hmelo L, Parsek MR, Whiteley M. 2013. Going local: technologies for exploring bacterial microenvironments. Nat. Rev. Microbiol. 11:337–48. [accessed 2013 Aug 10]. http://www.ncbi.nlm.nih.gov/pubmed/23588251

Wilder CN, Diggle SP, Schuster M. 2011. Cooperation and cheating in Pseudomonas aeruginosa: the roles of the las, rhl and pqs quorum-sensing systems. ISME J. 5:1332–43. [accessed 2012 Jul 19]. http://www.ncbi.nlm.nih.gov/pubmed/21368905

Williams BJ, Du R-H, Calcutt MW, Abdolrasulnia R, Christman BW, Blackwell TS. 2010. Discovery of an operon that participates in agmatine metabolism and regulates biofilm formation in Pseudomonas aeruginosa. Mol. Microbiol. 76:104–19. [accessed 2010 Oct 6]. http://www.ncbi.nlm.nih.gov/pubmed/20149107

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

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

Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock REW, Brinkman FSL. 2011. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res. 39:D596–600. [accessed 2013 Oct 29]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013766&tool=pmcentrez&rendertype=abstract

Wintermute EH, Silver P a. 2010. Dynamics in the mixed microbial concourse. Genes Dev. 24:2603–14. [accessed 2010 Dec 4]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2994034&tool=pmcentrez&rendertype=abstract

Wolkenhauer O. 2014. Why model? Front. Physiol. 5 JAN:1–5.

Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild H-H, Nielsen C, Brunak S, Knudsen S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol. 3:research0048.

Worlitzsch D, Tarran R, Ulrich M, Schwab U, Cekici A, Meyer KC, Birrer P, Bellon G, Berger J, Weiss T, et al. 2002. Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. J. Clin. Invest. 109:317–325.

Wurtzel O, Yoder-Himes DR, Han K, Dandekar A a, Edelheit S, Greenberg EP, Sorek R, Lory S. 2012. The single-nucleotide resolution transcriptome of Pseudomonas

aeruginosa grown in body temperature. PLoS Pathog. 8:e1002945. [accessed 2013 Aug 10]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3460626&tool=pmcentrez&rendertype=abstract

Xavier JB, Martinez-Garcia E, Foster KR. 2009. Social evolution of spatial patterns in bacterial biofilms: when conflict drives disorder. Am. Nat. 174:1–12. [accessed 2014 Dec 13]. http://www.ncbi.nlm.nih.gov/pubmed/19456262

Yang J, Chen L, Sun L, Yu J, Jin Q. 2008. VFDB 2008 release: An enhanced web-based resource for comparative pathogenomics. Nucleic Acids Res. 36:539–542.

Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, et al. 2011. Evolutionary dynamics of bacteria in a human host environment. Proc. Natl. Acad. Sci. 108:7481–7486. [accessed 2011 Apr 26]. http://www.pnas.org/cgi/doi/10.1073/pnas.1018249108

Yang L, Nilsson M, Gjermansen M, Givskov M, Tolker-Nielsen T. 2009. Pyoverdine and PQS mediated subpopulation interactions involved in Pseudomonas aeruginosa biofilm formation. Mol. Microbiol. 74:1380–92. [accessed 2013 Aug 25]. http://www.ncbi.nlm.nih.gov/pubmed/19889094

Yang Y, Gilbert D, Kim S. 2009. Annotation confidence score for genome annotation: A genome comparison approach. Bioinformatics 26:22–29.

Yeterian E, Martin LW, Guillon L, Journet L, Lamont IL, Schalk IJ. 2010. Synthesis of the siderophore pyoverdine in Pseudomonas aeruginosa involves a periplasmic maturation. Amino Acids 38:1447–59. [accessed 2015 Jan 21]. http://www.ncbi.nlm.nih.gov/pubmed/19787431

Yoder-Himes DR, Konstantinidis KT, Tiedje JM. 2010. Identification of potential therapeutic targets for Burkholderia cenocepacia by comparative transcriptomics. PLoS One 5:e8724. [accessed 2012 Jul 13]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2806911&tool=pmcentrez&rendertype=abstract

Yudistira H, McClarty L, Bloodworth R a M, Hammond S a, Butcher H, Mark BL, Cardona ST. 2011. Phenylalanine induces Burkholderia cenocepacia phenylacetic acid catabolism through degradation to phenylacetyl-CoA in synthetic cystic fibrosis sputum medium. Microb. Pathog. 51:186–93. [accessed 2013 Sep 9]. http://www.ncbi.nlm.nih.gov/pubmed/21511027

Yuhara S, Komatsu H, Goto H, Ohtsubo Y, Nagata Y, Tsuda M. 2008. Pleiotropic roles of iron-responsive transcriptional regulator Fur in Burkholderia multivorans. Microbiology 154:1763–74. [accessed 2013 Sep 23]. http://www.ncbi.nlm.nih.gov/pubmed/18524931

Zeise L, Bois FY, Chiu W a., Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environ. Health Perspect. 121:23–31.

Zengler K, Palsson BO. 2012. A road map for the development of community systems (CoSy) biology. Nat. Rev. Microbiol. 10:366–372. [accessed 2012 Mar 27]. http://www.nature.com/doifinder/10.1038/nrmicro2763

Zhang WC, Shyh-Chang N, Yang H, Rai A, Umashankar S, Ma S, Soh BS, Sun LL, Tai BC, Nga ME, et al. 2012. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. Cell 148:259–72. [accessed 2014 Jun 3]. http://www.ncbi.nlm.nih.gov/pubmed/22225612

Zhao J, Schloss PD, Kalikin LM, Carmody L a, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, et al. 2012a. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc. Natl. Acad. Sci. U. S. A. 109:5809–14. [accessed 2014 Jan 29]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3326496&tool=pmcentrez&rendertype=abstract

Zhao J, Schloss PD, Kalikin LM, Carmody L a, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, et al. 2012b. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc. Natl. Acad. Sci. U. S. A. 109:5809–14. [accessed 2014 Nov 16]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3326496&tool=pmcentrez&rendertype=abstract

Zlosnik JE a, Speert DP. 2010. The role of mucoidy in virulence of bacteria from the Burkholderia cepacia complex: a systematic proteomic and transcriptomic analysis. J. Infect. Dis. 202:770–81. [accessed 2012 Jul 13]. http://www.ncbi.nlm.nih.gov/pubmed/20670172

Zomorrodi AR, Islam MM, Maranas CD. 2013. d - OptCom: Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities.

Zomorrodi AR, Maranas CD. 2012. OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. Rao C V., editor. PLoS Comput. Biol. 8:e1002363. [accessed 2012 Feb 2]. http://dx.plos.org/10.1371/journal.pcbi.1002363

Zulianello L, Canard C, Köhler T, Caille D, Lacroix J-S, Meda P. 2006. Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by Pseudomonas aeruginosa. Infect. Immun. 74:3134–47. [accessed 2013 May 23]. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1479292&tool=pmcentrez&rendertype=abstract

Zur H, Ruppin E, Shlomi T. 2010. iMAT: an integrative metabolic analysis tool. Bioinformatics 26:3140–2. [accessed 2013 Feb 19]. http://www.ncbi.nlm.nih.gov/pubmed/21081510