ESSENTIAL GENES AND GENOMES OF THE BURKHOLDERIA CEPACIA COMPLEX

By

Ruhullah Alec Michael Bloodworth

A Thesis Submitted to the Faculty of Graduate Studies

of

The University of Manitoba

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Microbiology

University of Manitoba

Winnipeg

Copyright © 2015 Ruhullah Bloodworth

Abstract

The *Burkholderia cepacia* complex (Bcc) are a group of closely related species known for their intrinsic multidrug resistance, large multipart genomes and ability to infect people with cystic fibrosis. The clinical relevance of the Bcc and their large multipart genomes make the study of their essential genes of broad interest. Essential genes are those required for survival in standard laboratory conditions this makes them potential targets for novel antibiotics against a group of species where few existing antibiotics are effective. Furthermore, while essential gene studies have been carried out in a number of bacterial species, only one of these species had multiple chromosomes and none had a genome as large as the Bcc. In my research I identified essential genes in *B. cenocepacia* K56-2, a member of the Bcc, by using transposon mutagenesis to deliver a rhamnose inducible promoter randomly into the genome and screening for a conditional growth (CG) phenotype. The utility of the CG mutant library was confirmed by showing that, when grown in suboptimal concentrations of rhamnose, only mutants that under-expressed the target of the antibiotic were hypersensitive. The CG mutant library included transposon insertions upstream from widely conserved, well-characterized essential genes suggesting that the system is capable of recovering essential gene mutants. A number of genes with either no or mixed records of essentiality in other microorganisms were also recovered. Among these was one of the three electron transfer flavoproteins (ETFs) in *B. cenocepacia*. The ETFs are a family of proteins found in a large number of eukaryotic, archaeal and bacterial species, which are required for the metabolism of specific substrates or for symbiotic nitrogen fixation in some bacteria. Despite these non-essential functions, high throughput screens have identified ETFs as

putatively essential in several species. I showed that ETF expression is required for both viability and growth both on complex media and on media containing a variety of single carbon sources. Furthermore, cells depleted of ETF were determined to be nonviable and the morphologic shape of the cells changed from short rods to small spheres. In depth studies of essential genes are only possible for organisms with sequenced genomes. Of the 18 named species that currently comprise the Bcc, only 7 have been sequenced limiting the possibility of cross species comparative genomics. Therefore, I have assembled the first draft genomes of *B. contaminans* isolates, a species that has emerged as the dominant Bcc species recovered from the CF populations of Argentina and Spain. Identifying and characterizing essential genes in the Bcc, and sequencing additional Bcc species for comparative genomics are important first steps in understanding these clinically important bacteria.

Acknowledgements

I am particularly grateful to my supervisor Dr Silvia T. Cardona for her mentorship and support. The importance she places not only on the work that we do but also the people carrying it out has made working in her lab a pleasure. Her audacity and intellectual curiosity have profoundly changed the way I approach science.

I would like to offer my special thanks to the members of my committee Drs Karen Brassinga, Ivan Oresnik and Michael Domaratzki your support and suggestions have been greatly appreciated.

Finally I would like to thank my labmates you have all enriched my time here; Robyn and Jason for welcoming me into the lab and patiently teaching me how to accomplish the most basic tasks, Harry for keeping things metal and Ijeme for the R&B and Fela Kuti, Tanya for giving me the chance to be the optimistic person for a change, Brijesh for an extraordinary willingness to help out and accommodate other, Tasia and Silvina for consistently entertaining diversions, and finally April and Jackie for working with me on a project through all the frustrating finicky troubleshooting.

Contributions of Authors

Sections 2, 3 and 4 are the work of multiple authors. In all cases, my supervisor Silvia T. Cardona was involved in the design and interpretation of experiments and had a significant role in editing and writing manuscripts. In addition, she carried out sensitization index experiments described in section 3.2.7. April S. Gislason is responsible for a significant fraction of the mutants in the conditional mutant library described in section 2.2.4. and much of the enhanced sensitivity assay described in section 2.2.8. Soumaya Zlitni taught Silvia T. Cardona the sensitization index protocol described in section 3.2.7 and assisted in carrying out this experiment. The original technique and experiment were carried out in the laboratory of Eric D. Brown. Carrie Selin isolated the genomes for the strains sequenced in section 4. The *B. contaminans* strains came from the Bcc culture collection of José Degrossi and species identification was performed by Ma. Agustina López De Volder.

List of Copyrighted Material for which Permission was Obtained

Chapter 2. MicrobiologyOpen. 2013 April; 2(2):243-258. doi: 10.1002/mbo3.71. Title: *Burkholderia cenocepacia* conditional growth mutant library created by random promoter replacement of essential genes. (Entire paper). Published under the Creative Commons Attribution License, allowing reproduction with attribution

Chapter 3. Microbiology. In Press. Title: An electron transfer flavoprotein (ETF) is essential for viability and its depletion causes a rod-to-sphere change in *Burkholderia cenocepacia*. (Entire

paper). Society for General Microbiology License to Publish grants the author of an article the right to reproduce the article in any work for which they are the author.

Table of Contents

Abstract ii

Acknowledgements iv

Contributions of Authors v

List of Copyrighted Material for which Permission was Obtained v

List of tables xii

List of figures xii

List of abbreviations xv

1 Introduction 1

1.1 The *Burkholderia cepacia* complex 1

1.1.1 Taxonomy of the Bcc 3

1.1.2 Human pathogenicity 4

1.1.3 Ecology of the Bcc 11

1.1.4 Genomes of the Bcc 13

1.2 Essential Genes 17

1.2.1 Definitions of essential genes 17

1.2.2 Experimental identification of essential genes 19

1.2.3 Characteristics of essential genes..... 25

1.2.4 Essential genes and evolution 31

1.2.5 Minimal cells in synthetic biology..... 34

1.2.6 Essential gene products as antibiotic targets..... 36

1.3 Essential genes in *Burkholderia cenocepacia*..... 41

1.3.1 The mini-Tn5 transposon..... 41

1.3.2 The rhamnose inducible promoter system 43

1.4 Thesis objectives..... 44

2 *Burkholderia cenocepacia* conditional growth mutant library created by random promoter replacement of essential genes..... 46

2.1 Introduction..... 46

2.2 Experimental Procedures 49

2.2.1 Bacterial strains and growth conditions..... 49

2.2.2 Molecular biology techniques..... 52

2.2.3 Vector constructions 52

2.2.4 CG mutant library construction 53

2.2.5 Determination of transposon insertion sites 54

2.2.6 Comparisons with essential genes in other bacteria 55

2.2.7 Functional characterization of genes in putatively essential operons 56

2.2.8 Enhanced sensitivity assay 56

2.3 Results 57

2.3.1 Building a CG mutant library in *B. cenocepacia* K56-2 57

2.3.2 Functional characterization of essential operons 70

2.3.3 Analyzing the rate at which new essential operons were discovered 85

2.3.4 Global analysis of essential gene expression levels. 87

2.3.5 Analysis of gene redundancy in *B. cenocepacia.* 92

2.3.6 CG mutants demonstrate selective hypersensitivity at low rhamnose concentrations. 96

2.4 Discussion 100

3 An electron transfer flavoprotein (ETF) is essential for viability and its depletion causes a rod-to-sphere change in *Burkholderia cenocepacia* 104

3.1 Introduction 104

3.2 Experimental Procedures 108

3.2.1 Bacterial strains and growth conditions 108

3.2.2 Molecular biology techniques 113

3.2.3 Construction of an unmarked *etfdh2* deletion 116

3.2.4 Construction of *etfBA* and *etfdh1* conditional mutants 116

3.2.5 Construction of complementing plasmids 117

3.2.6 Microscopy analysis 117

3.2.7 Bacterial two-hybrid assay 118

3.2.8 Sensitization index experiments 118

3.2.9 Biolog assays 119

3.2.10 Bioinformatics 119

3.3 Results 120

3.3.1 The *B. cenocepacia etfBA* operon is essential 120

3.3.2 The essentiality of ETF is independent of the carbon source utilized 127

3.3.3 Depletion of EtfBA renders cells nonviable 128

3.3.4 Lack of ETF causes a rod-to-sphere change in cell morphology that is independent of MreB 134

3.3.5 The putative partner of ETF, ETF dehydrogenase also induces a rod-to-sphere change in morphology..... 140

3.4 Discussion..... 144

3.5 Acknowledgements..... 149

4 Draft genome Sequences of the *Burkholderia contaminans* strains LMG 23361 and FFH2055 and *Burkholderia cenocepacia* K56-2..... 150

4.1 Introduction..... 150

4.2 Experimental Procedures..... 152

4.2.1 Genome sequencing and annotation..... 152

4.2.2 Nucleotide accession numbers:..... 153

4.2.3 Assembly comparison with a Bcc reference genome..... 153

4.3 Results..... 153

4.4 Discussion..... 165

5 Conclusions and Future directions..... 165

6 References..... 172

List of tables

Table 2-1 Bacterial strains and plasmids 50

Table 2-2 Primers..... 51

Table 2-3 CG mutants included in the CG mutant library..... 65

Table 2-4 Genes found in this study with essential orthologs in *E. coli* and *P. aeruginosa* 73

Table 2-5 Burkholderia species-specific putative essential operons 76

Table 3-1 List of strains and plasmids used in this study 110

Table 3-2 Primers..... 114

Table 3-3 Abs600nm after 24h growth on Biolog PM1 sole carbon sources..... 125

Table 3-4 Change in viability of conditional growth mutants over 24 hours 131

Table 4-1 Bacterial strains 152

Table 4-2 Summary of assemblies..... 156

List of figures

Figure 1-1 Identifying essential genes via transposon mutagenesis 22

Figure 2-1. Construction of a *B. cenocepacia* CG mutant library. 61

Figure 2-2. The number of essential genes as a function of genome size. 62

Figure 2-3. Histogram of growth for 115 CG mutants in the absence of rhamnose. 64

Figure 2-4. Mutant growth defects and approximate location of insertion sites. 81

Figure 2-5. Transposon insertions relative to the start of relevant coding sequence. 82

Figure 2-6. Functional categories of genes in putative essential operons. 84

Figure 2-7. The rate at which new operons were discovered. 86

Figure 2-8. The distribution of gene expression in putatively essential and nonessential genes.. 91

Figure 2-9 Genetic duplication in *Burkholderia cenocepacia* J2315, *Pseudomonas aeruginosa* PA01, and *Escherichia coli* K12 as a function of DNA sequence similarity. 94

Figure 2-10 Conditional growth mutants show selective hypersensitivity. 98

Figure 3-1 A composite figure of electron paths through ETFs. 107

Figure 3-2 The *etfBA* operon is essential for growth on both complex media and single carbon sources 122

Figure 3-3 EtfA and EtfB interact with each other but there is no evidence that either interacts with themselves. 124

Figure 3-4 *etfBA* expression is required for viability. 130

Figure 3-5 Cells depleted of EtfBA maintain intact membranes. 133

Figure 3-6 Lack of *etfBA* expression causes a rod-to-sphere change in cell shape. 137

Figure 3-7 Lack of *etfBA* or *etfdh1* expression is not equivalent to loss of MreB function. 138

Figure 3-8 The effect of low levels of ETF on sensitivity to growth inhibitors. 139

Figure 3-9 Lack of putative ETF dehydrogenases causes a growth defect. 141

Figure 3-10 There is little change in cell morphology after depleting other electron transport proteins. 142

Figure 3-11 Depletion of cytochrome bc (29-3B1) or ATPAse (30-4C6) does not lead to smaller rounder cells. 143

Figure 3-12 Phylogentic trees of EtfA (a) and EtfB (b) proteins in *Burkholderia* spp. containing putatively essential ETF genes..... 148

Figure 4-1 Distribution of contig lengths..... 157

Figure 4-2 Alignments of assemblies against the closest completed genome sequence. 162

Figure 4-3 Coverage of PacBio reads across select contigs 164

List of abbreviations

Bcc	*Burkholderia cepacia* complex
Bp	Base pair
cAMP	Cyclic adenosine monophosphate
CDS	Coding DNA sequence
CF	Cystic fibrosis
CG	Conditional growth
CGD	Chronic granulomatous disease
DNA	Deoxyribonucleic acid
G	Giga
HGAP	Hierarchical Genome Assembly Process
IS	Insertion sequence
K	Kilo
M	Mega

ORF	Open reading frame
RAPD	Random Amplified Polymorphic DNA
RBH	Reciprocal Best Hit
wt	Wild Type

1 Introduction

1.1 The *Burkholderia cepacia* complex

The *Burkholderia cepacia* complex (Bcc) is a group of gram negative β-proteobacteria that are characterized by ambiguity, complexity and malleability. They belong to the genus *Burkholderia* which contains over 60 species ranging from benign nitrogen fixing plant symbiotes to *B. pseudomallei*, a dangerous human pathogen and select agent (Coenye & Vandamme., 2003). Bcc is the umbrella term used for a group of closely related *Burkholderia* species that exist at the intersection between nonpathogenic environmental isolates, plant growth enhancing rhizosphere isolates, plant pathogens and human opportunistic pathogens. Like all *Burkholderia* species the Bcc have large genomes with multiple circular chromosomes (Wattam *et al*., 2014). This large genomic capability gives the Bcc enormous metabolic versatility (Mahenthiralingam *et al*., 2005), but also harbors large numbers of insertion sequences and is subject to frequent genomic rearrangements (Ussery *et al*., 2009). This genomic repertoire and malleability may contribute to the ambiguous relationship the Bcc have with plants, capable of causing disease but also promoting plant growth often by different isolates of the same species (Parke & Gurian-Sherman., 2001, Vandamme & Dawyndt., 2011). Despite their potential for promoting plant growth and degrading pollutants (Mahenthiralingam *et al*., 2005), the biotechnological potential of the Bcc is limited by their ability to cause disease in the immunocompromised especially people with CF (cystic fibrosis) and CGD (Chronic Granulomatous Disease). For people with both diseases, infections with the Bcc are associated

with worsened clinical outcomes than infections with other bacterial species, but outcomes vary widely even among patients infected with apparently identical isolates.

This work primarily concerns the identification of essential genes in *B. cenocepacia* K56-2, followed by characterization of one of those genes in detail. Essential genes are those required for growth usually in some standard permissive laboratory conditions, though there have been some recent efforts to looking at environments that more closely mimic the human host,(Lee *et al*., 2015, Turner *et al*., 2015). Their absolute indispensability for growth makes essential genes potential targets for novel antibiotics. Members of the Bcc are intrinsically resistant to many antibiotics dramatically limiting treatment options and making the development of effective anti *Burkholderia* drugs clinically relevant. At the time this thesis research started, no essential gene studies had been carried out in bacteria with large multipart genomes so little was known about how they compare to commonly studied bacteria which typically have a single smaller chromosome. *B. cenocepacia* K56-2 in particular was chosen because unlike other CF pathogens, Bcc bacteria are capable of patient-to-patient spread, and K56-2 is a CF isolate belonging to the ET-12 linage which was responsible for the largest Bcc outbreak in CF patients in Canada and the United Kingdom (Mahenthiralingam *et al*., 2000, Mahenthiralingam *et al*., 2005). Since these initial outbreaks, strict infection control procedures have been introduced leading to a change in the epidemiology of Bcc infections of CF patients. In most countries the proportion of *B. multivorans* has been growing at the expense of *B. cenocepacia*, but in Argentina (Martina *et al*., 2013) and Spain (Medina-Pascual *et al*., 2015) *B. contaminans* has become the most prevalent species of the Bcc found in CF sputum. Then, a second aim was to

sequence the first *B. contaminans* isolates with the long-term goal of essential gene characterization in this emerging Bcc pathogen.

1.1.1 Taxonomy of the Bcc

The heterogeneity and genetic complexity of the Bcc has made accurate identification of isolates difficult and resulted in a number of taxonomic reorganizations. The type species for *Burkholderia*, then called *Pseudomonas cepacia*, was discovered in 1950 by W.H. Burkholder as the causative agent of sour skin rot in onions (Burkholder, 1950). By 1992, the availability of 16S rRNA sequence data made it increasingly apparent that many of the species in the *Pseudomonas* genus were not closely related, leading to a reorganization and the creation of the new genus *Burkholderia*, with *B. cepacia* as the type species (Yabuuchi *et al*., 1992, Yabuuchi *et al*., 1995). Over the same period of time, *Burkholderia* infections among CF patients were becoming a serious concern, but identification and classification was complicated by high levels of heterogeneity among the supposed "*B. cepacia*" isolates (Simpson *et al*., 1994, Tabacchioni *et al*., 1995). This prompted a polyphasic taxonomic study of clinical and environmental isolates by Vandamme *et al.* (Vandamme *et al*., 1997) which determined that the "*B.* cepacia" isolates actually belonged to at least 5 distinct genomovars, where genomovars refers to genetically distinct species which cannot yet be reliably distinguished phenotypically. This collection of genomovars was then collectively termed the *Burkholderia cepacia* complex. Subsequent research on the Bcc has identified phenotypic characteristics to distinguish the genomovars, and they have been reclassified as full species, in particular *B. cepacia* genomovar III was classified

as *B. cenocepacia* in 2003 (Vandamme *et al*., 2003). The number of species in the Bcc has also increased and there are currently 17 named species (Vandamme & Dawyndt., 2011). The plasticity of the Bcc genome means that while the species have very high levels of 16S rRNA (> 98%), and recA (94–95%) sequence similarity, whole genome comparisons and DNA-DNA hybridization shows only moderate similarity. Recently Vandamme *et al.* (Vandamme & Dawyndt., 2011) have proposed that, given the biochemical similarities between species, as well as the existence of many intermediate values for DNA-DNA hybridization and Multi Locus Sequence Typing types, the Bcc is not a complex collection of defined genetic species but rather a genetic continuum without sharp divisions. The currently named species would then represent fixed points on this underlying continuum allowing for the rational characterization and comparison of new isolates rather than representing discrete biological entities. Additional whole genome sequencing and comparative genomics especially of environmental isolates are necessary to examine the full genetic diversity of the Bcc.

1.1.2 Human pathogenicity

While the first Bcc was identified as a plant pathogen and was then known for having an extraordinary metabolic versatility (Mahenthiralingam *et al*., 2005). Today the Bcc are best known as opportunistic human pathogens. Somewhat ironically this change is almost exclusively due to improved medical care. Unlike *Burkholderia pseudomallei* which is capable of causing serious disease in healthy humans (Galyov *et al*., 2010), the Bcc are only capable of infecting people with compromised immune systems. These vulnerable populations are protected by

preventative measures (e.g. disinfectants) to block direct contact with infectious bacteria and the use of antibiotics to treat infections once they occur. The Bcc are difficult to control by both these methods. They are intrinsically resistant to many antibiotics (Nzula *et al*., 2002). In addition disinfectants meant to kill bacteria have been determined to serve as a source of Bcc contamination (Garcia-Erce *et al*., 2002, Kutty *et al*., 2007).

Outbreaks of Bcc infections in hospitals have been documented since at least the 1970s when "*P. cepacia"* was identified as a source of hospital-acquired pneumonia (Speller *et al*., 1971, Weinstein *et al*., 1973). The source of the contamination was eventually traced to contaminated deionized water used to prepare anesthetics (Schaffner *et al*., 1973). Subsequent Bcc outbreaks have generally followed to the same pattern. While patient-to-patient transmission has been reported (Vardi *et al*., 2013), the majority of infections appear to be due to contamination of a common product. These include medical solutions like nebulizer solution (Balkhy *et al*., 2005), compounded medications like the aforementioned anesthetics or fentanyl (Moehring *et al*., 2014) as well as disinfectants like alcohol free mouthwash (Kutty *et al*., 2007), and chlorhexidine (Garcia-Erce *et al*., 2002). Mortality rates vary widely depending on the condition of the patients and the method of contamination. There were no reported fatalities in the first cases of contaminated anesthetics reported in the 1970s (Schaffner *et al*., 1973), but an outbreak among mechanically ventilated intensive care unit patients in Italy resulted in significantly higher mortality rates than the overall population (Righi *et al*., 2013). While these hospital-acquired Bcc infections seem to be due to occasional failures of infection control

especially involving the use of non-sterile water, the Bcc are a constant and serious threat to people with chronic granulomatous disease (CGD) (Maignan *et al*., 2013) or cystic fibrosis (CF) (Lipuma, 2010). These two genetic illnesses are associated with infections by a relatively defined set of pathogens including the Bcc, which are especially virulent in these populations.

1.1.2.1 Chronic Granulomatous Disease

CGD is a group of rare genetic disorders of the innate immune system affecting 1/200,000 to 1/250,000 live births in the USA (Winkelstein *et al*., 2000). These genetic disorders result in low to nonexistent levels of activity by NADPH oxidase. As functional NADPH oxidase is required by phagocytes to generate pathogen-killing superoxide bursts, people with CGD are highly susceptible to bacterial and fungal infections (Heyworth *et al*., 2003). These recurring bacterial and fungal infections are the leading cause of morbidity and mortality among people with CGD (Winkelstein *et al*., 2000). Improved and more aggressive infection management has improved patient outcomes over the past decade but the median age at death is still just 28.1 years (Marciano *et al*., 2014).

Despite having a generalized defect in the innate immune system, patients with CGD are primarily infected with a relatively narrow range of pathogens, which includes *Burkholderia* species (Winkelstein *et al*., 2000). The taxonomy and epidemiology of *Burkholderia* infections in CGD are not well studied. Infections are usually caused by members of the Bcc and *Burkholderia gladioli* (Greenberg *et al*., 2009), though *Burkholderia glumae* (Weinberg *et al*.,

2007) and *Burkholderia pseudomallei* (Renella *et al*., 2006) infections have been reported. Species and strains vary widely between patients, with no particular species dominating (Greenberg *et al*., 2009). For example an 11 year study of 18 patients revealed that *B. multivorans* was the most commonly identified species, comprising 26 of the 50 *Burkholderia* isolates. Nevertheless in all patients with *Burkholderia* positive cultures for more than 1 year, multiple species were recovered. Infections appear to be acute rather than chronic with 26% of patients being reinfected with *Burkholderia* after testing negative. Interestingly 90% of these reoccurrences are due to new strains suggesting continuous acquisition from some presumably environmental source (Guide *et al*., 2003).

Burkholderia infections in CGD patients are particularly virulent and often present as pneumonia with concurrent septicemia (Winkelstein *et al*., 2000). This extreme virulence means that even though the occurrences of *Burkholderia* infections among CGD patients is not particularly high (1.06 infections per 100 patient years), they are still the leading bacterial cause of death (Marciano *et al*., 2014). The Bcc alone are responsible for 18% of the total CGD deaths on the American CGD registry (Winkelstein *et al*., 2000). The reason for the extreme virulence of *Burkholderia* in people with CGD is not well understood. One clue comes from a 2005 study of *B. cenocepacia*, which showed that it is capable of inducing necrosis in neutrophils lacking superoxide production, but not in healthy neutrophils (Bylund *et al*., 2005).

1.1.2.2 Cystic Fibrosis

Given the rarity of CGD, and the emphasis on infection control in hospitals, the vast majority of people infected by Bcc have CF. CF is the most common genetic disease in Caucasians occurring in approximately 1 in 2,500 live births (Dupuis *et al*., 2005). CF is caused by a variety of loss-of-function mutations in the gene that encodes for the cystic fibrosis transmembrane conductance regulator (CFTR) (Cuthbert, 2011), an ATP dependent anion channel primarily found in epithelial cells. Lack of functional CFTR impairs both salt transport and fluid secretion disrupting the osmotic balance of mucus layers on epithelial cells resulting in a build-up of thick viscous mucous on the surface of the airway, digestive system, reproductive tracts and other internal epithelial surfaces (Goss & Burns., 2007). With proper nutritional support, CF is not usually the direct cause of death. Rather CF patients have chronic and persistent airway infections probably due to impaired ciliary function and slow mucus clearance (Randell *et al*., 2006). These infections in turn induce periodic episodes of inflammation, termed exacerbations, which damage the lung (Goss & Burns., 2007). Repeated exacerbations result in a cumulative loss of lung function and eventually death.

Like CGD, CF infections are caused by a relatively small group of fungal and bacterial pathogens that includes *Burkholderia* (Lipuma, 2010). *Burkholderia* infections of the CF airway were identified as an emerging concern in the mid-1980s (Isles *et al*., 1984). It was quickly recognized that unlike other CF pathogens, Bcc infections could be particularly virulent, with approximately 10% of patients experiencing a rapid loss of lung function accompanied by

necrotizing pneumonia and septicemia known as ‘cepacia syndrome’ (Isles *et al*., 1984, Weidmann *et al*., 2008). The spread of Bcc through the CF population along with the potentially severe consequences of Bcc infections was a major factor driving taxonomy and epidemiology of the Bcc. Also, unlike other CF pathogens, members of the Bcc are capable of patient-to-patient spread, with a few epidemic strains, primarily of *B. cenocepacia,* responsible for many of the infections (Govan *et al*., 1993, LiPuma *et al*., 1990). The implementation of stringent infection control practices including segregating *Burkholderia* positive patients has greatly limited the spread of Bcc infections. The current prevalence of Bcc in CF patients range from approximately 3.1% of the patients in the USA CF registry (Razvi *et al*., 2009) to 11% in Canada (Stephenson *et al*., 2014); However Bcc infections are still associated with worsened patient outcomes (Courtney *et al*., 2004, Ledson *et al*., 2002).

The vast majority of *Burkholderia* infections in CF patients are due to members of the Bcc, all of which have been recovered from CF sputum samples (Campana *et al*., 2004, Reik R *et al*., 2005). Of the non-Bcc *Burkholderia* species, only *B. gladioli* is identified in a significant proportion of cases (Kennedy *et al*., 2007, Lipuma, 2010), though *B. fungorum* (Coenye *et al*., 2001) and *B. pseudomallei* (Corral *et al*., 2008) infections have also been reported. While all members of the Bcc are capable of colonizing the CF airway studies in the USA (Reik R *et al*., 2005), Canada (Speert *et al*., 2002) and Europe (Bevivino *et al*., 2002, Brisse *et al*., 2004, Cunha *et al*., 2003, Drevinek *et al*., 2003, Norskov-Lauritsen *et al*., 2010) have consistently identified *B. cenocepacia* and *B. multivorans* as the most common species, with the exception of Argentina

(Martina *et al*., 2013) and Spain (Medina-Pascual *et al*., 2015) where *B. contaminans* predominates. At least in the case of Spain, the prevalence of *B. contaminans* is relatively recent. Before 2008 the distribution of Bcc species in Spanish CF patients was similar to that seen in most other countries with *B. cenocepacia* and *B. multivorans* being the most common Bcc isolates. Since 2008, the incidence of *B. contaminans* has steadily increased and it now accounts for 36.5% of Bcc infections which is approximately twice that of the next most common species, *B. cenocepacia* (17.7%). In Argentina, the incidence of *B. contaminans* is even higher than it is in Spain, accounting for 57.6% of Bcc isolates (Martina *et al*., 2013) and its prevalence appears to have started earlier, perhaps in the years 2002-2006 (Jorda-Vargas *et al*., 2008, Minan *et al*., 2009). Elsewhere, the prevalence of *B. cenocepacia* is thought to be due to patient-to-patient transmission (Baldwin *et al*., 2005, Govan *et al*., 1993, LiPuma *et al*., 1990, Mahenthiralingam *et al*., 2005) with disproportionate representation by three epidemic strains, ET-12 (Johnson *et al*., 1994), Midwest (Kumar *et al*., 1997) and PHDC (Chen *et al*., 2001). *B. cenocepacia* ET-12 is primarily found in Canada and the United Kingdom where it was responsible for arguably the largest Bcc outbreak in CF patients (Johnson *et al*., 1994, Sun *et al*., 1995).

Given its prevalence and the early emergence of epidemic strains, *B. cenocepacia* is one of the most studied members of the Bcc. The outcome of colonization by *B. cenocepacia* ranges from apparently asymptomatic carriage to 'cepacia syndrome' (Weidmann *et al*., 2008). The combination of pathogen and host factors responsible for the outcomes of individual patients is unknown as clonal isolates have been recovered from patients presenting with a wide range of

symptoms (Courtney *et al*., 2004, Govan *et al*., 1993, Jones *et al*., 2004, Ledson *et al*., 2002). However, on aggregate infections with *B. cenocepacia*, are associated with greater loss of lung function, require more frequent interventions and have higher mortality than patients without *Burkholderia* infections (Courtney *et al*., 2004, Ledson *et al*., 2002). In addition, *B. cenocepacia* and to a lesser extent *B. gladioli* infections are associated with heightened post lung transplant mortality (Alexander *et al*., 2008, Murray *et al*., 2008), leading to patients infected with these species being excluded from lung transplants in many centers (Kotloff & Thabut., 2011). Treatment options for Bcc infections are limited as most CF isolates are resistant to most clinically relevant antibiotics (Leitao *et al*., 2008, Nzula *et al*., 2002). Minocycline is the most effective antibiotic inhibiting growth of just 38% of strains and 2 antibiotic combinations demonstrate synergy in less than 15% of strains (Zhou *et al*., 2007). This urgent clinical need for antibiotics effective against the Bcc is major reason for my essential gene study in *B. cenocepacia.*

1.1.3 Ecology of the Bcc

While most interest in the Bcc is driven by their roles as opportunistic human pathogens, they are only capable of infecting the immunocompromised. As modern medical advances increasingly protect this vulnerable population from more immediate threats, less immediately virulent bacteria like the Bcc become an increasing concern. Understanding their natural ecology is important for identifying the source of the Bcc that infect humans, but it is also the context in which the Bcc evolved the features that preadapted them for colonizing humans.

Members of the Bcc are found in diverse environments including, water from streams (Fang *et al*., 2011b, Olapade *et al*., 2005, Vermis *et al*., 2003), coastal sea water (Maravic *et al*., 2012), and industrial disinfectants (Garcia-Erce *et al*., 2002); however, the Bcc are primarily soil bacteria (Coenye & Vandamme., 2003, Vial *et al*., 2011). Bcc species are routinely identified in bulk soil samples but their distribution is uneven and they are not particularly abundant (Ramette *et al*., 2005). For example, in an investigation of soil from 91 urban sites (Miller *et al*., 2002) PCR on DNA isolated from soil, identified members of the Bcc in 82% of the samples but Bcc colonies could only be cultured from 15%. The Bcc are substantially more abundant in the rhizosphere, or the surface of and soil immediately adjacent to roots, where they are typically found at 10^3 to 10^7 cfu/g of root (Pirone *et al*., 2005, Ramette *et al*., 2005, Tsuchiya *et al*., 1995). Culture dependent (Pallud *et al*., 2001) and independent (Chelius & Triplett., 2001) studies with maize have confirmed that Bcc species are not just present in greater numbers near plant roots but they also form a greater proportion of the total bacterial population, becoming the most common β-proteobacteria (Chelius & Triplett., 2001). While on aggregate the Bcc can comprise a substantial portion of the bacteria within the rhizosphere, like in bulk soil samples, the prevalence of Bcc in the can vary dramatically between samples. In a study of maize the total counts viable Bcc isolates isolated from adjacent plants varied 140 fold (Ramette *et al*., 2005).

As might be expected given their close proximity, the Bcc interact closely with plants. The nature of these interactions is both host and strain dependent. Plant pathogenesis, especially

the ability to induce onion rot (Jacobs *et al*., 2008), is observed from some but not all field isolates. Plant beneficial interactions are also seen in the Bcc including suppression of fungal pathogen growth (Bevivino *et al*., 2005), and increased plant yields (Van *et al*., 2000). The deliberate use of Bcc to protect and enhance plant growth is unadvised, as there appears to be no clear distinction between environmental and human pathogenic Bcc. Many traits thought to contribute to human pathogenesis appear to have evolved in the context of this close association with plants and competition with other bacteria including intrinsic resistance to multiple antibiotics for which no difference can be found between environmental and clinical isolates (Nzula *et al*., 2002). Furthermore, strains isolated from CF sputum commonly match environmental isolates supporting the theory that the rhizosphere is the natural reservoir of the Bcc strains that infect humans (Baldwin *et al*., 2007, Jacobs *et al*., 2008, LiPuma *et al*., 2002). Current infection control practices in the United States and Canada have dramatically reduced the patient-to-patient spread of highly transmissible Bcc strains. The majority of infected CF patients now harbor genetically distinct strains (Lipuma, 2010), indicating that they have been acquired independently probably from some environmental source.

1.1.4 Genomes of the Bcc

The adaptability of the Bcc and their metabolic versatility is underpinned by unusually large and complex genomes. *Escherichia coli* and *Bacillus subtilis* are among the two most studied bacterial species and thus have become models have become models for what constitutes a normal bacterial genome. In both species, the genome consists of between 4 and 5 Mb of

encoding approximately 4,000 protein coding genes and organized in a single circular chromosome (Blattner *et al*., 1997, Kunst *et al*., 1997). In contrast, the genomes of the Bcc range from 7 to 9 Mb and are among the largest known bacterial genomes (Agnoli *et al*., 2012). These large genomes are divided into 3 circular chromosomes (Winsor *et al*., 2008). The first and largest chromosome is typically between 3 and 4 Mb, the second is closer to 3 Mb and the third is significantly smaller usually around 1 Mb. This multipartite genomic structure is seen in other bacteria, especially members of the α-proteobacteria, with the smaller replicons being termed "megaplasmids" or chromosomes (Mackenzie *et al*., 2004). Since the size and types of genes found on "megaplasmids" and "accessory chromosomes" can overlap, if they are to be rigorously distinguished it must be on the basis of their dispensability. Under this definition megaplasmids can be successfully deleted in at least some conditions while chromosomes are indispensible for bacterial growth. In practice, the nomenclature used seems to depend on tradition within a given research community. For instance pSymB of *Sinorhiboium melilotti* is termed a megaplasmid despite containing two essential genes (Milunovic *et al*., 2014) whereas chromosome 3 of the Bcc has been shown to be dispensable in at least 7 species (Agnoli *et al*., 2012).

Genome sequences are available for 35 strains of the Bcc covering 7 distinct species. These have revealed that in addition to their large size and multipart architecture, a significant portion of Bcc genomes consists of genomic islands. Genomic islands are relatively large segments of DNA that are transmitted horizontally, often differing between closely related strains (Juhas *et al*., 2009). They often carry genes involved in functions such as pathogenicity, symbiosis, accessory metabolism and antibiotic resistance. In the case of *B. cenocepacia* J2315, a member of the same ET-12 linage as the *B. cenocepacia* K56-2 strain used in this research,

approximately 10% of the genome is occupied by 14 genomic islands distributed among all three chromosomes (Holden *et al*., 2009). Genomic islands comprise a similar portion of the genomes of other Bcc species, for example the genomic island predictor island viewer (Dhillon *et al*., 2015) predicts that genomic islands occupy approximately 9% of the *B. multivorans* ATCC 17616 genome (Winsor *et al*., 2008). A study of marine bacteria found that 0-12% of their genomes are comprised of genomic islands (Fernandez-Gomez *et al*., 2012), so the 10% seen in *Bcc* species is high but certainly not unique. The mobile nature of these elements means that while they are often found in multiple *Burkholderia* species, strains within the same species often differ significantly. The genomic islands identified in *B. cenocepacia* J2315 are missing from other sequenced non-epidemic *B. cenocepacia* strains and include a pathogenicity island already known to have a role in Bcc virulence (Baldwin *et al*., 2004). The relatively large pool of mobile pan-*Burkholderia* elements transferred horizontally between bacteria supports the notion that the Bcc exist on a genomic continuum (Vandamme & Dawyndt., 2011) with significant variability both within and between species.

In addition to horizontally acquired genes, Bcc genome plasticity is driven by the large number of insertion sequence elements (IS) they contain (Drevinek *et al*., 2010, Graindorge *et al*., 2012, Holden *et al*., 2009, Lessie *et al*., 1996). These IS elements have been shown to promote genome rearrangement, inactivate genes, or upregulate their expression (Lessie *et al*., 1996). In *B. cenocepacia* IS movement is induced in response to oxidative stress (Drevinek *et al*., 2010), which may account for the genomic rearrangements seen in sequential isolates from

CF. Epidemic strains of *B. cenocepacia* are particularly rich in IS elements. For example, the J2315 genome includes 84 IS elements belonging to 7 distinct families *cenocepacia* (Graindorge *et al*., 2012). This is more than twice the number of IS elements found in other sequences *B. cenocepacia* strains. This may not be coincidental as IS elements may play a role in promoting epidemic spread of *B. cenocepacia*, with preliminary evidence of convergent genomic localization of IS elements among epidemic strains (Graindorge *et al*., 2012).

The large genome sizes of the Bcc is not accompanied by a drop in gene density and they are estimated to contain between 7,000 and 8,000 protein coding genes, approximately twice that seen in *E. coli* (Winsor *et al*., 2008). An attempt to estimate the *Burkholderia* pan-genome identified over 40,000 gene families, with approximately 20,000 found within the Bcc (Ussery *et al*., 2009). This is almost certainly an underestimate as only 16 of the 80 currently named *Burkholderia* species, and only 6 of the 17 Bcc species have been sequenced. Of these genes only 649 have orthologs in all species of the order Burkholderiales (Juhas *et al*., 2012). This core genome consists of 454 genes with clear homology to essential genes in other species, 111 genes with weak homology and 84 genes that are either completely uncharacterized or known to be dispensable in other species. Given the large genetic repertoire of the Bcc, it is unclear how much of the genome is redundant, how much is used only in certain conditions and how much is truly indispensible.

1.2 Essential Genes

As originally conceived genes were abstract entities used for calculating the inheritance patterns of observable phenotypes (Johansen, 1909, Portin, 2002). When genes became identified with linear segments of DNA that are expressed in some way, the intrinsic link between gene and phenotype was broken and it became possible to identify, mutate and even delete genes without identifying a corresponding phenotypic effect. One of the most basic phenotypes is whether an organism can survive when a specific gene is inactivated. The necessity of these essential genes for proper cellular function makes their identification and study of both theoretical and practical interest. Knowledge of essential genes is an important first step in linking genotype to phenotype (D'Elia *et al*., 2009), sheds light on fundamental bacterial physiology (Fang *et al*., 2005), allows the study of evolution under very strong constraints (Lerat *et al*., 2005), is important for discovering new targets for novel antibiotics (Chalker & Lunsford., 2002) and is required for the construction of "chassis" cells in synthetic biology (Luisi, 2002).

1.2.1 Definitions of essential genes

Differing experimental methodologies and approaches have produced a variety of subtlety different interpretations of what it means for a gene to be essential. In its most expansive form, a gene is essential if its inactivation completely abrogates growth of a specific strain completely in all possible environments (Fang *et al*., 2005, Koonin, 2003). Under this definition, genes can be conclusively proven to be non-essential by recovering a deletion mutant but the inability to inactivate a gene does not show that it is essential as it is impossible to test every possible

environmental condition. To enable systematic study on a genomic scale, essential genes are often defined as genes that cannot be inactivated in a specific "permissive" environment (Akerley *et al*., 2002, Baba *et al*., 2006, Kobayashi *et al*., 2003). It is thought that genes that are required for growth on a nutritionally rich media with minimal environmental stress will in general also be required for growth in more stringent conditions. Under either definition of essentiality, genes only needed in specific contexts like the biosynthesis of missing amino acids or specific electron transport chains are "conditionally essential" (Gerdes *et al*., 2006).

In practice, the distinction between essential and nonessential genes is not clear-cut. In particular, the loss of gene products can produce significant growth defects without necessarily completely preventing cell division. These genes are "essential for fitness" (Fang *et al*., 2005, Griffin *et al*., 2011) and mutants will be lost if grown in mixed cultures but can be isolated if grown clonally. Genome-wide essential gene studies carried out in *E. coli* using growth of both clonal (Baba *et al*., 2006) and mixed populations (Gerdes *et al*., 2003) found that over twice as many genes were required for growth in mixed populations.

It is important to note that while there is substantial overlap, the set of essential genes in a strain is not synonymous with the minimal set of genes required for a free-living replicating cell (Juhas *et al*., 2014, Xavier *et al*., 2014). Functions that are essential to the cell, such as carbon assimilation or energy generation, may be performed by different independent genes. Other genes are essential only when other nonessential genomic elements are present and would not be

required in a minimal cell. These include the anti-toxins in some toxin-antitoxin systems, which are dispensable after the cognate toxin has been cleared from the cell.. While most biologists believe that all cellular organisms evolved from a common ancestor, or LUCA for Last Universal Common Ancestor (Kyrpides *et al*., 1999), or Last Universal Cellular Ancestor (Philippe & Forterre., 1999), very little of this ancestral genome remains. Only 80 genes have orthologs in all currently sequenced organisms and only 50 of those have the same phylogenetic tree as rRNA indicating that they have been passed down from LUCA without horizontal gene transfer (Harris *et al*., 2003). Since bacteria have approximately 300 to 700 essential genes (Luo *et al*., 2014), there is not a single set of essential genes but rather a diverse repertoire reflecting the evolutionary history and environment of each species.

1.2.2 Experimental identification of essential genes

Broadly speaking, there are two main approaches to identifying essential genes. The negative approach identifies non-essential genes and essential genes are inferred by default. In this approach, attempts are made to knock out genes either systematically or using high-density random transposon mutagenesis and genes for which mutants cannot be recovered are presumed to be essential. In contrast, the positive approach attempts to identify essential genes directly. Inducible promoters are used to control the production of gene products, either by directly controlling gene expression or by driving expression of using antisense mRNA to knockdown protein production. In either case, the resulting mutants are screened in inducing and non-

inducing conditions for mutants with a conditional growth phenotype indicating that some essential operon is being controlled.

Putative gene essentiality is often discovered on an ad hoc basis when attempts to create knockout mutants fail (Dicker & Seetharam., 1992), but systematic gene-by-gene inactivation screens on a genomic scale are rare. Designing, constructing and verifying the thousands of mutants necessary to cover all the genes in a bacterial genome is an expensive and time-consuming undertaking requiring mature high-throughput mutagenesis techniques. To date it has only been carried out in four bacterial species (Baba *et al*., 2006, de Berardinis *et al*., 2008, Kobayashi *et al*., 2003, Xu *et al*., 2011). In species for which homologous recombination of linear DNA is possible (Baba *et al*., 2006, de Berardinis *et al*., 2008, Xu *et al*., 2011), each gene is replaced by an antibiotic resistance cassette. In *Bacillus subtilis* (Kobayashi *et al*., 2003) genes were disrupted by the integration of a suicide plasmid containing an outward facing IPTG inducible promoter to drive expression of downstream genes in the same operon. In either case, inability to recover mutants of a targeted gene was taken as evidence of the gene's essentiality.

A common alternative to designing and constructing knockout mutants for every gene individually is to use a transposon that inserts relatively randomly and uniformly into the genome, pool the mutants and map the transposon insertion sites present in the pool (Judson & Mekalanos., 2000b, Reznikoff & Winterberg., 2008). Mutants with insertions in essential regions of the genome will be lost during a period of outgrowth so when the number of insertions

mapped is high enough to saturate the genome, essentiality can be inferred for genes containing fewer insertions than expected based on a variety of statistical models (DeJesus *et al*., 2013, Gerdes *et al*., 2003, Zhang *et al*., 2012) (Figure 1-1 A). Transposon mutagenesis can be easily modified to recover conditional growth mutants of essential genes by including an outward-facing inducible promoter (Bloodworth *et al*., 2009, Cardona *et al*., 2006, Judson & Mekalanos., 2000a). If the transposon inserts upstream from an essential gene, then the mutant will have a conditional growth phenotype. The mutants are replicated into inducing and non-inducing, conditions and mutants with a conditional growth phenotype are recovered and their insertion sites are sequenced (Figure 1-1 B). The transposon used in either inactivation or promoter replacement studies is critical. It must have a relatively high frequency of transposition to generate large numbers of mutants but be genetically stable once integrated into the genome. Furthermore, it must also either insert into any sequence in a relatively unbiased manner or have a specific target sequence that is common enough to occur multiple times in every gene. Transposons commonly used in essential gene studies in bacteria include derivatives of the *Tn5* (Gallagher *et al*., 2013, Martinez-Garcia *et al*., 2014, Remmele *et al*., 2014, Winterberg *et al*., 2005) and *Tn7* (Metris *et al*., 2011, Salama *et al*., 2004) transposons which have little sequence specificity, and the mariner transposon (Akerley *et al*., 1998, Akerley *et al*., 2002, Griffin *et al*., 2011) which inserts at thymine-adenine (‘TA’) motifs.

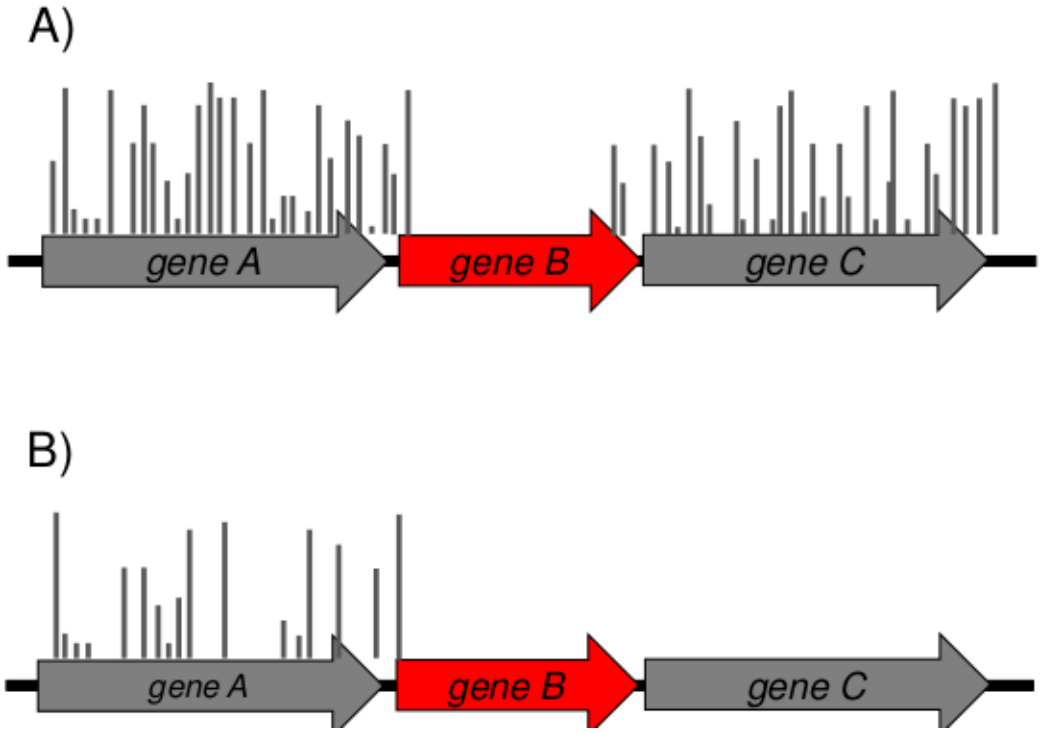


Figure 1-1 Identifying essential genes via transposon mutagenesis

A) To identify essential genes via high density transposon mutagenesis a pool of random transposon mutants with an outward facing promoter is created and their insertion sites mapped by sequencing. Only mutants with insertions into non-essential regions are viable and so essential genes e.g. *gene B* show up as gaps in a map of insertions. B) If the transposon includes an inducible promoter, individual mutants can be isolated and screened for a conditional growth phenotype, These mutants will have insertions upstream from essential genes e.g. *gene B*

Instead of directly controlling gene transcription, antisense RNA is widely used to down regulate gene expression post-transcriptionally (Brantl & Wagner., 2000, Caldelari *et al*., 2013, Nakashima *et al*., 2012). Since control of gene expression by the native promoter is unaffected, antisense RNA allow for native regulation in non-inducing conditions. Several studies have used inducible expression of antisense RNA to identify essential genes by isolating conditional-growth mutants (Forsyth *et al*., 2002, Ji *et al*., 2001, Meng *et al*., 2012, Rusmini *et al*., 2014, Wang & Kuramitsu., 2003). In these studies, the genome is fragmented into small pieces, typically 200-800 bp, which are then cloned into a plasmid under the control of an inducible promoter. The bacteria are then transformed with the library of plasmids, and each colony is replicated into inducing and non-inducing conditions. Clones with antisense fragments for an essential gene will have severely impaired growth in inducing conditions, and therefore can be isolated for sequencing and further study. The inducible promoter chosen to drive antisense expression is critical. It must have very tight regulation in non-inducing conditions while still being capable of high levels of expression when induced. Leaky expression in non-inducing conditions could silence genes transcribed at low levels, even completely inhibiting growth if the gene is essential. The first comprehensive antisense screen for essential genes in a Gram-negative species (Meng *et al*., 2012), *E. coli*, demonstrated a systematic bias towards highly expressed essential genes that is thought to be due to leaky expression from the promoter. A later study in *Pseudomonas aeruginosa* utilizing a more stringent promoter found a wider range of essential genes (Rusmini *et al*., 2014).

Since bacterial genes are typically organized into operons that are transcribed together, all essential gene identification methods must consider whether the expression of other genes in the operon are affected. When deleting a gene, the downstream ribosome binding site and start codon of the downstream gene must be preserved in order to allow translation. Overlapping genes are relatively common (17% of *E. coli* genes overlap (Baba *et al*., 2006)), so accurate gene annotation and careful selection of regions to delete are critical. Insertional inactivation , whether by site-directed or transposon mutagenesis, normally include antibiotic selection markers with transcriptional stops. Therefore an outward facing promoter is often included to drive expression of downstream genes (Jacobs *et al*., 2003, Kobayashi *et al*., 2003). In general, conditional growth mutants identify essential operons rather than essential genes per se. Inserted inducible promoters control the expression of all downstream genes, while antisense RNA appears to degrade mRNA (Meng *et al*., 2012), impairing expression of the entire operon.

If the experimental objective is confined to the identification of essential genes, high density transposon inactivation is by far the most common option (Luo *et al*., 2014). In the last 6 years, a variety of methods have been developed that use next generation sequencing technologies to simultaneously identify insertion sites in large pools of mutants (Gallagher *et al*., 2011, Gawronski *et al*., 2009, Goodman *et al*., 2009, Langridge *et al*., 2009, Opijnen & Camilli., 2010). These methods have dramatically lowered cost, allowing the construction of libraries containing more transposon mutants which improves transposon density and tripling the number

of essential gene studies published (Luo *et al*., 2014). Despite its popularity, the use of high-density transposon inactivation has several important drawbacks. First, the categorization of gene essentiality is probabilistic. Insertions in non-essential genes may be missing solely by chance and essential genes may tolerate insertions in non-essential domains. Secondly, mutants for essential genes are not recovered for further study. Recent studies have improved the confidence in identifying essential genes by increasing transposon mutant density but at the time of writing this thesis no studies have been published featuring the generation of conditional growth mutants for all the essential genes identified by transposon inactivation. *E. coli* and *P. aeruginosa* both have essential genes identified by transposon inactivation (Gerdes *et al*., 2003, Jacobs *et al*., 2003), but this information was not leveraged to create conditional growth mutant libraries. Instead, antisense RNA screens were run using random genome fragmentation (Meng *et al*., 2012, Rusmini *et al*., 2014) just as if the essential genomes were uncharacterized.

1.2.3 Characteristics of essential genes

Essential genes have diverse functions and are grouped together only by their indispensability for growth in some bacterial species. Despite these diverse functions, the availability of the essential genomes of a variety of bacterial species has shown (Luo *et al*., 2014) that essential genes are more likely to be widely conserved, highly expressed and encoded on the leading strand of the first replicon than genes in general. These characteristics, while generally true, do not allow the unambiguous identification of essential genes as highly expressed nonessential genes and species specific essential genes are not uncommon.

As might be expected given the severe phenotype for loss of function mutants, essential genes are more highly conserved than nonessential genes (Jordan *et al*., 2002, Kang *et al*., 2014, Kang *et al*., 2014). For example, in *E. coli* 61% of essential genes have orthologs in all Gamma-proteobacteria, while only 6.5% of all genes are similarly conserved (Fang *et al*., 2005). It is important to note that while essential genes are more widely distributed than average, not all widely distributed genes are essential. In *E. coli*, the majority (64%) of genes with orthologs in all Gamma-proteobacteria are dispensable in laboratory conditions (Fang *et al*., 2005). This discrepancy and the conservation of essential genes in general will be discussed in more detail in section 1.2.4 on using essential genes to study evolution.

In addition to being more conserved, essential genes also tend to have higher levels of expression than nonessential genes. This tendency has been observed by measuring optimized codon composition (Rocha & Danchin., 2003a, Rocha & Danchin., 2004), microarray hybridization (Bloodworth *et al*., 2009) and both single cell proteomics and transcriptomics (Taniguchi *et al*., 2010). Part of the difference in expression levels between essential and nonessential genes is due to the consistent expression of essential genes across different conditions while some nonessential genes are only used, and so expressed, in specific conditions. However, other factors must be involved, since even when comparing genes which are all expressed in the test condition, essential genes have significantly higher mean levels of mRNA and proteins than nonessential genes (Taniguchi *et al*., 2010). For example, genes encoding

enzymes in general are expressed at higher levels than genes encoding non-enzymes, and a disproportionate number of enzymes are essential (Gao & Zhang., 2011). Finally, given the extreme cost of not having sufficient quantities of an essential protein, essential genes may be expressed at higher levels to buffer against the random segregation of proteins between daughter cells, which could result in one daughter cell having less than the required levels of essential protein.

It is often common to analyze genomes as unordered collections of genes, but genomes have physical structure and essential genes are unevenly distributed over this structure. The orientation of essential genes in bacterial genomes is highly biased, with 95% of *B. subtilis* and 97% of *E. coli* essential genes found on the leading strand (Rocha & Danchin., 2003a, Rocha & Danchin., 2003b). While essential genes do tend to be more highly expressed than nonessential genes, strand bias is not driven by gene expression levels. Of the most highly expressed *E. coli* genes, only 64% of non-essential genes are found on the leading strand compared with 90% of the essential genes. It is hypothesized that this bias is due to the effects of DNA polymerase – RNA polymerase collisions when the processes of replication and transcription are both occuring in actively growing cells. When both polymerases are moving in the same direction (gene on the leading strand), the DNA polymerase slows down until transcription completes and in the case of head on collisions (gene on the lagging strand), transcription aborts. This not only reduces transcription levels but also potentially produces truncated nonfunctional proteins (Rocha & Danchin., 2003b).

Many bacterial species possess multiple stable replicons with the smaller replicons called either “megaplasmids” or “accessory chromosomes” (Mackenzie *et al*., 2004) or collectively referred to as chromids (Harrison *et al*., 2010). In these species, computational functional annotations consistently find that core metabolic genes and genes with predicted essential function are overrepresented on the first and largest chromosome (Egan *et al*., 2005). Traditional essential genes studies have been carried out in 3 species with multiple chromosomes *Vibrio cholera* (Cameron *et al*., 2008), *Burkholderia pseudomallei* (Moule *et al*., 2014) and *Burkholderia thailandensis* (Baugh *et al*., 2013). In the *V. cholera* experiment, of the 789 ORFs not interrupted by at least one transposon, 573 were on chromosome 1 and 216 were on chromosome 2 (Cameron *et al*., 2008). This distribution roughly corresponds with the relative sizes of the chromosomes with chromosome 1 being approximately twice as large as chromosome 2 (Heidelberg *et al*., 2000). However, saturation of the genome was relatively low and 74% ORFs less than 500bp long and 20% of ORFs larger than 1kb are expected to be missing purely due to chance. Of the 228 ORFs that were missing from the transposon library and also had essential orthologs in *E. coli*, only 5 were found on chromosome 2. This significant underrepresentation on chromosome 2 could be due to only considering genes with *E. coli* homologs regardless of essentiality. Across multiple species, genes on secondary replicons are less likely to have homologs in *E. coli* than genes on chromosome 1 regardless of putative function (Mackenzie *et al*., 2004). The higher transposon insertion density of the libraries used for identifying essential genes in *B. pseudomallei* (Moule *et al*., 2014) and *B. thailandensis* (Baugh *et al*., 2013) resolve the ambiguity when calling gene essentiality. In *B. pseudomallei,*

only 16.4% of the 505 putatively essential genes were found on chromosome 2; a noticeable underrepresentation as the second chromosome contains 40.5% of the annotated genes (Moule *et al*., 2014). Results in *B. thailandensis* were similar with only 10% of essential genes found on chromosome 2, which contains 40.9% of genes (Baugh *et al*., 2013). Beyond disrupting individual genes, the general dispensability of the chromids in *Sinorhizobium melilotti* has been confirmed. *S. meliloti* has a 3.7 Mb chromosome and two megaplasmids: the 1.4 Mb pSymA and 1.7Mb pSymB (Milunovic *et al*., 2014). The majority of both pSymA and pSymB have been deleted except for their *oriV* regions. This required supplying only 5 genes *in trans* two classical essential genes from pSymA, the GTPase *engA* and the only copy of a $tRNA^{Arg}$ gene and 3 essential antitoxin genes from toxin-antitoxin loci (2 on pSymA and 1 on pSymB). It is unclear how representative these results are as the gene composition of chromids varies greatly. In *S. meliloti* and the *Burkholderias,* the smaller chromids appear to have originally been dispensable plasmids that have accumulated accessory genes. There are, however, species like *Brucella melitensis* where a single original chromosome appears to have fractured into two independent chromosomes (DelVecchio *et al*., 2002, Mackenzie *et al*., 2004). No essential gene studies have been carried out in these species. In Bcc with three chromids, essential genes appear to be biased in order of increasing chromid size. It is unclear whether this is a general fact in multipart genomes or whether it reflects the particular evolutionary history of the Bcc.

There have been several attempts to use the differences between known essential and nonessential genes to try and predict gene essentiality of new species computationally (Arigoni

et al., 1998, Deng *et al*., 2011, Plaimas *et al*., 2010, Silander & Ackermann., 2009). The specificity of these tests is complicated by the small portion of essential genes in any genome. Bacterial essential gene studies typically find between 300 and 600 essential genes (Luo *et al*., 2014). Consider a bacterial genome with 4,000 genes of which 400 are truly essential. If these genes are classified according to some feature found in 100% of essential genes and only 10% of nonessential genes, then 760 genes will be identified as essential but 360 or 47% of them will be false positives. This phenomenon is clearly seen when using gene conservation across phylogenetic groups to predict essentiality. While most essential genes are highly conserved, when 26 *E. coli* genes with homologs in the very compact genome of the distantly related *Mycoplasma genitalium* were studied, less than 25% were essential in *E. coli* (Arigoni *et al*., 1998). More sophisticated models that take into account the phylogenetic relationship of the species being compared are more accurate (Silander & Ackermann., 2009). Still, when a gene conservation cutoff is chosen that includes 80% of the experimentally confirmed *E. coli* essential genes, 64% of the genes returned are actually nonessential. Classification accuracy can be improved by integrating multiple independent gene features using machine learning techniques (Deng *et al*., 2011, Plaimas *et al*., 2010). A training set of data is prepared where genes have already been classified as essential or nonessential along with their feature scores. Computer programs are then run on this data to learn how the different features can best be combined and weighted to predict gene essentiality. In one example, after been trained on data from *E. coli*, the sensitivity (proportion of essential genes predicted to be essential) was set to 80%. The program has a false positive rate (proportion of nonessential genes predicted to be essential) of less than 10% when run on its training data, 32% when predicting essential genes in *Acinetobacter baylyi*

and 60% in both *Bacillus subtilis* and *P. aeruginosa.* The high error rate in *P. aeruginosa* probably reflects a systematic biases in the experimental data. In general about 50% of the essential genome can be predicted computationally with very high specificity, with the remainder being more idiosyncratic.

1.2.4 Essential genes and evolution

The neutral theory of evolution posits that at the molecular level most variation is due primarily to mutations with no effect on fitness, constrained by purifying selection that removes mutations with a high fitness cost (Kimura, 1991). This was later extended to the nearly neutral theory of evolution (Ohta, 2011, Razeto-Barry *et al*., 2012) which claims that most mutations have a slight but non-zero effect on fitness. Under these models, the rate at which a gene evolves depends on the probability that a random mutation will interfere with gene function and the probability that the organism can reproduce without the gene (gene dispensability) (Wilson *et al*., 1977). Since essential genes are by definition completely indispensible it has been proposed that they are under more intense purifying selection and so evolve more slowly than nonessential genes (Hurst & Smith., 1999, Kimura & Ohta., 1974, Wilson *et al*., 1977).

The rate of gene evolution over shorter timescales can be tested by comparing the Ka/Ks ratio of orthologous genes in different strains or closely related species, where Ka is the number of nonsynonymous substitutions (nucleotide changes that alter the amino acid sequence of the corresponding protein) per nonsynonymous site and Ks is the number of synonymous

substitutions (nucleotide changes that have no effect on the amino acid sequence of the corresponding protein) per synonymous site (Jordan *et al*., 2002, Rocha & Danchin., 2004, Wall *et al*., 2005). Since mutations that lead to altered proteins are more likely to have fitness effects than mutations that do not, the lower the ratio the greater the selective pressure to maintain existing function. An initial study in *E. coli* found that the Ka/Ks ratio was 12x lower for essential genes than nonessential genes (Jordan *et al*., 2002), suggesting a stronger selective pressure for maintaining essential protein sequence. However, later studies found that after controlling for the higher levels of expression seen with many essential genes, essentiality per se had little impact on the rates of gene evolution in yeast (Wall *et al*., 2005), *E. coli* and *B. subtilis* (Rocha & Danchin., 2004). More recently, Wei *et al.* (Wei *et al*., 2013) found that while levels of expression are the primary factor influencing the rate of gene evolution in yeast, different biological features cooperatively influence the rates of gene evolution in bacteria and that their relative contributions vary between different species.

Over longer timescales, the Ka/Ks ratio becomes inaccurate as the orthologous genes are saturated with mutations. In those cases, gene conservation, or the extent of gene orthologs amongst diverse bacterial linages provide a better measure of molecular evolution (Jordan *et al*., 2002). Several comparative genomic studies have shown that essential genes are more conserved than nonessential genes (Jordan *et al*., 2002, Rocha & Danchin., 2004, Silander & Ackermann., 2009) with the rate at which orthologs are lost along branches of phylogenetic trees, having an especially high correlation with gene essentiality (Silander & Ackermann., 2009). Despite this

broad conservation, genes that are essential in one species are often nonessential or missing from other closely related species (Silander & Ackermann., 2009). This difference in conservation among essential genes appears to be driven in part by how frequently suppressor mutations arise that can compensate for the essential function (Bergmiller *et al*., 2012).

One hypothesis to explain the weak effect of gene essentiality on the short-term rate of evolution is that many genes identified as dispensable in laboratory conditions are required to survive in the natural environment of the microorganism. Comprehensive studies in yeast have found that while 80% of genes could be deleted without impacting growth on rich media, 97% of deletion mutants featured a quantifiable growth defect in at least one condition when subjected to chemical and environmental stresses (Hillenmeyer *et al*., 2008). Though using this dataset, the majority of differences in gene evolutionary rate were still due to differences in expression rather than essentiality in any of the tested conditions (Wang & Zhang., 2009). A similar quantitative genome study performed in *E. coli* identified growth defects for deletion of 49% of genes in at least one of the 324 conditions studied (Nichols *et al*., 2011); but these results have not been correlated with the rate of gene evolution. It is also unclear whether the contribution of genes to growth in the conditions tested accurately measures their contribution to fitness in the natural environment. Of note is the fact that all these assays study the growth of clonal populations even though bacteria are usually found in polymicrobial communities (Philippot *et al*., 2013, Wolcott *et al*., 2013).

In summary, gene essentiality is a weak predictor of the Ka/Ks ratio of orthologs genes but a good predictor of gene conservation across phylogenetically distant species. It seems likely that while these metrics are often thought of as measuring the same underlying process on different timescales, they are in fact governed by separate factors. A low Ka/Ks ratio implies that there is strong selective pressure for a gene present in a genome to produce a protein with a specific amino acid sequence while a high level of gene conservation implies that many genomes are under a strong selective pressure to have some version of that gene. If alterations to the amino acid sequence of highly expressed genes are toxic then both highly expressed essential and nonessential genes present in a genome should have a highly conserved sequence, but non-essential genes would still be more likely to be lost from a genome.

1.2.5 Minimal cells in synthetic biology

Synthetic biology is an interdisciplinary effort to systematically construct useful biological systems using modular well-characterized parts (Khalil & Collins., 2010). In order to function, these genetic circuits depend on an existing genome to provide common functions such as DNA replication, transcription and energy metabolism (Daszczuk *et al*., 2014, Layton & Trinh., 2014). Therefore, a major stand within synthetic biology is the construction of minimal cells containing just essential genes and a defined set of accessory genes. It is thought that these cells would provide a simple "chassis" that are easier to understand, less likely to divert carbon sources to undesired products and have more predictable regulation (Foley & Shuler., 2010, Vickers *et al*., 2010).

One approach towards the construction of chassis cells is to reduce existing genomes by successive large-scale deletions of nonessential chromosomal regions. Genomic reductions of 15 to 22% have already been carried out in *E. coli* (Mizoguchi *et al*., 2008, Posfai *et al*., 2006) and *B. subtilis* (Morimoto *et al*., 2008, Tanaka *et al*., 2013) . An *E. coli* strain with a 15% reduction in genome size, targeting mainly genomic islands and insertion sequences, demonstrated improved electroporation efficiency and maintenance of previously unstable recombinant genes (Posfai *et al*., 2006). Accumulating further genetic deletions of non-essential secondary metabolites produced a strain (Mizoguchi *et al*., 2008) that reaches 1.5x higher cell density than the wild type when grown on minimal media and could produced 2.4x more L-threonine. Similar results were seen in *B. subtilis* where a 20% reduction in genome size was accompanied by a 2 fold increase in protein expression along with extensive reconfiguration of the transcriptional regulatory network (Morimoto *et al*., 2008). While these results are promising, the reduced genomes are still far from minimal. Tanka *et al.* recently demonstrated that some 30% of the *B. subtilis* genome is dispensable (Tanaka *et al*., 2013). Their work also uncovered previously unknown essential regions and unpredicted reductions in growth rate demonstrating that even in this well characterized species the contribution of genes to growth phenotype is not completely understood.

1.2.6 Essential gene products as antibiotic targets

One of the major motivations for studying essential genes is their potential as new targets for novel antibiotics (Chalker & Lunsford., 2002, Pucci, 2006). Most classes of antibiotics in current use were discovered between 1940 to 1970 by systematically screening soil bacteria for the ability to inhibit the growth of target bacteria, and then isolating the antimicrobial compounds responsible (Bush, 2010). Widely adopted and highly successful for 20 years, this drug discovery methodology was ultimately abandoned by pharmaceutical companies due to the continuous rediscovery of known compounds (Lewis, 2013) combined with the unfavorable economics of developing antibiotics (Projan & Shlaes., 2004). The outcome of these factors is that only 3 new classes of antibiotics have entered the market since 2000, of which one can only be used topically (Spellberg *et al*., 2008). This dramatic reduction in rate at which new antibiotics are discovered has been coupled with the spread of antibiotic (Rossolini *et al*., 2014) and multiple antibiotic resistance (Gould, 2008) rendering our current arsenal of antibiotics less effective.

Since essential genes are by definition required for cellular growth, an attractive alternative to blindly screening for growth inhibition is to attempt to develop inhibitors for specifically chosen essential enzymes (Chalker & Lunsford., 2002, Pucci, 2006). In this approach, a specific target gene is chosen based on its essentiality in a species of interest. A biochemical test is designed to assay for enzyme function *in vitro* and a library of small molecules is screened for the ability to inhibit purified protein. While information on preclinical research and development by pharmaceutical companies is difficult to come by, it appears that

developing antibiotics by screening for inhibitors of essential enzymes has been attempted with limited success (Chan *et al*., 2004, Gwynn *et al*., 2010, Payne *et al*., 2007, Silver, 2011) due primarily to two major issues. First, selective inhibitors of antibiotic targets were found at rates approximately 5-fold less than drugs in other therapeutic areas (Payne *et al*., 2007). This is being addressed through changes to the medicinal chemical characteristics of the small molecules being screened (Galloway *et al*., 2009, Jacoby & Mozzarelli., 2009, Payne *et al*., 2007). The other problem has been poor choice of the target enzymes. For example, methionyl tRNA synthetase (*metS*) is required for protein synthesis in *Staphylococcus aureus*, and was thought to be essential in *S. pneumoniae* and *E. coli* due to high levels of sequence conservation (Payne *et al*., 2007). After screening and optimizing lead compounds, it was discovered that approximately 46% of *S. pneumoniae* isolates possess an alternative distantly related methionyl tRNA synthetase probably acquired through horizontal gene transfer (Gentry *et al*., 2003). Even when a gene is essential in the target species, a good understanding of the physiological and genetic context is necessary to validate it as a target. Peptidyl deformylase (*pdf*) is widely essential in bacteria and selective inhibitors of the Pdf protein have been created. However, resistance appears at very high frequencies in some species (10^{-7} in *E. coli* (Apfel *et al*., 2001)), due to a loss of function mutation in *fmt*, a transformylase. Normally, bacterial translation is initiated with the formylation of methionine by Fmt. This initial methionine must later be removed from some peptides to allow for proper function which cannot be done until it is first deformylated by Pdf making *pdf* essential. However, in some species including *E. coli*, *P. aeruginosa* and *S. aureus*, the loss of a functional *fmt* allows for initiation of translation using an alternative non-formylated process (Mazel *et al*., 1994) bypassing the need for *pdf*. These examples demonstrate that proper

validation of putative antimicrobial targets requires an in depth understanding of gene essentiality across a range of bacterial species.

One major problem in looking for effective antibiotics by measuring enzyme inhibition *in vitro* is that even a potent and selective enzyme inhibitor of an ideally chosen essential enzyme cannot be used as an antibiotic unless it can enter the bacterial cell and avoid efflux. This is especially problematic in Gram negative species like *B. cenocepacia*, where intrinsic multidrug resistance is often due to relatively impermeable membranes and an efficient array of efflux pumps (Bolla *et al*., 2011). Screening for inhibition of bacterial growth avoids this problem, but blind screening has been proven to be unsustainable due to the time and effort spent on nonspecific growth inhibitors or already known antibiotics (Lewis, 2013). An attractive solution to this dilemma is to screen conditional expression mutants which can be made hypersensitive or hyper-resistant to inhibitors of specific targets (Cardona *et al*., 2014). This approach combines the ability of whole cell screenings to interrogate inhibitor-protein interactions in their natural context while still avoiding additional time spent studying molecules with undesired mechanisms of action.

Target specific hypersensitivity is achieved by depleting the concentration of the essential target until it produces a mild growth defect. The theory is that the target now limits cellular growth rate and consequently, cell growth rate will respond strongly to any inhibitors of that target but will maintain normal sensitivity to small molecules with other mechanisms of action.

Hypersensitivity has been demonstrated using antisense inhibition in *S. aureus* where inhibition of *fabF,* which encodes a β-ketoacyl carrier protein synthase, caused hypersensitivity to cerulenin (Forsyth *et al*., 2002). This approach was later adapted to use an agar-diffusion assay where zones of clearing around the wild type and knockdown mutant were used to identify FabF/FabH inhibitors (Young *et al*., 2006) in a library of natural compounds. The screen identified both known inhibitors including cerulenin, thiolactomycin and novel inhibitors of FabF identified as phomallenic acids. Use of genetic constructs with inducible promoters has also been successful in a panel of *E. coli* mutants with essential genes placed under the control of the arabinose inducible promoter (DeVito *et al*., 2002). The minimal inhibitory concentration or MIC of a *murA* strain to fosfomycin changed 100 fold between low and high levels of expression. In the same study, a high throughput screen of the wild type strain against a library of small molecules identified three growth inhibitors. Screening the same small molecule library against the *murA* strain with low levels of expression identified the three inhibitors found by the wild type plus four additional growth inhibitors. Running the library against the *metG* depletion strain again identified the same three growth inhibitors found by the wild type and *murA* strain plus five additional growth inhibitors specific to *metG*.

Specific hyper-resistance has been achieved by using a second copy of the essential gene under the control of a strong inducible promoter resulting in excess target production,, which titrates the effect of inhibitors. The first use of multicopy suppressors for identifying antimicrobial targets by Li *et al.* (Li *et al*., 2004) involved and initial screening of 8,640

chemicals for growth inhibition of a hyperpermeable strain of *E. coli*. For each of the 49 hits, a random library of *E. coli* 3-4kb fragments under the control of the T7 promoter was screened for suppressors. Suppressors were identified for 33 compounds; However 32 clones were identified to contain acrB encoding the inner membrane component of a multidrug efflux pump, with only 2 clones for *folA* which encodes the dihydrofolate reductase, the putative target of the small molecules. To avoid recovering large numbers of clones for non-specific resistance, an ordered panel of clones overexpressing essential genes was later used (Kitagawa *et al*., 2005). This method identified a novel antimicrobial compound that blocks lipoprotein targeting by inhibiting LolA, a chaperone responsible for translocation of lipoproteins from the inner to outer membranes.

In addition to screening a single differentially sensitive mutant against the chemical compound library to identify inhibitors of a specific target, known growth inhibitors can be screened against a panel of mutants to identify chemogenomic interactions (Cardona *et al*., 2014). This approach has been confirmed to identify the direct targets of antibiotics such as trimethoprim (*folA*) (Kitagawa *et al*., 2005), cycloserine (*ddlB*) and indolmycin (*trpS*) (Xu *et al*., 2010). Exposure to growth inhibitors also reveals additional, often mechanistically relevant, chemical-genome interactions. For example, in the case of inolymicin both the direct target of the antibiotic and the targets substrate were identified. Inolymicin inhibits tryptophantyl-tRNA synthetase (TrpS) blocking the linking of tryptophan to its tRNA. Exposing a panel of *S. aureus*

antisense expressing strains to inolmycin found hypersensitivity upon depletion of either *trpS*, or an operon containing the gene for tryptophantyl-tRNA synthetase (Xu *et al*., 2010).

1.3 Essential genes in *Burkholderia cenocepacia*

Given their clinical relevance and the complexity of their genomes, I was interested in studying the essential genes in one of the Bcc. At the time, the only essential genome study in any *Burkholderia* species was a low throughput study that demonstrated the feasibility of generating *B. cenocepacia* conditional growth mutants by using a mini-Tn5 transposon derivative to deliver an outward facing rhamnose inducible promoter (Cardona *et al*., 2006). Identifying essential genes by generating conditional growth mutants would allow them to be studied and would provide a library for enhanced sensitivity studies with antimicrobials. However, the generation of conditional growth mutants by transposon mutagenesis has never been attempted on a genomic scale.

1.3.1 The mini-Tn5 transposon

Any attempt to identify essential genes by transposon mutagenesis requires that the transposon insert relatively evenly across the genome. The potential insertion sites must also be close enough together that no essential genes are missed. This is especially important when generating conditional growth mutants as the growth of all transposon mutants must be screened. Tn5

derived transposon systems are commonly used tools for genetic analysis due to their simplicity, relatively unbiased insertion sites, and broad host range (Reznikoff, 2008).

The native Tn5 is a composite transposon consisting of antibiotic resistance genes for kanamycin, bleomycin and streptomycin flanked by two almost identical inverted sequences IS50L and IS50R (Mazodier *et al*., 1985), though only the IS50R encodes an active transposase. During transposition, the transposase recognizes the inverted repeats in the IS elements excising and pasting the DNA bounded by the repeats into a new site without duplication(Reznikoff, 2008). However, working with the Tn5 transposon is unwieldy. It is large, has multiple antibiotic resistance markers and mutants are potentially unstable due to the presence of the transposase in the genome. To work around these issues, minitransposon derivatives have been created with transposase outside of IS50 inverted repeats which flank an antibiotic resistance cassette and multiple restriction enzyme sites (Dennis & Zylstra., 1998). This arrangement allows for the stable insertion of the inverted repeats and any intervening DNA into the genome.

The popularity of miniTn5 transposons has lead to several studies looking at the distribution of insertion sites. Sequencing 1955 Tn5 insertions into cDNA clones showed that insertion sites were slightly non-random (Shevchenko *et al*., 2002). In particular, the 9bp recognition sequence usually contains an A/T flanked by GC rich regions. Another study (Lodge & Berg., 1990) showed that insertions into highly expressed regions are favored. Despite these small biases, insertions into any site are possible and Tn5 derivatives are widely used for high

density transposon mutagenesis experiments (Gallagher *et al*., 2013, Martinez-Garcia *et al*., 2014, Remmele *et al*., 2014, Winterberg *et al*., 2005)

1.3.2 The rhamnose inducible promoter system

Achieving a conditional growth phenotype via promoter replacement requires that the following criteria are met: that expression is tightly regulated with very low or non-existent expression in the absence of induction, that the range of expression is wide enough to accommodate highly transcribed genes, and that induction does not alter other cell phenotypes. Most of the conditional expression systems tried in *B. cenocepacia* have been unacceptable. The lactose and *tac* promoter systems are reported to be nonfunctional (Lefebre & Valvano., 2002). A plasmid based arabinose system has been shown to function in *B. cenocepacia*, but the high levels of arabinose required for full expression cause a change in cell volume probably due to high osmolarity (Cardona & Valvano., 2005), and chromosomal integrations were unable to drive detectable levels of gene expression (Cardona *et al*., 2006). The rhamnose inducible promoter system meets these requirements and has been successfully used for conditional expression of essential genes in *B. cenocepacia* (Bloodworth *et al*., 2013, Cardona *et al*., 2006, Mohamed & Valvano., 2014, Ortega *et al*., 2007) (Figure 2-1).

The rhamnose inducible promoter system from *E. coli* consists of the divergent P_{rhaB} promoter and *rhaSR* operon (Brautaset *et al*., 2009). Both RhaS and RhaR are transcription factors belonging to the AraC/XylS family and sharing 30% identity (Egan & Schleif., 1993).

Both are activated by L-rhamnose. When exposed to rhamnose, RhaR induces transcription of the *rhaSR* operon producing more RhaR and RhaS, the latter of which in turn induces expression from PrhaB. The rhamnose regulon is also subject to catabolite repression via CRP, the cAMP (cyclic AMP) receptor protein (Wickstrum *et al*., 2005). Full activation of the regulon requires that CRP bound to cAMP, and intracellular cAMP rise in the absence of a preferred carbon source such as glucose. The rhamnose system maintains its tight regulation and wide range of expression levels in *B. cenocepacia* (Cardona & Valvano., 2005). It also has the added advantage that *B. cenocepacia* does not natively catabolize rhamnose (Fang *et al*., 2011a) so low concentrations of rhamnose produce a consistent level of induction.

1.4 Thesis objectives

This project consisted of the following objectives, two of which involve the identification and characterization of essential genes in *B. cenocepacia*. Due to my interest in *Burkholderia* genomes, a third objective involving the sequencing of a new species of *Burkholderia* was included towards the end of my PhD studies.

1) To construct a library of *B. cenocepacia* K56-2 conditional growth mutants on a genomic scale and demonstrate the utility of the library of mutants for identifying the mechanism of action of growth inhibitors. This would be the first attempt to identify the essential genome of one of the Bcc and the first attempt to use transposon delivered promoter replacement to identify essential genes on a genomic scale in any species. The intrinsic antibiotic resistance of the Bcc and their pathogenicity in susceptible subpopulations

makes the development of tools to ease the discovery of novel antibiotics effective against the Bcc a priority.

2) To study the physiological role of one of the *B. cenocepacia* essential genes commonly considered dispensable in other bacterial genera. Characterization and in-depth study of essential genes lags behind their identification. While essential genes tend to be widely conserved, genes that are known to be essential in one species can be nonessential in or missing completely from the genomes of other species. It is easy to imagine that genes essential in bacteria with less genetic capacity might be redundant in the *Bcc* with their very large genome size. Understanding the reverse is interesting, why would a gene be essential in the *Bcc* if its homologs are dispensable in other studied bacteria?

3) To sequence and assemble draft genome sequences for two *B. contaminans* strains LMG23361 and FFH2055 as well as *B. cenocepacia* K56-2. In most countries, *B. contaminans* forms a very small percentage of Bcc isolates from patients with CF but in Spain and Argentina, *B. contaminans* has become the most commonly isolated Bcc species. This will be the first *B. contaminans* isolates sequenced, an important early step in the analysis of this emerging pathogen. *B. cenocepacia* K56-2 will also be sequenced; K56-2 is commonly used for studies the *B. cenocepacia* ET12 linage but there is no complete sequence for this strain so most labs use the genome sequence of the related *B. cenocepacia* J2315. While the two strains are closely related they have different RAPD types and so must have some genomic differences.

2 *Burkholderia cenocepacia* conditional growth mutant library created by random promoter replacement of essential genes.

This chapter was published in *MicrobiologyOpen*, 2013: Ruhi A. M. Bloodworth, April S. Gislason and Silvia T. Cardona. (2013). *Burkholderia cenocepacia* conditional growth mutant library created by random promoter replacement of essential genes. MicrobiologyOpen. (2)-2:243-258. doi: 10.1002/mbo3.71

2.1 Introduction

The *Burkholderia cepacia* complex (referred to here as Bcc) is a group of closely related Gram negative bacteria that are widely distributed in natural and man-made environments (Vandamme & Dawyndt., 2011). Burkholderia species, included Bcc, are of great interest because of their large multipartite genomes, their great metabolic versatility and the wide array of ecological niches they occupy (Sousa *et al*., 2011). Also of interest, is the "dual personality" of Bcc as these environmental bacteria, which were initially considered harmless, are now known to cause human infections (Chiarini *et al*., 2006). While strains can be exploited for biocontrol, bioremediation and plant growth promotion purposes, safety issues arise regarding human infections, as many Bcc strains have emerged as opportunistic pathogens in patients with cystic fibrosis (CF), chronic granulomatous disease, and other medical conditions associated with a compromised immune system (Loutet & Valvano., 2010, Mahenthiralingam *et al*., 2005, Valvano *et al*., 2005). The relevance of defining the genetic components that are key for Bcc growth in all environments is then twofold: Bacterial components to which small molecules can

be directed to in order to inhibit growth can be exploited to control both human infections and biotechnological processes.

Bacterial genes that are required for growth in rich, undefined media are regarded as essential and hence their encoded products are potential targets of growth inhibitory molecules (Brown & Wright., 2005). Essential genes have been identified on a genomic scale by high-density transposon knockout mutagenesis (Akerley *et al*., 2002, Gerdes *et al*., 2003, Hutchison *et al*., 1999, Jacobs *et al*., 2003, Sassetti *et al*., 2003) or systematic gene-by-gene inactivation (Kang *et al*., 2004, Thanassi *et al*., 2002) where genes for which mutants could not be recovered are assumed to be essential. Alternatively, identification of essential genes by construction of conditional knock-outs with inducible promoters adds the value of obtaining conditional growth (CG) mutants that are then available for further studies (DeVito *et al*., 2002, Forsyth *et al*., 2002, Judson & Mekalanos., 2000). A number of inducible promoters have been used to express essential genes through the construction of CG mutants (Carroll *et al*., 2005, Wong & Akerley., 2003) with the *Escherichia coli* arabinose-inducible promoter (PBAD, P*araB*) being one of the best characterized (Guzman *et al*., 1995, Judson & Mekalanos., 2000). A great challenge is however, to achieve genomic representation of essential genes with conditional mutagenesis probably because of the different range of required expression levels. Promoters with very low uninduced expression levels are necessary to obtain mutants with a conditional growth phenotype. Yet, highly induced expression levels may be necessary for highly expressed

essential genes. Promoters that are inducible to such high levels may show uninduced levels of essential gene expression that are tolerable to bacterial growth (Bugrysheva *et al*., 2011).

The *E. coli* rhamnose-inducible promoter (P*rhaB*) is controlled by a cascade of two transcriptional regulators and provides a more tightly regulated transcription than P*araB* (Haldimann *et al*., 1998). We previously demonstrated that P*rhaB* is suitable for tightly regulated gene expression in the Bcc clinical isolate *B. cenocepacia* K56-2, and that essential genes can be identified by transposon-based delivery of P*rhaB* throughout the bacterial chromosome followed by screening for absence or growth without rhamnose (Cardona *et al*., 2006).

In the current work, we asked whether saturation of a genome with P*rhaB* allows identification of essential genes at the genomic level with representation of such genes in a library of CG mutants. Using a large-scale mutagenesis approach and robotic screening of more than 200,000 transposon mutants for rhamnose-dependent growth, we constructed a library of CG mutants (CG mutant library) and analysed the contribution of promoter expression levels and gene redundancy in the identification of essential genes. We demonstrate that screening for conditional growth in one condition identifies CG mutants of similar conditional growth phenotypes, which makes them suitable for chemogenomic experiments.

2.2 Experimental Procedures

2.2.1 Bacterial strains and growth conditions.

Bacterial strains and plasmids are listed in Table 2-1, and the identified CG mutants are listed in Table 2-3. All mutants were made in a *B. cenocepacia* K56-2 background and were grown in Luria Bertani (LB) media at 37°C supplemented as required with different concentrations of rhamnose, 100 μg/ml or 50 μg/ml trimethoprim (Tp) for *B. cenocepacia* or *E. coli*, respectively, 50 μg/ml gentamicin (Gm) and 40 μg/ml kanamycin (Km). All chemicals were purchased from Sigma Chemical Co., St. Louis, MO. unless otherwise indicated. To prepare standardized glycerol stocks, overnight cultures were washed twice with LB, adjusted to a final OD_{600nm} of 0.2 in LB 20% glycerol and aliquoted into PCR tubes for storage at -70°C. In assays involving bacterial growth, cultures were diluted to give a theoretical final OD_{600nm} of 0.001, and arranged in 96-well format. Plates were sealed with parafilm and incubated at 37°C with shaking at 200RPM in a New Brunswick Scientific E24 shaking incubator. OD_{600nm} were read using BioTek Synergy 2 plate reader.

Table 2-1 Bacterial strains and plasmids

	Features	Source
Strains		
Burkholderia cenocepacia K56-2 (LMG18863)	ET12 linage, CF isolate	(Mahenthiralingam *et al*., 2000)
B. cenocepacia gyrB	Promoter replacement of *gyrB* P*gyrB*:: pRB6	This study
E. coli SY327	*araD* Δ(*lac pro*) *argE* (Am) *recA*56 Rif^r *nalA* λ *pir*	(Miller & Mekalanos., 1988)
E. coli DH5α	F$^-$, ϕ 80 *lacZ*ΔM15 *endA1 recA1 hsdR*17(r_K^- m_K^+)*supE*44 *thi*-1 Δ*gyrA*96 (*ΔlacZYA-argF*)U169 *relA*1	Invitrogen
Plasmids		
pRK2013	ori_{colE1}, RK2 derivative, Km^r *mob*$^+$ *tra*$^+$	(Figurski & Helinski., 1979)
pRB-rham	pSCrhaboutgfp derivative (Cardona *et al*., 2006), *ori* $_{R6K,}$ *rhaR rhaS PrhaB* e-*gfp*,	This study
pSC200	pGpΩTp derivative (Flannagan *et al*., 2007), *ori* $_{R6K,}$ *rhaR rhaS PrhaB* e-*gfp*	(Ortega *et al*., 2007)
pRB6	pSC200 containing *gyrB* 5'end	This study

Table 2-2 Primers

Name	Oligonucleotide sequence, 5'-3'[a]	Purpose
1	AACACCCGGGATGGTGAGCAAGGGCGAG	Amplification of eGFP
2	CCGCTCTAGATTACTTGTACAGCTCGTCCA	Amplification of eGFP
34	CCAGAGGGTACCAAAATTGCTTTGAGAGGCTC	Amplification of *ori* $_{R6K}$
35	CTGTTAGGTACCCGT ACTAAGCTCTCATGTTT	Amplification of *ori* $_{R6K}$
178	GGCCACGCGTCGACTAGTCAGNNNNNNNNNNACGCC	Reverse primer for 1st round of arbitrary PCR
179	GGCCACGCGTCGACTAGTCAG	Reverse primer for 2nd round of arbitrary PCR
187	ACATCACCACAATTCAGCAA	Forward outer transposon specific primer
188	ACCCAGTCCGCCCTGAGCAA	Forward inner transposon specific primer
189	CCCTGAGCAAAGACCCCAAC	Sequencing insertion site
06	AATTAACATATGATGAGTGAACAGCACAATTCGC	Amplification of 300bp 5' end of *gyrB*
207	ATATAATCTAGACTTCGGCTCGTGCTTGTCG	Amplification of 300bp 5' end of *gyrB*
249	CCCGACACCTTGTAGCTGTTCTG	Reverse external primer for confirming *gyrB* insertion

[a] Restriction sites are underlined

2.2.2 Molecular biology techniques

DNA ligase and restriction enzymes (New England Biolabs) were used as recommended by the manufacturers. *E. coli* SY327 cells were transformed using the Z-competent buffer kit protocol (Zymo Research). Conjugation into *B. cenocepacia* K56-2 was accomplished by triparental mating (Craig *et al*., 1989) with *E. coli* DH5α carrying the helper plasmid pRK2013 (Figurski & Helinski., 1979)**.** DNA was amplified using a PTC-221 DNA engine (MJ Research) or an Eppendorf Mastercycler ep gradient S thermal cycler with either Taq DNA polymerase (Qiagen) or Phusion High-Fidelity PCR Kit (New England Biolabs). Amplification conditions were optimized for each primer pair. PCR products and plasmids were purified with QIAquick PCR Purification Kit (Qiagen) and QIAprep Spin Miniprep Kit (Qiagen), respectively. DNA sequencing was performed by The Center for Applied Genomics (TCAG) at The Hospital for Sick Children, Toronto, ON).

2.2.3 Vector constructions

pRB-rham (Figure 2-1A) is a derivative of pSCrhaBoutgfp(Cardona *et al*., 2006), in which the pMB1 origin of replication was replaced with that of the R6K plasmid (*ori* $_{R6K}$) to avoid possible reversion to a replicative plasmid in *B. cenocepacia. ori* $_{R6K}$-dependent replication has the absolute requirement of the Pir protein (Stalker *et al*., 1979) and thus plasmids bearing this replication origin do not replicate in bacteria not harboring *pir.*

To construct pRB-rham, pTnMod-RTp (Dennis & Zylstra., 1998) and pSCrhaBoutgfp (Cardona *et al*., 2006) were digested with *SpeI/KpnI* and the *ori* $_{R6K}$ and *dhfr* cassette from pTnMod-RTp

was ligated to the backbone of pSCrhaBoutgfp. To construct plasmid pRB6, a 300bp-DNA fragment containing the 5' end of *gyrB,* flanked by *XbaI* and *NdeI* restriction sites, was cloned into pSC200 immediately downstream from the rhamnose inducible promoter. The resulting plasmid was conjugated into *B. cenocepacia* K56-2 by triparental mating. Integration of pRB6 and replacement of the *gyrB* natural promoter was confirmed by colony PCR for the *ori* $_{R6K}$ and PCR amplification of the insertion interface.

2.2.4 CG mutant library construction

pRB-rham was introduced into *B. cenocepacia* K56-2 via triparental mating (Craig *et al*., 1989). The exconjugates were selected for by plating onto 500 cm^2 QTrays (Genetix, X6023) containing LB agar with 0.2% rhamnose and the appropriate antibiotics and incubated for 48h at 37°C. The resulting colonies were robotically picked with a Genetix, QPix2 XT colony picker into master plates containing liquid LB medium with 0.1% rhamnose and Tp100 and were incubated overnight. While initial picking and replicating was performed into 96-well plates (Greiner Bio-One 655185), the majority of the library was produced in 384 well microplates (Greiner Bio-One 781186). The master plates were robotically replicated into secondary plates containing LB and LB 0.1% rhamnose and incubated overnight. Bacterial growth was estimated by measuring OD_{600nm} of the cultures with a BioTek Synergy 2 plate reader equipped with a BioTek Bio-Stack automated plate stacker and the ratio of growth without and with rhamnose was calculated for each mutant. Transposon mutants showing at least a 50% decrease in OD_{600nm} in the absence of rhamnose in comparison to OD_{600nm} in the presence of rhamnose were manually rescreened for

growth in LB with or without 0.1% rhamnose. Mutants showing at least a 50% decrease in OD_{600nm}, after 16h of incubation were stored as glycerol stocks and included in the CG mutant library.

2.2.5 Determination of transposon insertion sites

Transposon insertion sites were identified either by arbitrary primed PCR (Das *et al*., 2005, Miller-Williams *et al*., 2006) or by self-cloning as previously described (Dennis & Zylstra., 1998). For each clone we first attempted arbitrary-PCR. A 1 µl aliquot of overnight culture was used directly as the template for an initial low stringency PCR reaction using a transposon specific primer and a degenerate arbitrary primer, which amplifies the transposon-genome junction as well as other random stretches of DNA. The products of this reaction were used as the template for a second PCR reaction using an inner transposon specific primer and a primer identical to the tail of the degenerate primer to preferentially amplify the transposon-genome junction (Table 2-2). The products were purified using a QIAquick PCR Purification Kit (Qiagen) and sequenced using a third transposon specific primer. Approximately 20% of the PCR products did not return a usable chromatogram. To determine if the unsuccessful sequencing was due to the presence of multiple insertions, we subjected these clones to Southern blot experiments. All clones showed only one restricted DNA fragment that hybridized with the transposon-complementary probe, demonstrating that the clones with a failed sequencing reaction harbored a single transposon insertion (data not shown). For these clones, insertions were determined successfully using the self-cloning procedure. DNA was digested with *NotI* or

NdeI and Southern blots for the *eGFP* on the end of the transposon were performed using an iBlot Gel Transfer System (Invitrogen) and DIG High-Prime DNA Labeling Kit (Roche). The location of the insertion site was determined using nucleotide BLAST against the genome of *B. cenocepacia* J2315 at www.burkholderia.com (Winsor *et al*., 2008). The distance from the insertion to the start site of the downstream open reading frame for insertions into putatively intergenic regions and the start site of the surrounding open reading frame for insertions within putative genes were also calculated.

2.2.6 Comparisons with essential genes in other bacteria

Essential *E. coli* genes were obtained from PEC (Profiling of *E. coli* Chromosome) (Hashimoto *et al*., 2005), and from Liberati *et al.* (Liberati *et al*., 2006) for *Pseudomonas aeruginosa.* Orthologs of *B. cenocepacia* J2315 in *E. coli* MG1655 and *P. aeruginosa* PAO1 were found using reciprocal best hit protein BLAST (Altschul *et al*., 1990) on the annotated open reading frames with an Expect cutoff of 10. *E. coli* microarray data was obtained from E-MEXP-3461 (Prieto *et al*., 2011), RNA-seq data from (Yi *et al*., 2011) and *B. cenocepacia* J2315 microarray data from (Bazzini *et al*., 2011). The expression levels of essential and non-essential genes were compared using the Mann-Whitney sum-rank test (Lehmann, 1975) assuming that the test-statistic U is normally distributed given the large sample size. This statistical test does not assume that the expression is normally distributed across genes nor that there is a linear relationship between the values and the underlying mRNA levels.

2.2.7 Functional characterization of genes in putatively essential operons

Since the rhamnose inducible promoter will control the expression of all downstream genes in the same transcriptional unit, all downstream genes in the same putative operon were included in our analysis. Genes were included if either OperonDB (confidence level of 50 or more) (Pertea *et al*., 2009) or DOOR (Database of Prokaryotic Operons) (Dam *et al*., 2007) placed them in the same putative operon.

2.2.8 Enhanced sensitivity assay

To calculate rhamnose concentrations that produced 30-60% of wild type growth, rhamnose dose-response curves of each mutant were run as follows: Mutants were grown for 22h in LB with a rhamnose gradient of 0% to 0.16%. The resulting OD_{600nm} were then fitted to the hill equation using GraphPad prism. To develop the enhanced sensitivity assay, all the mutants were grown in 96 well format plates with 200 µl of LB medium containing rhamnose concentrations required to achieve between 30-60% of wild type growth without the addition of antibiotics. Novobiocin (1 µg/ml) or chloramphenicol (2 µg/ml) were added as required. Plates were incubated for 22 h at 37°C with shaking. For each rhamnose concentration fold reduction was measured as OD_{600nm} without antibiotic/OD_{600nm} with antibiotic and mutant sensitivity was defined as $\log_{10}$ of the fold reduction in growth due to the antibiotic

2.3 Results

2.3.1 Building a CG mutant library in *B. cenocepacia* K56-2

We previously defined essential operons as transcriptional units containing at least one essential gene (Cardona *et al*., 2006). However, the total number of essential genes and their organization into transcriptional units in *B. cenocepacia* large genome is unknown. To estimate the number of *B. cenocepacia* essential operons expected to hit to achieve genome coverage, we first analyzed 14 bacterial genomes with experimental data on gene essentiality and compared the number of essential genes versus genome size (Figure 2-2). No correlation was found between the size of the essential genome with the overall number of genes with most of the bacterial essential genomes ranging from 300 to 700 genes (Christen *et al*., 2011, Gerdes *et al*., 2006, Griffin *et al*., 2011, Langridge *et al*., 2009). To estimate the number of operons that may contain essential genes in a model organism, we analyzed the *E. coli* genome *in silico*, and based on the 306 experimentally confirmed essential genes and a computational estimate of operons (Zhang *et al*., 2006) we found 180 *E. coli* operons that include one or more essential genes (data not shown). Thus, if *B. cenocepacia* contained 300-700 essential genes as in other bacterial genomes, with a similar distribution to that of *E. coli*, approximately 200-400 essential operons would be expected.

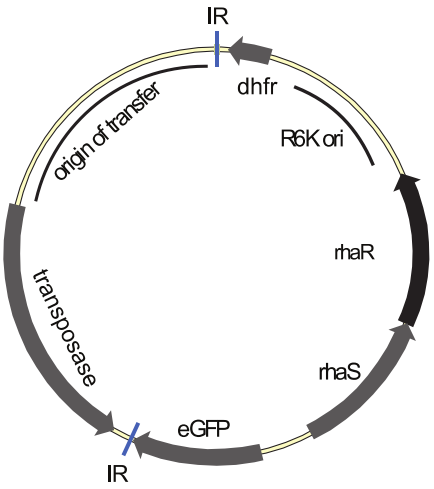
The construction of the *B. cenocepacia* CG mutant library is outlined in Figure 2-1. The previously developed method for delivering P*rhaB* by transposon mutagenesis and selection by replica plating on LB agar with and without rhamnose (Cardona *et al*., 2006) was modified to

achieve high throughput levels. After successful transposon mutagenesis and robotic screening in LB liquid media in the presence and absence of rhamnose, transposon mutants with at most 50% growth in the absence of rhamnose were isolated and rescreened before inclusion in the CG library. We reasoned that the permissive 50% cutoff would allow the inclusion of CG mutants with slow growth in the absence of rhamnose due to either low-levels of residual expression or remaining essential gene products from growth of the parental culture in rhamnose. This decision also means that the library would include mutants for genes that are important for, but not absolutely required for growth. By screening 200,000 transposon mutants with this single condition, 134 CG mutants were initially isolated and their growth phenotypes confirmed, representing a hit frequency of approximately 1/1,500. However, the conditional growth phenotype of 19 of these mutants could not be observed when the original glycerol stocks were further used, leaving 115 CG mutants in the library.

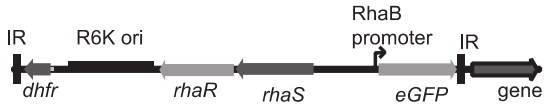
We identified the location and orientation of the delivered P*rhaB* by arbitrary primed PCR (Das *et al*., 2005, Miller-Williams *et al*., 2006) or self-cloning (Dennis & Zylstra., 1998) and by aligning the obtained DNA sequences with the genome sequence of the clonally related strain J2315 (Holden *et al*., 2009). All K56-2 DNA sequences could be matched to equivalent genomic regions in J2315. However, we found that a 57-kb duplication on chromosome 1 of J2315 does not exist in the K56-2 genome. Insertion sites in some mutants were mapped to the 3' end of *lepA1*/*lepA2*, near the middle of the 57-kb duplication on chromosome 1 of J2315. As our system works by usurping the native promoter to control expression of downstream genes, it should be

impossible to observe CG defect from an insertion into only one copy of large identical repeats. After confirming the CG phenotypes in clones that had transposon insertions in the K56-2 putative duplicated region, we verified that the duplication was not present in the genome of K56-2 by PCR amplification of the two duplication-genome interfaces (data not shown). Clones with insertions in this site were the most common ones found with up to 10 clones recovered towards completion of the screening procedure. Despite the permissive conditions for inclusion into the library the vast majority of the mutants (101 out of 115) showed less than 35% of wild-type growth over 22 h in the absence of rhamnose (Figure. 2-3). Eighty-two mutants in the library showed less than 20% of wild type growth in the absence of rhamnose, with 42 showing less than 5% of wild type growth over 22h.

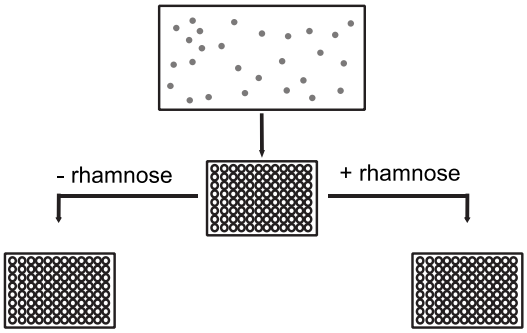
a) pRBrhaboutgfp



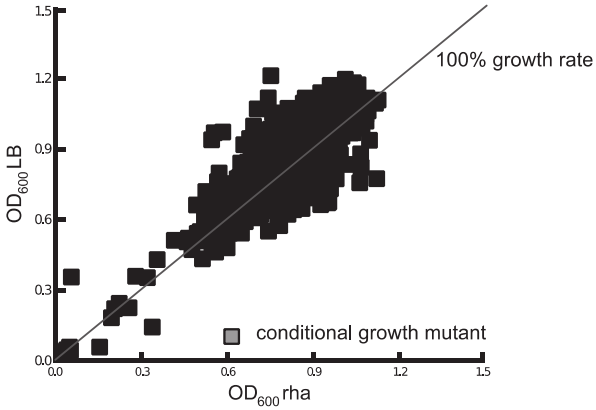
b) Delivery of the rhamnose system into the chromosome of *B. cenocepacia*



c) Screening for rhamnose-dependent growth strains



d) Screening results for 1,920 transposon mutants



e) Determining transposon insertion site with arbitrary PCR

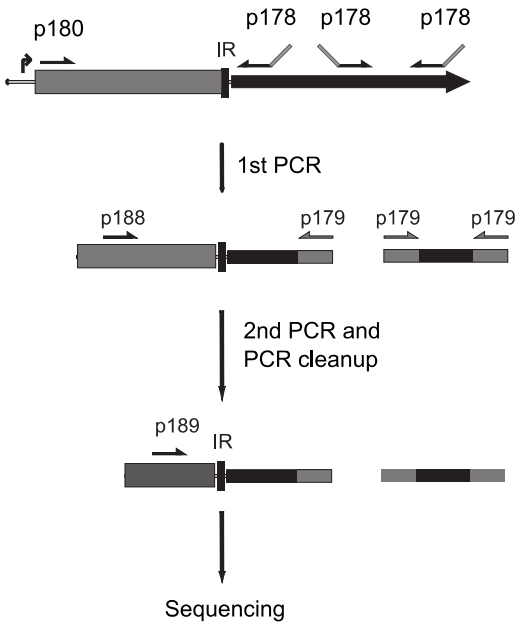


Figure 2-1. Construction of a *B. cenocepacia* CG mutant library.

A) Transposon vector pBR-rham (See material and methods and (Cardona *et al*., 2006) for details on plasmid construction B) Trimethoprim resistance provided by the dhfr cassette was used to select for the transconjugants containing an outward facing rhamnose inducible promoter PrhaB. C) The transconjugants were robotically picked into 96 or 384 well master plates before being robotically replicated into 96 or 384 well secondary plates containing LB with and without rhamnose. D) The OD600nm of the plates were read after 16h and mutants showing at least 50% less growth in the absence of rhamnose were included in the library. E) The insertion sites of the mutants were primarily determined using arbitrary primed PCR to preferentially amplify the transposon-genome junction and sequenced using a transposon specific primer.

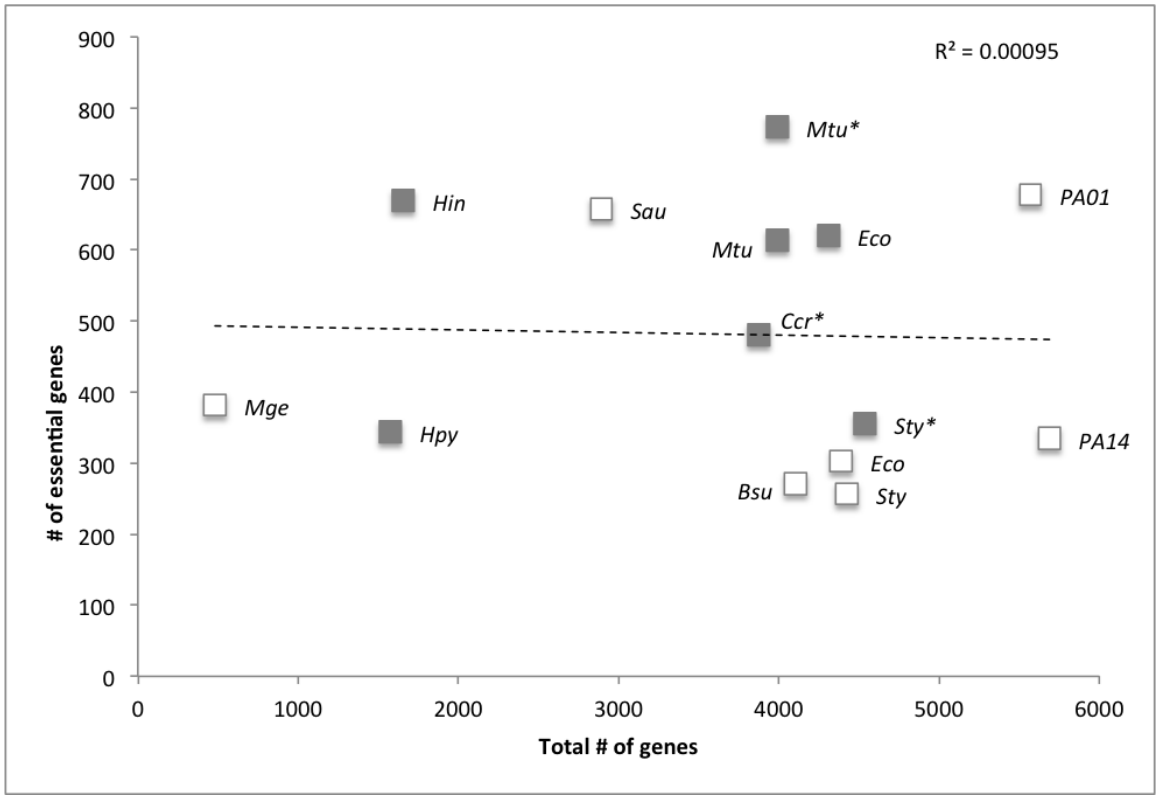


Figure 2-2. The number of essential genes as a function of genome size.

The overall number of essential genes in 14 bacterial genomes is plotted against the total number of genes in the same genome. *Mge*, *Mycoplasma genitalium*; *Hin*, *Haemophilus influenzae* Rd; *Hpy*, *Helicobacter pylori* G27; *Sau*, *Staphylococcus aureus* RN4220; *Mtu*, *Mycobacterium tuberculosis* H37Rv; *Eco*, *Escherichia coli*; *Sty*, *Salmonella Typhi*; *PAO1*, *Pseudomonas aeruginosa* PAO1; *PA14*, *P. aeruginosa* PA14; *Ccr*, *Caulobacter crescentus*. Data on gene essentiality for all genomes with the exception of data noted with asterisks was collected from Gerdes *et al.*, 2006. Gene essentiality of *C. crescentus*, *M. tuberculosis* and *S. Typhi* were

obtained from Christen *et al*., 2011, Griffin *et al*., 2011 and Landgridge *et al*., 2009, respectively. Mutants were generated by random or targeted transposon mutagenesis, and mutants were propagated. Mutant lack of survival was considered criteria for defining the interrupted gene as essential. Grey and white squares indicate that the data was obtained from propagating the mutants within a population or clonally, respectively.

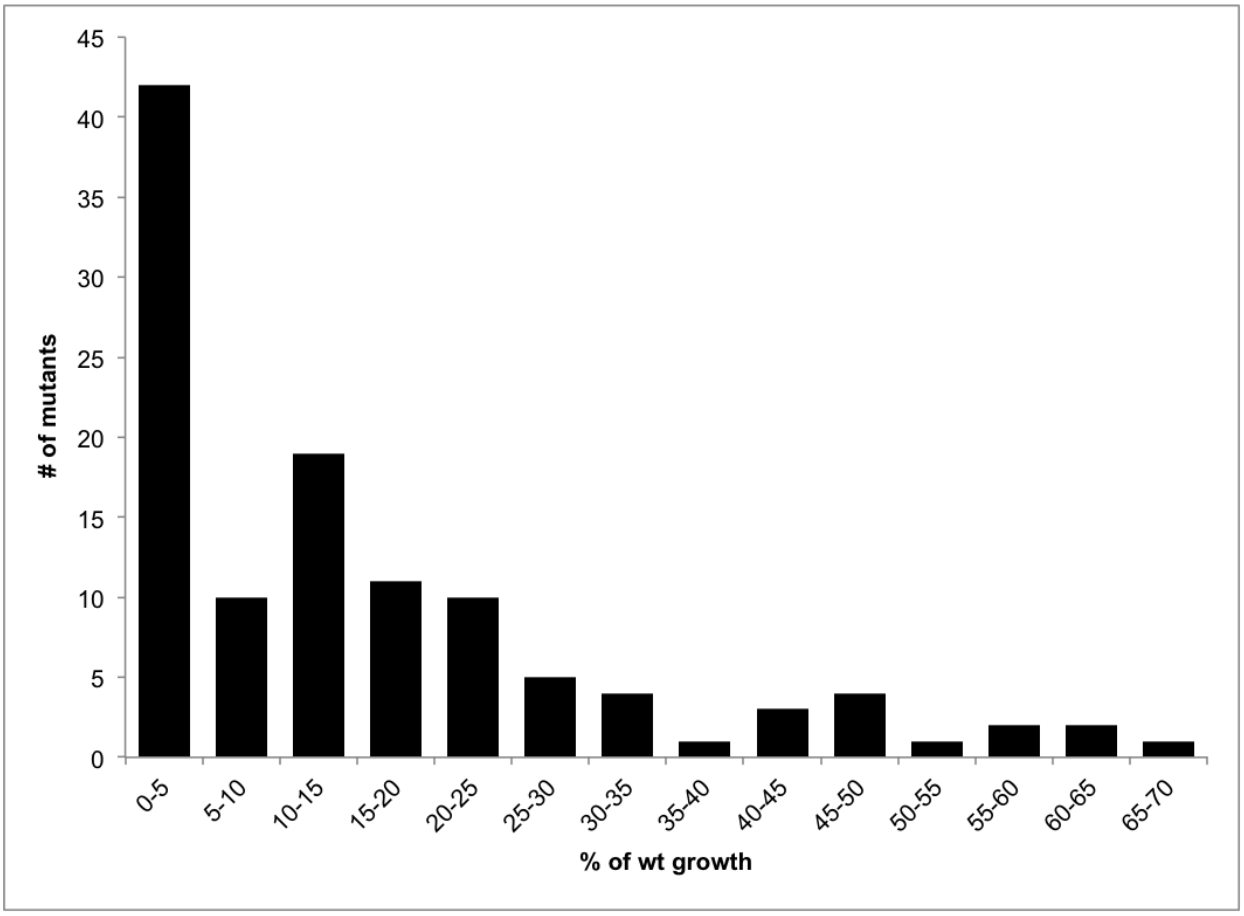


Figure 2-3. Histogram of growth for 115 CG mutants in the absence of rhamnose.

B. cenocepacia K56-2 (wild type) and CG mutants were grown in LB without rhamnose and OD600 was measured after 22h. Percent of wild type growth was defined as the growth of CG mutants relative to that of *B. cenocepacia* K56-2. Bars represent the total number of CG mutants with percent of wild type growth within the range indicated by flanking numbers.

Table 2-3 CG mutants included in the CG mutant library

Locus	Gene	OperonDB (a)	DOOR (b)	*E. coli* (c)	*P. aeruginosa* (d)	Strains	Average % growth (e)
BCAL0029		Y	Y	NA	NA	30-4C6	1.02
BCAL0030	atpB	Y		F	F		
BCAL0031	atpE	Y		F	T		
BCAL0032	atpF	Y		F	F		
BCAL0033	atpH	Y		F	F		
BCAL0034	atpA	Y		F	T		
BCAL0035	atpG	Y		F	F	67-10H9	2.57
BCAL0036	atpD	Y		F	F		
BCAL0037	atpC	Y		F	F		
BCAL0272		Y	Y	F	NA	81-26F1	5.23
BCAL0330	petC	Y	Y	NA	F	29-3B1, 72-11A3, 73-24D8, 79-38C1	23.69
BCAL0331		Y		F	F	76-1F3	12.76
BCAL0332		Y		F	F		
BCAL0422	dnaN	Y	Y	T	F	9-16E11, 80-11H9	0.53
BCAL0421	gyrB	Y		T	F	48-44F12, 58-14E1	0.84
BCAL0425		Y	Y	F	F	67-17C2, 79-11E11	11.39
BCAL0426		Y	Y	T	F		
BCAL0465	def	Y	Y	T	T	4-20B10	14.09
BCAL0466	fmt	Y	Y	T	T		
BCAL0467				NA	NA		
BCAL0468		Y		F	F		
BCAL0469		Y		F	F		
BCAL0470		Y		NA	NA		
BCAL0471		Y	Y	NA	NA	73-14C5, 79-36H5, 81-36C9	39.25
BCAL0472		Y	Y	NA	NA		
BCAL0653		Y	Y	NA	NA	34-24C5	19.09
BCAL0654			Y	NA	NA		
BCAL0691		Y	Y	NA	NA	8-15C6	24.26
BCAL0804		Y	Y	NA	F	69-17C4, 69-18F6,	16.21

						73-27E6	
BCAL0803	lolB	Y	Y	T	T	29-10G4, 30-12H3	7.79
BCAL0875		Y	Y	F	F	47-17H12, 82-25B9	0.84
BCAL0876	ubiB	Y	Y	F	T		
BCAL0877			Y	NA	F		
BCAL0879		Y	Y	NA	NA	32-30H3	2.17
BCAL0880	aspS	Y		T	F		
BCAL0881	ntpA	Y		F	NA		
BCAL0882		Y		F	F		
BCAL0895		Y	Y	F	F	29-17B9, 58-31B5, 67-4F2, 83-13G9	18.79
BCAL0896	pdxA	Y	Y	F	F		
BCAL0897	ksgA	Y		F	F		
BCAL1003	lepA2	Y	Y	F	NA	27-10B12, 28-13B11, 31-19E8,31-20G12, 31-21H6, 35-5G12, 35-6A9, 47-29C11, 61-16A9, 69-5C7, 72-29A1	13.17
BCAL1004	lepB2	Y	Y	T	T		
BCAL1255	rnhA	Y	Y	F	T	67-4B6	10.81
BCAL1254	dnaQ	Y		F	T		
BCAL1291		Y	Y	NA	F	29-15E2	0.80
BCAL1299		Y	Y	NA	NA	28-9C11, 64-4F1	14.27
BCAL1300			Y	NA	NA		
BCAL1506	nusA	Y	Y	T	F	4-7D1	6.51
BCAL1507	infB	Y	Y	T	T	67-5E12	2.66
BCAL1508	rbfA	Y		F	F		
BCAL1509	truB	Y		F	F		
BCAL1515	sucA	Y	Y	F	T	28-10H8	2.84
BCAL1516	sucB	Y	Y	F	T		
BCAL1517	odhL	Y	Y	NA	F		
BCAL1518		Y	Y	F	F		
BCAL1941	dnaB	Y	Y	T	F	31-7D3	0.84
BCAL2068				F	NA	32-32F10	2.84
BCAL2153	ppiB	Y	Y	F	F	61-12C7, 94-10O20	20.57
BCAL2154		Y	Y	T	T		
BCAL2199		Y	Y	F	F	58-18F5, 72-9A2, 83-18C2	1.95
BCAL2198	iscS	Y		F	F		

BCAL2197		Y		F	T		
BCAL2196		Y		F	T		
BCAL2195	hscB	Y		F	T		
BCAL2194	hscA	Y		F	F		
BCAL2193	fdx	Y		F	T		
BCAL2192		Y		F	F		
BCAL2343	nuoB	Y	Y	F	F	67-5H10	27.16
BCAL2342	nuoC	Y	Y	NA	NA		
BCAL2341	nuoD	Y	Y	F	F		
BCAL2340		Y	Y	F	F	30-5D10	11.07
BCAL2339	nuoF	Y	Y	F	F		
BCAL2338		Y	Y	F	F		
BCAL2337	nuoH	Y	Y	F	F		
BCAL2336		Y	Y	F	F		
BCAL2335	nuoJ	Y	Y	F	F		
BCAL2334	nuoK	Y	Y	F	F		
BCAL2333	nuoL	Y	Y	F	F		
BCAL2332	nuoM	Y	Y	F	F		
BCAL2331	nuoN	Y	Y	F	F		
BCAL2330		Y	Y	NA	NA		
BCAL2329		Y	Y	F	F		
BCAL2328		Y		NA	NA		
BCAL2388		Y	Y	F	F	34-16F6	15.55
BCAL2389	purD	Y	Y	F	F		
BCAL2390	hemF	Y		F	T		
BCAL2391		Y		T	F		
BCAL2392		Y		F	F		
BCAL2393		Y		F	F		
BCAL2394		Y		F	F		
BCAL2395	cafA	Y		F	F		
BCAL2408	msbA			T	T	48-31G4	5.45
BCAL2407				NA	NA		
BCAL2676	pepA	Y	Y	F	F	16-2C5, 29-5G1	3.46
BCAL2675		Y	Y	F	F		
BCAL2674			Y	NA	NA		
BCAL2677		Y	Y	T	F	16-6G10, 16-11C6	9.19
BCAL2678		Y	Y	F	T		
BCAL2679			Y	NA	NA		
BCAL2736		Y	Y	F	F	46-32G1	15.46
BCAL2838	purC	Y	Y	F	F	79-36A7, 84-37D12	41.18

BCAL2837	purE	Y	Y	F	F	58-23E8	34.91
BCAL2836	purK	Y	Y	F	F	79-27B9	20.95
BCAL2835		Y	Y	NA	T		
BCAL2934	etfA	Y	Y	F	F	77-16C10	0.44
BCAL2959	ubiG	Y	Y	F	T	70-1E1	47.98
BCAL2960		Y	Y	NA	F		
BCAL3013		Y	Y	F	F	57-20B6, 77-4D9	11.76
BCAL3012	gmk	Y	Y	T	F		
BCAL3011	rpoZ	Y		F	F		
BCAL3010	spoT	Y		T	F		
BCAL3035	trxB	Y	Y	F	F	51-6G10	13.25
BCAL3053	ribB	Y	Y	NA	NA	78-16D7, 81-14H5, 83-15D6	15.62
BCAL3054	ribH	Y	Y	T	F	8-8G4	16.39
BCAL3055	nusB	Y	Y	T	T		
BCAL3111	wbxY	Y	Y	NA	NA	58-11E9, 79-12B10	9.95
BCAL3110	waaA	Y	Y	T	T		
BCAL3142		Y	Y	F	F	4-22D3, 90-6H2	17.41
BCAL3141		Y	Y	T	T		
BCAL3140	pyrR	Y	Y	NA	F		
BCAL3139	pyrB	Y	Y	NA	F		
BCAL3138	pyrX	Y	Y	NA	F		
BCAL3137	plsC		Y	NA	F		
BCAL3266		Y	Y	NA	NA	72-10F11, 88-10A4	33.31
BCAL3267	panB	Y	Y	F	F		
BCAL3274	hemH			T	T	99-8N4	30.08
BCAL3273							
BCAL3306	secD	Y	Y	T	F	69-21A5	7.44
BCAL3307	secF	Y	Y	T	T		
BCAL3335	fis	Y	Y	F	F	64-10H7	49.62
BCAL3336	purH	Y	Y	F	F		
BCAL3337	ruvC	Y		F	F		
BCAL3338	ruvA	Y		F	F		
BCAL3339	ruvB	Y		F	F		
BCAL3351	pyrC	Y	Y	F	F	76-5B4	57.37
BCAL3350		Y	Y	NA	NA		
BCAL3419	aroQ1	Y	Y	NA	F	83-9E5, 83-9H10	0.87
BCAL3420	accB	Y	Y	T	T	89-10K20	1.46
BCAL3421	accC	Y	Y	T	T		
BCAL3422	prmA	Y	Y	F	F		

BCAL3423			Y	NA	F		
BCAL3424	tpx		Y	F	F		
BCAL3425			Y	NA	NA		
BCAL3426			Y	F	F		
BCAL3433	ffh	Y	Y	T	F	34-17G4	5.98
BCAL3471	mraZ	Y	Y	F	F	31-17A3, 81-23H12	2.82
BCAL3470	mraW	Y	Y	F	F	48-43E1	0.49
BCAL3469	ftsL		Y	T	T		
BCAL3468	ftsI	Y	Y	T	F	47-5A10	29.81
BCAL3467	murE	Y	Y	T	T		
BCAL3466	murF	Y	Y	T	T		
BCAL3465	mraY	Y	Y	T	T		
BCAL3464	murD	Y	Y	T	F		
BCAL3463	ftsW	Y	Y	T	T		
BCAL3462	murG	Y	Y	T	T		
BCAL3461	murC	Y	Y	T	T		
BCAL3460	ddl	Y	Y	F	F		
BCAL3459	ftsQ	Y	Y	T	F		
BCAL3458	ftsA	Y	Y	T	T		
BCAL3457	ftsZ	Y	Y	T	F		
BCAL3456		Y		NA	NA		
BCAL3455	lpxC	Y		T	F		
BCAM0909		Y	Y	F	F	87-8E1	51.19
BCAM0910		Y	Y	T	F	73-18C4 (Grows very slowly)	
BCAM0911	dxs	Y	Y	T	T	51-13E10	20.42
BCAM0912		Y		NA	NA		
BCAM0916		Y	Y	NA	F	48-43H3, 48-44A4	34.11
BCAM0917	dnaG	Y		T	F		
BCAM0918		Y		T	T		
BCAM0967		Y	Y	F	T	57-31D10	16.08
BCAM0968		Y	Y	F	T		
BCAM0969	sdhA	Y	Y	F	T		
BCAM0970	sdhB	Y	Y	F	T		
BCAM0971			Y	NA	NA		
BCAM0972	gltA	Y	Y	F	F	58-21H7	34.74
BCAM1881		Y	Y	NA	NA	96-1K12	46.12
BCAM1908		Y	Y	NA	NA	86-3D16	67.56
BCAM1909			Y	NA	NA		
BCAM1910			Y	NA	NA		

a) the gene is part of an operon with the previous gene as called by OperonDB, gaps occur when a *B. cenocepacia* gene is found between genes which are adjacent in other genomes.

b) the gene is part of an operon with the previous gene as called by DOOR

c) Homolog in *E. coli* is essential, T true, F false, N.A. no homolog found by RBH

d) Homolog in *P. aeruginosa* is essential,

e) % of growth with rhamnose after incubation for 22h in LB

2.3.2 Functional characterization of essential operons

Of the 115 CG mutants in the library, 106 CG mutants were successfully sequenced and had insertion sites in the same orientation as interrupted and/or adjacent downstream genes (Figure 2-4, Table 2-3). As P*rhaB* controls the expression of downstream genes in the same operon, insertion sites identified 50 unique putative essential operons (Figure 2-4). These operons contained 179 genes, which we organized into functional groups using the COG (Cluster of Orthologous Genes) (Tatusov *et al*., 2003) and GO (Gene Ontology) (Ashburner *et al*., 2000) annotations from the *Burkholderia* Genome Database (Winsor *et al*., 2008) (Figure 2-5). Genes involved in core metabolic functions such as energy production, cell envelope biosynthesis and DNA replication were overrepresented compared with the entire genome (Figure 2-5). Conversely, genes involved in transcription, which are peripheral to growth in the permissive conditions tested, were underrepresented. Similarly, genes involved in carbohydrate metabolism were also underrepresented. This is expected given the range of nutrients available in rich media such as LB. Genes of unknown function were also underrepresented, but still comprised the

second largest category after energy production, indicating the large number of uncharacterized genes involved in core *B. cenocepacia* processes.

A BLASTp reciprocal best hit (RBH) revealed 239 *B. cenocepacia* J2315 genes with essential orthologs in *E. coli* (Baba *et al*., 2008) and 249 with essential orthologs in *P. aeruginosa* PAO1 (Jacobs *et al*., 2003) of which 131 were essential in both (data not shown). Of the 179 genes found in essential operons in our study, 63 had essential orthologs in either *E. coli* or *P. aeruginosa*, of which 25 were essential in both (Figure 2-4, Tables 2-3 and 2-4). The identical genomic duplication in *B. cenocepacia* J2315 caused genes in *P. aeruginosa* or *E. coli* that matched to a *B. cenocepacia* J2315 gene in the duplicated region to return 2 identical "best hits". As we determined that the aforementioned duplication is not present in the *B. cenocepacia* K56-2 strain actually used for the experiments, the genes downstream of transposon insertions found in the region by our study were manually matched against *E. coli* and *P. aeruginosa* and added as orthologs to Table 2-4. The identification of clones with transposon insertions upstream of genes that are orthologous to essential genes related to cell division (*ftsA*, *ftsZ*, *ftsW*, etc.), and peptidoglycan biosynthesis and assembly (*murD*, *mraF*, *mraY*, etc.), shows that our method of large-scale screening for CG mutants does discover operons containing essential genes.

Of the 179 genes found in this study 117 do not have essential orthologs in *E. coli* or *P. aeruginosa*. Of these 66 are located in operons with at least one essential ortholog in the aforementioned genomes, for example BCAL1508 and BCAL1509 (Figure 2-4). The remaining

51 genes were organized in 19 operons that did not match to any essential orthologs in either *P. aeruginosa* or *E. coli*. We reasoned that if these genes were essential in *B. cenocepacia* they might be conserved at least across closely related species. We therefore examined the distribution of these open reading frames across the *Burkholderia* genus using Burkholderia Ortholog groups from the Burkholderia Genome Database (Winsor *et al*., 2008). Of the 51 genes, 38 were present in all the sequenced genomes of *Burkholderia* species and *B. cenocepacia* strains (Table 2-5). This is in agreement with the assumption that gene conservation among *Burkholderia* genomes is related to essentiality (Juhas *et al*., 2012). However, a few poorly conserved genes that were found in our study also suggest that species or strain-specific requirements for essentiality are also possible. The functions of many of these genes can be inferred from similar genes in other species but the reasons why they may be required for growth by *B. cenocepacia* K56-2 remains elusive. Seventy-one mutants had insertions inside of putative coding sequences (Figure 2-4 and Figure 2-5). The essentiality of the downstream genes then, could then be conditional to the absence of the product encoded by the disrupted gene.

Table 2-4 Genes found in this study with essential orthologs in *E. coli* and *P. aeruginosa*

Locus	*P.aeruginosa* ortholog	*E.coli* ortholog	Function	Essential in[a]	
				P. aerugiosa	*E. coli*
BCAL0031	PA5559	b3737	atpE, ATP synthase	E	N
BCAL0034	PA5556	b3734	atpA, ATP synthase	E	N
BCAL0421	PA0004	b3699	gyrB, DNA gyrase	N	E
BCAL0422	PA0002	b3701	dnaN, DNA polymerase III, beta subunit	N	E
BCAL0426	PA5568	b3705	yidC, membrane protein insertase	N	E
BCAL0465	PA0019	b3287	def, peptide deformylase	E	E
BCAL0466	PA0018	b3288	fmt, Aminoacyl tRNA synthetases	E	E
BCAL0803	PA4668	b1209	lolB, localization of lipoproteins	E	E
BCAL0876	PA5065	b3835	ubiB, 2-octaprenylphenol hydroxylase	E	N
BCAL0880	PA0963	b1866	aspS, aspartyl-tRNA synthetase	N	E
BCAL1004	PA0768	b2568	lepB2, signal peptidase I 2	N	E
BCAL1254	PA1816	b0215	dnaQ, DNA polymerase III epsilon subunit	E	N
BCAL1255	PA1815	b0214	rnhA, degrades RNA of DNA-RNA hybrids	E	N
BCAL1506	PA4745	b3169	nusA, transcription termination	N	E
BCAL1507	PA4744	b3168	infB, Translation inititation	E	E
BCAL1515	PA1585	b0726	sucA, oxoglutarate dehydrogenase	E	N
BCAL1516	PA1586	b0727	sucB, 2-oxoglutarate dehydrogenase	E	N
BCAL1941	PA4931	b4052	dnaB, replicative DNA helicase	N	E
BCAL2154	PA1792	b0524	lpxH, Lipid A biosynthesis	E	E
BCAL2193	PA3809	b2525	fdx, [2Fe-2S] ferredoxin	E	N
BCAL2195	PA3811	b2527	hscB, chaperone specific for IscU	E	N
BCAL2196	PA3812	b2528	iscA, FeS cluster assembly	E	N
BCAL2197	PA3813	b2529	iscU, FeS cluster assembly	E	N
BCAL2390	PA0024	b2436	hemF, coproporphyrinogen III oxidase	E	N

BCAL2391	PA4006	b0639	nadD, NAD biosynthesis	N	E
BCAL2408	PA4997	b0914	msbA, ATP-binding transport protein	E	E
BCAL2677	PA3828	b4261	lptF, lipopolysaccharide export permease	N	E
BCAL2678	PA3827	b4262	lptG, lipopolysaccharide export permease	E	E
BCAL2835	PA0022	None	conserved hypothetical protein	E	NA
BCAL2959	PA3171	b2232	ubiG, 3-demethylubiquinone-93-methyltransferase/ 2-octaprenyl-6-hydroxy phenol methylase	E	N
BCAL3010	PA5338	b3650	spoT, (p)ppGpp synthetase	N	E
BCAL3012	PA5336	b3648	gmk, guanylate kinase	N	E
BCAL3054	PA4053	b0415	ribE, riboflavin synthase	N	E
BCAL3055	PA4052	b0416	nusB, transcription antitermination	E	E
BCAL3110	PA4988	b3633	waaA, KDO transferase	E	E
BCAL3141	PA0404	b2949	yqgF, predicted Holliday junction resolvase	E	E
BCAL3274	PA4655	b0475	hemH, ferrochelatase	E	E
BCAL3306	PA3821	b0408	secD, protein translocase auxillary subunit	N	E
BCAL3307	PA3820	b0409	secF, protein translocase auxillary subunit	E	E
BCAL3420	PA4847	b3255	accB, acetyl CoA carboxylase	E	E
BCAL3421	PA4848	b3256	accC, acetyl-CoA carboxylase	E	E
BCAL3433	PA3746	b2610	ffh, Signal Recognition Particle	N	E
BCAL3455	PA4406	b0096	lpxC, lipidA biosynthesis	N	E
BCAL3457	PA4407	b0095	ftsZ, cell division	N	E
BCAL3458	PA4408	b0094	ftsA, recruitment of FtsK to Z ring	E	E
BCAL3459	PA4409	b0093	ftsQ, ingrowth of wall at septum	N	E
BCAL3461	PA4411	b0091	murC, UDP-N-acetylmuramate:L-alanine ligase	E	E
BCAL3462	PA4412	b0090	murG, peptidoglycan biosynthesis	E	E
BCAL3463	PA4413	b0089	ftsW, stabilizing FstZ ring during cell division	E	E
BCAL3464	PA4414	b0088	murD, peptidoglycan biosynthesis	N	E
BCAL3465	PA4415	b0087	mraY, peptidoglycan biosynthesis	E	E
BCAL3466	PA4416	b0086	murF, peptidoglycan biosynthesis	E	E
BCAL3467	PA4417	b0085	murE, peptidoglycan biosynthesis	E	E

BCAL3468	PA2272	b0084	ftsI, peptidoglycan biosynthesis	N	E
BCAL3469	PA4419	b0083	ftsL, ingrowth of wall at septum	E	E
BCAM0910	PA4043	b0421	putative geranyltranstransferase	N	E
BCAM0911	PA4044	b0420	dxs, 1-deoxyxylulose-5-phosphate synthase	E	E
BCAM0917	PA0577	b3066	dnaG, DNA primase	N	E
BCAM0918	PA0576	b3067	rpoD, RNA polymerase, sigma 70	E	E
BCAM0967	PA1581	b0721	sdhC, succinate dehydrogenase	E	N
BCAM0968	PA1582	b0722	sdhD, succinate dehydrogenase	E	N
BCAM0969	PA1583	b0723	sdhA, succinate dehydrogenase	E	N
BCAM0970	PA1584	b0724	sdhB, succinate dehydrogenase	E	N

[a] E – Essential, N – Nonessential, NA not applicable due to lack of a homolog

Table 2-5 Burkholderia species-specific putative essential operons

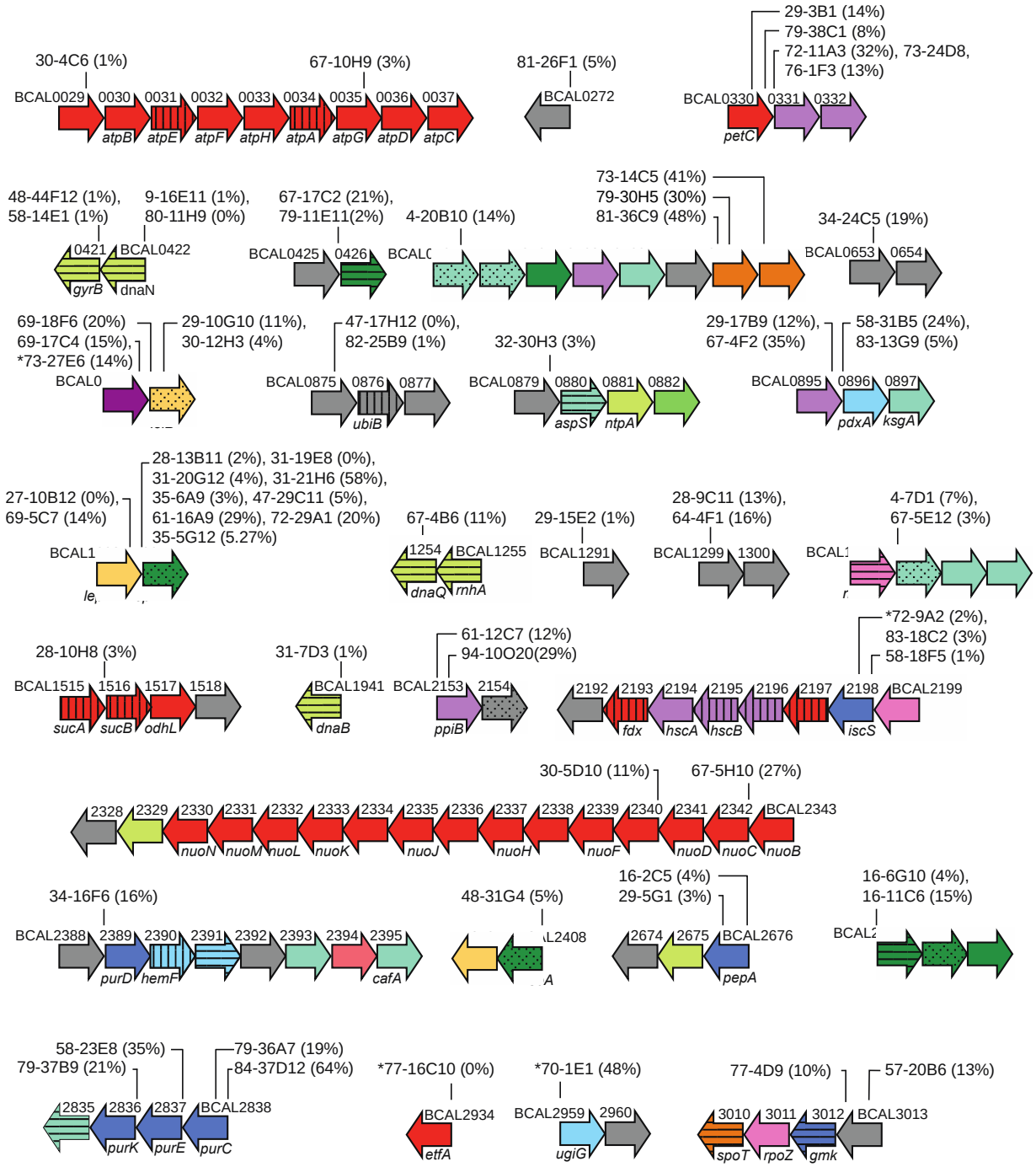
Locus	Gene name or function	Mutants with insertion site at the beginning of or immediately upstream from the locus	Contains at least one gene in the same ortholog group											
			Bce AU 1054	Bce H12424	Bce MC0-3	Bam	Bla	Bma	Bmu	Bph	Bps	Bth	Bvi	Bxe
BCAL0272	Putative lipoprotein	81-26F1	•	•	•	•			•	•			•	•
BCAL0330	petC	29-3B1, 73-24D8, 72-11A3, 73-24D8, 76-1F3	•	•	•	•	•	•	•	•	•	•	•	•
BCAL0331	Putative stringent starvation protein A		•	•	•	•	•	•	•	•	•	•	•	•
BCAL0332	ClpXP		•	•	•	•	•	•	•	•	•	•	•	•
BCAL0653		34-24C5	•	•	•		•	•			•	•		
BCAL0654			•	•	•									
BCAL0691	putative cytidylyltransferase	8-15C6	•	•	•	•	•	•	•	•	•	•	•	•
BCAL0895	Putative surA	29-17B9, 67-4F2	•	•	•	•	•	•	•	•	•	•	•	•
BCAL0896	pdxA	58-31B5, 83-13G9	•	•	•	•		•	•	•	•	•	•	•
BCAL0897	ksgA		•	•	•	•		•	•	•	•	•	•	•
BCAL1291		29-15E2												
BCAL1299		28-9C11, 64-4F1												
BCAL1300			•			•								
BCAL2068		32-32F10	•	•	•	•	•	•	•	•	•	•	•	•
BCAL2343	nuoB	67-5H10	•	•	•	•	•	•	•	•	•	•	•	•
BCAL2342	nuoC		•	•	•	•	•	•	•	•	•	•	•	•
BCAL2341	nuoD		•	•	•	•	•	•	•	•	•	•	•	•
BCAL2340	Putative nuoE	30-5D10	•	•	•	•	•	•	•	•	•	•	•	•
BCAL2339	nuoF		•	•	•	•	•	•	•	•	•	•	•	•
BCAL2338	Putative nuoG		•	•	•	•	•	•	•	•	•	•	•	•
BCAL2337	nuoH		•	•	•	•	•	•	•	•	•	•	•	•
BCAL2336	Putative nuoI		•	•	•	•	•	•	•	•	•	•	•	•

BCAL2335	nuoJ		• • • • • • • • • • • •
BCAL2334	nuoK		• • • • • • • • • • • •
BCAL2333	nuoL		• • • • • • • • • • • •
BCAL2332	nuoM		• • • • • • • • • • • •
BCAL2331	nuoN		• • • • • • • • • • • •
BCAL2330			• • • • • • • • • • • •
BCAL2329	NUDIX hydrolase		• • • • • • • • • • • •
BCAL2328			• • • • • • • • • • • •
BCAL2676	pepA	16-2C5	• • • • • • • • • • • •
BCAL2675	DNA polymerase III chi subunit		• • • • • • • • • • • •
BCAL2674			• • • • • • • • • • • •
BCAL2736	isocitrate dehydrogenase	46-32G1	• • • • • • • • • • • •
BCAL2934	etfA	77-16C10	• • • • • • • • • • • •
BCAL2959	ugiG	70-1E1	• • • • • • • • • • • •
BCAL2960			• • • • • • • • • • • •
BCAL3035	trxB	51-6G10	• • • • • • • • • • • •
BCAL3266	Putative deoxynucleotide kinase	72-10F11, 88-10A4	• • • • • • • • • • • •
BCAL2367	panB		• • • • • • • • • • • •
BCAL3335	fis	64-10H7	• • • • • • • • • • • •
BCAL3336	purH		• • • • • • • • • • • •
BCAL3337	ruvC		• • • • • • • • • • • •
BCAL3338	ruvA		• • • • • • • • • • • •
BCAL3339	ruvB		• • • • • • • • • • • •
BCAL3351	pyrC	76-5B4	• • • • • • • • • • • •
BCAL3350			• • • • • • • • • • • •
BCAM1881		96-1K12	
BCAM1908		86-3D16	

BCAM1909

BCAM1910

Bce, *B. cenocepacia;* Bam, *B. ambifaria;* Bla, *B. lata*; Bma, *B. mallei*; Bmu, *B. multivorans;* Bph, *B. phymatum;* Bps, *B. pseudomallei;* Bth, *B. thailandensis;* Bvi, *B. vietnamiensis;* Bxe, *B. xenovorans*.


30-4C6 (1%)
67-10H9 (3%)
81-26F1 (5%)
29-3B1 (14%)
79-38C1 (8%)
72-11A3 (32%), 73-24D8,
76-1F3 (13%)
BCAL0029 0030 0031 0032 0033 0034 0035 0036 0037
atpB atpE atpF atpH atpA atpG atpD atpC
BCAL0272
BCAL0330 0331 0332
petC
48-44F12 (1%),
58-14E1 (1%)
9-16E11 (1%),
80-11H9 (0%)
67-17C2 (21%),
79-11E11(2%)
4-20B10 (14%)
73-14C5 (41%)
79-30H5 (30%)
81-36C9 (48%)
34-24C5 (19%)
0421 BCAL0422
gyrB dnaN
BCAL0425 0426
BCAL0653 0654
69-18F6 (20%)
69-17C4 (15%),
*73-27E6 (14%)
29-10G10 (11%),
30-12H3 (4%)
47-17H12 (0%),
82-25B9 (1%)
32-30H3 (3%)
29-17B9 (12%),
67-4F2 (35%)
58-31B5 (24%),
83-13G9 (5%)
BCAL0
BCAL0875 0876 0877
ubiB
BCAL0879 0880 0881 0882
aspS ntpA
BCAL0895 0896 0897
pdxA ksgA
28-13B11 (2%), 31-19E8 (0%),
31-20G12 (4%), 31-21H6 (58%),
35-6A9 (3%), 47-29C11 (5%),
61-16A9 (29%), 72-29A1 (20%)
35-5G12 (5.27%)
27-10B12 (0%),
69-5C7 (14%)
67-4B6 (11%)
29-15E2 (1%)
28-9C11 (13%),
64-4F1 (16%)
4-7D1 (7%),
67-5E12 (3%)
1254 BCAL1255
dnaQ rnhA
BCAL1291
BCAL1299 1300
*72-9A2 (2%),
83-18C2 (3%)
58-18F5 (1%)
28-10H8 (3%)
31-7D3 (1%)
61-12C7 (12%)
94-10O20(29%)
BCAL1515 1516 1517 1518
sucA sucB odhL
BCAL1941
dnaB
BCAL2153 2154
ppiB
2192 2193 2194 2195 2196 2197 2198 BCAL2199
fdx hscA hscB iscS
30-5D10 (11%)
67-5H10 (27%)
2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 BCAL2343
nuoN nuoM nuoL nuoK nuoJ nuoH nuoF nuoD nuoC nuoB
34-16F6 (16%)
48-31G4 (5%)
16-2C5 (4%)
29-5G1 (3%)
16-6G10 (4%),
16-11C6 (15%)
BCAL2388 2389 2390 2391 2392 2393 2394 2395
purD hemF cafA
2674 2675 BCAL2676
pepA
58-23E8 (35%)
79-37B9 (21%)
79-36A7 (19%)
84-37D12 (64%)
*77-16C10 (0%)
*70-1E1 (48%)
77-4D9 (10%)
57-20B6 (13%)
2835 2836 2837 BCAL2838
purK purE purC
BCAL2934
etfA
BCAL2959 2960
ugiG
3010 3011 3012 BCAL3013
spoT rpoZ gmk

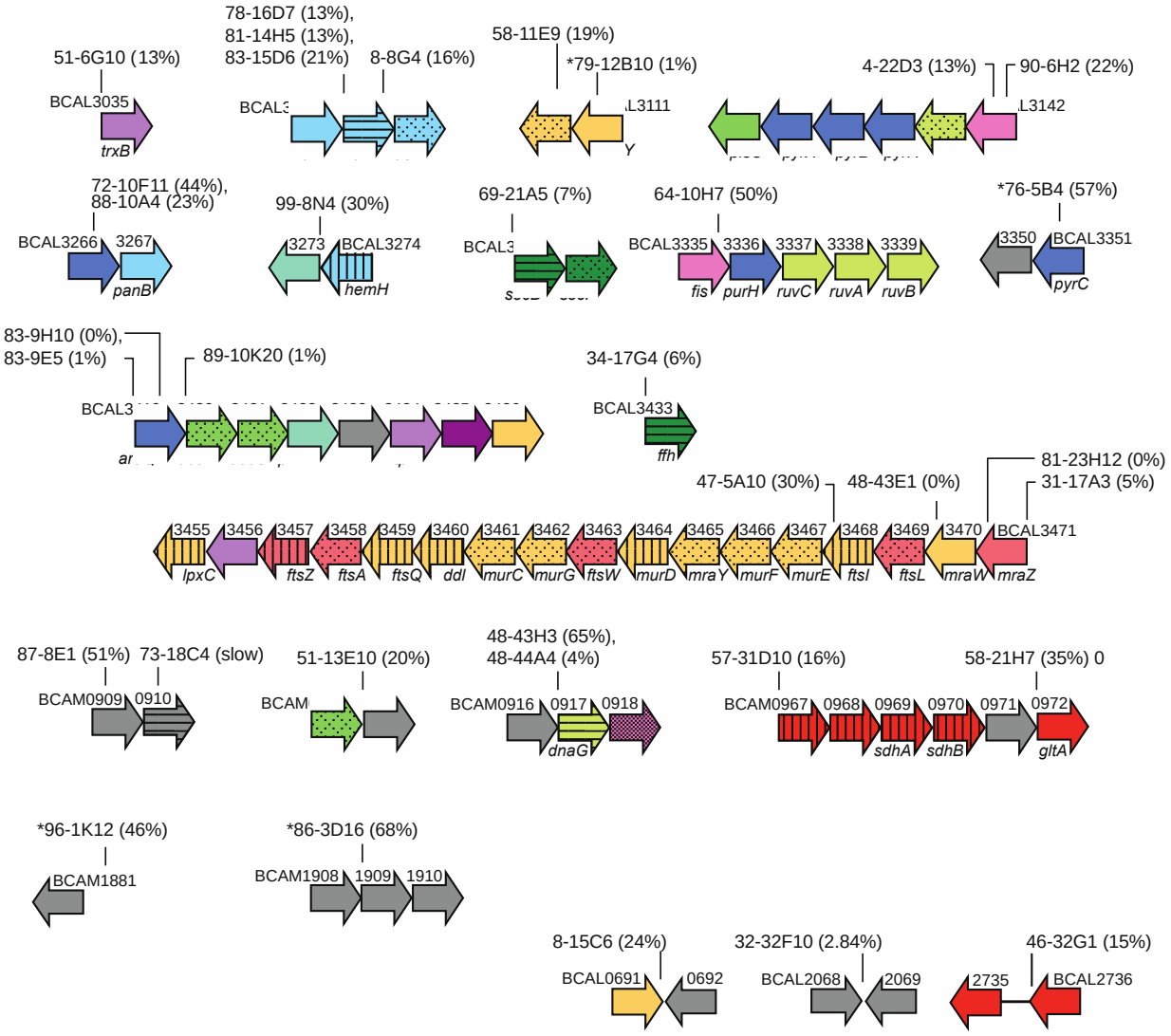

51-6G10 (13%)
BCAL3035
trxB
78-16D7 (13%),
81-14H5 (13%),
83-15D6 (21%)
8-8G4 (16%)
58-11E9 (19%)
*79-12B10 (1%)
4-22D3 (13%)
90-6H2 (22%)
L3142
72-10F11 (44%),
88-10A4 (23%)
BCAL3266
3267
panB
99-8N4 (30%)
3273
BCAL3274
hemH
69-21A5 (7%)
64-10H7 (50%)
BCAL3335
3336
3337
3338
3339
fis
purH
ruvC
ruvA
ruvB
*76-5B4 (57%)
3350
BCAL3351
pyrC
83-9H10 (0%),
83-9E5 (1%)
89-10K20 (1%)
34-17G4 (6%)
BCAL3433
ffh
47-5A10 (30%)
48-43E1 (0%)
81-23H12 (0%)
31-17A3 (5%)
3455
3456
3457
3458
3459
3460
3461
3462
3463
3464
3465
3466
3467
3468
3469
3470
BCAL3471
lpxC
ftsZ
ftsA
ftsQ
ddl
murC
murG
ftsW
murD
mraY
murF
murE
ftsI
ftsL
mraW
mraZ
87-8E1 (51%)
73-18C4 (slow)
BCAM0909
0910
51-13E10 (20%)
48-43H3 (65%),
48-44A4 (4%)
BCAM0916
0917
0918
dnaG
57-31D10 (16%)
58-21H7 (35%) 0
BCAM0967
0968
0969
0970
0971
0972
sdhA
sdhB
gltA
*96-1K12 (46%)
BCAM1881
*86-3D16 (68%)
BCAM1908
1909
1910
8-15C6 (24%)
BCAL0691
0692
32-32F10 (2.84%)
BCAL2068
2069
46-32G1 (15%)
2735
BCAL2736
Functional Categories
Energy production and conversion
Function unknown
Nucleotide / Amino acid transport and metabolism
Cell envelope biogenesis
DNA replication, recombination, and repair
Translation, ribosomal structure and biogenesis
Coenzyme / Lipid metabolism
Posttranslational modification, protein turnove r, chaperones
Transcription
Cell division and chromosome partioning
Intracellular trafficking and secretion
Lipid metabolism
Signal Transduction
Carbohydrate transport and metabolism
Essential homologs
None
E. coli
P. aeruginosa
Both

Figure 2-4. Mutant growth defects and approximate location of insertion sites.

Vertical lines indicate the quarter of the open reading frame containing the insertion site. Operons include genes downstream from the mutant insertion sites called as part of the same operons by either OperonDB or DOOR. Genes are color coded according to putative function and black patterns indicate essentiality of homologs in *E. coli* (horizontal), *P. aeruginosa* (vertical), or both (hash). The percentage in brackets behind every mutant name is the percent of wild type growth in the absence of rhamnose. Operons are arranged in ascending order of locus with the exception of three mutants (46-32G1, 8-15C6 and 32-32F10) whose insertion sites at the end of the last gene of the putative operons.

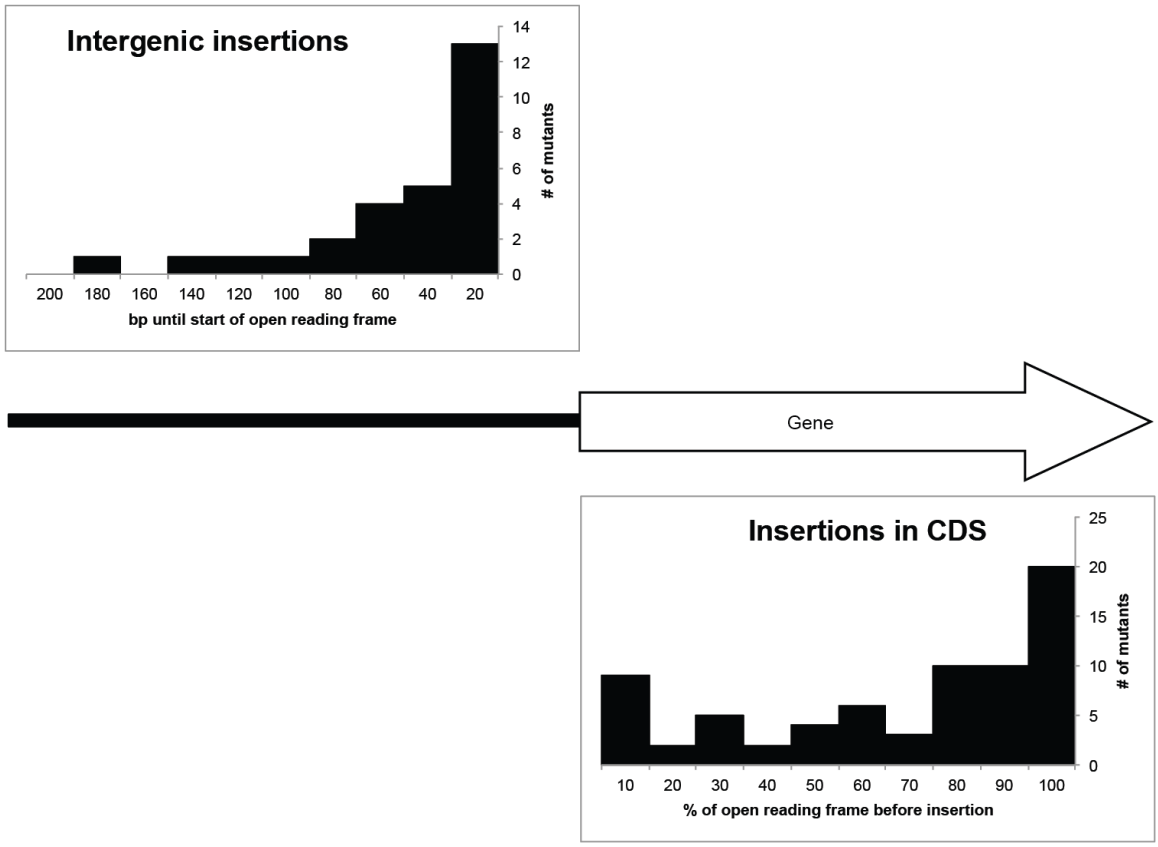


Figure 2-5. Transposon insertions relative to the start of relevant coding sequence.

A) Histogram of the distance from insertions into putative intergenic regions to the putative start codon of the downstream gene measured in base pairs. B) Histogram of the distance from insertions inside of putative coding sequences to the start codon of the surrounding gene measured as a percentage of total gene length.

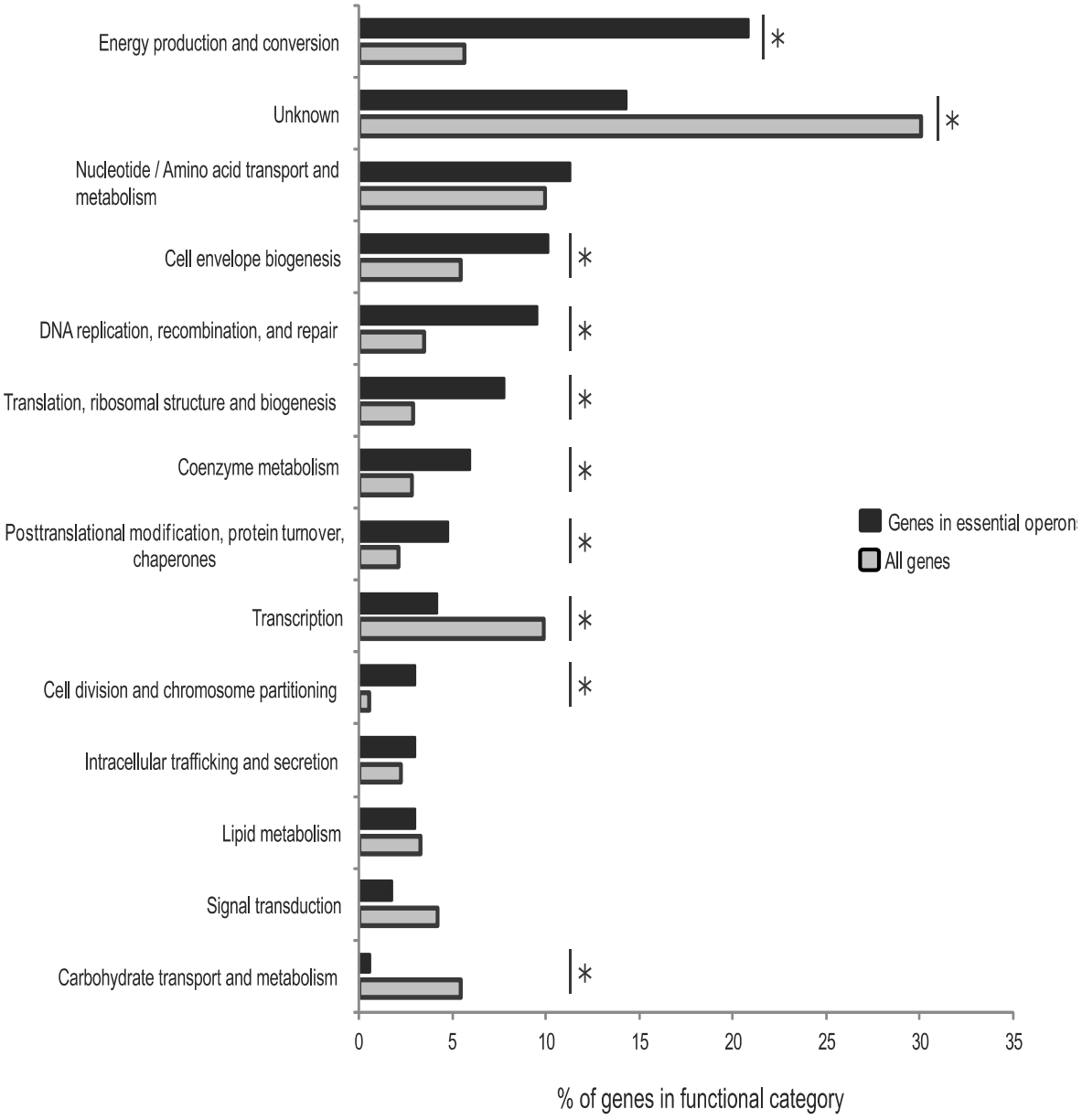

Energy production and conversion
Unknown
Nucleotide / Amino acid transport and metabolism
Cell envelope biogenesis
DNA replication, recombination, and repair
Translation, ribosomal structure and biogenesis
Coenzyme metabolism
Posttranslational modification, protein turnover, chaperones
Transcription
Cell division and chromosome partitioning
Intracellular trafficking and secretion
Lipid metabolism
Signal transduction
Carbohydrate transport and metabolism
*
Genes in essential operon
All genes
0
5
10
15
20
25
30
35
% of genes in functional category

Figure 2-6. Functional categories of genes in putative essential operons.

Putative gene function is based on the GO (Gene Ontology) and COG (Cluster of Orthologous Genes) annotations. For each functional category, Pearson's chi-squared test was used to determine whether the occurrence of the category in the entire genome differs statistically from its occurrence in the putative essential genes identified in this study. A star indicates a p-value of less than 0.05.

2.3.3 Analyzing the rate at which new essential operons were discovered

Progress toward the identification of new essential operons was monitored by plotting the number of unique operons discovered against the total number of CG mutants sequenced (Figure 2-7). We assumed that as more mutants were sequenced, the proportion of new operons was expected to fall at a rate proportional to the fraction of unique operons already discovered. In addition, the proportion of new operons could also depend on whether all essential operons have the same probability of being discovered. We ran computer simulations of randomly chosen essential operons out of a pool of 200 using either an equal chance of picking every operon or applying the frequency distribution of the experimentally identified essential operons to the entire pool. The actual rate at which new operons were being discovered fell below the theoretical predictions (Figure 2-7). When we reduced the theoretical number of essential operons to 70, the simulations matched the experimental results. This means that either that the probability of finding certain essential genes with our methodology is unequal or that *B. cenocepacia* has a significantly lower number of essential genes than other bacterial genomes. We then analyzed two scenarios that could explain the rate at which new essential operons were being found assuming that *B. cenocepacia* has the same number of essential genes as other bacterial species: limited expression range of the promoter and biased transposon insertion. Alternatively, we analyzed *B. cenocepacia* large gene redundancy as a factor that would render fewer essential genes than expected due to gene duplication.

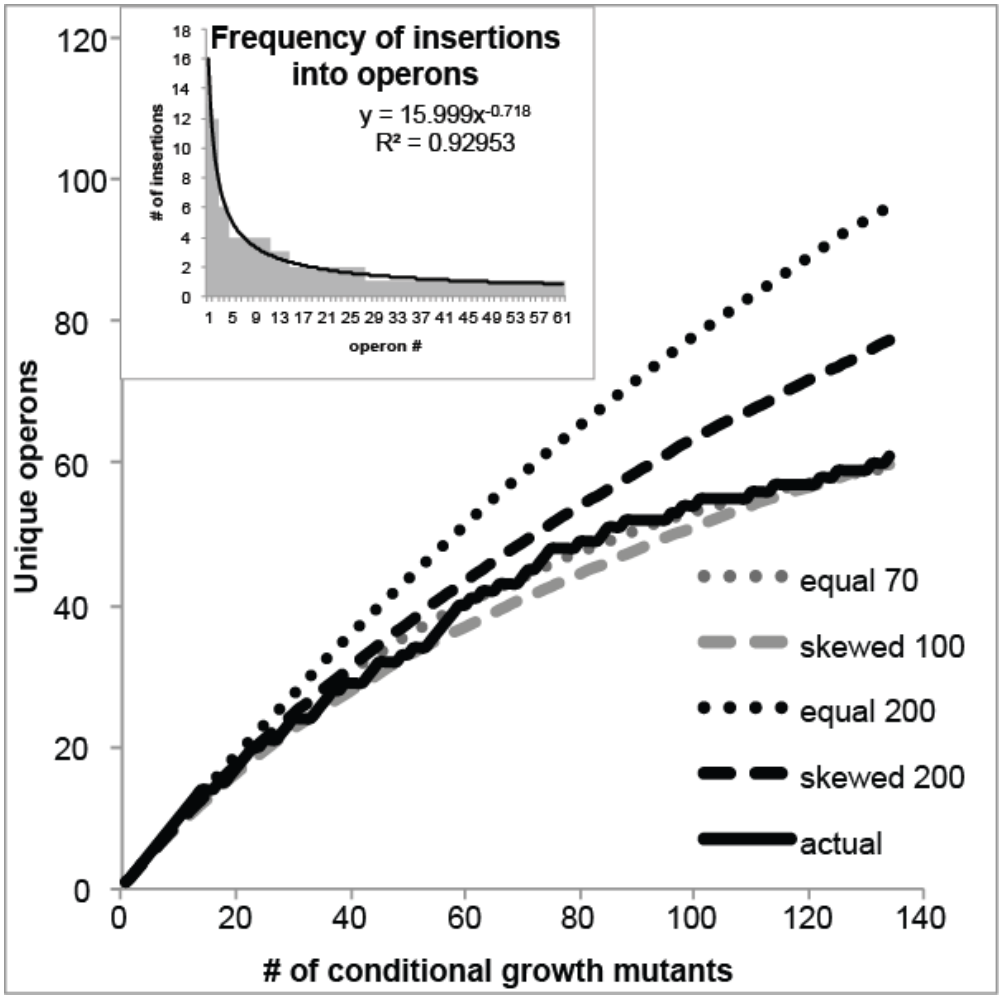


Figure 2-7. The rate at which new operons were discovered.

The insert shows the frequency at which essential operons were discovered experimentally. The rate at which unique operons were discovered experimentally (actual) is compared with 4 simulations averaging 100 trials of 200, 90 or 70 essential operons assuming either that every essential operon is equally likely to be detected (equal) or that the observed frequency distribution applies to all essential operons (skewed).

2.3.4 Global analysis of essential gene expression levels.

It is reasonable to expect that essential genes require strong promoters as essential genes tend to be highly expressed (Dotsch *et al*., 2010). In previous attempts to modulate essential gene expression via inducible promoters, achieving a conditional growth phenotype has been complicated by the requirement for the chosen system to allow high levels of expression required by essential genes while simultaneously providing tight regulation (Xu *et al*., 2010). We previously demonstrated that the PrhaB is tightly regulated and thus can be used to identify essential genes (Cardona 2005 and 2006). However, we did not rule out the possibility of missing highly expressed essential genes due to a narrow dynamic range of expression levels driven by PrhaB. To examine whether our procedure was excluding highly expressed essential genes, we examined whether global gene expression analysis can show that essential genes are on average more highly expressed than nonessential genes, whether the operons identified by our study showed a similar bias in expression, and finally, whether there was a correlation between frequency at which essential operons were recovered and the level of operon expression.

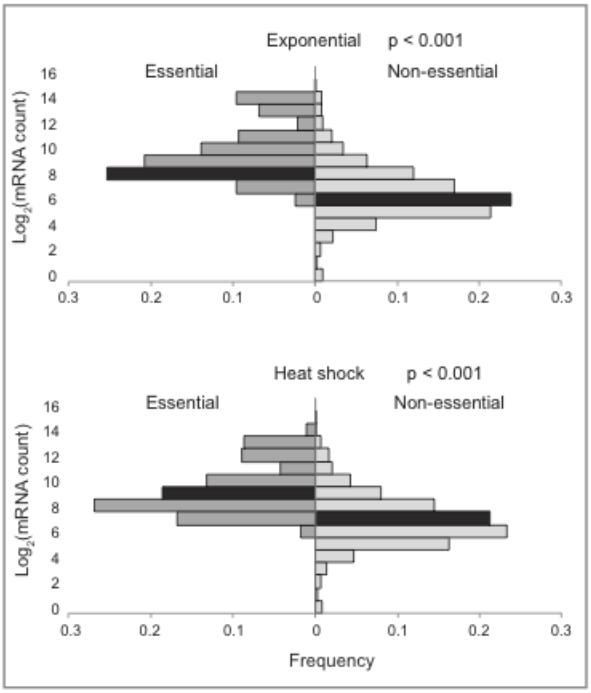
RNA-Seq measures gene expression by sequencing single molecules of mRNA and is thought to provide the most accurate and unbiased absolute quantitation of transcription on a genomic level (Fu *et al*., 2009). Therefore we used previously published RNA-Seq data to examine whether essential genes are more likely to be highly expressed in *E. coli* (Prieto *et al*., 2011). Our analysis showed that *E. coli* essential genes are more highly expressed than nonessential genes in cultures

harvested during exponential growth (Figure 2-8A), after heat-shock treatment (Figure 2-8A), and anaerobic growth (data not shown), with only stationary-phase cells showing no statistically significant difference between essential and nonessential genes (data not shown). As there are no RNA-Seq data available for *B. cenocepacia* J2315 grown in LB, we looked at whether the normalized fluorescence from cDNA microarrays would show similar differences. As microarrays rely on hybridization of labeled cDNA, the fluorescent intensity for any probe depends not only on the number of transcripts, but also on the hybridization efficiency and the possibility of off-target hybridization to other transcripts (Fu *et al*., 2009). These biases make comparisons of expression levels between different genes based on differences in microarray fluorescence questionable. As we are interested in the difference in expression between essential and nonessential genes in general, these biases should be equally present in both classes of genes. Therefore, we hypothesized that microarrays could substitute for RNA-Seq for our purposes. We confirmed that previously published cDNA microarray data (DeVito *et al*., 2002) showed similar differences in expression between essential and nonessential *E. coli* genes (Figure 2-8B), confirming that there is a bias toward higher levels of expression by essential genes and that this difference can be observed using cDNA microarrays. To determine whether the operons in the CG mutant library were similarly biased toward highly expressed genes, we repeated this analysis using previously published cDNA microarray data for *B. cenocepacia* J2315 grown in LB (Bazzini *et al*., 2011). As the expression of genes within operons is highly correlated, only the first genes in the identified operons were included in the analysis. There was a significant bias towards high expression for genes from putative essential operons (Figure 2-8C). In addition, only 43 genes not represented in our CG mutant library showed higher

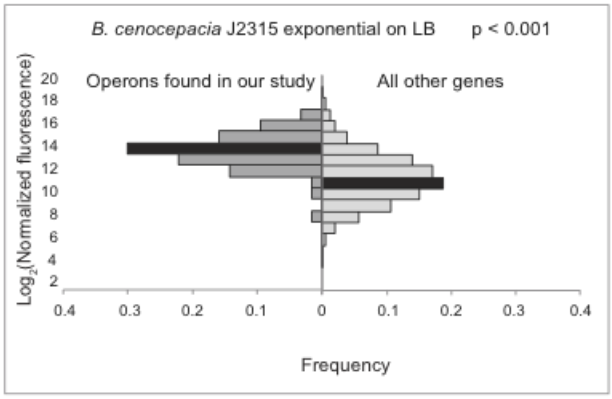
expression than the most highly expressed operons identified experimentally. Of these, 15 genes in 5 putative operons have essential orthologs in *E. coli* or *P. aeruginosa,* and all had fluorescence within 2.1 fold of that of the experimentally identified genes. These data suggest the P*rhaB* can drive the levels of expression needed by the vast majority of essential genes.

We next asked if the higher frequency of finding insertions upstream of certain operons was related to hotspots of the Tn*5*-mini transposon. These hotspots have been associated with negatively super-coiled regions (Lodge & Berg., 1990) that could be created upstream of highly transcribed genes (Rovinskiy *et al*., 2012). We then analyzed the relationship between the frequency of recovering CG mutants of certain operons and transcription levels. However we were unable to find a correlation between a high number of CG mutants recovered for essential operons and high level of expression for those operons (Figure 2-8D). This suggests that regions of negatively super-coiled DNA that are usually upstream highly expressed genes are not hotspots for Tn*5*-mini insertions. Taken together, neither the range of expression levels of the P*rhaB* promoter nor hotspots for transposon mutagenesis due to the presence of highly expressed genes can completely explain the observed rate at which new essential operons were discovered.

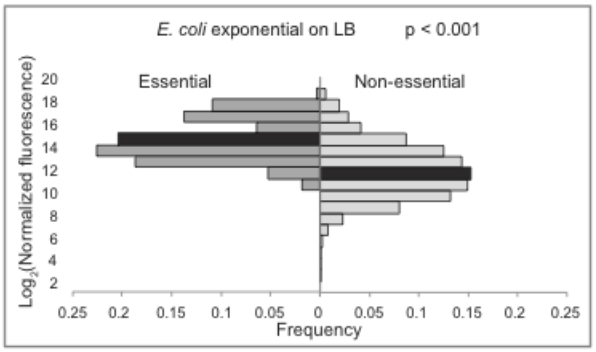
a) *E. coli*, RNA-Seq



c) *B. cenocepacia* J2315, Microarray



b) *E. coli*, Microarray



d) Correlation between insertions and gene expression

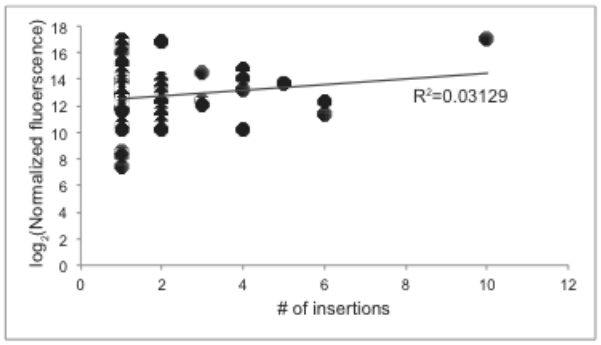


Figure 2-8. The distribution of gene expression in putatively essential and nonessential genes.

The black bars contain the median level of gene expression for each class. The distribution of expression between essential and nonessential genes for each species/condition/methodology was compared using a Mann–Whitney test. Median values, ***U*** statistic, and number (***n***) of essential and nonessential genes are as follows. (A) Exponentially growing cells (median mRNA transcripts: 364.7 essential, 51.8 nonessential, $\boldsymbol{U} = 893{,}584$, $\boldsymbol{P} < 0.001$), heat-shock treatment (median mRNA transcripts: 294 essential, 69.3 nonessential, $\boldsymbol{U} = 786{,}268$, $\boldsymbol{P} < 0.001$) $\boldsymbol{n}_{\text{nonessential}} = 4038$, $\boldsymbol{n}_{\text{essential}} = 280$; (B) cDNA microarray for exponentially growing cells (median fluorescence: 16,700.09 essential, 3328.65 nonessential, $\boldsymbol{U} = 701{,}538$, $\boldsymbol{n}_{\text{essential}} = 280$, $\boldsymbol{n}_{\text{nonessential}} = 3868$, $\boldsymbol{P} < 0.001$); (C) cDNA microarray for *Burkholderia cenocepacia* J2315 (median fluorescence: 10,448.6 operons in library, 1864.7 other genes, $\boldsymbol{U} = 346{,}475$, $\boldsymbol{n}_{\text{essential}} = 6875$, $\boldsymbol{n}_{\text{nonessential}} = 63$, $\boldsymbol{P} < 0.001$); (D) There is no correlation between levels of gene expression and the frequency at which insertions into an operon were recovered.

2.3.5 Analysis of gene redundancy in *B. cenocepacia.*

To explain the lower than expected rate at which we were discovering essential operons (Figure 2-7), we reasoned that a higher number of genes encoding for essential proteins have to be duplicates in comparison with other essential genomes. To address this possible gene-redundancy effect, we first analyzed the criteria for two genes to be considered duplicated. While two exact copies of a gene are undoubtedly duplicates, duplicated genes that diverged after the duplication event may only share a certain DNA sequence similarity over a partially alignable region. To analyze the presence of duplicate genes in *B. cenocepacia,* considering not only identical copies but also duplicate genes that may have further differentiated, we estimated the presence of such genes at different stringency cutoff levels. Similar to other studies of gene duplication (Gu *et al*., 2002), two parameters were considered for defining stringency: DNA sequence identity and the proportion of alignable sequence over a gene. We compared gene duplications of *B. cenocepacia* with those of *P. aeruginosa* and *E. coli* at all levels of stringency. For example, two genes were considered duplicates with a stringency cutoff of 40 if they shared a DNA sequence identity and percent of alignable sequence equal or higher than 40%. The *B. cenocepacia* genome showed higher gene duplication than the *P. aeruginosa* and *E. coli* genomes at all stringency cutoff levels, but the proportion of duplicates varied greatly with stringency cutoff (Figure 2-9). When exact copies were considered (stringency cutoff of 100) approximately 2% and 1% of the genes came out as duplicates for *B. cenocepacia* and *E. coli*, respectively (Figure 2-9, inset). This twofold difference was also observed at a stringency cutoff of 60, where the percentage of duplicates increased to 7.6 % and 3.5% in *B. cenocepacia* and *E.*

coli, respectively. Thus, if the same proportion of duplicated genes observed in whole genomes is present in essential genomes, then gene redundancy could explain the previously observed lower than expected hit frequency.

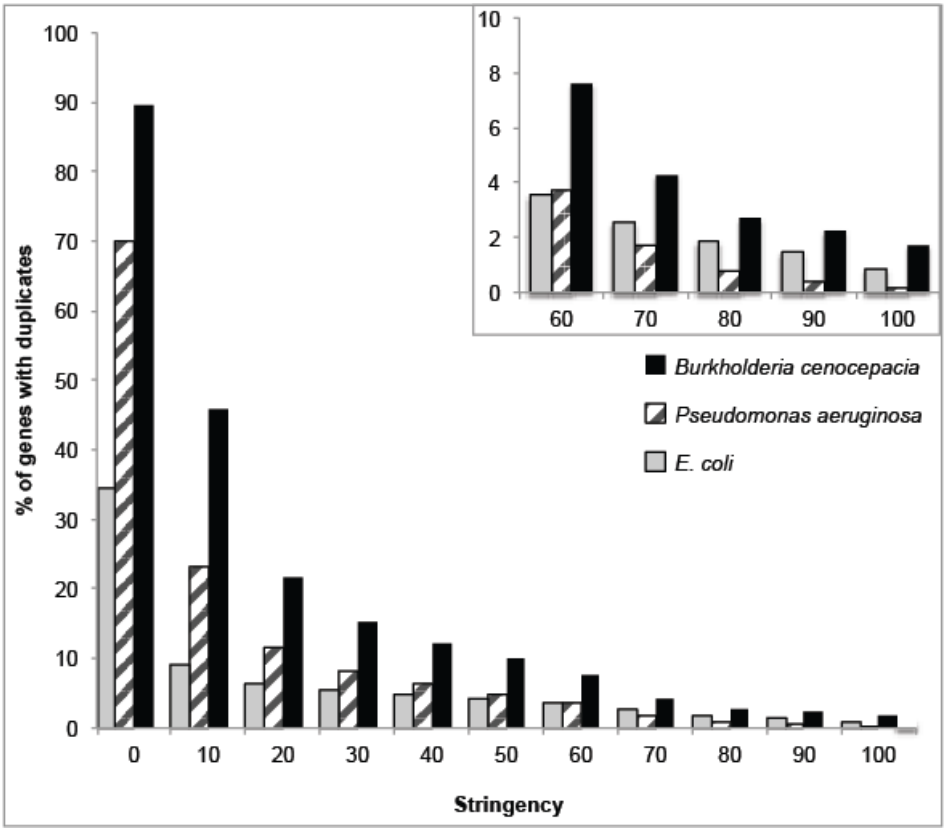


Figure 2-9 Genetic duplication in *Burkholderia cenocepacia* J2315, *Pseudomonas aeruginosa* PA01, and *Escherichia coli* K12 as a function of DNA sequence similarity.

For each strain, the genome was downloaded from the Genome directory of NCBI and a BLAST database was built containing all annotated coding regions. For each gene, similar genes within the same genome were identified using blastn with an expect cutoff of 0.1. Stringency cutoff was defined by percent coverage in the gene alignment and % sequence identity. A stringency cutoff of 40%, for example, means that both percent coverage and sequence identity between two genes are equal or higher than 40%. A gene was included in the same paralogous group when percent

coverage and percent identity with at least one member of the group satisfied the stringency cutoff. The inset shows a scaled-up figure of the 60–100 stringency cutoff. Note that paralogous group denotes genes that meet the required conditions for inclusion within the group without any reference to gene history or gene evolution.

2.3.6 CG mutants demonstrate selective hypersensitivity at low rhamnose concentrations.

If a small molecule with antibacterial activity exerts its effect by binding and inhibiting an essential protein, then underexpressing this essential gene should cause cells to become more sensitive to that small molecule (DeVito *et al*., 2002, Donald *et al*., 2009). This hypersensitivity should allow growth inhibitors to be matched to their specific molecular targets. We then reasoned that only a CG mutant of the *gyrB* gene should show enhanced sensitivity to the antibiotic novobiocin and other CG mutants should not. Novobiocin exerts its inhibitory action by binding the GyrB subunit of DNA gyrase (Lewis *et al*., 1996). Using 12 different concentrations of rhamnose and a novobiocin concentration that inhibits 30% of wild type growth (IC_{30}), we established that a *B. cenocepacia gyrB* mutant was hypersensitive at rhamnose concentrations that produced less than 60% of wild-type growth (data not shown). Next, we tested the sensitivity to novobiocin of the CG mutant library at various rhamnose concentrations. We reasoned that while mutants under too little stress may not show any hypersensitivity, severely stressed cells could be generally hypersensitive to growth inhibitors. Reducing growth to 30% or less of wild-type growth caused hypersensitivity and high variability across experiments for most of the CG mutants (data not shown). We then screened 25 randomly chosen mutants at rhamnose concentrations that produced 30-60% of wild-type growth (Figure 2-10). Chloramphenicol, which inhibits protein synthesis by binding to the ribosome (Wilson, 2011), was used as a negative control at its IC_{30}. Only mutants with transposon insertions upstream of *gyrB* were hypersensitive to novobiocin (Figure 2-10A) and none of the CG mutants

were hypersensitive to chloramphenicol (Figure 2-10B).

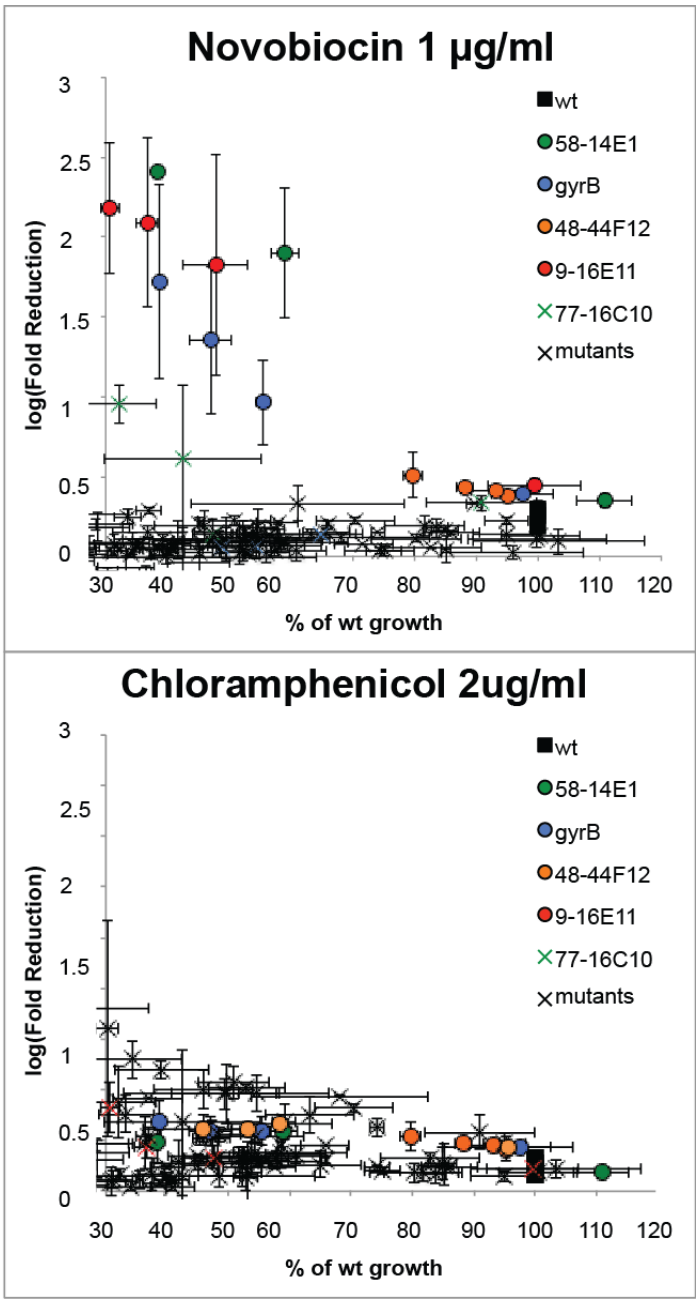


Figure 2-10 Conditional growth mutants show selective hypersensitivity.

Mutants were grown in rhamnose concentration gradients estimated to produce more than 30% of wild type growth and challenged with either novobiocin or chloramphenicol at the IC_{30} of the wild type. Circles represent CG mutants of the direct target of novobiocin, gyrB. Green crosses represent a CG mutant of the electron transfer flavoprotein gene (*etfA).* Black crosses correspond to non-sensitive mutants. Mutants of *gyrB* show hypersensitivity to novobiocin when grown in rhamnose concentrations producing 30 to 60% of wild type growth but not when grown in rhamnose concentrations producing 80 to 100% of wild type growth. A CG mutant of *etfA* shows intermediate hypersensitivity. None of the mutants showed hypersensitivity to chloramphenicol. Error bars represent 1 standard deviation calculated from 2 biological replicates.

2.4 Discussion

In this work, we demonstrate that random transposon mutagenesis with an outward-facing promoter P*rhaB* followed by screening for a rhamnose-dependent CG phenotype can identify putative essential operons at a large scale while simultaneously constructing CG mutants. The recovery of genes with essential orthologs in other species known to be involved in essential cellular processes (e.g. *ftsZ*) shows that this methodology is capable of recovering mutants of truly essential genes on a large scale. While secondary site mutations that produce a rhamnose-conditional phenotype cannot be ruled out, we consider this possibility to be unlikely. First, rhamnose is not used as a carbon source in *B. cenocepacia* (data not shown), and it is only known to be part of the O-antigen, a nonessential component of the lipopolysacharide (Ortega *et al*., 2005, Ortega *et al*., 2007). Second, we recovered independent transposon mutants that contain the rhamnose promoter controlling the same gene or gene cluster. Finally, independent CG mutants of *gyrB* were constructed showing the same phenotype as recovered by transposon mutagenesis.

If the number of essential genes in B. cenocepacia essential genome is not substantially different from what was found in other genomes, our study could identify 20–35% of the B. cenocepacia essential operons using a single inducible promoter and single screening condition. A possible

explanation for the lower than expected hit frequency could be that the *B. cenocepacia* genome is not uniformly available to the transposon system used. The Tn*5*-mini transposon has been shown to insert into a highly degenerate consensus sequence (Shevchenko *et al*., 2002) allowing for almost unbiased insertions based on sequence. However, hotspots are common, and are associated with local DNA topology with highly transcribed negatively supercoiled regions being more favorable (Lodge & Berg., 1990). This suggests that operons with high levels of expression should be more accessible to the transposon and hence insertions in these operons should be recovered more frequently. However, among the recovered CG mutants, there was no correlation between the frequency of insertions into an operon and the operon's level of expression (Figure 2-6D). Conversely, the presence of DNA regions that may never be targeted by a transposon should be considered. Previously, two independent studies performed in *P. aeruginosa* PAO1 and PA14 used two different transposons, a Tn*5*–based system and a mariner transposon, respectively, to identify putative essential genes (Jacobs *et al*., 2003, Liberati *et al*., 2006). Approximately half of the 678 PA14/PAO1 orthologs not hit with Tn5 were disrupted by the mariner-based transposon in the PA14 library. These 343 genes account for 6.7% of P. aeruginosa PAO1/PA14 orthologs, suggesting that only 6.7% of PAO1 genome was missed by the Tn5 system due to cold spots. Then, a similar distribution of cold spots in B. cenocepacia could not account for the rate of essential operon discovery in our study.

The promoters for essential genes differ from those of nonessential genes by having high levels of expression and lower levels of noise (Silander *et al*., 2012). In a previous study, genome-wide

promoter replacement of essential genes in *S. aureus* was attempted by site-directed delivery of tetracycline-inducible promoters with different expression levels. Of 150 essential genes that were targeted with three different promoter variants in more than 400 different constructs, only 64 essential genes were found to have a CG phenotype (Xu *et al*., 2010). An analysis of the expression levels of the *B. cenocepacia* K56-2 recovered operons showed that the most highly expressed genes in *B. cenocepacia* J2315 were included and that only five predicted essential operons had higher levels of expression than the most highly expressed operons recovered by our study. This suggests that the rhamnose-inducible promoter used is capable of driving the expression required by the vast majority of *B. cenocepacia* essential genes when the promoter is delivered randomly and CG mutants are isolated by their conditional phenotype.

The presence of duplicated essential genes in *B. cenocepacia* could lower the hit frequency of our methodology. We then analyzed the possible effect of gene redundancy on the rate at which we were finding new essential operons. Our analysis predicted that for this effect to account for the lower than expected frequency of finding essential operons, gene duplication in *B. cenocepacia* should be at least twice as high as in *E. coli*. Our analysis of gene redundancy of *B. cenocepacia* J2315 in comparison with the ones of *E. coli* and *P. aeruginosa* showed that *B. cenocepacia* gene redundancy indeed seems to be twice as high. However, essential genes tend to have fewer paralogs on average than nonessential genes (Deng *et al*., 2011). Therefore, it is unknown if the observed gene redundancy in *B. cenocepacia* also occurs among essential genes.

In summary, the reasons for the modest success in promoter replacement of essential genes are currently unknown and could be due to effects of the selection conditions or loss of necessary regulatory regions that account for the genomic organization of essential operons in regulons. The frequency at which CG mutants for different essential operons are recovered could also depend on the length of the region into which the transposon can insert and still produce a phenotype (Christen *et al*., 2011). Promoter replacement of essential genes that occur in clusters or have very short promoters may require high-throughput approaches such as those performed with high density transposon mutagenesis and next generation sequencing technology (Gawronski *et al*., 2009, Langridge *et al*., 2009).

This study also shows that the *B. cenocepacia* library of CG mutants can be used to screen for the targets of specific growth inhibitors. However, one limitation of this methodology is the inability to detect the target of antibiotics that inhibit growth by forming toxic complexes, rather than by target inactivation. During an antisense RNA study (Xu *et al*., 2010) a *S. aureus gyrA* mutant was not hypersensitive to ciprofloxacin, an observation we were able to confirm with a *B. cenocepacia* CG mutant of *gyrA* (data not shown). Identifying the target of a small molecule with growth-inhibitory characteristics using the developed assay requires screening the whole library against each compound of interest. Since the library is sensitive over a broad range of rhamnose concentrations and sublethal concentrations of novobiocin, we predict that it will be possible to screen pools of CG mutants with broadly similar rhamnose sensitivities. This will allow for the

simultaneous screening of many targets against small molecules with growth-inhibitory properties.

3 An electron transfer flavoprotein (ETF) is essential for viability and its depletion causes a rod-to-sphere change in *Burkholderia cenocepacia*

This chapter has been submitted to *Microbiology*, 2015 as: Ruhi A.M. Bloodworth, Soumaya Zlitni, Eric D. Brown and Silvia T. Cardona. An electron transfer flavoprotein (ETF) is essential for viability and its depletion causes a rod-to-sphere change in *Burkholderia cenocepacia* Microbiology. In press

3.1 Introduction

The advent of next generation sequencing has led to a wealth of genetic knowledge but the ability to assign phenotypes to these genotypes has lagged. One basic phenotype is whether genes are absolutely required for growth in standard conditions. The search for these so called essential genes is given added urgency by their potential as novel targets for antibiotics in an era of increasing antibiotic resistance (Cardona *et al*., 2014, Fang *et al*., 2005, Gerdes *et al*., 2006, Juhas *et al*., 2012). High-throughput screens (HTS) for essential genes normally involve high-density transposon mutagenesis followed by mapping transposon insertion sites in the population to identify the non-essential genes (Langridge *et al*., 2009). These screens provide an accurate and cost effective method to identify essential genes on a genomic scale. However, the

confidence of assigning essentiality to a gene depends on the transposon insertion density. At lower densities, non-essential genes may not have insertions due to chance or insertion biases, while essential genes may tolerate insertions that do not disrupt required domains. While this limitation can partially explain why some genes appear to be essential when they encode products of demonstrated dispensability, HTS screens for essential genes also uncover novel essential functions for the products of genes with already characterized non-essential homologs in other species.

One example where HTS essentiality results and known function of gene products conflict is that of the electron transfer flavoproteins (ETFs). These soluble FAD-containing heterodimeric proteins are distributed across all domains of life and are responsible for funneling electrons from dehydrogenases to the membrane bound respiratory chain of bacteria and mitochondria (Toogood *et al*., 2007) or to nitrogen fixation in some bacteria (Scott & Ludwig., 2004). ETFs consist of a large (ETF-α) and small (ETF-β) subunit and based on comparative amino acid sequence analysis, they are traditionally divided into 3 groups, which have different biological functions (Tsai & Saier., 1995). Group I ETFs, found in eukaryotes (Ghisla & Thorpe., 2004), and some bacteria including *Clostridium kluyveri* (Li *et al*., 2008), *Paracoccus denitrificans* (Roberts *et al*., 1999) and *Bacillus subtilis* (Matsuoka *et al*., 2007) funnel electrons from the oxidation of fatty acids, branched chain amino acids (Li *et al*., 2008), lysine and tryptophan (Roberts *et al*., 1999) to the electron transport chain (Figure 3-1). Group II ETF genes (*fixB*, *fixA*) are found in nitrogen fixing bacteria where they divert electrons from dehydrogenases to

nitrogenases bypassing the ETC (Scott & Ludwig., 2004) (Figure 3-1). The only studied Group III ETF is found in *E. coli* where it is required for the anaerobic reduction of carnitine (Walt & Kahn., 2002). Multiple ETFs are present in some bacterial genomes, as in the case of *Bradyrhizobium japonicum*, which contains both a group I ETF expressed in aerobic conditions, and a group II ETF that is expressed during anaerobic growth (Weidenhaupt *et al*., 1996). Intriguingly, HTS have identified ETFs as putatively essential in *Acinetobacter baylyi* (de Berardinis *et al*., 2008), *Caulobacter crescentus* (Christen *et al*., 2011), *Porphyromonas gingivalis* (Klein *et al*., 2012), *Mycobacterium tuberculosis* (Griffin *et al*., 2011), *Sphingomonas wittichii* RW1 (Roggo *et al*., 2013) and *Burkholderia pseudomallei* (Moule *et al*., 2014). The existence of ETFs with an essential role is unexpected because nitrogen fixation is nonessential on rich media and metabolism of many carbon sources is expected to bypass ETFs. For example, both NADH and succinate dehydrogenases transfer electrons to ubiquinone without the involvement of ETFs (Figure 3-1).

We previously reported a screen for essential genes in *Burkholderia cenocepacia* K56-2 (Bloodworth *et al*., 2013), a clinical isolate recovered from the sputum of a cystic fibrosis patient (Mahenthiralingam *et al*., 2000). *B. cenocepacia* belongs to the *Burkholderia cepacia* complex, a group of multiple-antibiotic resistant Gram-negative species (Vandamme & Dawyndt., 2011) known to infect the airways of people with CF (Drevinek & Mahenthiralingam., 2010). In our study, we identified a group I ETF-α coding gene, *etfA*, which its conditional expression led to a complete inhibition of growth. This finding, together with the inconsistency of this phenotype

with previously described functions, led us to further characterize the effect of depleting one of the three ETFs in *B. cenocepacia*. In this article, we confirm that the *etfBA* operon (BCAL2394, BCAL2395) is essential in *B. cenocepacia* K56-2. We further demonstrate that cells depleted of the corresponding protein (EtfBA) lose viability and redox potential. These cells also transition from the short rods characteristic of *B. cenocepacia* to small spheres. We propose that the lack of aerobic respiration and small sphere phenotype of cells depleted of EtfBA are related, with the metabolic defect inducing the morphologic change.

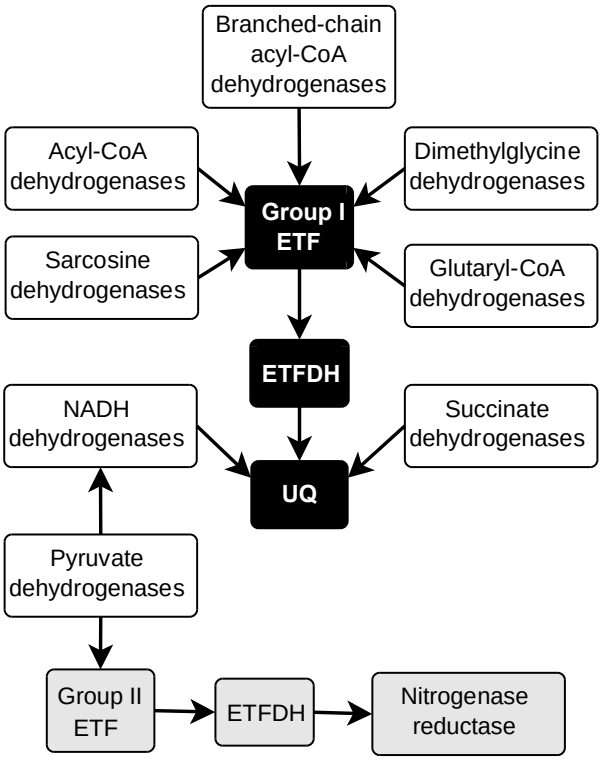


Figure 3-1 A composite figure of electron paths through ETFs.

Group I ETFs funnel electrons from a variety of species-specific primary dehydrogenases (additional partners may be present in some species) to the ETF dehydrogenase (ETFDH) from which they enter the electron transport chain (ETC) at the ubiquinone pool (UQ) (Ghisla & Thorpe., 2004). Group II ETFs divert electrons away from primary dehydrogenases towards nitrogenase reductase. Single knockouts of primary dehydrogenases do not impair symbiotic nitrogen fixation so it is expected that group II ETFs have multiple electron donors, though the only confirmed partner is pyruvate dehydrogenase in *Azorhizobium caulinodans* (Scott & Ludwig., 2004). The only studied group III ETF is known to be required for the anaerobic reduction of carnitine though the direction of electron flow is unknown (Walt & Kahn., 2002). NADH dehydrogenases and succinate dehydrogenases transport electrons to UQ independently of ETFs (Anraku, 1988). Arrows represent experimentally demonstrated electron flow. White, black and grey boxes denote primary dehydrogenases, proteins in the ETF I pathway leading to the ETC, and proteins in the ETF II pathway leading to nitrogenase, respectively.

3.2 Experimental Procedures

3.2.1 Bacterial strains and growth conditions

Strains and plasmids used in this study are listed in Table 3-1. Unless otherwise indicated all strains were grown in Luria Bertani (LB) media at 37°C supplemented as required with 0.2% (w/v) rhamnose, 100 μg ml^{-1} or 50 μg ml^{-1} trimethoprim (Tp) for *B. cenocepacia* or *E. coli* respectively, 50 μg ml^{-1} gentamicin (Gm) and 40 μg ml^{-1} kanamycin (Km). Growth was monitored by optical density at 600 nanometers (OD_{600nm}) using a BioTek Synergy 2 plate

reader. Growth experiments were inoculated with a 10^{-3} dilution of a 1 OD_{600nm} culture, while viability experiments used a 10^{-2} dilution. Strains grown longer than 24h in a single condition were diluted 10^{-3} into fresh media every 24h. All chemicals were purchased from Sigma Chemical Co., St. Louis, MO. unless otherwise indicated.

Table 3-1 List of strains and plasmids used in this study

Strain or plasmid	Relevant information	Source
E. coli		
BTH101	F$^-$, *cya-99, aaraD139, galE15, galK16, rpsL1 (Str^r), hsdR2, mcrA1, mcrB1*	Euromedex
SY327	*araD* Δ(*lac pro*) *argE* (Am) *recA*56 Rif^r *nalA* λ *pir*	(Miller & Mekalanos., 1988)
Burkholderia cenocepacia		
K56-2	(LMG18863), ET12 linage, CF isolate	(Mahenthiralingam *et al*., 2000)
CgyrB	K56-2; *rhaP::*BCAL0421; Tp^r, rhamnose dependent *gyrB* expression	(Bloodworth *et al*., 2013)
CetfBA	K56-2; *rhaP*::*e*BCAL2935; Tp^r, rhamnose dependent *etfBA* expression	This study
Detfdh2	K56-2; ΔBCAS0609, deletion of putative ETF dehydrogenase	This study
Cetfdh1	K56-2; *rhaP*::BCAL1468; Tp^r, rhamnose dependent *etfdh1* expression in K56 background	This study
Cetfdh	K56-2; *rhaP*::BCAL1468,ΔBCAS0609; Tp^r, rhamnose dependent *etfdh1* expression in Δ*etfdh2* background	This study
16-2C5	K56-2; Tn insertion at chr1:2944874 reverse strand, upstream from BCAL2676 (*pepA*)	(Bloodworth *et al*., 2013)
16-11C6	K56-2; Tn insertion at chr2: 2944852, upstream from BCAL2677 (putative permease)	(Bloodworth *et al*., 2013)

29-3B1	K56-2; Tn insertion at chr1: 358492, upstream from BCAL0330 (*petC*).	(Bloodworth *et al*., 2013)
30-4C6	K56-2; Tn insertion at chr1:33394, upstream from BCAL0029 (F0F1 ATPase)	(Bloodworth *et al*., 2013)
77-16C10	K56-2; Tn insertion at chr1: 3213629 reverse strand, upstream from BCAL2934 (*etfA)*	(Bloodworth *et al*., 2013)
78-16D7	K56-2; Tn insertion at chr1: 3347313, upstream from BCAL3053 (*ribB*)	(Bloodworth *et al*., 2013)
89-10K20	K56-2; Tn insertion at chr1: 3745902, upstream from BCAL3420 (*accB)*	(Bloodworth *et al*., 2013)
Plasmids		
pSC200	pGpΩTp derivative (Flannagan *et al*., 2007), *ori* $_{R6K,}$ *rhaR rhaS PrhaB* e-*gfp*	(Ortega *et al*., 2007)
pRB10	pSC200 containing 5' end of BCAL2935	This study
pRB18	pSC200 containing 5' end of BCAL1468	This study
pAP20	*ori*$_{PBBR1}$P$_{DHFR}$ Cm^r	(Law *et al*., 2008)
pAP20-etfBA	pAP20 derivative containing the BCAL2935-BCAL2934 operon	This study
pAP20-etfB	pAP20 derivative containing the BCAL2935 locus	This study
pAP20-etfA	pAP20 derivative containing the BCAL2934 locus	This study
pGPI-SceI	SceI recognition sequence, Tp^r	(Flannagan *et al*., 2008)
pDAI-SceI	DHFR promoter expressing e-ISce-I, Tc^r	(Flannagan *et al*., 2008)
pRB17	Derivative of pGPI-SceI containing the upstream and downstream flanking regions of BCAS0609	This study

pKT25	*p15* ori Km^r; vector for N-terminal fusion with T25	Euromedex
pUT18	ColE1 *ori* Ap^r; vector for N-terminal fusion with T18	Euromedex
pU18c	ColE1 *ori* Ap^r; vector for C-terminal fusion with T18	Euromedex
pKT25-zip	Leucine zipper motif cloned into pKT25	Euromedex
pUT18c-zip	Leucine zipper motif cloned into pU18c	Euromedex
pUT18-etfA	Derivative of pUT18 containing a truncation of BCAL2934 (missing 12 bp from 3' end)	This study
pKT25-etfA	Derivative of pKT25 containing a truncation of BCAL2934 (missing 12 bp from 3' end)	This study
pUT18-etfB	Derivative of pUT18 containing BCAL2935 locus	This study
pKT25-etfB	Derivative of pKT25 containing BCAL2935 locus	This study

Ap^r Ampicillin resistance, Cm^r Chloramphenicol resistance, Km^r Kanamycin resistance, Tp^r Trimethoprim resistance, Tc^r Tetracycline resistance

3.2.2 Molecular biology techniques

The primers used in this study are listed in Table 3-2. DNA ligations and restriction digestions were performed as recommended by the manufacturer (New England Biolabs). PCR amplification was carried out on an Eppendorf Mastercycler ep gradient S thermal cycler using either Taq or HotStarTaq DNA polymerase (Qiagen). PCR conditions were optimized for each primer pair and products were purified using the QIAquick PCR purification Kit (Qiagen). DNA sequencing was carried out at the Manitoba Institute of Cell Biology DNA Sequencing Facility. *E. coli* cells were transformed using the Z-competent buffer kit protocol (Zymo Research). Conjugation into *B. cenocepacia* K56-2 was accomplished by triparental mating (Craig *et al*., 1989) with *E. coli* DH5α carrying the helper plasmid pRK2013 (Figurski & Helinski., 1979).

Table 3-2 Primers

Primer number	Sequence 5'-3'	Restriction Sites
171	catctttccctggttgccaatgg	
288	tataacatatgacgattctggtgattgc	NdeI
377	tcactatctagacggctcgtactcgacgagcttcatc	XbaI
395	atgtcacatatgaaaatcctggtgccagtgaaaag	NdeI
396	atataatctagatgtcgttcgactcgacgaggatc	XbaI
397	atgaattctagatttctttccgctccctcttacag	XbaI
398	aataattctagattctcgtgtgcttaccgctgaag	XbaI
445	gtcatctagagatgaaaatcctggtgccagtg	XbaI
446	atatggtacccgcagcaccttggcttcggtcttc	KpnI
447	gtcatctagagatgacgattctggtgattgcag	XbaI
448	aattggtaccttaccgctgaagcatcacac	KpnI
449	aattggtacccgcacgagttccggcacgag	KpnI
482	cagtcagatatcgtcggtgacgatcatccatc	EcoRV

485	cagtcatctagaacaccggcttgagtttcgtg	XbaI
520	ggaagaccagttctgacgcaacgcatccccgc	
521	gcggggatgcgttgcgtcagaactggtcttcc	
522	aaggaacgcgtgcaaagccg	
523	cgcgagcaagtcctgtgtggtcaa	
524	cagaacatcgtgtgggtcac	
525	atacccatgtacgccggatc	
526	atctacatatgacccccgcaagcctcatc	NdeI
527	tgacatctagaggacaggaacaggaagcggtc	XbaI
552	tgccgtgattcttgaagttg	

- Restriction sites are underlined.

3.2.3 Construction of an unmarked *etfdh2* deletion

An unmarked deletion of *etfdh2* (BCAS0609) was made using the method described by Flannagan *et al.* (Flannagan *et al*., 2008) with the following modifications. Upstream and downstream flanking regions were amplified using primer sets 482/521 and 485/520 respectively (Table 3-2). The flanking regions have an overlapping region allowing for amplification of the fusion of the two regions using primers 482 and 485. The fusion product was ligated into pGPI-SceI after digestion with *XbaI* and *EcoRV* (pRB17, Table 3-1). The resulting suicide plasmid was moved into K56-2 by triparental mating and integration was confirmed by PCR with primer sets 522/523 for the upstream region and primer set 524/525 for the downstream region. After introduction of pDAI-SceI, loss of the flanking regions was confirmed by PCR with primer sets 522/523 and 524/525. The deletion mutant (Detfdh2, Table 3-1) was then cured of pDAI-SceI by repeated subculturing and the mutation was confirmed by sequencing the flanking regions.

3.2.4 Construction of *etfBA* and *etfdh1* conditional mutants

The 5' ends of *etfB* (BCAL2395) and *etfdh1* (BCAL1468) were amplified using the primer pairs 395/396 and 526/527 respectively (Table 3-2). After digestion with both *XbaI* and *NdeI* the 5' fragments were cloned into pSC200 (Table 3-1) immediately downstream from the rhamnose inducible promoter. The resulting plasmids, pRB10 and pRB18 (Table 3-1) were moved into K56-2 and Detfdh2, respectively by triparental mating. Integration was confirmed by PCR-amplification of the plasmid-genome junction using the transposon specific primer 171 and primers 377 and 552 for *PrhaB*::*etfBA* and *PrhaB*::*etfdh1*, respectively.

3.2.5 Construction of complementing plasmids

The putative *etfBA* operon (BCAL2395, BCAL2394) was amplified using primers 395 and 398 (Table 3-2), *etfB* with primers 395 and 397 and *etfA* with primers 288 and 398. After digestion with both *XbaI* and *NdeI* the fragments were cloned into pAP20 (Table 3-1) immediately downstream from the dhfr promoter. The resulting plasmids (Table 3-1) were moved into CetfBA by triparental mating.

3.2.6 Microscopy analysis

10 µl of cultures were spotted onto glass microscope slides covered with a film of 1% agarose. Images were taken with an AxioCamMR attached to an Axio Imager Z1 (Carl Zeise Canada) at 1000x magnification using bright field and DIC filters. Bright field images were exported as TIFF files and image analysis was done in ImageJ (http://imagej.nih.gov/ij/). Images were thresholded to distinguish cells and backgrounds. Large clumps of cells were excluded from further analysis and attached cells that had completed septation were digitally separated. The integral "Analyze Particles" function of ImageJ was used to determine the major and minor axes of bounding ellipses for each cell, which were in turn used to calculate area, and roundness (major/minor axis). Cell area and roundness were compared using the Mann-Whitney sum-rank test (Lehmann, 1975) assuming that the test-statistic U is normally distributed given the large sample size.

3.2.7 Bacterial two-hybrid assay

Protein-protein interactions were tested using the BACTH system (Euromedex, Mundolsheim, France) according to the manufactures directions. *etfB* and *etfA* ORFs were amplified using primer pairs 445/446 and 447/448 respectively (Table 3-2). After digestion with both *XbaI* and *KpnI* the fragments were cloned into pUT18c and pKT25 to form translational fusions with the T18 and T25 fragments respectively. The plasmids were purified and co-transformed in all combinations into the reporter strain (BTH101, Table 3-1). Successful co-transformation was confirmed by colony PCR for *etfB* and $etfA_t$. For each transformant 10µl of overnight culture was spotted onto LB X-Gal (40 µg ml^{-1}), IPTG (0.5 mM) plates containing Ampicillin and Kanamycin. The plates were incubated for 3-days at 25°C.

3.2.8 Sensitization index experiments

Suspensions of K56-2 and mutant colonies in LB were adjusted to an OD_{600nm} of 0.1 and diluted 10^{-2}. 180 µl of cell suspensions were aliquoted into 96 well plates with 0.12%, 0.10% - 0.08%, 0.06%, 0.05%, 0.04%, 0.03%, 0 % (w/v) final concentrations of rhamnose and 2 fold serial dilutions of the antibiotic to be tested starting with Novobiocin 32 µg ml^{-1}, Chloramphenicol 125 µg ml^{-1}, S-(3,4-Dichlorobenzyl) isothiourea (A22) 125 µg ml^{-1}, H_2O_2 1.1 mM, and Meropenem 8 µg ml^{-1} to a total volume of 200 µl. Plates were then incubated for 16 hours at 37°C with no shaking. The sensitization index for a given rhamnose concentration was calculated as the ratio between the minimal inhibitory concentration (MIC) of K56-2 and the MIC of the mutant.

3.2.9 Biolog assays

Biolog plates and accessories were obtained from Biolog Inc. Hayward, CA. The procedure for growth on PM1 plates was modified as follows. Overnight cultures were grown in LB supplemented with rhamnose and antibiotics as needed. The cultures were washed twice with inoculating fluid for Gram negative/positive (IF-0 GN/GP) and diluted to a final theoretical OD_{600nm} of 0.001. Rhamnose and antibiotics were added as appropriate. Respiration was measured in LB using reduction of the tetrazolium Biolog redox dye A as a proxy. Reduction of the dye was calculated by subtracting the absorbance at 600nm of the cultures with redox dye A from the absorbance of the cultures without the dye.

3.2.10 Bioinformatics

etfA and *etfB* amino acid sequences from *Burkholderia spp.* were downloaded from www.burkholderia.com (Winsor *et al*., 2008) while sequences from other species were retrieved from NCBI. PhyML (Guindon *et al*., 2010) with the WAG substitution model and the approximate likelihood ratio test was used to construct a maximum-likelihood phylogenetic tree. Sequences were aligned using MUSCLE (Edgar, 2004) and regions with more than 8 nonconserved amino acids or present in less than 85% of the sequences were removed using GBLOCKS (Castresana, 2000) with no gaps and a minimum block length of 10 amino acids.

3.3 Results

3.3.1 The *B. cenocepacia etfBA* operon is essential

Previously, we identified putatively essential genes in *B. cenocepacia* K56-2 by screening for rhamnose-dependent growth in transposon mutants after delivery of an outward facing rhamnose-inducible promoter (Bloodworth *et al*., 2013). Mutants showing a rhamnose dependent growth phenotype suggested that the rhamnose-inducible promoter had replaced the native promoter of an essential gene. Mutants of well-known essential genes often showed a reduction in growth but not complete growth arrest in the absence of rhamnose. In contrast, mutant 77-16C10, in which the transposon had inserted between the *etfB* and *etfA* genes of a putative *etfBA* operon, had no detectable growth in the absence of rhamnose (Figure 3-2a). A newly constructed mutant strain with the rhamnose inducible promoter inserted immediately upstream from the *etfBA* operon (CetfBA, Table 3-1) likewise had no measurable growth in the absence of rhamnose. When we complemented CetfBA with the *etfBA* operon constitutively expressed from a plasmid, the rhamnose dependent growth phenotype was abolished. Complementation with *etfA* or *etfB* alone, however, showed different effects. Expressing *etfA* from the plasmid during rhamnose-induced expression of the chromosomal copy of *etfBA* had no effects on the growth phenotype. In contrast, expression of *etfB* in similar conditions resulted in decrease of growth. In the absence of rhamnose, expressing *etfB* but not *etfA* abolished the conditional growth phenotype, although not to the growth levels of CetfBA in rhamnose-inducing conditions. We then hypothesized that overexpression of *etfB* from a plasmid could trigger the formation of EtfB homodimers that were somehow functional. Although most ETFs function as heterodimers, in *Clostridium perfringens* only EtfA is required for the biosynthesis of dipicolinic acid (Orsburn *et*

al., 2010). To explore the possibility of homodimer formation, we used the BACTH (bacterial adenylate cyclase two-hybrid) complementation assay (Battesti & Bouveret., 2012). *etfB* and *etfA* were cloned into vectors containing the T25 and T18 fragments of *Bordetella pertussis* adenylate cyclase respectively. In agreement with the reported heterodimeric nature of ETFs (Chen & Swenson., 1994), co-transformation into the reporter strain demonstrated that EtfA and EtfB interact with each other but there was no evidence that either interacts with itself (Figure 3-3).

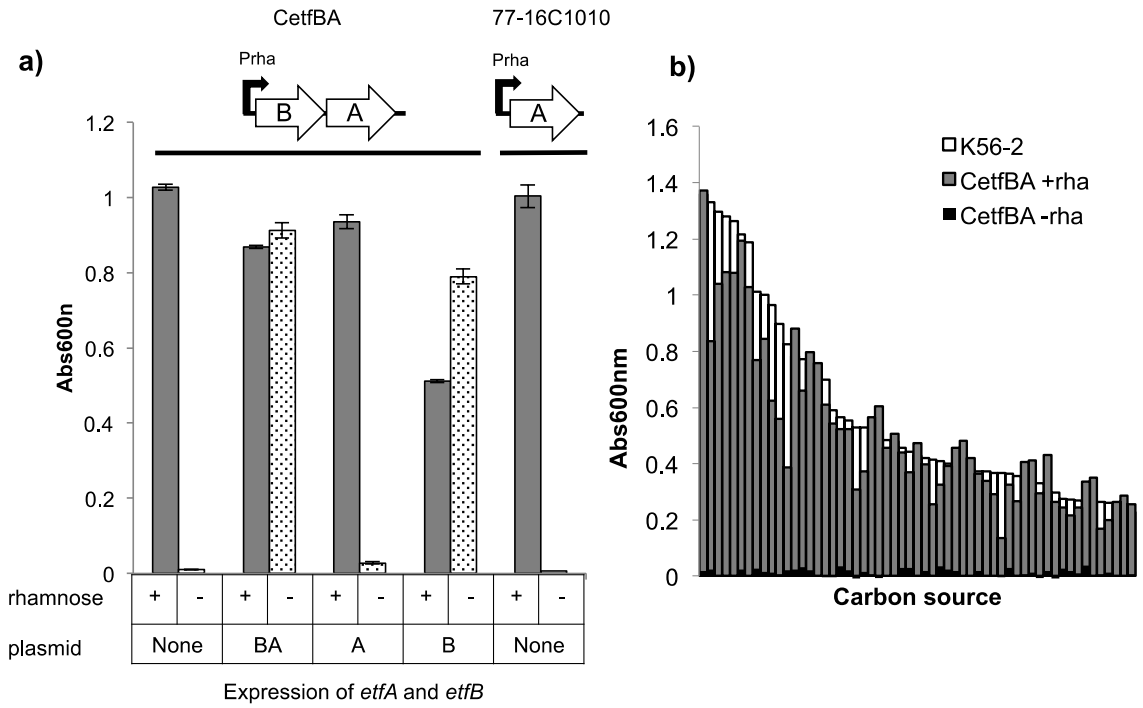


Figure 3-2 The *etfBA* operon is essential for growth on both complex media and single carbon sources.

a) Cultures of the rhamnose inducible expression mutant of the *etfBA* operon (CetfBA) and *etfA* (77-16C10) were grown on LB with and without chromosomal expression of the genes. CetfBA was complemented with constitutive expression of *etfBA, etfA* and *etfB* from a plasmid. Reads were taken at 24h and error bars represent 1 standard deviation of three independent experiments. b) K56-2 and the CetfBA conditional expression mutant were grown for 72h in

Biolog PM1 plates, which provide 95 different sole carbon sources. The bars are representative of the absorbance of two trials.

pKT25-zip	pKNT25	etfA	etfA	etfB
pUT18c-zip	pUT18	etfB	etfA	etfB

Figure 3-3 EtfA and EtfB interact with each other but there is no evidence that either interacts with themselves.

BACTH two-hybrid co-transformants grown on LB supplemented with X-Gal and IPTG for 3 days at 25°C

Table 3-3 Abs600nm after 24h growth on Biolog PM1 sole carbon sources

	K56-2				CetfBA	
	-rha	-rha	+rha	+rha	-rha	-rha
Carbon source	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
D-Sorbitol	1.36	0.90	1.37	0.91	0.02	0.03
m-Inositol	1.33	1.21	0.84	1.31	0.02	0.02
D-Trehalose	1.30	1.24	1.04	0.79	0.00	0.03
Adonitol	1.28	1.18	1.08	1.06	0.00	-0.01
L-Fucose	1.26	1.16	1.08	0.75	0.00	-0.01
D-Galactose	1.22	1.11	1.19	1.02	0.02	0.00
Sucrose	1.19	1.40	1.03	1.06	0.00	0.01
D-Mannitol	1.01	1.01	0.77	0.85	0.03	0.00
alpha-D-Glucose	1.00	1.39	0.85	0.84	0.01	-0.01
Glycerol	0.96	0.75	0.63	0.86	0.01	0.02
D-Fructose	0.90	1.21	0.56	1.20	0.01	-0.01
L-Proline	0.82	0.55	0.39	0.42	0.02	0.00
Dulcitol	0.78	1.08	0.88	0.83	0.02	0.01
Phenylethyl-amine	0.77	0.64	0.66	0.68	0.03	0.01
m-Hydroxy Phenyl Acetic Acid	0.76	0.58	0.80	0.65	0.02	0.01
p-Hydroxy Phenyl Acetic Acid	0.71	0.69	0.76	0.79	0.00	0.00
D-Gluconic Acid	0.70	0.42	0.61	0.51	0.00	-0.02
L-Serine	0.59	0.70	0.54	0.72	0.00	0.03
D-Mannose	0.57	0.56	0.52	0.65	0.04	0.01
D-Glucosaminic Acid	0.56	0.69	0.52	0.58	0.02	0.02
D-Glucose-6-phosphate	0.53	0.49	0.31	0.40	-0.02	0.00
L-Glutamine	0.53	0.44	0.37	0.33	0.01	0.03
D-Galactonic Acid-gamma-Lactone	0.52	0.56	0.56	0.57	0.01	0.02
L-Threonine	0.52	0.66	0.60	0.52	-0.01	0.02
L-Alanine	0.48	0.50	0.46	0.57	0.00	0.01
D-Alanine	0.48	0.55	0.51	0.71	0.00	0.01
D,L-Malic Acid	0.46	0.45	0.44	0.42	0.03	0.01
D-Galacturonic Acid	0.44	0.43	0.37	0.34	0.03	0.04
L-Aspargine	0.43	0.44	0.47	0.34	0.00	0.02
Tween 40	0.42	0.51	0.40	0.44	0.02	0.00

Succinic Acid	0.41	0.37	0.26	0.32	0.00	0.00
Mono Methyl Succinate	0.41	0.36	0.32	0.38	0.03	0.00
Mucic Acid	0.40	0.34	0.39	0.41	0.02	0.02
D-Saccharic Acid	0.40	0.55	0.46	0.49	0.02	0.01
Tricarballylic Acid	0.38	0.40	0.48	0.35	0.00	0.00
2-Aminoethanol	0.38	0.45	0.42	0.39	0.00	0.02
Citric Acid	0.37	0.35	0.36	0.30	0.02	0.02
L-Aspartic Acid	0.37	0.32	0.34	0.38	0.00	-0.03
alpha-keto-Glutaric Acid	0.37	0.33	0.29	0.24	0.00	0.01
L-Lactic Acid	0.37	0.32	0.14	0.34	0.00	0.02
L-Glutamic Acid	0.36	0.31	0.33	0.35	0.03	0.02
D-Malic Acid	0.36	0.43	0.27	0.23	0.00	0.01
alpha-hydroxy Glutaric Acid-gamma-Lactone	0.35	0.33	0.41	0.29	0.01	0.00
Glycyl-L-Proline	0.34	0.42	0.41	0.33	0.01	-0.01
Tyramine	0.33	0.23	0.30	0.28	-0.01	-0.01
Pyruvic Acid	0.30	0.46	0.43	0.25	0.02	0.02
Methyl Pyruvate	0.30	0.33	0.26	0.27	-0.02	0.01
alpha-keto-Butyric Acid	0.28	0.37	0.24	0.34	0.03	0.01
D,L-alpha-Glycerol-Phosphate	0.27	0.22	0.22	0.20	0.02	0.03
Fumaric Acid	0.27	0.30	0.24	0.28	0.01	-0.01
Tween 80	0.27	0.26	0.34	0.14	0.04	0.02
D-Glucuronic Acid	0.27	0.42	0.35	0.44	0.00	0.01
2-Deoxy Adenosine	0.26	0.23	0.17	0.23	0.01	0.00
D-Fructose-6-Phosphate	0.26	0.30	0.20	0.23	0.01	0.00
L-Malic Acid	0.26	0.33	0.26	0.19	0.00	0.02
Propionic Acid	0.23	0.32	0.29	0.33	0.00	-0.01
Bromo Succinic Acid	0.23	0.24	0.25	0.25	0.01	-0.02
Tween 20	0.17	0.09	0.30	0.11	0.00	0.04
Glucuronamide	0.16	0.18	0.12	0.12	0.03	0.01
Glycolic Acid	0.16	0.17	0.26	0.19	0.02	0.00
Acetic Acid	0.16	0.14	0.23	0.17	0.00	0.02
L-Alanyl-Glycine	0.13	0.12	0.14	0.11	0.01	-0.01
Glycyl-L-Aspartic Acid	0.13	0.14	0.11	0.12	0.01	0.02
Formic Acid	0.12	0.09	0.07	0.06	0.02	0.02
alpha-hydroxy Butyric Acid	0.11	0.10	0.09	0.11	0.01	0.00
L-Galactonic Acid-gamma-Lactone	0.08	0.08	0.06	0.06	0.02	0.01
D-Serine	0.08	0.09	0.09	0.09	0.00	0.00
Thymidine	0.08	0.10	0.10	0.07	0.00	0.00
Adenosine	0.08	0.08	0.05	0.08	0.00	0.00
D-Threonine	0.07	0.07	0.05	0.07	0.00	0.02

glycyl-L-Glutamic Acid	0.07	0.07	0.06	0.06	-0.01	-0.01
D-Glucose-1-Phosphate	0.07	0.11	0.08	0.10	0.01	0.01
m-Tartaric Acid	0.07	0.08	0.06	0.06	0.00	0.01
Glyoxylic Acid	0.04	0.06	0.04	0.06	0.01	-0.02
Uridine	0.04	0.04	0.03	0.04	0.04	0.01
D-Psicose	0.04	0.03	0.03	0.02	0.01	0.01
Inosine	0.03	0.03	0.03	0.03	0.02	0.01
Maltose	0.03	0.02	0.03	0.03	0.02	0.01
D-Cellobiose	0.03	0.04	0.04	0.03	0.00	0.03
D-Aspartic Acid	0.03	0.03	0.03	0.02	0.01	0.01
alpha-methyl-D-Galactoside	0.03	0.03	0.03	0.02	-0.01	0.02
D-Melibiose	0.02	0.02	0.02	0.02	0.02	0.01
Lactulose	0.02	0.02	0.03	0.02	0.02	0.02
beta-methyl-D-Glucoside	0.02	0.02	0.02	0.02	0.03	-0.01
alpha-D-Lactose	0.02	0.03	0.03	0.03	0.02	0.00
L-Rhamnose	0.02	0.02	0.02	0.01	0.02	-0.01
1,2-Propanediol	0.01	0.01	0.01	0.02	0.00	0.02
Acetoacetic Acid	0.01	0.02	0.01	0.02	-0.01	0.02
N-acetyl-beta-D-Mannosamine	0.01	0.01	0.01	0.01	0.01	0.02
D-Ribose	0.01	0.01	0.01	0.01	0.00	0.03
D-Xylose	0.01	0.01	0.01	0.01	0.00	-0.01
L-Arabinose	0.01	0.01	0.01	0.01	0.02	0.00
L-Lyxose	0.01	0.01	0.01	0.01	0.03	0.01
Maltotriose	0.01	0.01	0.01	0.01	0.00	0.01
N-Acetyl-D-Glucosamine	0.01	0.01	0.00	0.00	0.01	0.00

3.3.2 The essentiality of ETF is independent of the carbon source utilized

In other organisms, ETFs function as electron shuttles, and are required to utilize specific substrates, like fatty acids in *Paracoccus denitrificans* (Roberts *et al*., 1999) or carnitine in *E. coli* (Walt & Kahn., 2002). If ETF has a comparable role in K56-2, the *etfBA* operon should be essential for the utilization of specific carbon sources and the conditional phenotype should be abolished when CetfBA is grown on an alternative substrate. While this seems unlikely given that the conditional growth phenotype of CetfBA is present in LB, a complex medium, it is not

unprecedented. In *Bacillus subtilis* the glycolytic genes *gapA, pgm,* and *eno* are required for growth on rich media for as yet undetermined reasons but mutants of all three genes are capable of growth on minimal media with malate and glucose as sole carbon sources and the *gapA* mutant can grow on glucose alone (Commichau *et al*., 2013). To investigate growth of CetfBA on different single carbon sources we used Biolog PM1 plates, which contain 95 individual carbon sources, including compounds which might require an ETF for metabolism like the short chain fatty acids propionic and acetic acid or the amino acid threonine (Table 3-3). Cultures of K56-2 and CetfBA grown in LB with rhamnose were washed and resuspended in IF-0 GN/GP media with and without rhamnose. After incubation for 72h both strains showed at least 15% of maximum growth in 57 out of 95 carbon sources when supplemented with rhamnose (Figure 3-2b, Table 3-3). However, the mutant was unable to grow on any of these carbon sources in the absence of rhamnose (Figure 3-2b), suggesting that the essential function of EtfBA is independent of the utilized carbon source.

3.3.3 Depletion of EtfBA renders cells nonviable

The lack of growth of CetfBA in the absence of *etfBA* expression raised the question of whether growth was merely arrested but could be restored if *etfBA* expression resumed or whether the cells had permanently lost viability. To measure the ability of CetfBA to resume growth, an overnight culture was washed and inoculated into LB media without rhamnose. Aliquotes of this culture were then taken at various time-intervals and transferred onto LB with rhamnose to determine culturability (Figure 3-4a) on both liquid media using the most probable number (MPN) method and on solid media by counting colony forming units (c.f.u). Both methodologies

showed that after 30 hours without *etfBA* expression, the number of viable cells fell below the limits of detection (7 MPN ml^{-1} and 1 c.f.u ml^{-1}) indicating that EtfBA depletion had irreversible effects. This loss of culturability was not common among conditional expression mutants of essential genes (Table 3-3) (Bloodworth *et al*., 2013). In addition to measuring culturability, we also used the BacLight Live/Dead kit to measure membrane integrity based on the ability of two dyes to enter the cell. Syto 9 stains all cells but propidium iodide is excluded from cells with intact membranes. These experiments showed that although CetfBA lost viability during incubation over 24 hours without rhamnose, cells maintained intact membranes (Figure 3-5).

Cells depleted of EtfBA were unable to resume growth despite their intact membranes. To examine whether respiration had also stopped, cultures were inoculated into LB containing Biolog redox dye A, which is reduced by cellular respiration forming a purple color (Tracy *et al*., 2002). After 24 hours, cultures were moved into fresh media supplemented with rhamnose to determine whether respiration resumed. An ATP synthase mutant, used as a negative control for respiration, showed no dye reduction, indicating that interruption of respiration was irreversible (Figure 3-4b). Likewise, a *gyrB* mutant continued to respire regardless of the absence of rhamnose, consistent with GyrB playing no role in respiration. Cells without *etfBA* expression showed a slight reduction of the dye by 24 hours but respiration did not resume after being switched to inducing conditions, indicating that EtfBA depletion leads to an irreversible loss of respiration.

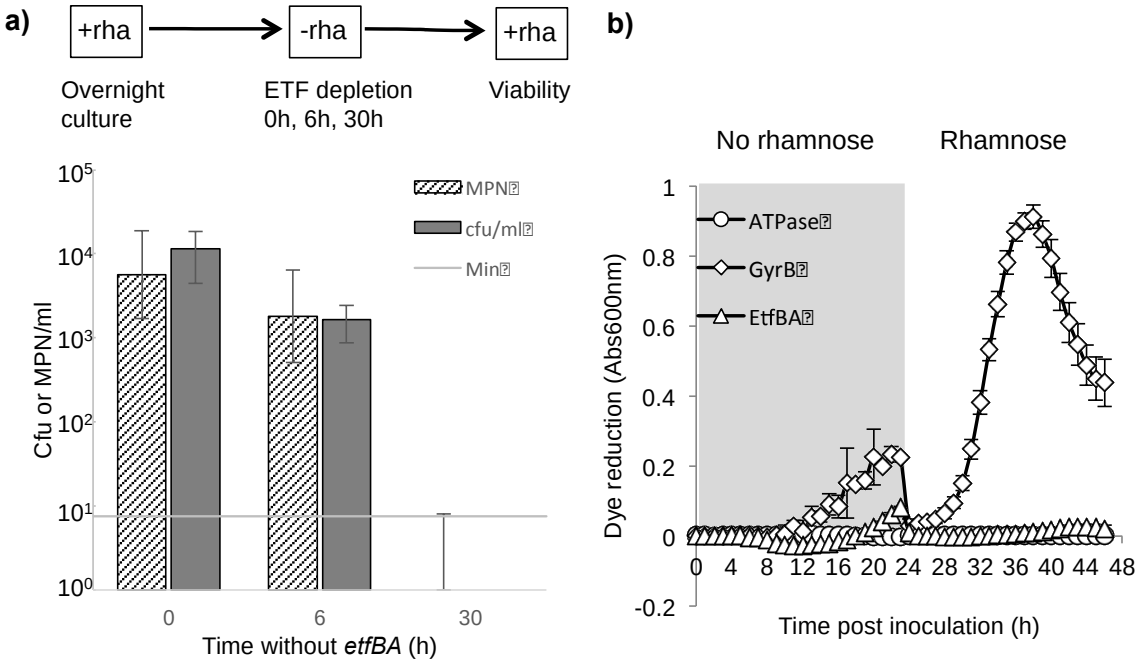


Figure 3-4 *etfBA* expression is required for viability.

a) An overnight culture of CetfBA was adjusted to an OD_{600nm} of 0.01. Viability at 0, 6 and 30 hours post inoculation was measured both on LB rhamnose plates by c.f.u. ml^{-1} and in liquid culture using MPN ml^{-1}. Error bars represent a 95% confidence interval. The horizontal grey line shows the minimum level of detection by MPN (7 MPN ml^{-1}), the c.f.u. method could detect 1 c.f.u. ml^{-1}. b) Reduction of Biolog redox dye A was used as a proxy to measure respiration of the conditional expression mutants of *gyrB, etfBA* and the ATP synthase after inoculation into LB without rhamnose.. At 24 hours post inoculation, the cultures were sub-cultured into new media with rhamnose . Reads were taken every 60 minutes of 4 replicates; error bars are 1 standard deviation.

Table 3-4 Change in viability of conditional growth mutants over 24 hours

Name	MPN ml^{-1} at 0h	MPN ml^{-1} at 24h	Log change in viable cells	Downstream genes
CetfBA	$5.6x10^3$	0	< -3	BCAL235, *etfB*
				BCAL234, *etfA*
CgyrB	$5.2x10^6$	$1.65x10^9$	2.5	BCAL0421, *gyrB*
16-2C5	$5.4x10^7$	$1.9x10^8$	0.5	BCAL2676, *pepA*
				BCAL2675, DNA polymerase III subunit
16-11C6	$8.7x10^7$	$1x10^9$	1	BCAL2677, putative permease
				BCAL2678, putative permease
				BCAL2679, putative cobaltochelatase
77-16C10	$1.5x10^9$	$2.2x10^5$	-3.8	BCAL2934, *etfA*
				BCAL3053, *ribB*
78-16D7	$6.3x10^7$	$9.2x10^7$	0.16	BCAL3054, *ribH*
				BCAL3055, *nusB*
82-10A9	$4.8x10^9$	$4.8x10^8$	-1	n.d.
89-10K20	$1.7x10^6$	$3.5x10^6$	0.31	BCAL3420, *accB*
				BCAL3421, *accC*
				BCAL3422, *prmA*
				BCAL3423, hypothetical protein

BCAL3424, *tpx*

BCAL3425, putative sugar kinase

BCAL3426, putative lipoprotein

n.d. not determined

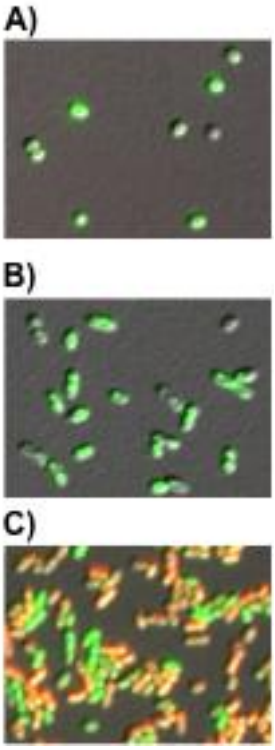


Figure 3-5 Cells depleted of EtfBA maintain intact membranes.

A BacLight Live/Dead kit was used to stain CetfBA cells grown for 24 hours A) without rhamnose B) with rhamnose and C) a 50:50 mixture of live K56-2 and cells killed with isoproponal which degrades the membrane.

3.3.4 Lack of ETF causes a rod-to-sphere change in cell morphology that is independent of MreB

When measuring cell viability, we noticed that even though all cultures were adjusted to the same initial OD_{600nm}, cultures of CetfBA contained approximately 3 log fewer viable cells than equivalent cultures of other mutants (Table 3-4). This lack of correspondence between $OD600_{nm}$ and c.f.u ml^{-1} suggested a change in CetfBA cell morphology. However, DIC microscopy of cells grown in the presence of 0.2% rhamnose showed normal morphology (Figure 3-6a, lower left). Instead, after 24h without *etfBA* expression, cells had transitioned from the normal short rod morphology of *B. cenocepacia* cells to small spheres (Figure 3-6a, lower right). In rod-shaped bacteria, cell diameter is controlled by the MreB complex, the loss of which is known to trigger a change from rods to spheres (Young, 2010). Treating K56-2 cells with 16μg ml^{-1} of A22, a known MreB complex inhibitor (Bean *et al*., 2009), also caused a transition from rods to spheres (Figure 3-6a, upper right). However, A22 treated cells appeared to be larger than those depleted of EtfAB. This observation was confirmed quantitatively by fitting a bounding oval to each cell in the microscopy images and calculating cell area and roundness (length/width). Both treatment of the K56-2 with A22 and depletion of EtfBA caused a statistically significant shift towards rounder cells (Figure 3-6a) but while EtfBA depleted cells were smaller, treatment with A22 actually increased cell size (Figure 3-6b).

To confirm that the effect of depleting ETF was different from MreB depletion, we used A22 in a sensitization index assay, where controlled under-expression of an essential gene results in enhanced sensitivity to growth inhibitors that target the encoded gene product (Cardona *et al*., 2014). This specific effect upon exposure of small molecules of known function can be used to profile an essential gene of unclear function. We exposed CetfBA and 77-16C10 to the action of A22 and other growth inhibitors, namely novobiocin (DNA replication, targets GyrB) chloramphenicol (protein synthesis, targets ribosome), meropenem (cell wall synthesis, targets penicillin binding proteins) and hydrogen peroxide (an oxidative stress inducer) at a range of rhamnose concentrations known to include those necessary to produce 30 to 60% of wild type growth (data not shown). We previously showed that this range of growth is necessary to sensitize the mutants to specific growth inhibitors (Bloodworth *et al*., 2013). When a control mutant under-expressing *gyrB* was exposed to its binding antibiotic novobiocin, a strong sensitization effect could be observed (Figure 3-7), resulting in a sensitization index of 8 (8X lower MIC than that of the wild type). An intermediate sensitization index of 4 could be observed for the *gyrB* mutant in response to hydrogen peroxide. This was reminiscent of the effect of fluoroquinolone-related antibiotics that target GyrB and generate reactive oxygen species (Goswami *et al*., 2006). Expression mutants of *lepA*, a gene coding for a translation elongation factor (Liu *et al*., 2011) and *lepB*, a gene which codes for a signal peptidase (Dalbey & Wickner., 1985), also showed intermediate levels of sensitivity to chloramphenicol which inhibits translation but does not directly interact with the products of either gene. As in previous studies (Bharat & Brown., 2014, Campbell *et al*., 2005), sensitization indexes lower than 4 were not considered indicative of sensitization. No hypersensitivity to A22 was observed with either

the *etfA* (2.7x) or *etfBA* (2x) mutants suggesting that the effect of ETF on cell shape is independent of MreB. Interestingly, while the *etfBA* mutant only showed intermediate levels of sensitivity to hydrogen peroxide the *etfA* mutant had intermediate sensitivity to both hydrogen peroxide and meropenem. This is not a general response to antibiotics as neither mutant was hypersensitive to either novobiocin or chloramphenicol. Taken together, these data suggest that in *B. cenocepacia* ETF triggers a rod-to-sphere change in cell morphology, which is independent of MreB.

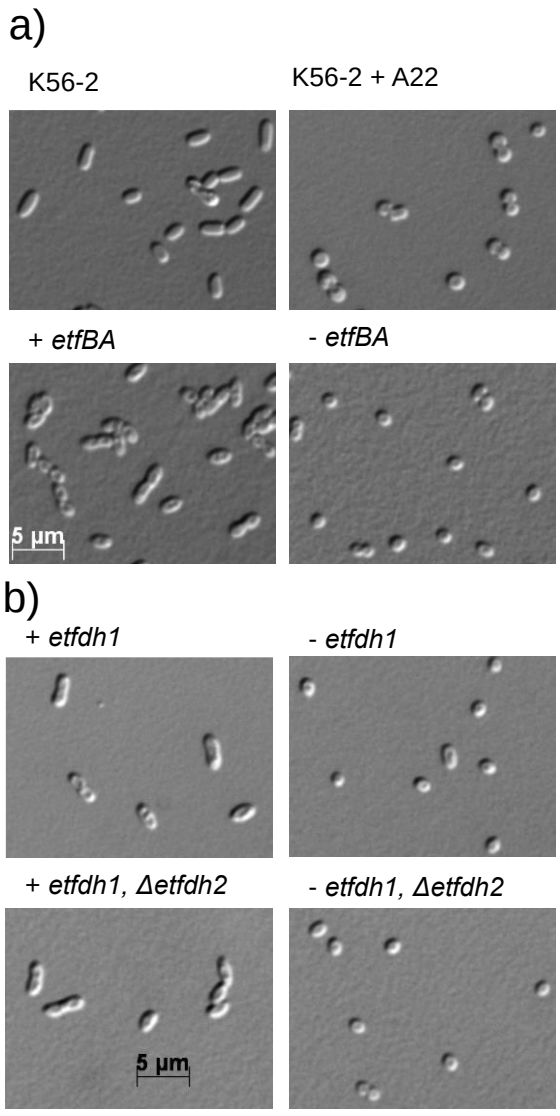


Figure 3-6 Lack of *etfBA* expression causes a rod-to-sphere change in cell shape.

DIC microscopy images of K56-2 and *etfBA* conditional mutants were taken after 24h of incubation and *etfdh1* conditional mutants after 48h of growth. K56-2 cells were grown with and without A22 while the conditional expression mutants were grown in rhamnose to induce

expression of the controlled genes (left column) and without rhamnose (no gene expression, right column)

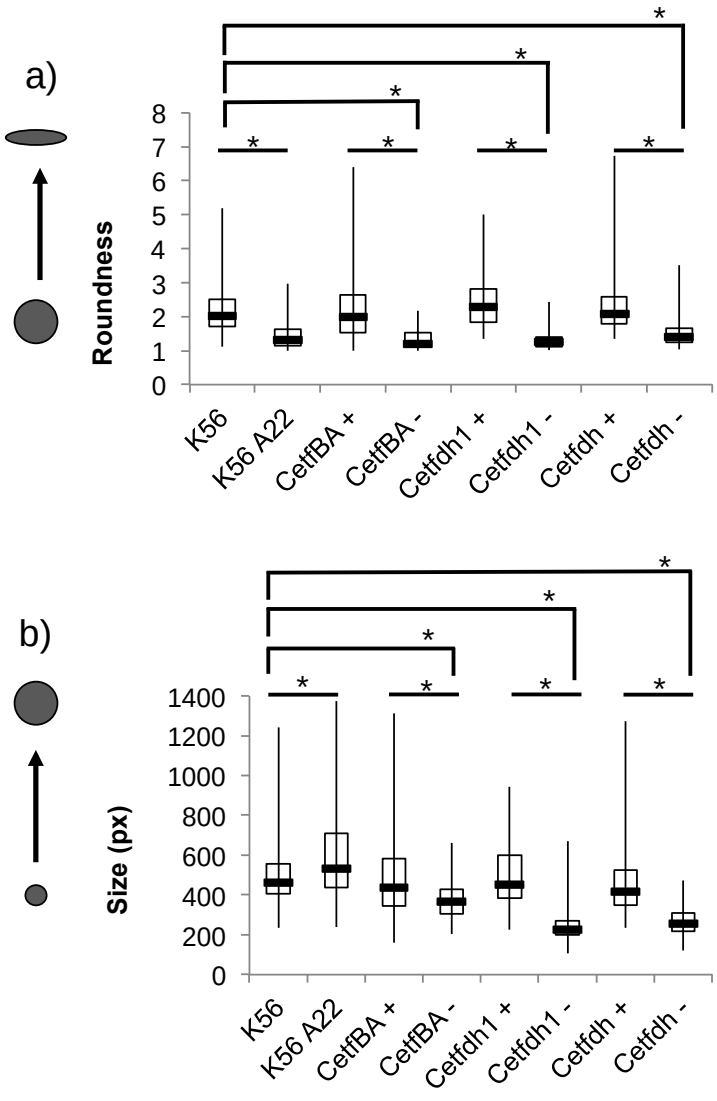


Figure 3-7 Lack of *etfBA* or *etfdh1* expression is not equivalent to loss of MreB function.

a) Cell roundness (length/width) and b) cell size (pixels) of K56-2 with and without A22 and CetfBA, Cetfdh1 and Cetfdh with (+) and without (-) rhamnose. For each strain and condition

the top of the vertical bar is the maximum value, the bottom the minimum value. The box bound the first through third quartiles and the black bar is the median. Statistically significant differences (Mann-Whitney U Test, two tailed $P < 0.01$) are marked by black over bars with an asterisk.

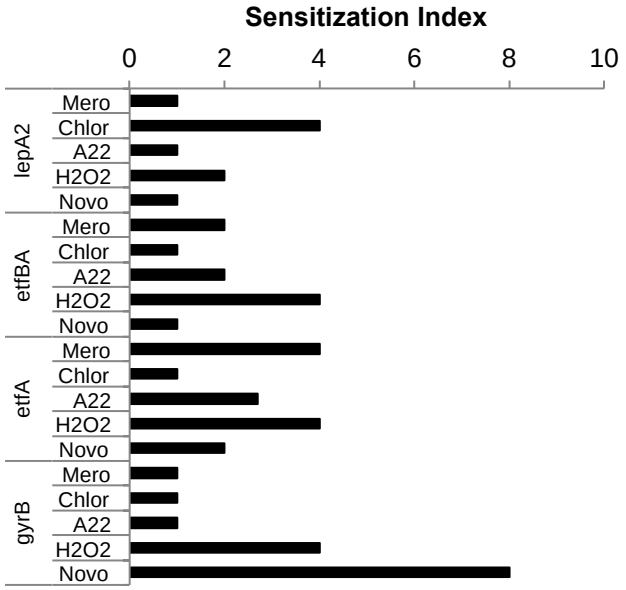


Figure 3-8 The effect of low levels of ETF on sensitivity to growth inhibitors.

Conditional growth mutants for *lepA2, gyrB, etfBA* and *etfA* were grown in a range of rhamnose concentrations while exposed to serial dilutions of known growth inhibitors (Mero: meropenem, Chlor: chloramphenicol, Novo: novobiocin, A22: S-(3,4-Dichlorobenzyl) isothiourea, H2O2: hydrogen peroxide). For each mutant/inhibitor the sensitization index was calculated as the highest fold change in MIC compared to K56-2 and are representative of two independent experiments.

3.3.5 The putative partner of ETF, ETF dehydrogenase also induces a rod-to-sphere change in morphology.

While ETFs are known to accept electrons from a variety of partners, in all known cases they donate electrons to corresponding ETF dehydrogenases (Ghisla & Thorpe., 2004, Roberts *et al*., 1999, Scott & Ludwig., 2004). If the essential function of EtfBA involves passing electrons through to ETF dehydrogenase, then depletion of ETF dehydrogenase should have a similar effect as depleting EtfBA. *B. cenocepacia* contains two putative ETF dehydrogenases BCAL1468 (*etfdh1*) and BCAS0609 (*etfdh2*) with 99% amino acid identity. Since the third chromosome of *B. cenocepacia* is dispensable (Agnoli *et al*., 2012), none of the genes found on the third chromosome including *etfdh2* are likely to be essential. We therefore created a site directed rhamnose inducible expression mutant of *etfdh1* in the K56-2 (Cetfdh1, Table 1) and Δ*etfdh2* (Cetfdh, Table 1) backgrounds. In the presence of *etfdh2,* lack of *etfdh1* expression reduced growth to 29% of growth with rhamnose (Figure 3-9). In the Δ*etfdh2* mutant, growth was unaffected as long as *etfdh1* was expressed but fell to 13% in the absence of rhamnose suggesting that *etfdh2* can partially complement *etfdh1*.

DIC microscopy of the ETF dehydrogenase mutants grown without rhamnose showed that after 48 hours cells had undergone a similar change in morphology to depletion of EtfBA (Figure 3-6b). Quantification of cell size and shape confirmed that depletion of ETF dehydrogenase resulted in significantly rounder (Figure 3-7a) and smaller cells (Figure 3-7b). To determine

whether this conversion to small round cells is a general result of metabolic disruption, rather than having any relation to ETF function, the morphology of conditional expression mutants of cytochrome c1 and the F0F1 ATPase were examined. In both mutants cell roundness was unaffected by incubation without rhamnose while cell size actually increased slightly in the ATPase mutant (Figure 3-10 and 3-11). These results suggest the morphology seen when EtfAB or ETF dehydrogenase are depleted is not due to a general defect of respiration.

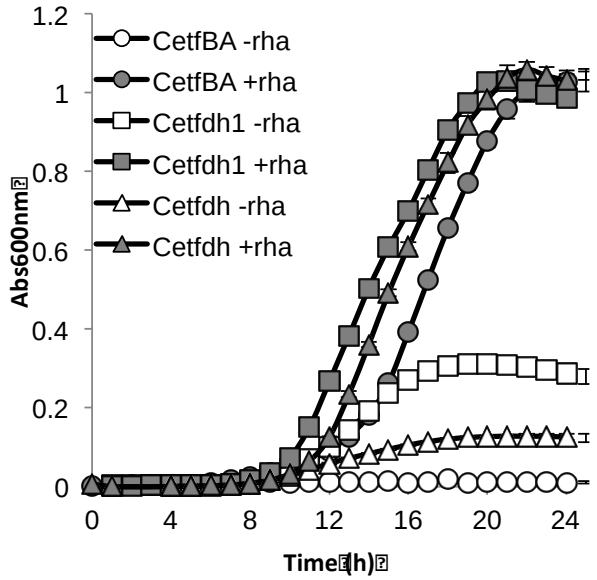


Figure 3-9 Lack of putative ETF dehydrogenases causes a growth defect.

Cultures of rhamnose inducible conditional expression mutants of *etfBA* (CetfBA) and *etfdh1* in the K56-2 (Cetfdh1) and Δ*etfdh2* (Cetfdh) backgrounds were grown for 24h with and without rhamnose. Readings are of 4 replicates; error bars show one standard deviation.

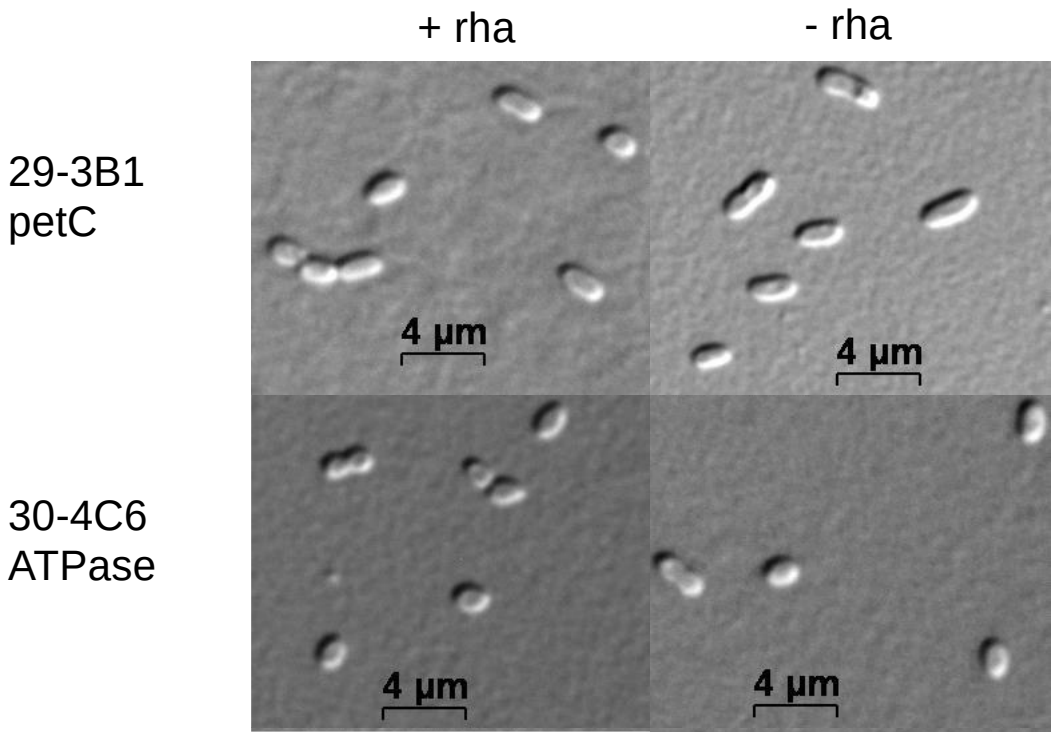


Figure 3-10 There is little change in cell morphology after depleting other electron transport proteins.

Conditional growth mutant of cytochrome bc (29-3B1) and ATPAse (30-4C6), were grown for 24h in LB without rhamnose, then diluted 10^{-3} into fresh LB or LB rhamnose and grown for a further 24h before DIC microscopy.

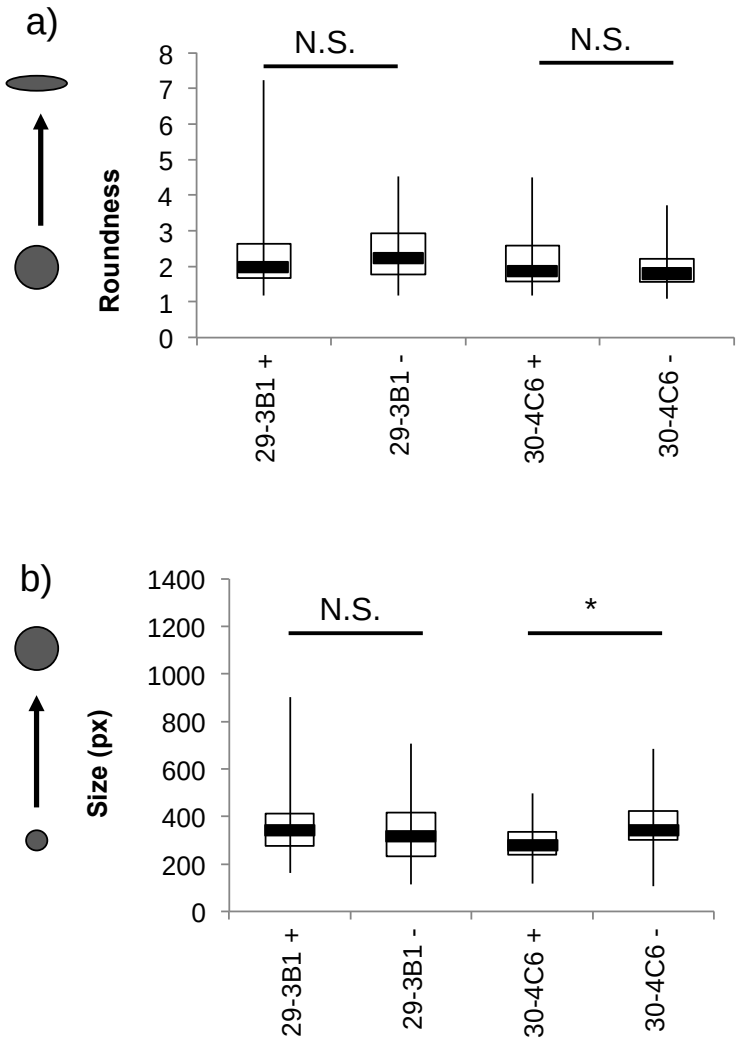


Figure 3-11 Depletion of cytochrome bc (29-3B1) or ATPAse (30-4C6) does not lead to smaller rounder cells.

Cell roundness (length/width) and b) cell size (pixels) were measured for 29-3B1 and 30-4C6 after 48h with (+) or without (-) rhamnose. For each strain and condition the top of the vertical bar is the maximum value, the bottom the minimum value. The box bound the first through third

quartiles and the black bar is the median. Statistically significant differences (Mann-Whitney U Test, two tailed $P < 0.01$) between a single strain in two different conditions are marked by black overbars with an asterisk, statistically non significant differences are marked with N.S.

3.4 Discussion

Bacteria belonging to the *B. cepacia* complex have high levels of intrinsic resistance to current antibiotics (Leitao *et al*., 2008, Maravic *et al*., 2012) and disinfectants (Rushton *et al*., 2013) , which warrants the investigation of novel essential genes as possible drug targets. ETFs have been identified as putatively essential in several high-throughput screens and so represent an intriguing case of an essential gene with no known essential functions. We constructed a rhamnose inducible *etfBA* expression mutant in *B. cenocepacia* K56-2 (CetfBA) and confirmed (through complementation assays) that expression of the operon is required for growth on both rich media and a variety of single carbon sources. This indicates that while traditionally ETF functions in the metabolism of specific carbon sources, the essential function in *B. cenocepacia* is carbon source independent. We did find that high levels of *etfB* expression alone could partially complement lack of *etfBA* expression. Since *etfA* is essential in the original transposon mutant (77-16C10) despite native expression of *etfB* from the original promoter (Bloodworth *et al*., 2013), the ability of *etfB* to restore growth of *etfBA*-depleted cells appears to be dose dependent. However, this effect does not appear to be related to homodimeratization as there is no evidence that EtfB interacts with itself (Figure 3-3). One possible explanation is that other EtfA subunits expressed by the paralogous genes BCAM2321 or BCAM1606 (Figure 3-12a) are

able to bind EtfB with low affinity, and this effect is evidenced when EtfB is overexpressed. In any case, it seems that optimal function of ETF is observed when *etfA* and *etfB* are expressed at the same level, as expression of *etfB* from both the chromosome and plasmid reduced growth to 65% of the wild type (Figure 3-2a).

In contrast to most of the conditional expression mutants, CetfBA suffered from a severe loss of viability, with the number of viable cells falling below the limits of detection after 24 hours (Figure 3-4a). The inability of CetfBA to resume growth after depletion of ETF is associated with an irreversible loss of reducing power (Figure 3-4b). This is consistent with a central role for EtfBA in respiration but it is also possible that EtfBA depletion instead causes severe cellular damage, which in turn leads to collapse of respiration. In all known cases ETFs function by transferring electrons from primary dehydrogenases to a corresponding ETF dehydrogenase (Ghisla & Thorpe., 2004, Roberts *et al*., 1999, Scott & Ludwig., 2004). In K56-2 lack of *etfdh1* expression leads to a reduction in cell growth (Figure 3-9), rather than the complete abrogation of growth seen in CetfBA. Since both *etfBA* and *etfdh1* expression in the mutants are driven by the same promoter the difference in the mutant phenotypes may be due to the length of time required to deplete protein concentration, suggesting that either less Etfdh1 is required or that being a membrane associated protein it is more stable (Nagata *et al*., 1998, St John *et al*., 1979).

In addition to inhibiting cell growth, depletion of either EtfBA or Etfdh results in a change in cell shape from rods to small spheres (Figure 3-6, 3-7). The uniform spherical cells could be due to

arrest of cell growth immediately after cell division. Coupled with the simultaneous arrest of respiration this suggests a metabolic connection. Growth in nutrient poor media (Donachie & Begg., 1989, Sargent, 1975) and long term starvation (Givskov *et al*., 1994) have been shown to cause a shift to shorter cells. Recently specific metabolic sensors have been identified in both *Bacillus subtilis* (Monahan *et al*., 2014, Weart *et al*., 2007) and *E. coli* (Hill *et al*., 2013) which couple concentration of UDP-glucose to FtsZ polymerization and hence cell size at division. We propose that in *Burkholderia* spp. the lack of EtfBA or Etfdh1 may induce a signal for low nutrient availability triggering early cell division. The lack of a shift to small round cells in electron transport chain mutants (Fig S4 and S5) is consistent with the signal being a yet unknown metabolic intermediate rather than energy availability in general.

The confirmation of ETF essentiality in *B. cenocepacia* and the strong phenotype suggest that it may warrant study in other important human pathogens, especially *Mycobacterium tuberculosis* and *Burkholderia pseudomallei.* An essential gene screen in *M. tuberculosis* (Griffin *et al*., 2011) identified its single ETF operon *fixBA* as putatively essential. In contrast all sequenced *Burkholderia* species have multiple group I ETF operons. Among the *Burkholderias* high-throughput essential gene studies have only been carried out in *B. pseudomallei* (Moule *et al*., 2014) a tier 1 select agent and the cause of melioidosis and *Burkholderia thailandensis* (Baugh *et al*., 2013), a closely related but far less pathogenic species. In both screens all the annotated ETF operons contained at least one putatively essential gene and in both species one operon clusters

closely with the *etfBA* operon in *B. cenocepacia* described here (Figure 3-12), suggesting a common essential function among *Burkholderia* ETFs.

Regardless of the precise nature of ETF essentiality in *B. cenocepacia,* the loss of viability in cells without *etfBA* expression suggests that ETF would be a strong target for killing *B. cenocepacia*, and potentially other *Burkholderias.* However, chemical inhibition of bacterial ETFs may be challenging. Inhibition of the human ETF causes metabolic disorders, as fatty-acid metabolism is important for energy hungry tissues (Frerman & Goodman., 2001), this makes selective inhibition of the bacterial ETF essential. In addition, ETFs interact with a number of different primary dehydrogenases via a dynamic interface (Frerman & Goodman., 2001) sampling a wide variety of structures before inducing a fit with its partner. This structural flexibility may reduce the possibility of an inhibitor binding to the bacterial but not human ETF. An attractive alternative is the development of antisense RNA silencing of ETF. The *etfBA* conditional expression mutant has a strong lethal phenotype suggesting that post-transcription inhibition would likewise be bactericidal.

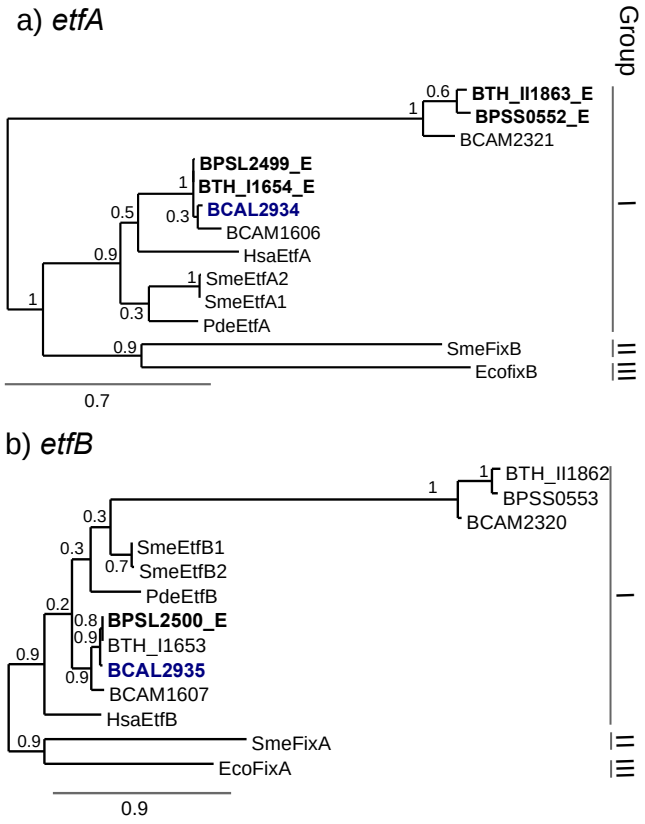


Figure 3-12 Phylogentic trees of EtfA (a) and EtfB (b) proteins in *Burkholderia* spp. containing putatively essential ETF genes.

Trees were constructed based on amino acid similarity; numbers on branches are the support values for each branch; branch length is proportional to the number of substitutions per site. ORFs annotated as being ETFs were included from *Burkholderia thailandensis* (Bth), *B. pseudomallei* (Bps) and *B. cenocepacia* (Bce). Known ETFs from *E. coli* K-12 substr. 3110 (Eco), *Paracoccus denitrificans* (Pde) *Sinorhizobium meliloti* 1021 (Sme), *Homo sapiens* (Hsa), were used to establish the division into 3 families (Tsai & Saier., 1995). Genes identified as

being putatively essential by high-density transposon mutagenesis are identified with an _E and are bolded in black. Genes examined in this paper are in bold blue.

3.5 Acknowledgements

This work was supported by awards from Cystic Fibrosis Canada (315046-352600-2000) and the Natural Sciences and Engineering Research Council of Canada (12500918).

4 Draft genome Sequences of the *Burkholderia contaminans* strains LMG 23361 and FFH2055 and *Burkholderia cenocepacia* K56-2

4.1 Introduction

The *Burkholderia cepacia* complex (Bcc) are a group of at least 18 closely related but phenotypically distinct species (Vandamme & Dawyndt., 2011), known primarily for infecting immunocompromised in general and people with cystic fibrosis in particular (CF) (Lipuma, 2010). While all members of the Bcc are capable of colonizing the CF respiratory tract their prevalence varies greatly. In most studies *B. cenocepacia* and more recently *B. multivorans* are by far the most commonly recovered species accounting together for more than 80% of isolates (Bevivino *et al*., 2002, Brisse *et al*., 2004, Cunha *et al*., 2003, Drevinek *et al*., 2003, Norskov-Lauritsen *et al*., 2010, Reik R *et al*., 2005, Speert *et al*., 2002). The two notable exceptions are Argentina (Martina *et al*., 2013) and Spain (Medina-Pascual *et al*., 2012) where *B. contaminans* is the most common Bcc species in CF patients. Before 2008 the distribution of Bcc species in Spanish CF patients was similar to that seen in most other countries with *B. cenocepacia* and *B. multivorans* being the most common Bcc isolates. Since 2008 the incidence of *B. contaminans* has steadily increased and it now accounts for 36.5% of Bcc infections approximately twice that of the next most common species (*B. cenocepacia* 17.7%) (Medina-Pascual *et al*., 2012). While recovery of *B. contaminans* from CF patients is relatively rare outside of Argentina and Spain, *B. contaminans* outbreaks among non CF patients have been reported in several hospitals, caused by contamination pharmaceuticals and industrial products including prepackaged moist wipes

(Martin *et al*., 2011), ventilation equipment (Peterson *et al*., 2013) and compounded medication (Moehring *et al*., 2014).

B. contaminans was first known as the indeterminate Bcc isolate "Taxon K" (Payne *et al*., 2005, Vermis *et al*., 2002). Later, taxon K was classified in two species, *B. lata* and *B. contaminans* (Vanlaere *et al*., 2009). The species name "contaminans" refers to an episode of sample contamination with this bacterium during a study of marine microbial populations (Venter *et al*., 2004). The so called "Burkholderia SAR-1" metagenome, recovered from the Sargasso Sea was further investigated and it was concluded that it represented a sample contaminant as Bcc isolates grow very little in sea water (Mahenthiralingam *et al*., 2006). This metagenome is the only available *B. contaminans* sequence and it is not tied to an isolated strain for experimentation. It is currently unclear whether *B. contaminans* in general has a particular ability to persist in some man-made environments and its prevalence in Spain and Argentina are due to higher levels of exposure due to local contamination or whether the local *B. contaminans* strains in Spain and Argentina have some genetic differences from strains in other countries making them particularly able to colonize the CF airway.

In this study we report the genome sequences of two *B. contaminans* strains LMG23361 and FFH2055. *B. contaminans* LMG23361 is the type strain for the species, and was isolated from the milk of a sheep with mastitis in Spain (Vanlaere *et al*., 2009). *B. contaminans* FFH2055 is an early Bcc isolate from a CF patient, in Buenos Aires, Argentina. In addition *B. cenocepacia* K56-

2, a member of the highly transmissible ET12 linage commonly used in studies of *B. cenocepacia* (Darling *et al*., 1998).

4.2 Experimental Procedures

4.2.1 Genome sequencing and annotation

Strains are listed in table 4-1. Cultures were grown in LB and genomic DNA was isolated using phenol-chloroform as per Sambrook *et al.* (Sambrook & Russell., 2001). Extracted genomes were set to at the Duke University Genome Sequencing & Analysis Core Resource for library prep and sequencing using the PacBio RS II system. The raw reads were then corrected and assembled using HGAP (PacBio SMRT Analysis software version 2.3), followed by correction of the contigs using Quiver (PacBio SMRT Analysis software version 2.3). Assembly contigs were aligned against the closest available reference genomes using local blast and alignments were visualized using an in house python script. Contigs were annotated using RAST (Aziz *et al*., 2008) and the NCBI prokaryotic genome annotation pipeline (Angiuoli *et al*., 2008).

Table 4-1 Bacterial strains

	Features	Source
Strains		
Burkholderia cenocepacia K56-2 (LMG18863)	ET12 linage, CF isolate	(Mahenthiralingam *et al*., 2000)
Burkholderia contaminans LMG23361	*B. contaminans* type strain	John J. LiPuma
Burkholderia contaminans FFH2055	*B. contaminans*, CF isolate	Hospital de Niños Ricardo Gutierrez Buenos Aires

4.2.2 Nucleotide accession numbers:

The *B. contaminans* LMG23361, FFH2055 and *B. cenocepacia* K56-2 draft genomes have been deposited at NCBI under the accession numbers GCA_000987075.1, GCA_000987055.1 and GCA_000981305.1 respectively

4.2.3 Assembly comparison with a Bcc reference genome

An in house python program was written to visualize similar regions between all the contigs in an assembly and all the replicons in the reference genome. The python libraries matplotlib (http://matplotlib.org) and seaborn (http://stanford.edu/~mwaskom/software/seaborn/) were used to output the comparison as a png file. Local blastn was used to identify local homology regions between each contig and all of the replicons in the reference genome.

4.3 Results

Second-generation sequencing platforms like Illumina have revolutionized sequencing of genomes with their very high throughput, low cost per read and high read accuracy but their read lengths are less than 150 bp off each side of a DNA oligomer (Chin *et al*., 2013, Ferrarini *et al*., 2013). *De novo* assemblies cannot correctly place repeats longer than read length. To span the hundreds of repeats longer than 150bp in a typical bacterial chromosome special libraries with larger defined insert sizes must be prepared and sequenced for inclusion in the assembly. The

PacBio RS system captures sequencing from single molecules in real-time resulting in reads that are significantly longer though also substantially more error prone. With PacBio reads having approximately a 14% error rate compared with the 0.1 to 1% error rate seen with other platforms (Chin *et al*., 2013). This high error rate precludes using raw PacBio reads directly for assembly, instead reads must first be corrected and it is these corrected reads that are used for assembly. Since the errors in PacBio reads are random (Ferrarini *et al*., 2013) given enough coverage (at least 60x) they can be corrected by taking the consensus of many overlapping reads. Given the large number of genomic duplications in *Burkholderia* genomes and their high GC content PacBio sequencing promised more complete assemblies without having to prepare multiple independent insert libraries.

Genomic preps of the two *B. contaminans* and one *B. cenocepacia* strains were sent for library prep and sequencing on 3 PacBio RS II SMRT cells. The sequenced genomes contained more than 200x read coverage, assuming a total genome size of approximately 8Mb (see Table 4-2) this allowed *de novo* assembly of the Pacbio reads without the use of additional sequencing. HGAP was set to error correct the longest reads necessary for 20x coverage of an 8Mb genome, or 160Mb with the remaining smaller reads. These error corrected reads were then assembled into contigs and all the reads were then used to error correct the contigs. The assemblies appeared reasonable containing between 8 and 18 contigs (Table 4-2, Figure 4-1), with the genomes consisting of 3 to 4 replicons (the plasmid in *B. cenocepacia* is 92kb and should be present in the genomic preps). The HGAP assembly process can produce spurious small low

coverage contigs. Only five of the LMG23361 contigs and none of the FFH2055 contigs were smaller than 20kb (Figure 4-1). In contrast eight of the contigs in the K56-2 assembly were less than 20kb and the 4 smallest contigs had a low (less than 50x coverage) and decidedly uneven coverage (See Figure 4-3 contig 10 for a representative example) making them unreliable.

Table 4-2 Summary of assemblies

	LMG23361	FFH2055	K56-2
# of reads	237,907	256,171	241,868
# of bp in reads	2.06 G	1.78 G	1.6 G
N50 of reads	15,036	14,060	11,908
# of contigs	18	8	17
bp in contigs	9.8 M	8.2 M	7.8 M
N50 of contigs	1.1 M	1.4 M	1.5 M

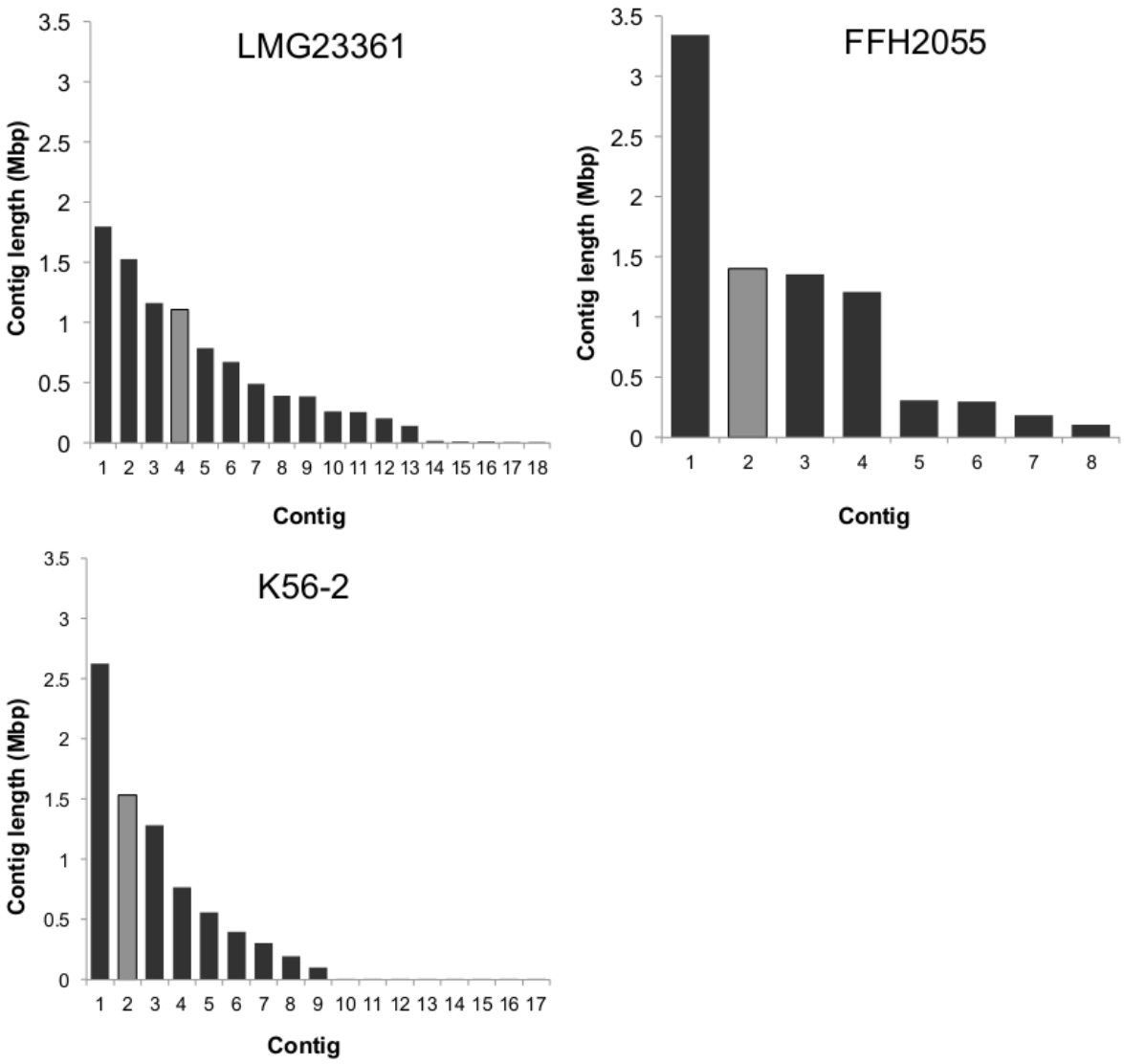


Figure 4-1 Distribution of contig lengths.

Contigs in each assembly are ordered from largest to smallest. The contig with the N50 length is in light grey.

The overall structure and coverage of the assemblies were checked by aligning them against the closest existing closed genomes, *B. lata* for the *B. contaminans* strains and *B. cenocepacia* J2315 for *B. cenocepacia* K56-2. There are a variety of tools for visualizing pairwise comparison of genetic sequences (Carver *et al*., 2005, Krzywinski *et al*., 2009) however they are not designed to display multiple comparisons between multiple replicons. ACT requires a pairwise comparison between the concatenated assembly and concatenated genomes without providing a way to display sequence boundaries (Carver *et al*., 2005). Circos allows chromosomes and contigs to be displayed as separate arc segments around a circle (Krzywinski *et al*., 2009). If the segments are drawn clockwise around the circle contigs and genomes on opposite sides of the circle are flipped relative to each other (eg. At the top of a circle clockwise points right but at the bottom of a circle it points left), making the visualization of inversions difficult. So an in house script was developed (Figure 4-2). Contigs are displayed on the top of the graph as rectangles in arbitrary colors while replicons are displayed as black rectangles on the bottom of the graph. Regions of similarity larger then 1kb are connected by a semi-transparent color of the corresponding contig. The x-axis is scaled to be the greater of the total assembly length or total genome length plus gaps between the replicons. All the assemblies appear to have relatively complete coverage of the reference genomes with chromosome 3 resolved to a single contig.

With the fewest contigs the FFH2055 assembly is also the best match to its reference genome having four contigs covering chromosome 2 and three contigs covering chromosome 1. The green contig that matches both the end and the beginning of chromosome 2 is an artifact of

displaying a linear instead of circular comparison and merely indicates that the contig includes the origin of replication. There does appear to be an inversion on chromosome 1 of a little over 1 Mbp in size bounded on both sides by rRNA operons.

The comparison of the LMG23361 assembly with the *B. lata* genome reveals two areas for concern. The contig shown as an orange box has homology regions split between both chromosome 1 and 2 is indicative of either an inter-chromosomal translocation or missassembly. Like the potential inversion in the FFH2055 assembly the homology regions for chromosome 1 is 200kb located between 2 rRNA operons. As all the rRNAs are all identical they are both hotspots for genomic rearrangements and potential sites for missassembly. In addition there are six contigs that have less than 2% sequence homology with *B. lata*. Four of these contigs are less than 18kb and appear to consist of a phage integrase operon including a Tn3 family transposase. A similar sequence is found in chromosome 3 of *B. vietnamiensis* G4. These contigs overlap with each other in complex ways and probably represent multiple inserts into *B. contaminans* LMG23361 chromsome 2 that could not be resolved. The other 2 contigs that have little sequence similarity with *B. lata* are significantly larger at 206 and 393 kbp respectively. Both are bounded by phage integrases. Similar integrases are found in chromosome 1 of *B. phytofirmans* PsJN but the interiors of the contigs don't have large scale similarity with each other or any other sequenced *Burkholderia* species.

The assembly of *B. cenocepacia* K56-2 resolves both chromosome 3 and the plasmid to a single contig. There appears to be a 117kbp deletion in the middle of chromosome 3. Mapping of PacBio reads against the contig doesn't reveal any dips or spikes in coverage and the region does not appear to have been translocated to another chromosome. Unlike the potential genomic rearrangements discussed so far this deletion is bounded by a putative ABC transporter and monoxygenase rather than identical rRNA operons. The other features of note are two regions of anomalously high read coverage corresponding to genomic islands. After assembly reads were mapped back onto the contigs, regions of unusually high coverage are either due to repetitive elements or indicate or indicate the incorrect collapse of a duplicate region in the assembly. In the *B. contaminans* assemblies coverage is relatively even across the contigs except for the every ends which often have sharp spikes and drops in coverage (Figure 4-3, LMG23361 contig 8) indicating that the contig could not reliably be extended. In contrast contig 1 and 6 in the K56-2 assembly show islands of significantly higher coverage (Figure 4-3, K56-2 contigs 1 and 6). The region of high coverage at the end of contig 6 corresponds with the ~47kb genomic island BcenGI12 found in chromosome 2 of *B. cenocepacia* J2315 (Holden *et al*., 2009). The other island of high coverage occurs in the interior of contig1 and corresponds to the ~ 38kb genomic island BcenGI7 found in chromosome 1 of *B. cenocepacia* J2315. Only a single copy of these genomic islands exist in J2315 suggesting K56-2 specific duplications (or conversely J2315 specific deletions). Consistent with the incorrect collapse of a repeat neither island is found anywhere else in the K56-2 assembly. This is not particularly problematic in the case of contig 6 as the presence of a collapsed repeat at the end of a contig only indicates that the assembly is still incomplete. In contrast the collapsed repeat internal to contig1 raises the possibility that contig1

is misassembled, though I have been unable to determine a putative location for the second copy of BcenGI7.

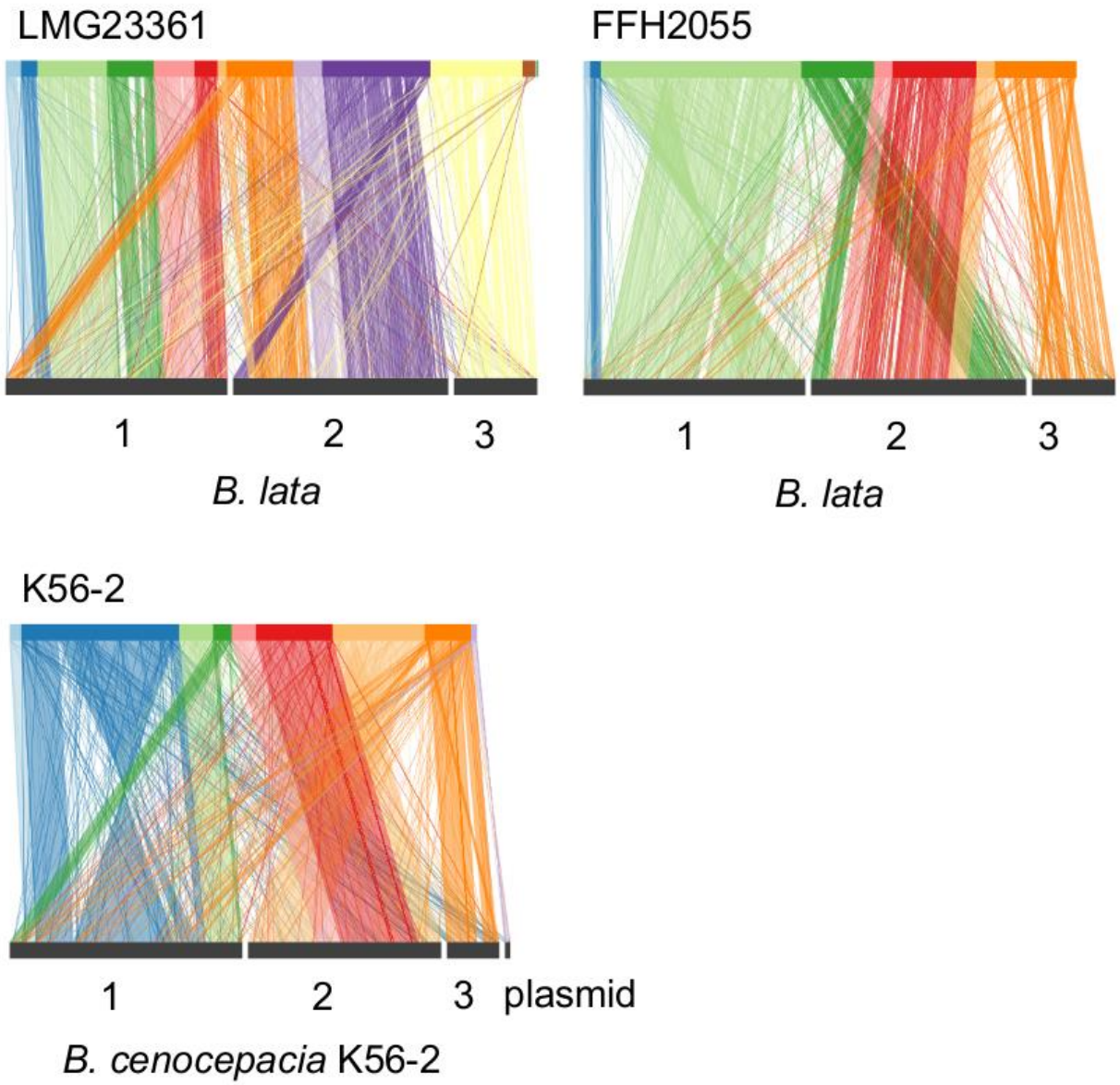


Figure 4-2 Alignments of assemblies against the closest completed genome sequence.

The contigs in each assembly appear at the top of each figure in an arbitrary color while the replicons of the reference appear as black bars at the bottom. Each contig is aligned against all

the replicons in the reference genome, lines between the contigs and replicons for each alignment larger than 1kb. Lengths within a given figure are proportionate to the size of the sequences.

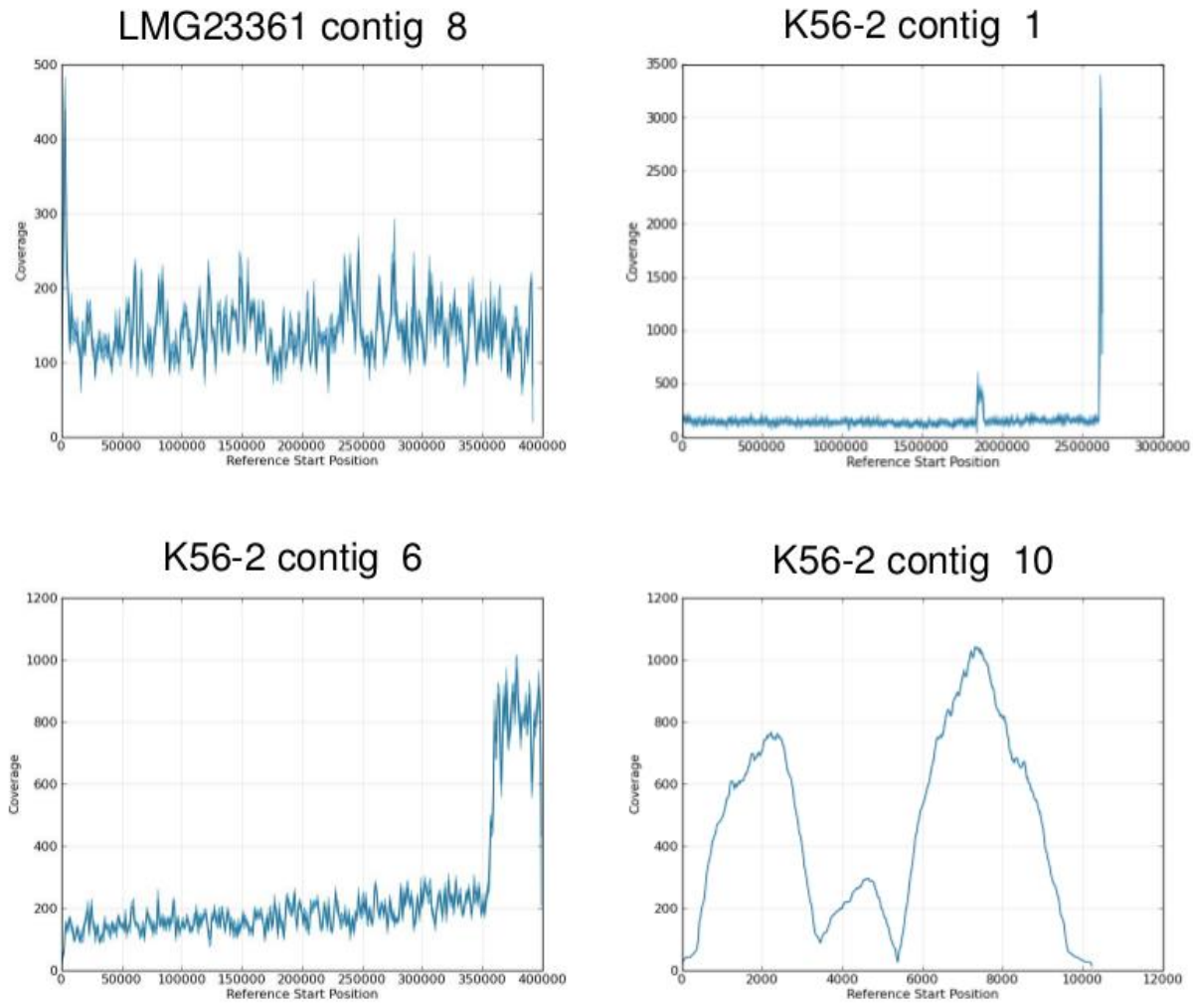


Figure 4-3 Coverage of PacBio reads across select contigs

After assembly all PacBio reads are mapped back against the contigs. Plots show the average coverage over windows of 20bp (K56-2 contig 10) to 5kb (K56-2 contig 1) across 4 select contigs.

4.4 Discussion

The whole genome sequences for *B. contaminans* LMG23361 and FFH2055 are the first whole genome sequences of isolated *B. contaminans* strains, with the existing *B. contaminans* genomic sequences coming from a metagenomics study (Venter *et al*., 2004). We have also sequenced *B. cenocepacia* K56-2, a commonly used lab strain that currently lacks a closed genome sequence. Using 3 Smrt cells per genome provided over twice the needed read coverage for a PacBio only assembly producing 8-18 contigs. Contig lengths were typically bounded by the presence of the ~5kb rRNA operons. In the corrected reads used for the FFH2055 assembly only 8 reads spanned an entire rRNA operon. Closing *Burkholderia* genomes then requires longer reads rather than more read depth. Increasing read lengths depends on large amounts of good quality DNA, aggressive size selection during library prep and improved PacBio read chemistry. We had issues with DNA fragmentation during isolation particularly of *B. contaminans* FFH2055 and have been advised to switch from phenol-chloroform extractions to gravity flow anion-exchange columns (Vaughn Cooper and Amy H. Lee, personal communication). The Duke University sequencing center has also recently switched to the new PacBio P6-C4 chemistry allowing for 20kb insert size rather than the 4-10kb insert size used for our assemblies. Despite not being closed the assemblies appear to be reasonably complete especially that of FFH2055.

5 Conclusions and Future directions

The Bcc exist at the intersection between nonpathogenic environmental isolates, plant growth enhancing rhizosphere isolates, plant and human opportunistic pathogens. Despite their

important ecological niche in the rhizosphere (Coenye & Vandamme., 2003, Vial *et al*., 2011), the vast majority of research on the Bcc is driven by their particular virulence towards people with CF (Isles *et al*., 1984, Weidmann *et al*., 2008) or CGD (Winkelstein *et al*., 2000). Effective treatment is limited by the intrinsic resistance of the Bcc to most clinically relevant antibiotics (Leitao *et al*., 2008, Nzula *et al*., 2002), and effective infection control is complicated by their ability to contaminate many common disinfectants (Garcia-Erce *et al*., 2002, Kutty *et al*., 2007). My research therefore focused primarily on the identification and characterization of essential genes in *B. cenocepacia* K56-2. As these genes are required for growth in non-stressful laboratory conditions, they represent potential targets for novel antimicrobials. Due to my interest in *Burkholderia* genomes, towards the end of my PhD studies I also undertook to sequence two isolates from a new *Burkholderia* species.

Several small scale experiments have demonstrated the feasibility of identifying essential genes by randomly inserting an outward facing inducible promoter and screening for a conditional growth phenotype (Cardona *et al*., 2006, Judson & Mekalanos., 2000). This includes a study in *B. cenocepacia,* which had success using the *E. coli* rhamnose inducible promoter system delivered by a mini-Tn5 transposon derivative (Cardona *et al*., 2006). However, there have been no published reports of attempting to identify the entire essential genome of a species by transposon mediated promoter replacement.

We modified the previously published protocol (Cardona *et al*., 2006) to allow the high-throughput identification of *B. cenocepacia* essential genes using a robotic colony picker and screening mutant growth in liquid micro-titer plates (Bloodworth *et al*., 2013). Approximately 1/1,500 mutants screened had a rhamnose dependent CG phenotype. Sequencing these mutants revealed transposon insertions upstream from some genes with essential orthologs in other species, which are also known to be required for essential cellular processes (e.g. *ftsZ* is required for cell division) demonstrating that this methodology is capable of recovering mutants of truly essential genes on a large scale. As screening progressed, the rate at which conditional growth mutants of novel essential operons was recovered fell substantially below what would be expected in the *B. cenocepacia* was the same size as the essential genome in other species and all essential operons could be recovered at an even frequency. One explanation is that the *B. cenocepacia* genome contains approximately 70 essential operons, 65 to 80% fewer than seen in other species (Luo *et al*., 2014). Alternatively the transposon used by not inserted uniformly across the genome. The Tn*5*-mini transposon has been shown to insert into a highly degenerate consensus sequence (Shevchenko *et al*., 2002) allowing for insertion into almost any sequence. Instead Tn5 hotspots are associated with highly transcribed negatively supercoiled regions of DNA (Lodge & Berg., 1990). Transposon bias has been seen in essential gene studies in *P. aeruginosa.* Two independent essential gene studies performed in *P. aeruginosa* PAO1 and PA14 used two different transposons, a Tn*5*–based system and a mariner transposon, respectively (Jacobs *et al*., 2003, Liberati *et al*., 2006). Approximately half of the orthologs present in both strains that were not hit with the Tn5-based system were disrupted by the mariner transposon. Finally, the chosen rhamnose inducible system may not be able to effectively

substitute for the native essential promoters. In *S. aureus,* an attempt to generate conditional growth mutants via site-directed delivery of a tetracycline-inducible promoter found that of the 150 essential genes targeted, only 64 produced a CG phenotype (Xu *et al*., 2010). In summary, the reasons for the relatively modest rate at which novel essential operons were recovered is currently unknown. Currently other members of the lab are using the same transposon system to carry out a traditional high-density transposon knockout screen for essential genes in *B. cenocepacia* K56-2 allowing a better comparison of the two methods of identifying essential genes.

One of the *B. cenocepacia* essential genes identified in our study belonged to the ETF family of proteins found in Eukaryotes and some bacteria (Scott & Ludwig., 2004, Toogood *et al*., 2007). ETFs are heterodimeric proteins containing a large (ETF-α) and small (ETF-β) subunit. In all studied cases, ETFs funnel electrons from primary dehydrogenases to either the membrane bound respiratory chain (Toogood *et al*., 2007) or to nitrogen fixation in some symbiotic bacteria (Scott & Ludwig., 2004). This putative essential ETF is an intriguing case as ETFs have no known essential functions but have been identified as putatively essential in several essential gene screens. I constructed a site-directed conditional expression mutant and was able to confirm that *etfBA* expression is required for the growth of *B. cenocepacia* K56-2. The essential function of *etfBA* is carbon source independent despite characterized ETFs transferring electrons only from a subset of primary dehydrogenases (Toogood *et al*., 2007). In addition to being essential for cell growth, cells depleted of ETF are depleted of reduced NADH dehydrogenase and are

unable to resume growth when returned to inducing conditions. This loss of cell viability is coupled with a transition from the short rods characteristic of *B. cenocepacia* to small spheres. These spheres are significantly smaller than the round cells produced when the MreB complex is inhibited. This indicates that rather than a loss of control over cell width, EtfBA-depleted cells have stopped elongating after cell division. Cell length is known to be coupled to metabolism with growth in nutrient-poor media (Donachie & Begg., 1989, Sargent, 1975) and long term starvation (Givskov *et al*., 1994) both of which have been shown to induce a shift to shorter cell lengths. In *Bacillus subtilis* (Monahan *et al*., 2014, Weart *et al*., 2007) and *E. coli* (Hill *et al*., 2013), cell length at division is linked to metabolism via sensors that inhibit FtsZ polymerization when bound to UDP-glucose. Therefore in energy rich conditions more copies of FtsZ are required to initiate cell division providing time for the cells to elongate further. I hypothesize that either in addition to, or as a consequence of, its essential function depletion of EtfBA triggers a low energy state in *B. cenocepacia.* This would cause cell division to occur as soon as physically possible resulting in the production of small uniform circular cells. This theory could be tested using some of the central metabolism and energy generation mutants in our CG mutant collection. We have mutants that allow for the selective depletion of most of the elements of the electron transport chain, ATP synthase, the cytochrome bc complex, succinate dehydrogenase and NADH dehydrogenase. We also have one conditional growth mutant for central metabolism, 28-10H8 where 2-oxoglutarate dehydrogenase and dihydrolipoamide dehydrogenase, which are required for the citrate cycle as well as tryptophan metabolism and lysine degradation, are regulated.

The final component of my thesis was the sequencing of the genomes for *B. contaminans* LMG23361 and FFH2055 as well as *B. cenocepacia* K56-2. The genomes were sequenced using PacBio and a 4-10kb insert library. This resulted in draft assemblies containing 8, 17 and 18 contigs for FFH2055, K56-2 and LMG23361 respectively. The contigs in each assembly were aligned against the closest fully sequenced genome, *B. lata* for the contaminans strains and *B. cenocepacia* J2315 for *B. contaminans* K56-2. These revealed several large-scale inversions, which are common in the Bcc, as well as a potential translocation from chromosome 1 to chromosome 2 of *B. contaminans* LMG23361. Both of the *B. contaminans* assembly had even read coverage over the length of the contigs but 2 of the contings in the *B. cenocepacia* K56-2 assembly had islands of anomalously high coverage. These regions of high coverage most likely come from the incorrect collapse of a repeat during assembly where a sequence is present multiple times in the genome but only once in the assembly. Both regions correspond to genomic islands present only once in *B. cenocepacia* J2315, but which I believe exists in multiple copies in K56-2. Using 3 SMRT cells per genome provided more than enough coverage and in future sequencing runs, could probably be reduced to 2. Rather than lack of coverage per se the length of contigs were limited by read length. Correctly assembling one contig per replicon requires multiple reads that are longer than the largest repeat in the genome. In a sequencing library with inserts of 4 to 10 kb, not enough reads span the rRNA operons which are typically 5 to 6 kbp long. Better quality genome preps and the larger insert size supported by newer PacBio chemistry should enable future genome assemblies to produce a single contig per replicon. While not complete given the small number of contigs in the draft assemlifes, they should be able to be

closed relatively easily using long range PCR across the ends of the contigs potentially with Sanger sequencing to identify any missing regions.

Beyond identifying and characterizing essential genes in *B. cenocepacia*, it would be interesting to study the evolution in response to their inhibition. Very few essential genes are universally conserved across bacterial species (Silander & Ackermann., 2009), with genes that are essential in one species being either nonessential or absent from other species. This requires that genomes regularly transition from a specific gene being absolutely required for viability to being superfluous. Development of otherwise promising antibiotic candidates have been abandoned due to this problem including a peptidyl deformylase inhibitor due to the frequent appearance of spontaneous mutants activating a normally cryptic alternative means of translation (Mazel *et al*., 1994) and a methionyl tRNA synthetase inhibitor that proved to be ineffective against a horizontally acquired gene present in some but not all strains of *Streptococcus pneumonia* (Gentry *et al*., 2003). A better understanding of how essential genes are displaced or rendered nonessential would help in choosing better potential targets for novel antibiotics.

6 References

Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D. F., Nguyen, D. T., Sokol, P. A., Carlier, A. & Eberl, L. (2012). Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol Microbiol* **83**, 362-378.

Akerley, B. J., Rubin, E. J., V, N., Amaya, K., Judson, N. & Mekalanos, J. J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc Natl Acad Sci U S A* **99**, 966-971.

Akerley, B. J., Rubin, E. J., Camilli, A., Lampe, D. J., Robertson, H. M. & Mekalanos, J. J. (1998). Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc Natl Acad Sci U S A* **95**, 8927-8932.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Alexander, B. D., Petzold, E. W., Reller, L. B., Palmer, S. M., Davis, R. D., Woods, C. W. & Lipuma, J. J. (2008). Survival after lung transplantation of cystic fibrosis patients infected with Burkholderia cepacia complex. *Am J Transplant* **8**, 1025-1030.

Anraku, Y. (1988). Bacterial electron transport chains. *Annu Rev Biochem* **57**, 101-132.

Apfel, C. M., Locher, H., Evers, S., Takacs, B., Hubschwerlen, C., Pirson, W., Page, M. G. & Keck, W. (2001). Peptide deformylase as an antibacterial drug target: target validation and resistance development. *Antimicrob Agents Chemother* **45**, 1058-1064.

Arigoni, F., Talabot, F., Peitsch, M., Edgerton, M. D., Meldrum, E., Allet, E., Fish, R., Jamotte, T., Curchod, M. L. & Loferer, H. (1998). A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* **16**, 851-856.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. & other authors (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. & other authors (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75-2164-9-75.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008.

Baba, T., Huan, H. C., Datsenko, K., Wanner, B. L. & Mori, H. (2008). The applications of systematic in-frame, single-gene knockout mutant collection of *Escherichia coli* K-12. *Methods Mol Biol* **416**, 183-194.

Baldwin, A., Sokol, P. A., Parkhill, J. & Mahenthiralingam, E. (2004). The *Burkholderia cepacia* epidemic strain marker is part of a novel genomic island encoding both virulence and metabolism-associated genes in *Burkholderia cenocepacia*. *Infect Immun* **72**, 1537-1547.

Baldwin, A., Mahenthiralingam, E., Thickett, K. M., Honeybourne, D., Maiden, M. C., Govan, J. R., Speert, D. P., Lipuma, J. J., Vandamme, P. & Dowson, C. G. (2005). Multilocus sequence typing scheme that provides both species and strain differentiation for the Burkholderia cepacia complex. *J Clin Microbiol* **43**, 4665-4673.

Baldwin, A., Mahenthiralingam, E., Drevinek, P., Vandamme, P., Govan, J. R., Waine, D. J., LiPuma, J. J., Chiarini, L., Dalmastri, C. & other authors (2007). Environmental Burkholderia cepacia complex isolates in human infections. *Emerg Infect Dis* **13**, 458-461.

Balkhy, H. H., Cunningham, G., Francis, C., Almuneef, M. A., Stevens, G., Akkad, N., Elgammal, A., Alassiri, A., Furukawa, E. & other authors (2005). A National Guard outbreak of Burkholderia cepacia infection and colonization secondary to intrinsic contamination of albuterol nebulization solution. *Am J Infect Control* **33**, 182-188.

Battesti, A. & Bouveret, E. (2012). The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. *Methods* **58**, 325-334.

Baugh, L., Gallagher, L. A., Patrapuvich, R., Clifton, M. C., Gardberg, A. S., Edwards, T. E., Armour, B., Begley, D. W., Dieterich, S. H. & other authors (2013). Combining

functional and structural genomics to sample the essential *Burkholderia* structome. *PLoS One* **8**, e53851.

Bazzini, S., Udine, C., Sass, A., Pasca, M. R., Longo, F., Emiliani, G., Fondi, M., Perrin, E., Decorosi, F. & other authors (2011). Deciphering the Role of RND Efflux Transporters in *Burkholderia cenocepacia* . *PLoS One* **6**, e18902.

Bean, G. J., Flickinger, S. T., Westler, W. M., McCully, M. E., Sept, D., Weibel, D. B. & Amann, K. J. (2009). A22 disrupts the bacterial actin cytoskeleton by directly binding and inducing a low-affinity state in MreB. *Biochemistry* **48**, 4852-4857.

Bergmiller, T., Ackermann, M. & Silander, O. K. (2012). Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLoS Genet* **8**, e1002803.

Bevivino, A., Peggion, V., Chiarini, L., Tabacchioni, S., Cantale, C. & Dalmastri, C. (2005). Effect of Fusarium verticillioides on maize-root-associated Burkholderia cenocepacia populations. *Res Microbiol* **156**, 974-983.

Bevivino, A., Dalmastri, C., Tabacchioni, S., Chiarini, L., Belli, M. L., Piana, S., Materazzo, A., Vandamme, P. & Manno, G. (2002). Burkholderia cepacia complex bacteria from clinical and environmental sources in Italy: genomovar status and distribution of traits related to virulence and transmissibility. *J Clin Microbiol* **40**, 846-851.

Bharat, A. & Brown, E. D. (2014). Phenotypic investigations of the depletion of EngA in *Escherichia coli* are consistent with a role in ribosome biogenesis. *FEMS Microbiol Lett* **353**, 26-32.

Blattner, F. R., Plunkett, G.,3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462.

Bloodworth, R. A. M., Park, M., Peters, A. & Cardona, S. T. (2009). A genome-scale identification of essential *Burkholderia cenocepacia* genes using transposon mediated insertion of a rhamnose inducible promoter. *CSM General Meeting, June 15-18, Montreal QB* .

Bloodworth, R. A., Gislason, A. S. & Cardona, S. T. (2013). *Burkholderia cenocepacia* conditional growth mutant library created by random promoter replacement of essential genes. *Microbiologyopen* **2**, 243-258.

Bolla, J. M., Alibert-Franco, S., Handzlik, J., Chevalier, J., Mahamoud, A., Boyer, G., Kiec-Kononowicz, K. & Pages, J. M. (2011). Strategies for bypassing the membrane barrier in multidrug resistant Gram-negative bacteria. *FEBS Lett* **585**, 1682-1690.

Brantl, S. & Wagner, E. G. (2000). Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. *Mol Microbiol* **35**, 1469-1482.

Brautaset, T., Lale, R. & Valla, S. (2009). Positively regulated bacterial expression systems. *Microb Biotechnol* **2**, 15-30.

Brisse, S., Cordevant, C., Vandamme, P., Bidet, P., Loukil, C., Chabanon, G., Lange, M. & Bingen, E. (2004). Species distribution and ribotype diversity of Burkholderia cepacia complex isolates from French patients with cystic fibrosis. *J Clin Microbiol* **42**, 4824-4827.

Brown, E. D. & Wright, G. D. (2005). New targets and screening approaches in antimicrobial drug discovery. *Chem. Rev.* **105**, 759-774.

Bugrysheva, J. V., Froehlich, B. J., Freiberg, J. A. & Scott, J. R. (2011). The histone-like protein Hlp is essential for growth of *Streptococcus pyogenes*: comparison of genetic approaches to study essential genes. *Appl Environ Microbiol* **77**, 4422-4428.

Burkholder, W. H. (1950). Sour skin, a bacterial rot of onion bulbs. *Phytopathol* **40**, 115-117.

Bush, K. (2010). The coming of age of antibiotics: discovery and therapeutic value. *Ann N Y Acad Sci* **1213**, 1-4.

Bylund, J., Campsall, P. A., Ma, R. C., Conway, B. A. & Speert, D. P. (2005). Burkholderia cenocepacia induces neutrophil necrosis in chronic granulomatous disease. *J Immunol* **174**, 3562-3569.

Caldelari, I., Chao, Y., Romby, P. & Vogel, J. (2013). RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harb Perspect Med* **3**, a010298.

Cameron, D. E., Urbach, J. M. & Mekalanos, J. J. (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.

Campana, S., Taccetti, G., Ravenni, N., Masi, I., Audino, S., Sisi, B., Repetto, T., Doring, G. & de Martino, M. (2004). Molecular epidemiology of Pseudomonas aeruginosa, Burkholderia cepacia complex and methicillin-resistant Staphylococcus aureus in a cystic fibrosis center. *J Cyst Fibros* **3**, 159-163.

Campbell, T. L., Daigle, D. M. & Brown, E. D. (2005). Characterization of the *Bacillus subtilis* GTPase YloQ and its role in ribosome function. *Biochem J* **389**, 843-852.

Cardona, S. T. & Valvano, M. A. (2005). An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid* **54**, 219-228.

Cardona, S. T., Selin, C. & Gislason, A. S. (2014). Genomic tools to profile antibiotic mode of action. *Crit Rev Microbiol* .

Cardona, S. T., Mueller, C. & Valvano, M. A. (2006). Identification of essential operons in *Burkholderia cenocepacia* with a rhamnose inducible promoter. *Appl Environ Microbiol* **72**, 2547-2555.

Carroll, P., Muttucumaru, D. G. & Parish, T. (2005). Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Appl Environ Microbiol* **71**, 3077-3084.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3423.

Chalker, A. F. & Lunsford, R. D. (2002). Rational identification of new antibacterial drug targets that are essential for viability using a genomics-based approach. *Pharmacol Ther* **95**, 1-

20.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-552.

Chan, P. F., Holmes, D. J. & Payne, D. J. (2004). Finding the gems using genomic discovery: antibacterial drug discovery strategies – the successes and the challenges. *Drug Discovery Today: Therapeutic Strategies* **1**, 519-527.

Chelius, M. K. & Triplett, E. W. (2001). The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microb Ecol* **41**, 252-263.

Chen, D. & Swenson, R. P. (1994). Cloning, sequence analysis, and expression of the genes encoding the two subunits of the methylotrophic bacterium W3A1 electron transfer flavoprotein. *J Biol Chem* **269**, 32120-32130.

Chen, J. S., Witzmann, K. A., Spilker, T., Fink, R. J. & LiPuma, J. J. (2001). Endemicity and inter-city spread of *Burkholderia cepacia* genomovar III in cystic fibrosis. *J Pediatr* **139**, 643-649.

Chiarini, L., Bevivino, A., Dalmastri, C., Tabacchioni, S. & Visca, P. (2006). *Burkholderia cepacia* complex species: health hazards and biotechnological potential. *Trends Microbiol* **14**, 277-286.

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J. & other authors (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10**, 563-569.

Christen, B., Abeliuk, E., Collier, J. M., Kalogeraki, V. S., Passarelli, B., Coller, J. A., Fero, M. J., McAdams, H. H. & Shapiro, L. (2011). The essential genome of a bacterium. *Mol Syst Biol* **7**, 528-535.

Coenye, T. & Vandamme, P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ Microbiol* **5**, 719-729.

Coenye, T., Laevens, S., Willems, A., Ohlen, M., Hannant, W., Govan, J. R., Gillis, M., Falsen, E. & Vandamme, P. (2001). Burkholderia fungorum sp. nov. and Burkholderia caledonica sp. nov., two new species isolated from the environment, animals and human clinical samples. *Int J Syst Evol Microbiol* **51**, 1099-1107.

Commichau, F. M., Pietack, N. & Stulke, J. (2013). Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Mol Biosyst* **9**, 1068-1075.

Corral, D. M., Coates, A. L., Yau, Y. C., Tellier, R., Glass, M., Jones, S. M. & Waters, V. J. (2008). Burkholderia pseudomallei infection in a cystic fibrosis patient from the Caribbean: a case report. *Can Respir J* **15**, 237-239.

Courtney, J. M., Dunbar, K. E., McDowell, A., Moore, J. E., Warke, T. J., Stevenson, M. & Elborn, J. S. (2004). Clinical outcome of Burkholderia cepacia complex infection in cystic fibrosis adults. *J Cyst Fibros* **3**, 93-98.

Craig, F. F., Coote, J. G., Parton, R., Freer, J. H. & Gilmour, N. J. (1989). A plasmid which can be transferred between *Escherichia coli* and *Pasteurella haemolytica* by electroporation and conjugation. *J Gen Microbiol* **135**, 2885-2890.

Cunha, M. V., Leitao, J. H., Mahenthiralingam, E., Vandamme, P., Lito, L., Barreto, C., Salgado, M. J. & Sa-Correia, I. (2003). Molecular analysis of *Burkholderia cepacia* complex isolates from a Portuguese cystic fibrosis center: a 7-year study. *J Clin Microbiol* **41**, 4113-4120.

Cuthbert, A. W. (2011). New horizons in the treatment of cystic fibrosis. *Br J Pharmacol* **163**, 173-183.

Dalbey, R. E. & Wickner, W. (1985). Leader peptidase catalyzes the release of exported proteins from the outer surface of the *Escherichia coli* plasma membrane. *J Biol Chem* **260**, 15925-15931.

Dam, P., Olman, V., Harris, K., Su, Z. & Xu, Y. (2007). Operon prediction using both genome-specific and general genomic information. *Nucleic Acids Res* **35**, 288-298.

Darling, P., Chan, M., Cox, A. D. & Sokol, P. A. (1998). Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia*. *Infect Immun* **66**, 874-877.

Das, S., Noe, J. C., Paik, S. & Kitten, T. (2005). An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. *J Microbiol Methods* **63**, 89-94.

Daszczuk, A., Dessalegne, Y., Drenth, I., Hendriks, E., Jo, E., van Lente, T., Oldebesten, A., Parrish, J., Poljakova, W. & other authors (2014). Bacillus subtilis biosensor engineered to assess meat spoilage. *ACS Synth Biol* **3**, 999-1002.

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G. & other authors (2008). A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol Syst Biol* **4**, 174-189.

DeJesus, M. A., Zhang, Y. J., Sassetti, C. M., Rubin, E. J., Sacchettini, J. C. & Ioerger, T. R. (2013). Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries. *Bioinformatics* **29**, 695-703.

D'Elia, M. A., Pereira, M. P. & Brown, E. D. (2009). Are essential genes really essential? *Trends Microbiol* **17**, 433-438.

DelVecchio, V. G., Kapatral, V., Redkar, R. J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A. & other authors (2002). The genome sequence of the facultative intracellular pathogen Brucella melitensis. *Proc Natl Acad Sci U S A* **99**, 443-448.

Deng, J., Deng, L., Su, S., Zhang, M., Lin, X., Wei, L., Minai, A. A., Hassett, D. J. & Lu, L. J. (2011). Investigating the predictability of essential genes across distantly related organisms using an integrative approach. *Nucleic Acids Res* **39**, 795-807.

Dennis, J. J. & Zylstra, G. J. (1998). Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. *Appl Environ Microbiol* **64**, 2710-2715.

DeVito, J. A., Mills, J. A., Liu, V. G., Agarwal, A., Sizemore, C. F., Yao, Z., Stoughton, D. M., Cappiello, M. G., Barbosa, M. D. & other authors (2002). An array of target-specific

screening strains for antibacterial discovery. *Nat Biotechnol* **20**, 478-483.

Dhillon, B. K., Laird, M. R., Shay, J. A., Winsor, G. L., Lo, R., Nizam, F., Pereira, S. K., Waglechner, N., McArthur, A. G. & other authors (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* **43**, W104-W108.

Dicker, I. B. & Seetharam, S. (1992). What is known about the structure and function of the Escherichia coli protein FirA? *Mol Microbiol* **6**, 817-823.

Donachie, W. D. & Begg, K. J. (1989). Cell length, nucleoid separation, and cell division of rod-shaped and spherical cells of *Escherichia coli*. *J Bacteriol* **171**, 4633-4639.

Donald, R. G. K., Skwish, S., Forsyth, R. A., Anderson, J. W., Zhong, T., Burns, C., Lee, S., Meng, X., LoCastro, L. & other authors (2009). A *Staphylococcus aureus* Fitness Test Platform for Mechanism-Based Profiling of Antibacterial Compounds. *Chem Biol* **16**, 826-836.

Dotsch, A., Klawonn, F., Jarek, M., Scharfe, M., Blocker, H. & Haussler, S. (2010). Evolutionary conservation of essential and highly expressed genes in Pseudomonas aeruginosa. *BMC Genomics* **11**, 234.

Drevinek, P., Baldwin, A., Lindenburg, L., Joshi, L. T., Marchbank, A., Vosahlikova, S., Dowson, C. G. & Mahenthiralingam, E. (2010). Oxidative stress of Burkholderia cenocepacia induces insertion sequence-mediated genomic rearrangements that interfere with macrorestriction-based genotyping. *J Clin Microbiol* **48**, 34-40.

Drevinek, P., Cinek, O., Melter, J., Langsadl, L., Navesnakova, Y. & Vavrova, V. (2003). Genomovar distribution of the Burkholderia cepacia complex differs significantly between Czech and Slovak patients with cystic fibrosis. *J Med Microbiol* **52**, 603-604.

Drevinek, P. & Mahenthiralingam, E. (2010). *Burkholderia cenocepacia* in cystic fibrosis: epidemiology and molecular mechanisms of virulence. *Clin Microbiol Infect* **16**, 821-830.

Dupuis, A., Hamilton, D., Cole, D. E. & Corey, M. (2005). Cystic fibrosis birth rates in Canada: a decreasing trend since the onset of genetic testing. *J Pediatr* **147**, 312-315.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Egan, E. S., Fogel, M. A. & Waldor, M. K. (2005). Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol Microbiol* **56**, 1129-1138.

Egan, S. M. & Schleif, R. F. (1993). A regulatory cascade in the induction of rhaBAD. *J Mol Biol* **234**, 87-98.

Fang, G., Rocha, E. & Danchin, A. (2005). How essential are nonessential genes? *Mol Biol Evol* **22**, 2147-2156.

Fang, K., Zhao, H., Sun, C., Lam, C. M., Chang, S., Zhang, K., Panda, G., Godinho, M., Martins dos Santos, V. A. & Wang, J. (2011a). Exploring the metabolic network of the epidemic pathogen Burkholderia cenocepacia J2315 via genome-scale reconstruction. *BMC Syst Biol* **5**, 83-0509-5-83.

Fang, Y., Xie, G. L., Lou, M. M., Li, B. & Muhammad, I. (2011b). Diversity analysis of Burkholderia cepacia complex in the water bodies of West Lake, Hangzhou, China. *J Microbiol* **49**, 309-314.

Fernandez-Gomez, B., Fernandez-Guerra, A., Casamayor, E. O., Gonzalez, J. M., Pedros-Alio, C. & Acinas, S. G. (2012). Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* **13**, 347-2164-13-347.

Ferrarini, M., Moretto, M., Ward, J. A., Surbanovski, N., Stevanovic, V., Giongo, L., Viola, R., Cavalieri, D., Velasco, R. & other authors (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics* **14**, 670-2164-14-670.

Figurski, D. H. & Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**, 1648-1652.

Flannagan, R. S., Linn, T. & Valvano, M. A. (2008). A system for the construction of targeted unmarked gene deletions in the genus *Burkholderia*. *Environ Microbiol* **10**, 1652-1660.

Flannagan, R. S., Aubert, D., Kooi, C., Sokol, P. A. & Valvano, M. A. (2007). *Burkholderia cenocepacia* requires a periplasmic HtrA protease for growth under thermal and osmotic stress and for survival *in vivo*. *Infect Immun* **75**, 1679-1689.

Foley, P. L. & Shuler, M. L. (2010). Considerations for the design and construction of a synthetic platform cell for biotechnological applications. *Biotechnol Bioeng* **105**, 26-36.

Forsyth, R. A., Haselbeck, R. J., Ohlsen, K. L., Yamamoto, R. T., Xu, H., Trawick, J. D., Wall, D., Wang, L., V, B. & other authors (2002). A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol Microbiol* **43**, 1387-1400.

Frerman F. E. & Goodman S. I. (2001). Defects of electron transfer flavoprotein and electron transfer flavoprotein-ubiquinone oxidoreductase: glutaric acidemia type II. In *The Metabolic and Molecular Bases of Inherited Disease*, 2nd edn, pp. 2357-2365. Edited by C. R. Scriver. New York: McGraw-Hill.

Fu, X., Fu, N., Guo, S., Yan, Z., Xu, Y., Hu, H., Menzel, C., Chen, W., Li, Y. & other authors (2009). Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* **10**, 161.

Gallagher, L. A., Shendure, J. & Manoil, C. (2011). Genome-Scale Identification of Resistance Functions in *Pseudomonas aeruginosa* Using Tn-seq. *Mbio* **2**, e00315-10.

Gallagher, L. A., Ramage, E., Patrapuvich, R., Weiss, E., Brittnacher, M. & Manoil, C. (2013). Sequence-defined transposon mutant library of Burkholderia thailandensis. *Mbio* **4**, e00604-13.

Galloway, W. R., Bender, A., Welch, M. & Spring, D. R. (2009). The discovery of antibacterial agents using diversity-oriented synthesis. *Chem Commun (Camb)* **(18):2446-62. doi**, 2446-2462.

Galyov, E. E., Brett, P. J. & DeShazer, D. (2010). Molecular insights into Burkholderia pseudomallei and Burkholderia mallei pathogenesis. *Annu Rev Microbiol* **64**, 495-517.

Gao, F. & Zhang, R. R. (2011). Enzymes are enriched in bacterial essential genes. *PLoS One* **6**, e21683.

Garcia-Erce, J. A., Grasa, J. M., Solano, V. M., Gimeno, J. J., Lopez, A., Hernandez, M. J., Marco, M. L., Arribas, J. L. & Giralt, M. (2002). Bacterial contamination of blood components due to Burkholderia cepacia contamination from clorhexidine bottles. *Vox Sang* **83**, 70-71.

Gawronski, J. D., Wong, S. M., Giannoukos, G., Ward, D. V. & Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci U S A* **106**, 16422-16427.

Gentry, D. R., Ingraham, K. A., Stanhope, M. J., Rittenhouse, S., Jarvest, R. L., O'Hanlon, P. J., Brown, J. R. & Holmes, D. J. (2003). Variable sensitivity to bacterial methionyl-tRNA synthetase inhibitors reveals subpopulations of Streptococcus pneumoniae with two distinct methionyl-tRNA synthetase genes. *Antimicrob Agents Chemother* **47**, 1784-1789.

Gerdes, S., Edwards, R., Kubal, M., Fonstein, M., Stevens, R. & Osterman, A. (2006). Essential genes on metabolic maps. *Curr Opin Biotechnol* **17**, 448-456.

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I. & other authors (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185**, 5673-5684.

Ghisla, S. & Thorpe, C. (2004). Acyl-CoA dehydrogenases. A mechanistic overview. *Eur J Biochem* **271**, 494-508.

Givskov, M., Eberl, L., Moller, S., Poulsen, L. K. & Molin, S. (1994). Responses to nutrient starvation in *Pseudomonas putida* KT2442: analysis of general cross-protection, cell shape, and macromolecular content. *J Bacteriol* **176**, 7-14.

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R. & Gordon, J. I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279-289.

Goss, C. H. & Burns, J. L. (2007). Exacerbations in cystic fibrosis {middle dot} 1: Epidemiology and pathogenesis. *Thorax* **62**, 360-367.

Goswami, M., Mangoli, S. H. & Jawali, N. (2006). Involvement of reactive oxygen species in the action of ciprofloxacin against *Escherichia coli*. *Antimicrob Agents Chemother* **50**, 949-954.

Gould, I. M. (2008). The epidemiology of antibiotic resistance. *Int J Antimicrob Agents* **32 Suppl 1**, S2-9.

Govan, J. R., Brown, P. H., Maddison, J., Doherty, C. J., Nelson, J. W., Dodd, M., Greening, A. P. & Webb, A. K. (1993). Evidence for transmission of *Pseudomonas cepacia* by social contact in cystic fibrosis. *Lancet* **342**, 15-19.

Graindorge, A., Menard, A., Monnez, C. & Cournoyer, B. (2012). Insertion sequence evolutionary patterns highlight convergent genetic inactivations and recent genomic island acquisitions among epidemic Burkholderia cenocepacia. *J Med Microbiol* **61**, 394-409.

Greenberg, D. E., Goldberg, J. B., Stock, F., Murray, P. R., Holland, S. M. & Lipuma, J. J. (2009). Recurrent Burkholderia infection in patients with chronic granulomatous disease: 11-year experience at a large referral center. *Clin Infect Dis* **48**, 1577-1579.

Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J. & Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**, e1002251.

Gu, Z., Cavalcanti, A., Chen, F. C., Bouman, P. & Li, W. H. (2002). Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol Biol Evol* **19**, 256-262.

Guide, S. V., Stock, F., Gill, V. J., Anderson, V. L., Malech, H. L., Gallin, J. I. & Holland, S. M. (2003). Reinfection, rather than persistent infection, in patients with chronic granulomatous disease. *J Infect Dis* **187**, 845-853.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-30.

Gwynn, M. N., Portnoy, A., Rittenhouse, S. F. & Payne, D. J. (2010). Challenges of antibacterial discovery revisited. *Ann N Y Acad Sci* **1213**, 5-19.

Haldimann, A., Daniels, L. L. & Wanner, B. L. (1998). Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon. *J Bacteriol* **180**, 1277-1286.

Harris, J. K., Kelley, S. T., Spiegelman, G. B. & Pace, N. R. (2003). The genetic core of the universal ancestor. *Genome Res* **13**, 407-412.

Harrison, P. W., Lower, R. P., Kim, N. K. & Young, J. P. (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol* **18**, 141-148.

Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y. & other authors (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol Microbiol* **55**, 137-149.

Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D. & other authors (2000). DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature* **406**, 477-483.

Heyworth, P. G., Cross, A. R. & Curnutte, J. T. (2003). Chronic granulomatous disease. *Curr Opin Immunol* **15**, 578-584.

Hill, N. S., Buske, P. J., Shi, Y. & Levin, P. A. (2013). A moonlighting enzyme links *Escherichia coli* cell size with central metabolism. *PLoS Genet* **9**, e1003663.

Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St Onge, R. P., Tyers, M. & other authors (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362-365.

Holden, M. T., Seth-Smith, H. M., Crossman, L. C., Sebaihia, M., Bentley, S. D., Cerdeno-Tarraga, A. M., Thomson, N. R., Bason, N., Quail, M. A. & other authors (2009). The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. *J Bacteriol* **191**, 261-277.

Hurst, L. D. & Smith, N. G. (1999). Do essential genes evolve slowly? *Curr Biol* **9**, 747-750.

Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. & Venter, J. C. (1999). Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165-2169.

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, J. L., Fasi, A. C., Ramette, A., Smith, J. J., Hammerschmidt, R. & Sundin, G. W. (2008). Identification and onion pathogenicity of Burkholderia cepacia complex isolates from the onion rhizosphere and onion field soil. *Appl Environ Microbiol* **74**, 3121-3129.

Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C. & other authors (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **100**, 14339-14344.

Jacoby, E. & Mozzarelli, A. (2009). Chemogenomic strategies to expand the bioactive chemical space. *Curr Med Chem* **16**, 4374-4381.

Ji, Y., Zhang, B., Van, S. F., Horn, Warren, P., Woodnutt, G., Burnham, M. K. & Rosenberg, M. (2001). Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**, 2266-2269.

Johansen, W. (1909). Elemente der exakten Erblichkeitslehre. *Molecular and General Genetics MGG* **2**, 136-137.

Johnson, W. M., Tyler, S. D. & Rozee, K. R. (1994). Linkage analysis of geographic and clinical clusters in Pseudomonas cepacia infections by multilocus enzyme electrophoresis and ribotyping. *J Clin Microbiol* **32**, 924-930.

Jones, A. M., Dodd, M. E., Govan, J. R., Barcus, V., Doherty, C. J., Morris, J. & Webb, A. K. (2004). *Burkholderia cenocepacia* and *Burkholderia multivorans*: influence on survival in cystic fibrosis. *Thorax* **59**, 948-951.

Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. (2002). Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* **12**, 962-968.

Jorda-Vargas, L., Degrossi, J., Castaneda, N. C., D'Aquino, M., Valvano, M. A., Procopio, A., Galanternik, L. & Centron, D. (2008). Prevalence of indeterminate genetic species of Burkholderia cepacia complex in a cystic fibrosis center in Argentina. *J Clin Microbiol* **46**, 1151-1152.

Judson, N. & Mekalanos, J. J. (2000a). TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat Biotechnol* **18**, 740-745.

Judson, N. & Mekalanos, J. J. (2000b). Transposon-based approaches to identify essential bacterial genes. *Trends Microbiol* **8**, 521-526.

Juhas, M., Eberl, L. & Church, G. M. (2012). Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol* **30**, 601-607.

Juhas, M., Reuss, D., Zhu, B. & Commichau, F. M. (2014). *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology* .

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W. & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33**, 376-393.

Juhas, M., Stark, M., von Mering, C., Lumjiaktase, P., Crook, D. W., Valvano, M. A. & Eberl, L. (2012). High confidence prediction of essential genes in *Burkholderia cenocepacia*. *PLoS One* **7**, e40064.

Kang, Y., Durfee, T., Glasner, J. D., Qiu, Y., Frisch, D., Winterberg, K. M. & Blattner, F. R. (2004). Systematic mutagenesis of the *Escherichia coli* genome. *J Bacteriol* **186**, 4921-4930.

Kang, Y., Gu, C., Yuan, L., Wang, Y., Zhu, Y., Li, X., Luo, Q., Xiao, J., Jiang, D. & other authors (2014). Flexibility and symmetry of prokaryotic genome rearrangement reveal lineage-associated core-gene-defined genome organizational frameworks. *Mbio* **5**, e01867-14.

Kennedy, M. P., Coakley, R. D., Donaldson, S. H., Aris, R. M., Hohneker, K., Wedd, J. P., Knowles, M. R., Gilligan, P. H. & Yankaskas, J. R. (2007). Burkholderia gladioli: five year experience in a cystic fibrosis and lung transplantation center. *J Cyst Fibros* **6**, 267-273.

Khalil, A. S. & Collins, J. J. (2010). Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367-379.

Kimura, M. (1991). The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet* **66**, 367-386.

Kimura, M. & Ohta, T. (1974). On some principles governing molecular evolution. *Proc Natl Acad Sci U S A* **71**, 2848-2852.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H. & Mori, H. (2005). Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res* **12**, 291-299.

Klein, B. A., Tenorio, E. L., Lazinski, D. W., Camilli, A., Duncan, M. J. & Hu, L. T. (2012). Identification of essential genes of the periodontal pathogen *Porphyromonas gingivalis*. *BMC Genomics* **13**, 578-595.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S. & other authors (2003). Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U S A* **100**, 4678-4683.

Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**, 127-136.

Kotloff, R. M. & Thabut, G. (2011). Lung transplantation. *Am J Respir Crit Care Med* **184**, 159-171.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645.

Kumar, A., Dietrich, S., Schneider, W., Jacobson, R., Downes, F. P., Robinson-Dunn, B. E., Honicky, R., Smith, J. & Martin, R. (1997). Genetic relatedness of Burkholderia (Pseudomonas) cepacia isolates from five cystic fibrosis centers in Michigan. *Respir Med* **91**, 485-492.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A. & other authors (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**, 249-256.

Kutty, P. K., Moody, B., Gullion, J. S., Zervos, M., Ajluni, M., Washburn, R., Sanderson, R., Kainer, M. A., Powell, T. A. & other authors (2007). Multistate outbreak of Burkholderia cenocepacia colonization and infection associated with the use of intrinsically contaminated alcohol-free mouthwash. *Chest* **132**, 1825-1831.

Kyrpides, N., Overbeek, R. & Ouzounis, C. (1999). Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* **49**, 413-423.

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E. & other authors (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* **19**, 2308-2316.

Law, R. J., Hamlin, J. N. R., Sivro, A., McCorrister, S. J., Cardama, G. A. & Cardona, S. T. (2008). A functional phenylacetic acid catabolic pathway is required for full pathogenicity of *Burkholderia cenocepacia* in the *Caenorhabditis elegans* host model. *J Bacteriol* **190**, 7209-7218.

Layton, D. S. & Trinh, C. T. (2014). Engineering modular ester fermentative pathways in Escherichia coli. *Metab Eng* **26C**, 77-88.

Ledson, M. J., Gallagher, M. J., Jackson, M., Hart, C. A. & Walshaw, M. J. (2002). Outcome of Burkholderia cepacia colonisation in an adult cystic fibrosis centre. *Thorax* **57**, 142-

145.

Lee, S. A., Gallagher, L. A., Thongdee, M., Staudinger, B. J., Lippman, S., Singh, P. K. & Manoil, C. (2015). General and condition-specific essential functions of Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* **112**, 5189-5194.

Lefebre, M. D. & Valvano, M. A. (2002). Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in *Burkholderia cepacia* complex isolates. *Appl Environ Microbiol* **68**, 5956-5964.

Lehmann E. L. (1975). Nonparametrics : statistical methods based on ranks : Holden-Day.

Leitao, J. H., Sousa, S. A., Cunha, M. V., Salgado, M. J., Melo-Cristino, J., Barreto, M. C. & Sa-Correia, I. (2008). Variation of the antimicrobial susceptibility profiles of *Burkholderia cepacia* complex clonal isolates obtained from chronically infected cystic fibrosis patients: a five-year survey in the major Portuguese treatment center. *Eur J Clin Microbiol Infect Dis* **27**, 1101-1111.

Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**, e130.

Lessie, T. G., Hendrickson, W., Manning, B. D. & Devereux, R. (1996). Genomic complexity and plasticity of Burkholderia cepacia. *FEMS Microbiol Lett* **144**, 117-128.

Lewis, K. (2013). Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**, 371-387.

Lewis, R. J., Tsai, F. T. & Wigley, D. B. (1996). Molecular mechanisms of drug inhibition of DNA gyrase. *Bioessays* **18**, 661-671.

Li, F., Hinderberger, J., Seedorf, H., Zhang, J., Buckel, W. & Thauer, R. K. (2008). Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*. *J Bacteriol* **190**, 843-850.

Li, X., Zolli-Juran, M., Cechetto, J. D., Daigle, D. M., Wright, G. D. & Brown, E. D. (2004). Multicopy suppressors for novel antibacterial compounds reveal targets and drug efflux

susceptibility. *Chem Biol* **11**, 1423-1430.

Liberati, N. T., Urbach, J. M., Miyata, S., Lee, D. G., Drenkard, E., Wu, G., Villanueva, J., Wei, T. & Ausubel, F. M. (2006). An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci U S A* **103**, 2833-8.

Lipuma, J. J. (2010). The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev* **23**, 299-323.

LiPuma, J. J., Spilker, T., Coenye, T. & Gonzalez, C. F. (2002). An epidemic Burkholderia cepacia complex strain identified in soil. *Lancet* **359**, 2002-2003.

LiPuma, J. J., Dasen, S. E., Nielson, D. W., Stern, R. C. & Stull, T. L. (1990). Person-to-person transmission of *Pseudomonas cepacia* between patients with cystic fibrosis. *Lancet* **336**, 1094-1096.

Liu, H., Chen, C., Zhang, H., Kaur, J., Goldman, Y. E. & Cooperman, B. S. (2011). The conserved protein EF4 (LepA) modulates the elongation cycle of protein synthesis. *Proc Natl Acad Sci U S A* **108**, 16223-16228.

Lodge, J. K. & Berg, D. E. (1990). Mutations that affect Tn5 insertion into pBR322: importance of local DNA supercoiling. *J Bacteriol* **172**, 5956-5960.

Loutet, S. A. & Valvano, M. A. (2010). A decade of *Burkholderia cenocepacia* virulence determinant research. *Infect Immun* **78**, 4088-4100.

Luisi, P. L. (2002). Toward the engineering of minimal living cells. *Anat Rec* **268**, 208-214.

Luo, H., Lin, Y., Gao, F., Zhang, C. T. & Zhang, R. (2014). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res* **42**, D574-80.

Mackenzie C., Kaplan S., Choudhary M. (2004). Multiple Chromosomes. In *Microbial Evolution: Gene Establishment, Survival, and Exchange*pp. 82-101. Edited by R. V. Miller & M. J. Day. Washington, D.C.: ASM Press.

Mahenthiralingam, E., Baldwin, A., Drevinek, P., Vanlaere, E., Vandamme, P., LiPuma, J. J. & Dowson, C. G. (2006). Multilocus sequence typing breathes life into a microbial metagenome. *PLoS One* **1**, e17.

Mahenthiralingam, E., Urban, T. A. & Goldberg, J. B. (2005). The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol* **3**, 144-156.

Mahenthiralingam, E., Coenye, T., Chung, J. W., Speert, D. P., Govan, J. R., Taylor, P. & Vandamme, P. (2000). Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia* complex. *J Clin Microbiol* **38**, 910-913.

Maignan, M., Verdant, C., Bouvet, G. F., Van Spall, M. & Berthiaume, Y. (2013). Undiagnosed Chronic Granulomatous Disease, Burkholderia cepacia complex Pneumonia, and Acquired Hemophagocytic Lymphohistiocytosis: A Deadly Association. *Case Rep Pulmonol* **2013**, 874197.

Maravic, A., Skocibusic, M., Sprung, M., Samanic, I., Puizina, J. & Pavela-Vrancic, M. (2012). Occurrence and antibiotic susceptibility profiles of *Burkholderia cepacia* complex in coastal marine environment. *Int J Environ Health Res* **22**, 531-542.

Marciano, B. E., Spalding, C., Fitzgerald, A., Mann, D., Brown, T., Osgood, S., Yockey, L., Darnell, D. N., Barnhart, L. & other authors (2014). Common Severe Infections in Chronic Granulomatous Disease. *Clin Infect Dis* .

Martina, P., Bettiol, M., Vescina, C., Montanaro, P., Mannino, M. C., Prieto, C. I., Vay, C., Naumann, D., Schmitt, J. & other authors (2013). Genetic diversity of Burkholderia contaminans isolates from cystic fibrosis patients in Argentina. *J Clin Microbiol* **51**, 339-344.

Martinez-Garcia, E., Aparicio, T., de Lorenzo, V. & Nikel, P. I. (2014). New transposon tools tailored for metabolic engineering of gram-negative microbial cell factories. *Front Bioeng Biotechnol* **2**, 46.

Matsuoka, H., Hirooka, K. & Fujita, Y. (2007). Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *J Biol Chem* **282**, 5180-5194.

Mazel, D., Pochet, S. & Marliere, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. *Embo j* **13**, 914-923.

Mazodier, P., Cossart, P., Giraud, E. & Gasser, F. (1985). Completion of the nucleotide sequence of the central region of Tn5 confirms the presence of three resistance genes. *Nucleic Acids Res* **13**, 195-205.

Medina-Pascual, M. J., Valdezate, S., Carrasco, G., Villalon, P., Garrido, N. & Saez-Nieto, J. A. (2015). Increase in isolation of Burkholderia contaminans from Spanish patients with cystic fibrosis. *Clin Microbiol Infect* **21**, 150-156.

Meng, J., Kanzaki, G., Meas, D., Lam, C. K., Crummer, H., Tain, J. & Xu, H. H. (2012). A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing Escherichia coli essential genes. *FEMS Microbiol Lett* **329**, 45-53.

Metris, A., Reuter, M., Gaskin, D. J., Baranyi, J. & van Vliet, A. H. (2011). In vivo and in silico determination of essential genes of Campylobacter jejuni. *BMC Genomics* **12**, 535-2164-12-535.

Miller, S. C., LiPuma, J. J. & Parke, J. L. (2002). Culture-based and non-growth-dependent detection of the Burkholderia cepacia complex in soil environments. *Appl Environ Microbiol* **68**, 3750-3758.

Miller, V. L. & Mekalanos, J. J. (1988). A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires toxR. *J Bacteriol* **170**, 2575-2583.

Miller-Williams, M., Loewen, P. C. & Oresnik, I. J. (2006). Isolation of salt-sensitive mutants of *Sinorhizobium meliloti* strain Rm1021. *Microbiology* **152**, 2049-2059.

Milunovic, B., diCenzo, G. C., Morton, R. A. & Finan, T. M. (2014). Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of Sinorhizobium meliloti. *J Bacteriol* **196**, 811-824.

Minan, A., Bosch, A., Lasch, P., Stammler, M., Serra, D. O., Degrossi, J., Gatti, B., Vay, C., D'aquino, M. & other authors (2009). Rapid identification of Burkholderia cepacia complex species including strains of the novel Taxon K, recovered from cystic fibrosis patients by intact

cell MALDI-ToF mass spectrometry. *Analyst* **134**, 1138-1148.

Mizoguchi, H., Sawano, Y., Kato, J. & Mori, H. (2008). Superpositioning of deletions promotes growth of Escherichia coli with a reduced genome. *DNA Res* **15**, 277-284.

Moehring, R. W., Lewis, S. S., Isaacs, P. J., Schell, W. A., Thomann, W. R., Althaus, M. M., Hazen, K. C., Dicks, K. V., Lipuma, J. J. & other authors (2014). Outbreak of Bacteremia Due to *Burkholderia contaminans* Linked to Intravenous Fentanyl From an Institutional Compounding Pharmacy. *JAMA Intern Med* .

Mohamed, Y. F. & Valvano, M. A. (2014). A Burkholderia cenocepacia MurJ (MviN) homolog is essential for cell wall peptidoglycan synthesis and bacterial viability. *Glycobiology* **24**, 564-576.

Monahan, L. G., Hajduk, I. V., Blaber, S. P., Charles, I. G. & Harry, E. J. (2014). Coordinating bacterial cell division with nutrient availability: a role for glycolysis. *Mbio* **5**, e00935-14.

Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., Ozawa, T., Kodama, T., Kakeshita, H. & other authors (2008). Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res* **15**, 73-81.

Moule, M. G., Hemsley, C. M., Seet, Q., Guerra-Assuncao, J. A., Lim, J., Sarkar-Tyson, M., Clark, T. G., Tan, P. B., Titball, R. W. & other authors (2014). Genome-wide saturation mutagenesis of *Burkholderia pseudomallei* K96243 predicts essential genes and novel targets for antimicrobial development. *Mbio* **5**, e00926-13.

Murray, S., Charbeneau, J., Marshall, B. C. & LiPuma, J. J. (2008). Impact of burkholderia infection on lung transplantation in cystic fibrosis. *Am J Respir Crit Care Med* **178**, 363-371.

Nagata, T., Fukuda, R., Koike, I., Kogure, K. & Kirchman, D. L. (1998). Degradation by bacteria of membrane and soluble protein in seawater. *Aquat Microb Ecol* **14**, 29-37.

Nakashima, N., Goh, S., Good, L. & Tamura, T. (2012). Multiple-gene silencing using antisense RNAs in Escherichia coli. *Methods Mol Biol* **815**, 307-319.

Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K. M., Lee, K. J. & other authors (2011). Phenotypic landscape of a bacterial cell. *Cell* **144**, 143-156.

Norskov-Lauritsen, N., Johansen, H. K., Fenger, M. G., Nielsen, X. C., Pressler, T., Olesen, H. V. & Hoiby, N. (2010). Unusual distribution of Burkholderia cepacia complex species in Danish cystic fibrosis clinics may stem from restricted transmission between patients. *J Clin Microbiol* **48**, 2981-2983.

Nzula, S., Vandamme, P. & Govan, J. R. W. (2002). Influence of taxonomic status on the in vitro antimicrobial susceptibility of the Burkholderia cepacia complex. *J Antimicrob Chemother* **50**, 265-269.

Ohta, T. (2011). Near-neutrality, robustness, and epigenetics. *Genome Biol Evol* **3**, 1034-1038.

Olapade, O. A., Gao, X. & Leff, L. G. (2005). Abundance of three bacterial populations in selected streams. *Microb Ecol* **49**, 461-467.

Opijnen, T. & Camilli, A. (2010). Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr Protoc Microbiol* **19**, 1E.3.1-1E.3.16.

Orsburn, B. C., Melville, S. B. & Popham, D. L. (2010). EtfA catalyses the formation of dipicolinic acid in *Clostridium perfringens*. *Mol Microbiol* **75**, 178-186.

Ortega, X., Hunt, T. A., Loutet, S., Vinion-Dubiel, A. D., Datta, A., Choudhury, B., Goldberg, J. B., Carlson, R. & Valvano, M. A. (2005). Reconstitution of O-specific lipopolysaccharide expression in *Burkholderia cenocepacia* strain J2315, which is associated with transmissible infections in patients with cystic fibrosis. *J Bacteriol* **187**, 1324-1333.

Ortega, X. P., Cardona, S. T., Brown, A. R., Loutet, S. A., Flannagan, R. S., Campopiano, D. J., Govan, J. R. & Valvano, M. A. (2007). A putative gene cluster for aminoarabinose biosynthesis is essential for *Burkholderia cenocepacia* viability. *J Bacteriol* **189**, 3639-3644.

Pallud, C., Viallard, V., Balandreau, J., Normand, P. & Grundmann, G. (2001). Combined use of a specific probe and PCAT medium to study Burkholderia in soil. *J Microbiol Methods*

47, 25-34.

Parke, J. L. & Gurian-Sherman, D. (2001). Diversity of the Burkholderia cepacia complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol* **39**, 225-258.

Payne, G. W., Vandamme, P., Morgan, S. H., Lipuma, J. J., Coenye, T., Weightman, A. J., Jones, T. H. & Mahenthiralingam, E. (2005). Development of a recA gene-based identification approach for the entire Burkholderia genus. *Appl Environ Microbiol* **71**, 3917-3927.

Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40.

Pertea, M., Ayanbule, K., Smedinghoff, M. & Salzberg, S. L. (2009). OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Res* **37**, D479-82.

Peterson, A. E., Chitnis, A. S., Xiang, N., Scaletta, J. M., Geist, R., Schwartz, J., Dement, J., Lawlor, E., Lipuma, J. J. & other authors (2013). Clonally related *Burkholderia contaminans* among ventilated patients without cystic fibrosis. *Am J Infect Control* .

Philippe, H. & Forterre, P. (1999). The rooting of the universal tree of life is not reliable. *J Mol Evol* **49**, 509-523.

Philippot, L., Raaijmakers, J. M., Lemanceau, P. & van der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* **11**, 789-799.

Pirone, L., Chiarini, L., Dalmastri, C., Bevivino, A. & Tabacchioni, S. (2005). Detection of cultured and uncultured Burkholderia cepacia complex bacteria naturally occurring in the maize rhizosphere. *Environ Microbiol* **7**, 1734-1742.

Plaimas, K., Eils, R. & Konig, R. (2010). Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Syst Biol* **4**, 56-0509-4-56.

Portin, P. (2002). Historical development of the concept of the gene. *J Med Philos* **27**, 257-286.

Posfai, G., Plunkett, G.,3rd, Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S. & other authors (2006). Emergent properties of reduced-genome Escherichia coli. *Science* **312**, 1044-1046.

Prieto, A. I., Kahramanoglou, C., Ali, R. M., Fraser, G. M., Seshasayee, A. S. & Luscombe, N. M. (2011). Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Res* .

Projan, S. J. & Shlaes, D. M. (2004). Antibacterial drug discovery: is it all downhill from here? *Clin Microbiol Infect* **Suppl 4**, 18-22.

Pucci, M. J. (2006). Use of genomics to select antibacterial targets. *Biochem Pharmacol* **71**, 1066-1072.

Ramette, A., LiPuma, J. J. & Tiedje, J. M. (2005). Species abundance and diversity of Burkholderia cepacia complex in the environment. *Appl Environ Microbiol* **71**, 1193-1201.

Randell, S. H., Boucher, R. C. & University of North Carolina Virtual Lung Group (2006). Effective mucus clearance is essential for respiratory health. *Am J Respir Cell Mol Biol* **35**, 20-28.

Razeto-Barry, P., Diaz, J. & Vasquez, R. A. (2012). The nearly neutral and selection theories of molecular evolution under the fisher geometrical framework: substitution rate, population size, and complexity. *Genetics* **191**, 523-534.

Razvi, S., Quittell, L., Sewall, A., Quinton, H., Marshall, B. & Saiman, L. (2009). Respiratory microbiology of patients with cystic fibrosis in the United States, 1995 to 2005. *Chest* **136**, 1554-1560.

Reik R, FAU - Spilker, T., Spilker T, FAU - LiPuma, J. J. & LiPuma JJ (2005). Distribution of *Burkholderia cepacia* complex species among isolates recovered from persons with or without cystic fibrosis. *Journal of Clinical Microbiology* **43:6**, 2926-2928.

Remmele, C. W., Xian, Y., Albrecht, M., Faulstich, M., Fraunholz, M., Heinrichs, E., Dittrich, M. T., Muller, T., Reinhardt, R. & Rudel, T. (2014). Transcriptional landscape and essential genes of Neisseria gonorrhoeae. *Nucleic Acids Res* **42**, 10579-10595.

Renella, R., Perez, J. M., Chollet-Martin, S., Sarnacki, S., Fischer, A., Blanche, S., Casanova, J. L. & Picard, C. (2006). Burkholderia pseudomallei infection in chronic granulomatous disease. *Eur J Pediatr* **165**, 175-177.

Reznikoff, W. S. (2008). Transposon Tn5. *Annu Rev Genet* **42**, 269-286.

Reznikoff, W. S. & Winterberg, K. M. (2008). Transposon-based strategies for the identification of essential bacterial genes. *Methods Mol Biol* **416**, 13-26.

Righi, E., Girardis, M., Marchegiano, P., Venturelli, C., Tagliazucchi, S., Pecorari, M., Borsari, L., Carluccio, E., Codeluppi, M. & other authors (2013). Characteristics and outcome predictors of patients involved in an outbreak of Burkholderia cepacia complex. *J Hosp Infect* **85**, 73-75.

Roberts, D. L., Salazar, D., Fulmer, J. P., Frerman, F. E. & Kim, J. J. (1999). Crystal structure of *Paracoccus denitrificans* electron transfer flavoprotein: structural and electrostatic analysis of a conserved flavin binding domain. *Biochemistry* **38**, 1977-1989.

Rocha, E. P. & Danchin, A. (2004). An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol Biol Evol* **21**, 108-116.

Rocha, E. P. & Danchin, A. (2003a). Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res.* **31**, 6570-6577.

Rocha, E. P. & Danchin, A. (2003b). Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* **34**, 377-378.

Roggo, C., Coronado, E., Moreno-Forero, S. K., Harshman, K., Weber, J. & van der Meer, J. R. (2013). Genome-wide transposon insertion scanning of environmental survival functions in the polycyclic aromatic hydrocarbon degrading bacterium *Sphingomonas wittichii* RW1. *Environ Microbiol* **15**, 2681-2695.

Rossolini, G. M., Arena, F., Pecile, P. & Pollini, S. (2014). Update on the antibiotic resistance crisis. *Curr Opin Pharmacol* **18C**, 56-60.

Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z. & Higgins, N. P. (2012). Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genet* **8**, e1002845.

Rushton, L., Sass, A., Baldwin, A., Dowson, C. G., Donoghue, D. & Mahenthiralingam, E. (2013). A key role for efflux in the preservative susceptibility and adaptive resistance of *Burkholderia cepacia* complex bacteria. *Antimicrob Agents Chemother* **57**, 2972-2980.

Rusmini, R., Vecchietti, D., Macchi, R., Vidal-Aroca, F. & Bertoni, G. (2014). A shotgun antisense approach to the identification of novel essential genes in *Pseudomonas aeruginosa*. *BMC Microbiol* **14**, 24-2180-14-24.

Salama, N. R., Shepherd, B. & Falkow, S. (2004). Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. *J Bacteriol* **186**, 7926-7935.

Sambrook J. & Russell D. W. (2001). Molecular cloning: A laboratory manual, 3rd ed.. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sargent, M. G. (1975). Control of cell length in *Bacillus subtilis*. *J Bacteriol* **123**, 7-19.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84.

Schaffner, W., Reisig, G. & Verrall, R. A. (1973). Outbreak of *Pseudomonas cepacia* infection due to contaminated anaesthetics. *Lancet* **1**, 1050-1051.

Scott, J. D. & Ludwig, R. A. (2004). *Azorhizobium caulinodans* electron-transferring flavoprotein N electrochemically couples pyruvate dehydrogenase complex activity to N2 fixation. *Microbiology* **150**, 117-126.

Shevchenko, Y., Bouffard, G. G., Butterfield, Y. S., Blakesley, R. W., Hartley, J. L., Young, A. C., Marra, M. A., Jones, S. J., Touchman, J. W. & Green, E. D. (2002). Systematic sequencing of cDNA clones using the transposon Tn5. *Nucleic Acids Res* **30**, 2469-2477.

Silander, O. K. & Ackermann, M. (2009). The constancy of gene conservation across divergent bacterial orders. *BMC Res Notes* **2**, 2-0500-2-2.

Silver, L. L. (2011). Challenges of antibacterial discovery. *Clin Microbiol Rev* **24**, 71-109.

Simpson, I. N., Finlay, J., Winstanley, D. J., Dewhurst, N., Nelson, J. W., Butler, S. L. & Govan, J. R. (1994). Multi-resistance isolates possessing characteristics of both Burkholderia (Pseudomonas) cepacia and Burkholderia gladioli from patients with cystic fibrosis. *J Antimicrob Chemother* **34**, 353-361.

Sousa, S. A., Ramos, C. G. & Leitao, J. H. (2011). *Burkholderia cepacia* Complex: Emerging Multihost Pathogens Equipped with a Wide Range of Virulence Factors and Determinants. *Int J Microbiol* **2011**, 607575. Epub 2010 Aug 3.

Speert, D. P., Henry, D., Vandamme, P., Corey, M. & Mahenthiralingam, E. (2002). Epidemiology of *Burkholderia cepacia* complex in patients with cystic fibrosis, Canada. *Emerg Infect Dis* **8**, 181-187.

Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., Edwards, J.,Jr & Infectious Diseases Society of America (2008). The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin Infect Dis* **46**, 155-164.

Speller, D. C., Stephens, M. E. & Viant, A. C. (1971). Hospital infection by *Pseudomonas cepacia* . *Lancet* **1**, 798-799.

St John, A. C., Jakubas, K. & Beim, D. (1979). Degradation of proteins in steady-state cultures of *Escherichia coli*. *Biochim Biophys Acta* **586**, 537-544.

Stalker, D. M., Kolter, R. & Helinski, D. R. (1979). Nucleotide sequence of the region of an origin of replication of the antibiotic resistance plasmid R6K. *Proc Natl Acad Sci U S A* **76**, 1150-1154.

Stephenson, A. L., Tom, M., Berthiaume, Y., Singer, L. G., Aaron, S. D., Whitmore, G. A. & Stanojevic, S. (2014). A contemporary survival analysis of individuals with cystic fibrosis: a cohort study. *Eur Respir J* .

Sun, L., Jiang, R. Z., Steinbach, S., Holmes, A., Campanelli, C., Forstner, J., Sajjan, U., Tan, Y., Riley, M. & Goldstein, R. (1995). The emergence of a highly transmissible lineage of cbl+ Pseudomonas (Burkholderia) cepacia causing CF centre epidemics in North America and Britain [see comments]. *Nature Medicine* **1**, 661-6.

Tabacchioni, S., Visca, P., Chiarini, L., Bevivino, A., Di Serio, C., Fancelli, S. & Fani, R. (1995). Molecular characterization of rhizosphere and clinical isolates of Burkholderia cepacia. *Res Microbiol* **146**, 531-542.

Tanaka, K., Henry, C. S., Zinner, J. F., Jolivet, E., Cohoon, M. P., Xia, F., Bidnenko, V., Ehrlich, S. D., Stevens, R. L. & Noirot, P. (2013). Building the repertoire of dispensable chromosome regions in Bacillus subtilis entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res* **41**, 687-699.

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A. & Xie, X. S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L. & other authors (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.

Thanassi, J. A., Hartman-Neumann, S. L., Dougherty, T. J., Dougherty, B. A. & Pucci, M. J. (2002). Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res* **30**, 3152-3162.

Toogood, H. S., Leys, D. & Scrutton, N. S. (2007). Dynamics driving function: new insights from electron transferring flavoproteins and partner complexes. *Febs j* **274**, 5481-5504.

Tracy, B. S., Edwards, K. K. & Eisenstark, A. (2002). Carbon and nitrogen substrate utilization by archival *Salmonella typhimurium* LT2 cells. *BMC Evol Biol* **2**, 14-20.

Tsai, M. H. & Saier, M. H., Jr (1995). Phylogenetic characterization of the ubiquitous electron transfer flavoprotein families ETF-alpha and ETF-beta. *Res Microbiol* **146**, 397-404.

Tsuchiya, K., Homma, Y., Komoto, Y. & Suzui, T. (1995). Practical detection of *Pseudomonas cepacia* from rhizosphere antagonistic to plant pathogens with a combination of selective medium and ELISA. *Ann Phytopathol Soc Japan* **61**, 318-324.

Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L. & Whiteley, M. (2015). Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc Natl Acad Sci U S A* **112**, 4110-4115.

Ussery, D. W., Kiil, K., Lagesen, K., Sicheritz-Ponten, T., Bohlin, J. & Wassenaar, T. M. (2009). The genus burkholderia: analysis of 56 genomic sequences. *Genome Dyn* **6**, 140-157.

Valvano, M. A., Keith, K. E. & Cardona, S. T. (2005). Survival and persistence of opportunistic *Burkholderia* species in host cells. *Curr Opin Microbiol* **8**, 99-105.

Van, V. T., Berge, O., Ke, S. N., Balandreau, J. & Heulin, T. (2000). Repeated beneficial effects of rice inoculation with a strain of Burkholderia vietnamiensison early and late yield components in low fertility sulphate acid soils of Vietnam. *Plant Soil* **218**, 273-284.

Vandamme, P. & Dawyndt, P. (2011). Classification and identification of the *Burkholderia cepacia* complex: Past, present and future. *Syst Appl Microbiol* **34**, 87-95.

Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthiralingam, E., LiPuma, J. J. & Govan, J. R. (2003). *Burkholderia cenocepacia* sp. nov.--a new twist to an old story. *Res Microbiol* **154**, 91-96.

Vandamme, P., Holmes, B., Vancanneyt, M., Coenye, T., Hoste, B., Coopman, R., Revets, H., Lauwers, S., Gillis, M. & other authors (1997). Occurrence of multiple genomovars of *Burkholderia cepacia* in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. *Int J Syst Bacteriol* **47**, 1188-1200.

Vanlaere, E., Baldwin, A., Gevers, D., Henry, D., De Brandt, E., LiPuma, J. J., Mahenthiralingam, E., Speert, D. P., Dowson, C. & Vandamme, P. (2009). Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. *Int J Syst Evol Microbiol* **59**, 102-111.

Vardi, A., Sirigou, A., Lalayanni, C., Kachrimanidou, M., Kaloyannidis, P., Saloum, R., Anagnostopoulos, A. & Sotiropoulos, D. (2013). An outbreak of Burkholderia cepacia bacteremia in hospitalized hematology patients selectively affecting those with acute myeloid leukemia. *Am J Infect Control* **41**, 312-316.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E. & other authors (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.

Vermis, K., Brachkova, M., Vandamme, P. & Nelis, H. (2003). Isolation of Burkholderia cepacia complex genomovars from waters. *Syst Appl Microbiol* **26**, 595-600.

Vial, L., Chapalain, A., Groleau, M. C. & Deziel, E. (2011). The various lifestyles of the *Burkholderia cepacia* complex species: a tribute to adaptation. *Environ Microbiol* **13**, 1-12.

Vickers, C. E., Blank, L. M. & Kromer, J. O. (2010). Grand challenge commentary: Chassis cells for industrial biochemical production. *Nat Chem Biol* **6**, 875-877.

Wall, D. P., Hirsh, A. E., Fraser, H. B., Kumm, J., Giaever, G., Eisen, M. B. & Feldman, M. W. (2005). Functional genomic analysis of the rates of protein evolution. *Proc Natl Acad Sci U S A* **102**, 5483-5488.

Walt, A. & Kahn, M. L. (2002). The *fixA* and *fixB* genes are necessary for anaerobic carnitine reduction in *Escherichia coli*. *J Bacteriol* **184**, 4044-4047.

Wang, B. & Kuramitsu, H. K. (2003). Assessment of the utilization of the antisense RNA strategy to identify essential genes in heterologous bacteria. *FEMS Microbiol Lett* **220**, 171-176.

Wang, Z. & Zhang, J. (2009). Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet* **5**, e1000329.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D. & other authors (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**, D581-91.

Weart, R. B., Lee, A. H., Chien, A. C., Haeusser, D. P., Hill, N. S. & Levin, P. A. (2007). A metabolic sensor governing cell size in bacteria. *Cell* **130**, 335-347.

Wei, W., Zhang, T., Lin, D., Yang, Z. J. & Guo, F. B. (2013). Transcriptional abundance is not the single force driving the evolution of bacterial proteins. *BMC Evol Biol* **13**, 162-2148-13-162.

Weidenhaupt, M., Rossi, P., Beck, C., Fischer, H. M. & Hennecke, H. (1996). *Bradyrhizobium japonicum* possesses two discrete sets of electron transfer flavoprotein genes: *fixA, fixB* and *etfS, etfL*. *Arch Microbiol* **165**, 169-178.

Weidmann, A., Webb, A. K., Dodd, M. E. & Jones, A. M. (2008). Successful treatment of cepacia syndrome with combination nebulised and intravenous antibiotic therapy. *J Cyst Fibros* **7**, 409-411.

Weinberg, J. B., Alexander, B. D., Majure, J. M., Williams, L. W., Kim, J. Y., Vandamme, P. & LiPuma, J. J. (2007). Burkholderia glumae infection in an infant with chronic granulomatous disease. *J Clin Microbiol* **45**, 662-665.

Weinstein, A. J., Moellering, R. C.,Jr, Hopkins, C. C. & Goldblatt, A. (1973). *Pseudomonas cepacia* pneumonia. *Am J Med Sci* **265**, 491-494.

Wickstrum, J. R., Santangelo, T. J. & Egan, S. M. (2005). Cyclic AMP receptor protein and RhaR synergistically activate transcription from the L-rhamnose-responsive rhaSR promoter in Escherichia coli. *J Bacteriol* **187**, 6708-6718.

Wilson, A. C., Carlson, S. S. & White, T. J. (1977). Biochemical evolution. *Annu Rev Biochem* **46**, 573-639.

Wilson, D. N. (2011). On the specificity of antibiotics targeting the large ribosomal subunit. *Ann N Y Acad Sci* **1241**, 1-16.

Winkelstein, J. A., Marino, M. C., Johnston, R. B.,Jr, Boyle, J., Curnutte, J., Gallin, J. I., Malech, H. L., Holland, S. M., Ochs, H. & other authors (2000). Chronic granulomatous disease. Report on a national registry of 368 patients. *Medicine (Baltimore)* **79**, 155-169.

Winsor, G. L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M. D. & Brinkman, F. S. (2008). The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. *Bioinformatics* **24**, 2803-2804.

Winterberg, K. M., Luecke, J., Bruegl, A. S. & Reznikoff, W. S. (2005). Phenotypic screening of Escherichia coli K-12 Tn5 insertion libraries, using whole-genome oligonucleotide microarrays. *Appl Environ Microbiol* **71**, 451-459.

Wolcott, R., Costerton, J. W., Raoult, D. & Cutler, S. J. (2013). The polymicrobial nature of biofilm infection. *Clin Microbiol Infect* **19**, 107-112.

Wong, S. M. & Akerley, B. J. (2003). Inducible expression system and marker-linked mutagenesis approach for functional genomics of *Haemophilus influenzae*. *Gene* **316**, 177-186.

Xavier, J. C., Patil, K. R. & Rocha, I. (2014). Systems biology perspectives on minimal and simpler cells. *Microbiol Mol Biol Rev* **78**, 487-509.

Xu, H. H., Trawick, J. D., Haselbeck, R. J., Forsyth, R. A., Yamamoto, R. T., Archer, R., Patterson, J., Allen, M., Froelich, J. M. & other authors (2010). *Staphylococcus aureus* TargetArray: comprehensive differential essential gene expression as a mechanistic tool to profile antibacterials. *Antimicrob Agents Chemother* **54**, 3659-3670.

Xu, P., Ge, X., Chen, L., Wang, X., Dou, Y., Xu, J. Z., Patel, J. R., Stone, V., Trinh, M. & other authors (2011). Genome-wide essential gene identification in Streptococcus sanguinis. *Sci Rep* **1**, 125.

Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H. & Nishiuchi, Y. (1995). Transfer of two Burkholderia and an Alcaligenes species to *Ralstonia* gen. Nov.: Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum* (Smith 1896) comb. Nov. and *Ralstonia eutropha* (Davis 1969) comb. Nov. *Microbiol Immunol* **39**, 897-904.

Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., Ezaki, T. & Arakawa, M. (1992). Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol* **36**, 1251-1275.

Yi, H., Cho, Y. J., Won, S., Lee, J. E., Jin Yu, H., Kim, S., Schroth, G. P., Luo, S. & Chun, J. (2011). Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Res* **39**, e140.

Young, K. D. (2010). Bacterial shape: two-dimensional questions and possibilities. *Annu Rev Microbiol* **64**, 223-240.

Young, K., Jayasuriya, H., Ondeyka, J. G., Herath, K., Zhang, C., Kodali, S., Galgoci, A., Painter, R., Brown-Driver, V. & other authors (2006). Discovery of FabH/FabF inhibitors from natural products. *Antimicrob Agents Chemother* **50**, 519-526.

Zhang, G. Q., Cao, Z. W., Luo, Q. M., Cai, Y. D. & Li, Y. X. (2006). Operon prediction based on SVM. *Comput Biol Chem* **30**, 233-240.

Zhang, Y. J., Ioerger, T. R., Huttenhower, C., Long, J. E., Sassetti, C. M., Sacchettini, J. C. & Rubin, E. J. (2012). Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. *PLoS Pathog* **8**, e1002946.

Zhou, J., Chen, Y., Tabibi, S., Alba, L., Garber, E. & Saiman, L. (2007). Antimicrobial susceptibility and synergy studies of *Burkholderia cepacia* complex isolated from patients with cystic fibrosis. *Antimicrob Agents Chemother* **51**, 1085-1088.