The Dissertation Committee for Peter James Enyeart Certifies that this is the approved version of the following dissertation:

STUDIES IN BACTERIAL GENOME ENGINEERING AND ITS APPLICATIONS

Committee:

Andrew D. Ellington, Supervisor

Hal Alper

Makkuni Jayaram

Alan M. Lambowitz

Edward M. Marcotte

STUDIES IN BACTERIAL GENOME ENGINEERING AND ITS APPLICATIONS

by

Peter James Enyeart, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2014

Dedication

To my family

Acknowledgements

It takes a village to raise a scientist, and I've certainly not been an exception. First I must of course acknowledge my advisor, Andy Ellington, who took a bit of a chance in letting me into the lab, and who has always been very generous in allowing me to follow my own interests while occasionally providing gentle (and not so gentle) redirection when I needed it. Andy is always good for ideas and advice, and despite his reputation for ego, he's always very quick to assign credit for success to his students. He wouldn't be the right advisor for everyone, but he was the right advisor for me.

My committee of Hal Alper, Makkuni Jayaram, Alan Lambowitz, and Edward Marcotte has also been invaluable in helping me shape my experiments into doable chunks. Professors Lambowitz and Marcotte in particular have been active collaborators on some of the projects described here, and I wouldn't have been able to do it without them.

I also owe a great debt to the members of the Ellington lab. When I started in the lab I had essentially no idea what I was doing and had to rely heavily on senior lab members to learn how to perform even basic protocols. In particular, Randy Hughes, Eric Davidson, Paulina Dlugosz, Xi Chen, Gwen Stovall, Amrita Singh, Na Li, Angel Syrett, and Brad Hall all took significant chunks of time out from their research to help me figure out how to do mine. Particular thanks are due to Jiri Perutka, who taught me the art of targetrons, designed many of targetrons used herein, and helped me find a productive research direction at a time when I was struggling. Even as I became more confident in my experimental abilities, I still frequently found myself relying on my lab mates and the administrative staff when issues came up, technical and otherwise. I'd like

to further thank Peter Allen, Sanchita Bhadra, Michelle Byrom, Tiffany Carmichael, Sarah Caton, Vlad Codrea, Meredith Corley, Yan Du, Jared Ellefson, Daniel Garry, Jimmy Gollihar, Alexandra Hubbard, Stephanie Huntzis Achebe, Tony Hwang, Sherry Jiang, Christien Kluwe, Michael Ledbetter, Bingling Li, Chunhua Liu, Wayne Lu, Oana Lungu, Andre Maranhao, Adam Meyer, Alex Miklos, John Milligan, Dorothy Podgornoff, Arti Pothukuchy, Christian Ramirez, Ashley Rasmussen, Caitlin Sanford, Sara Stewart, Drew Tack, Ross Thyer, and Jorge Villafana, for all their help and support. Again particular thanks are due to Steven Chirieleison, who did a ridiculous amount of work for the recombination-mediated cassette exchanges in the targetron/*lox* paper, described in Chapter 3.

I also received a lot of help from others on specific projects. Savitri Mandapati gave a great deal of help in developing targetrons to carry *Ter* sites, and the project would never have happened without Patrick Schaeffer and Morgane Moreau at James Cook University. Jun Yao and Travis Whitt were kind enough to let me use their *B. subtilis* introns, and Adrian Keatinge-Clay was indispensable in dealing with the polyketide synthases. Erik Quandt let me use the rDNA-targeting intron he developed, and also taught me the dark art of desalting, which turned out to be invaluable. Jeff Barrick helped me design the transposon selections, and Scott Hunicke-Smith, Heather Deiderick, Jessica Podnar, Daniel Deatherage, Craig Barnhart, Dhivya Arasappan, and Taejoon Kwon were invaluable in helping me figure out how to sequence them and what to do with the sequences once I had them. I'd also like to thank Martin Krzywinski of Canada's Michael Smith Genome Sciences Centre for creating the Circos visualization package and being readily available to answer questions about it. In my work in extracellular electron transfer, Eun Jeong Cho taught me the basics of bio-electrochemical measurements, Keith Stevenson and Ben Hahn did their best to teach me more, and

Caroline Ajo-Franklin and Heather Jensen at Lawrence Berkeley Labs were kind enough to help me figure out some alternate research possibilities after they scooped me. Jagannath Swaminathan helped out in determining the operon network for MtrCAB. For the bactonomics project, I'd like to thank Zack Simpson for giving me the idea and helping me troubleshoot, and Jonathan Dingwell for teaching me non-linear dynamics and Matlab. Finally, thanks to Josh Russell for general moral support.

Additionally, I'd like to thank CMB Graduate Coordinators, Barbara Welch and Jolie Cota Flink, for helping me navigate all the bureaucratic hoops a graduate student must leap through.

And of course I'd be nowhere without my family. My parents, Tom and Sandy, nurtured my interest in science from a young age and always supported me in whatever I wanted to do. My brother John and my sister Rebecca have always been there for me. Aunt Donna was ever ready to lend a sympathetic ear as someone who'd been there. Thank you.

And then Mai Dao is in a class by herself.

STUDIES IN BACTERIAL GENOME ENGINEERING AND ITS APPLICATIONS

Peter James Enyeart, PhD.

The University of Texas at Austin, 2014

Supervisor: Andrew D. Ellington

Many different approaches exist for engineering bacterial genomes. The most common current methods include transposons for random mutagenesis, recombineering for specific modifications in *Escherichia coli*, and targetrons for targeted knock-outs. Site-specific recombinases, which can catalyze a variety of large modifications at high efficiency, have been relatively underutilized in bacteria. Employing these technologies in combination could significantly expand and empower the toolkit available for modifying bacteria.

Targetrons can be adapted to carry functional genetic elements to defined genomic loci. For instance, we re-engineered targetrons to deliver *lox* sites, the recognition target of the site-specific recombinase, Cre. We used this system on the *E. coli* genome to delete over 100 kilobases, invert over 1 megabase, insert a 12-kilobase polyketide-synthase operon, and translocate a 100 kilobase section to another site over 1 megabase away. We further used it to delete a 15-kilobase pathogenicity island from *Staphylococcus aureus*, catalyze an inversion of over 1 megabase in *Bacillus subtilis*, and simultaneously deliver nine *lox* sites to the genome of *Shewanella oneidensis*. This represents a powerful, versatile, and broad-host-range solution for bacterial genome engineering.

We also placed *lox* sites on *mariner* transposons, which we leveraged to create libraries of millions of strains harboring rearranged genomes. The resulting data represents the most thorough search of the space of potential genomic rearrangements to date. While simple insertions were often most adaptive, the most successful modification found was an inversion that significantly improved fitness in minimal media. This approach could be pushed further to examine swapping or cutting and pasting regions of the genome, as well.

As potential applications, we present work towards implementing and optimizing extracellular electron transfer in *E. coli*, as well as mathematical models of bacteria engineered to adhere to the principles of the economic concept of comparative advantage, which indicate that the approach is feasible, and furthermore indicate that economic cooperation is favored under more adverse conditions. Extracellular electron transfer has applications in bioenergy and biomechanical interfaces, while synthetic microbial economics has applications in designing consortia-based industrial bioprocesses. The genomic engineering methods presented above could be used to implement and optimize these systems.

Table of Contents

List of Tables

List of Figures

Chapter 1: Background

1.1 Introduction

Bacteria have long played important roles in human food and health, and in recent decades genetically modified bacteria have become useful for the industrial synthesis of a large variety of proteins and other useful chemical compounds, among other applications (Glazer & Nikaido, 2007; Lee, 2006). There is hope that in the near future designed bacteria will play a large role in transitioning from an economy powered by fossil fuels toward one based on renewable resources, in addition to helping solve other problems that humanity faces (Arora, 2012; Keasling, 2008; Khalil & Collins, 2010; Savage et al, 2008; Stephanopoulos, 2008). Thus there is great potential utility in developing better methods for reprogramming bacteria to perform human-directed tasks. Plasmids have been the method of choice for modifying bacteria in recent decades, since they are relatively easy to manipulate. The main drawback of plasmids is that the amount of DNA that can be put on a single plasmid is limited. For making global changes in the biosynthetic machinery of bacteria or for adding large amounts of foreign DNA to perform a new function, as is becoming increasingly necessary as the scope and ambition of microbial biotechnology continues to expand, genome-scale techniques are needed. Following is a survey of different techniques for engineering bacterial genomes, presented approximately in the order of their development.

1.2 Serial culture

Humans have been unconsciously engineering bacterial genomes for millennia. In addition to *Saccharomyces cerevisiae* and other fungi, domesticated strains of bacteria have long been used to make such fermented food products as yogurt, cheese, vinegar, Japanese *natto* (fermented soybeans), Korean *kimchi* (fermented cabbage), and a number

of other traditional foods and beverages. The origins of most of these foods are mysterious, but vinegar, which typically results from bacteria of the genus *Acetobacter* and *Gluconacetobacter* fermenting ethanol to acetic acid (Gullo et al, 2006; Kittelmann et al, 1989; Nakayama, 1959; Sievers et al, 1992; Sokollek et al, 1998), is likely the oldest, having in all likelihood arrived immediately after the development of alcohol (Adams, 1985). By continuously picking particularly delicious batches of fermented foods to use as starter cultures for further fermentations, ancient chefs in effect used the laboratory technique of serial growth and dilution to gradually design bacteria that were more suitable to their needs. The evolution of the bacteria inhabiting the human digestive tract can be considered along similar lines (Walter & Ley, 2011).

In addition to *Acetobacter*, some well-studied examples of bacteria use to produce fermented food products include *Lactococcus lactis*, which is used to make cheese and other fermented milk products, *Lactobacillus delbrueckii* subsp. *bulgaricus*, which is widely used in making yogurt, and *Bacillus natto*, which is used to make its namesake food in Japan. The genomes of food-isolates of *L. lactis* and *L. delbrueckii* subsp. *bulgaricus* indicate size reductions and loss of functions, consistent with an evolutionary transition from life in complex environments to life in the relatively simple environment of milk (Kelly et al, 2010; van de Guchte et al, 2006). Others, such as *Acetobacter pasteurianus*, have been found to be adapted for extremely rapid evolution. The genome of *A. pasteurianus* contains 280 transposons (nearly 10% of the genome) and five genes with hyper-mutable tandem repeats (Azuma et al, 2009).) *B. natto* is a variant of *Bacillus subtilis* that secretes 15-20 times more proteases than standard *B. subtilis* strains (Nagami & Tanaka, 1986; Uehara et al, 1974), including a fibrinolytic enzyme dubbed "nattokinase" (Sumi et al, 1987) that has been investigated for medical applications (Hsia et al, 2009; Kim et al, 2008; Pais et al, 2006). Other unique species isolated from

fermented foods include *Gluconacetobacter kombuchae* (Dutta & Gachhui, 2007) and *Lactobacillus kimchii* (Yoon et al, 2000). Such strains can be considered as examples of initial efforts by humans to engineer bacterial genomes.

These traditional approaches tend to be avoided in modern biotechnological settings due to their slow and unpredictable nature, but they may still be employed in industry to avoid a "GMO" label and are continuously used in academic settings as a model of natural evolution. The most well-known example of the latter is the *E. coli* long-term evolution experiment run by the Lenski lab at Michigan State University. In 1988, twelve separate inoculations of *Escherichia coli* were made into minimal media, and every day since then 1% of the previously grown population has been reinoculated into fresh media, with frozen stocks being made every 500 generations. Adaptation to the new environment was rapid during the first 2000 generations and then began to slow (Lenski et al, 1991). Differences in morphology and fitness between the different lineages were established by generation 10,000 (Lenski & Travisano, 1994), though even at 20,000 generations, mutations and changes in gene-expression profiles were found to be quite similar between the different lineages (Cooper et al, 2003; Pelosi et al, 2006; Woods et al, 2006). Despite the slowing of adaptation, changes continued to occur in the genomes, and at least one of the lineages developed a mutator phenotype between generations 20,000 and 30,000 (Barrick et al, 2009), which was later attenuated (Wielgoss et al, 2013). Perhaps most surprisingly, around generation 31,000 one of the lineages evolved the ability to consume the citrate in the media, and had taken over the population by generation 33,000 (Blount et al, 2012; Blount et al, 2008). The experiment has now covered more than 50,000 generations of bacteria (Lenski, 2011) and has been extremely valuable in testing hypotheses about evolution. Thus these age-old methods of genome engineering still play an important role in the cutting-edge science of today.

1.3 MUTAGENS

Though Louis Pasteur demonstrated that microorganisms were responsible for fermentation in the 19th century (Pasteur, 1866; Pasteur, 1876), there were still no known genetic-engineering technologies other than the traditional method of applying a selective pressure and waiting for new mutations to be fixed in the population. This began to change in the 1920s, when Hermann Muller showed that X-rays cause heritable genetic changes in fruit flies (Muller, 1927; Muller, 1928). The field of bacterial genetics took some time to catch up, however, and for many years thereafter X-rays were more frequently used for killing bacterial cells than for inducing mutations (Lincoln & Gowen, 1942; Wyckoff, 1930), though some changes in cell morphology were noted following irradiation (Haberman & Ellsworth, 1940). Gray and Tatum finally succeeded in using X-rays to generate *E. coli* auxotrophs in the mid-1940s (Gray & Tatum, 1944; Tatum, 1945).

Mustard gas was the first demonstrated chemical mutagen in 1947 (Auerbach et al, 1947). The ease of use of chemical mutagens as opposed to ionizing radiation contributed significantly to the subsequent flowering of bacterial genetics, and many key studies made use of them. For instance, the fact that constitutive expression mutants are easily generated in many inducible systems was a key clue in helping Jacob and Monod identify the regulatory motif of gene repression (Jacob & Monod, 1961). The use of chemical mutagens to produce temperature-sensitive mutations in essential genes was important for identifying the genes involved in DNA replication (Carl, 1970; Fangman & Novick, 1968; Hirota et al, 1970; Wechsler & Gross, 1971). Many other examples of the use of mutagens to generate knock-outs or other mutations could be given.

While mutagens were certainly useful for the study of bacteria, bacteria have been important in the study of mutagens, as well. One of the most frequently used assays for

determining the mutagenicity of a given compound is the Ames test, which involves adding the compound to a culture of *Salmonella* bacteria auxotrophic for histidine and assessing the rate at which the ability to synthesize histidine is recovered (Ames et al, 1973b; Mortelmans & Zeiger, 2000). Some of the findings resulting from the use of this test include the fact that many carcinogens are also mutagens, and thus that mutation can cause cancer (Ames et al, 1973a).

Nowadays mutagens are not frequently used as a genomic-engineering tool when other techniques are available, because they can be dangerous to handle, and off-target mutations are common. However, they are still useful for quickly generating strains with genomic markers. For instance, bacterial cells resistant to streptomycin can be easily obtained after overnight growth in the presence of a weak mutagen such as 2-aminopurine (Miller, 1992).

1.4 F PLASMID

Lederberg and Tatum, in studying the mutant strains of *E. coli* they had made using X-rays and ultraviolet light, found that certain strains could swap mutations merely by being cultured together (Lederberg & Tatum, 1946a; Lederberg & Tatum, 1946b). This was the discovery of conjugation, by which bacteria can temporarily fuse their membranes and trade genetic material. The genetic element responsible in this particular case was later identified as the F ("fertility") factor (Cavalli et al, 1953; Hayes, 1952), which was the first identified plasmid. Strains that harbor the F factor are called F+, and those without are F-. The F plasmid contains machinery for transferring itself from F+ strains to other strains. The ability of the F plasmid to facilitate inter-genome recombination comes from the fact that it occasionally integrates into a random location in the genome of its host (strains harboring such an integration are called "Hfr" (high-

frequency recombination) strains). Upon mating, the integrated F plasmid will then begin transferring the entire genome of its host into the partner cell. A depiction of the mechanism by which this occurs is shown in **Figure 1.1**.

A key discovery was when it was found that the transfer of genetic material (mating) between Hfr strains can be interrupted, and that by timing how long the mating must be for a given pair of mutations to both be transferred, the relative distance between the two can be assessed (Wollman et al, 1956). This led to the ability to map the locations of the genes on the chromosome, and even today locations in the *E. coli* genome are sometimes referred to in terms of "minutes" as a result of this practice. By 1964 over one hundred E. coli genes had been mapped, almost exclusively by creating mutations as described in **Section 1.3** and then locating those mutations using this method (Taylor & Thoman, 1964).

Though the F plasmid has been employed less frequently as a genetic tool since the advent of large-scale sequencing technology and the publication of the *E. coli* genome sequence (Blattner et al, 1997), in recent years interest has resurged in using it to collect desired mutations together for planned and combinatorial genomic recoding. For instance, in the recently reported creation of an *E. coli* strain completely lacking the UAG stop codon, the F plasmid was used in combination with positive and negative selectable markers to combine the modified sections from partially recoded genomes into larger recoded segments, culminating in the creation of a completely recoded genome (Isaacs et al, 2011; Lajoie et al, 2013). Winkler and Kao have also demonstrated the utility of the F plasmid for speeding the rate of evolution of genetically diverse libraries of *E. coli*, (Winkler & Kao, 2012). Meanwhile, Quandt and coworkers recently performed almost the reverse operation in using the F plasmid to parcel out the mutations acquired by the citrate-utilizing strain generated in the long-term evolution experiment (see **Section 1.2**)

in order to determine exactly which are necessary for efficient citrate consumption (Quandt et al, 2014). Automated approaches to generating and mating Hfr libraries have also been developed in recent years (Typas et al, 2008).

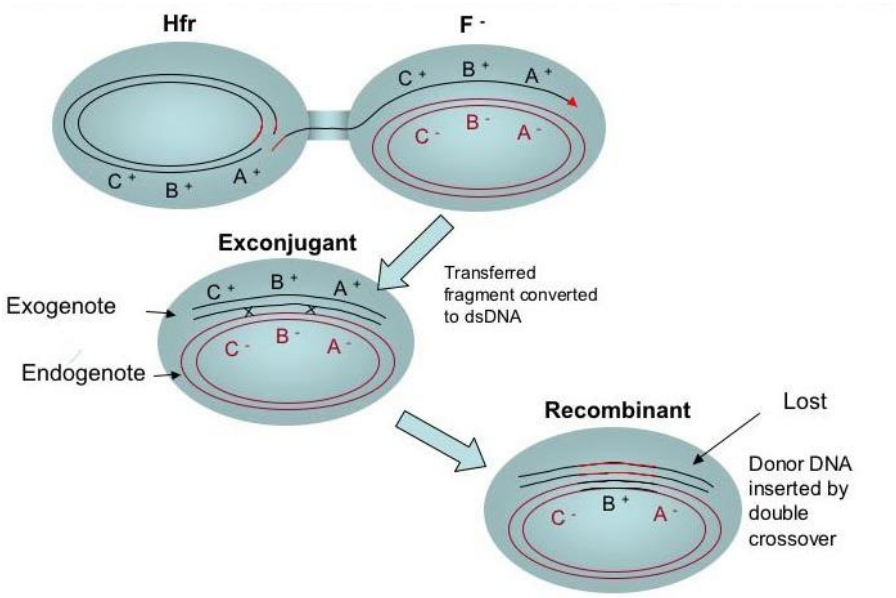

Figure 1.1. Mechanism of F plasmid transfer and subsequent recombination by Hfr strains.

(Image from http://www.bio.miami.edu/dana/pix/Hfr_transfer.jpg.)

For strains other than *E. coli*, the IncP plasmids, including RK2 and RP4, can function similarly to the F plasmid and work well in most Gram-negative strains (Thomas & Smith, 1987). Such plasmids can also be used to facilitate inter-species transfer of DNA.

1.5 PHAGE

Credit for the discovery that viruses also infect bacteria is typically given to Frederick Twort, who first reported observations of phenomena we now understand to be caused by such viruses (Twort, 1915), and Félix d'Hérelle, who was the first to

definitively identify these phenomena as caused by an infective agent dubbed "bacteriophage" (d'Herelle, 1917). (See Duckworth (1976) for a discussion of the issues surrounding assignment of credit for this discovery.) Phage subsequently played a key role in a number of important experiments, such as the work by Hershey and Chase that demonstrated that DNA was the carrier of genetic information (Hershey & Chase, 1952).

The utility of phage for engineering bacterial genomes first became clear in 1951, when Lederberg and Zinder carried out a search for a *Salmonella* equivalent of the F plasmid discussed in **Section 1.4**, but instead discovered that certain phage could assume a similar role (Lederberg et al, 1951; Zinder & Lederberg, 1952). They dubbed this process "transduction," and the particular version of it they observed is now called generalized transduction. During the process of replicating themselves, generalized transducing phage fragment the genome of the host bacterium, and these genomic fragments are occasionally packaged into phage capsids. When the newly produced phage leave to infect other cells, a fraction of the new hosts will receive the fragment of genomic DNA from the previous hosts, which can be recombined into the new genome (Miller, 1992).

The size of the packaged fragments corresponds to about 2 minutes of the *E. coli* genome, and thus phage proved to be a welcome complement to the F factor for mapping chromosomes in finer detail (Miller, 1992; Neidhardt & Curtiss, 1996; Taylor & Thoman, 1964). Like the F factor, generalized transducing phage are used less often now in the era of high-throughput genome sequencing, but are occasionally employed to transfer desired chromosomal mutations between strains. For instance, Yu and coworkers used such a phage to shuttle transposon-delivered *lox* sites (see **Sections 1.6** and **1.8**) between strains to generate planned chromosomal deletions, as well as to combine deletions from different strains (Yu et al, 2002). The recent development of automated approaches

(Donath et al, 2011) may cause a resurgence in the use of this technique, as for the F plasmid.

The λ phage was discovered by Esther Lederberg in 1950 (Lederberg, 1950; Lederberg, 1951) and became extremely important to the study of both virology and molecular genetics (Ptashne, 2004). After the discovery of generalized transduction in *Salmonella*, M. Laurance Morse and the Lederbergs discovered that λ is also capable of transferring genetic material between strains of *E. coli*. However, it was always fragments from the same part of the genome that were transferred, and thus this became known as specialized transduction (Morse et al, 1956). This results from the fact that λ phage integrates into a specific site in the host genome and occasionally takes DNA from that site with it when it excises itself. The fact that so-called lysogenic phage insert themselves into the genome at a defined site without harming the host has led to development of such phage as cloning vectors for delivering DNA to the chromosome. These can carry 5 to 11 kilobases of DNA to the genome, with more possible if the genes for lysogeny are replaced and the phage is used as an extra-chromosomal vector (Preston, 2003; Short et al, 1988).

A number of enzymes originally isolated from bacteriophage have also proven to be extremely useful for genetic manipulations, which are discussed in **Sections 1.8** and **1.10** below.

1.6 TRANSPOSONS

Barbara McClintock was the first to notice that certain genes are capable of moving (McClintock, 1950), but the scientific community was slow in accepting this discovery. Genetic studies in *Escherichia coli* came to the rescue in the late 1960s, when a new class of strong polar mutations (mutations that affect the expression of a gene other

than the one in which the mutation is found) were encountered in well-studied operons such as *lac* and *gal* (Jordan et al, 1967; Malamy, 1966). These mutations were different from previous polar mutations, which had been found to be the result of single-base-pair changes causing frame shifts or premature stop codons, in that they were both stronger than such mutations and failed to revert back to wild-type upon addition of chemical mutagens or nonsense suppressor mutations. The cause of these new mutations turned out to be the insertion of extra DNA from elsewhere (Jordan et al, 1968; Malamy, 1970). These so-called "insertion elements" (later more commonly called "transposable elements" or "transposons," though these latter terms are sometimes differentiated from the term "insertion element" as containing identifiable genes) soon became the subject of intense study and were found to be ubiquitous in nearly all forms of life (Calos & Miller, 1980). It has been found, for instance, that transposon genes account for at least half of the genomic DNA of many plant species (Li et al, 2004; SanMiguel et al, 1996), as well as approximately half of the human genome sequence (Lander et al, 2001).

The basic transposon structure consists of a pair of inverted repeats flanking DNA of variable content, often but not always including a gene for the transposase (the enzyme that catalyzes the movement of the transposon) as well as other genes, such as antibiotic resistance, that may provide a benefit to the host and aid in the genetic maintenance of the transposon (Calos & Miller, 1980). In fact, many of the antibiotic resistance markers in use today were originally isolated from transposons. The transposon/transposase complex recognizes a sequence of DNA with varying degrees of specificity as a site for integration. Since many of these sequences are either quite short or not stringently defined, a typical genome contains numerous potential insertion sites, and many transposons can be considered to insert essentially randomly (Calos & Miller, 1980). Transposition in bacteria may involve a copy mechanism, where the transposon remains

in both the old and new loci (Shapiro, 1979), or a cut-and-paste mechanism, where the transposon is removed from the previous locus and reinserted at a new one (Bender & Kleckner, 1986).

The utility of transposons as a genetic tool was quickly realized (Kleckner et al, 1977). Compared to the mutagens discussed in **Section 1.3**, transposons are safer to handle, and have strong and comparably consistent effects as a result of interrupting rather than simply changing DNA sequences. Further, the use of transposons carrying antibiotic-resistance genes makes it much easier to maintain deleterious mutations through antibiotic selection, and in mapping studies, marker-carrying transposons allowed loci to be mapped that otherwise have phenotypes that are difficult or impossible to observe. It is also typically not difficult to find a transposon that will function in any given species (one example of an exceptionally broad-host-range class of transposon, the *mariner* type, is discussed in more detail below).

Due to these advantages, transposons remain a very widely used tool for random mutagenesis in bacteria even today, and developments and improvements in transposon technology have been ongoing. Far more work has been done than can be discussed here, but some examples follow. For instance, transposons have been widely used in studies to determine essential genes and potentially define a minimal genome (Akerley et al, 1998; Gerdes et al, 2003; Glass et al, 2006; Goryshin et al, 2003; Hare et al, 2001; Judson & Mekalanos, 2000; Sassetti et al, 2001). Transposon mutagenesis has also been used to determine "conditional essentiality" in, for instance, determining genes required in minimal versus rich media (Badarinarayana et al, 2001), or genes not needed in culture media but essential for infection (Hensel et al, 1995) or survival in the digestive tract (Goodman et al, 2009). Transposons have naturally also been used to search for genes required for specific processes, such as sporulation in *Bacillus subtilis* (Jaacks et al, 1989;

Youngman et al, 1983) or extracellular electron transport in *Shewanella oneidensis* (Beliaev & Saffarini, 1998; Bouhenni et al, 2005), and they have even been used to mutagenize individual genes that code for catalytically active RNAs in order to determine important structural determinants (Belhocine et al, 2008). For biotechnological purposes, transposons have also been used to determine genes useful for increasing the metabolic yield of desired chemicals (Alper et al, 2005; Elischewski et al, 1999; Smith & Liao, 2011). Systems such as "transposomes" have also been developed to allow transposon/transposases complexes to be formed *in vitro*, stored indefinitely, and then used to induce mutagenesis upon electroporation without the need for expressing the transposase gene in the target organism (Goryshin et al, 2000).

The utility of transposons has also been increased by using them to deliver other genetic elements to the chromosome. For instance, transposons can deliver genes encoding reporter proteins such as LacZ or GFP to allow transcriptional rates at the site of insertion to be monitored or simply to tag modified cells (Casadaban & Cohen, 1979; Lambertsen et al, 2004; Youngman et al, 1984). Protein-purification tags can be delivered on transposons to allow easy capture of gene products identified in mutagenesis screens (Chiang & Rubin, 2002). Inducible promoters have been added in order to search for cryptic genes (Bordi et al, 2008; Salipante et al, 2003) or aid in subsequent identification of the site of insertion (Badarinarayana et al, 2001). Yu and coworkers used transposons to deliver site-specific recombinase recognition sites (see **Section 1.8**) to the genome in order to create targeted deletions (Yu et al, 2002), and in a similar vein, Goryshin and coworkers used a nested transposon structure, in which the entire chromosome is essentially treated as a transposon that inserts into itself, in order create deletions and inversions (Goryshin et al, 2003).

Most of the transposons employed in the work cited above, such as the Tn5, Tn7, Tn10, and Mu transposons, have been in use since the 1970s, but the *mariner* class of transposons, which was discovered relatively recently, merits special mention. The *mariner* transposons were originally discovered in *Drosophila* (Jacobson et al, 1986). The variant used in bacteria is *Himar1*, originally isolated from the fly, *Haematobia irritans*, and was shown to function *in vitro* in 1996 (Lampe et al, 1996) and in bacteria in 1999 (Rubin et al, 1999). The particular utility of *mariner* transposons comes from the fact that they work very efficiently in a wide variety of organisms, both eukaryotic and prokaryotic; they have very little insertion bias other than the recognition site, "TA"(Lampe et al, 1996); they do not rely on host factors for transposition and are active *in vitro* (Lampe et al, 1996); and they are more efficient than most other available transposons. In the case of *Himar1*, a hyperactive transposase mutant was developed that inserts at 50 times the efficiency of the wild-type (Lampe et al, 1999). The frequency of transposition is on the order of 3×10^{-3} cells that harbor the transposon in a given generation (Chiang & Rubin, 2002), whereas the more traditional transposons are typically in the range of 10^{-4} to 10^{-7} (Kleckner, 1981; Kleckner et al, 1977).

Finally, some mention should be made of the methods by which the location of a transposon insertion can be determined. The mapping approaches discussed in **Sections 1.3** and **1.4** can of course be employed, but these are both laborious and imprecise by modern standards. Most approaches currently in use make use of modified polymerase chain reaction (PCR) techniques. In inverse PCR (Ochman et al, 1988), for instance, genomic DNA containing transposon insertions is digested with a restriction enzyme or otherwise fragmented, and the pieces are ligated into circular DNA molecules. PCR is then performed using primers that anneal to transposon sequence and face outward toward the genome, resulting in amplification of the ligated genomic DNA sequences

outside. The amplified fragments can then be sequenced to identify the genomic regions flanking the transposon insertions.

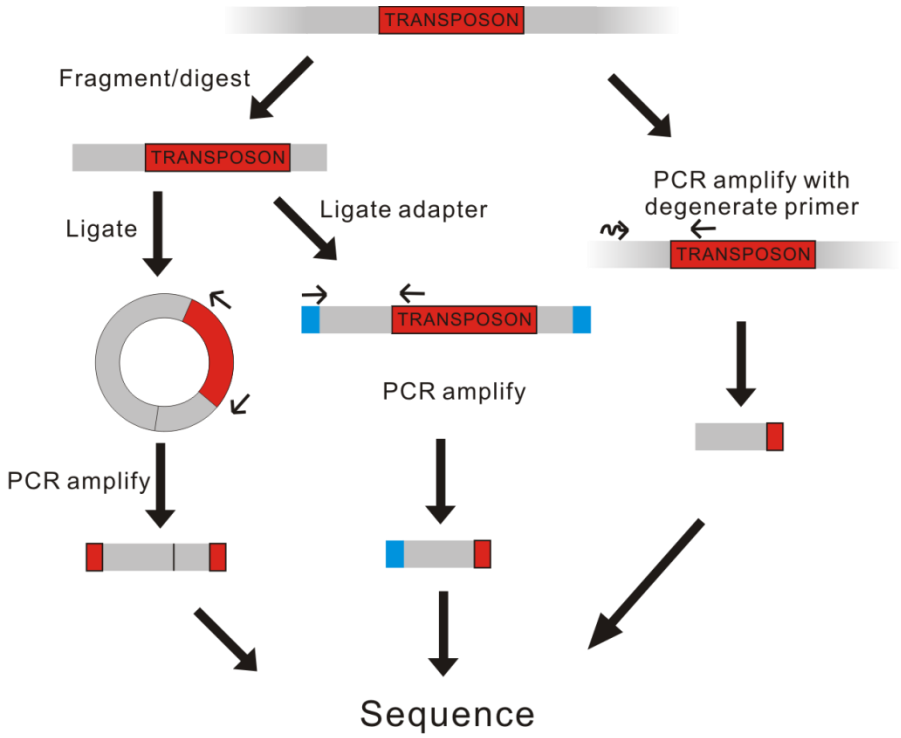

Figure 1.2. Outline of methods for determining transposon integration sites.

PCR can also be performed directly on unmodified genomic DNA using one primer that binds to the transposon and a second set of non-specific primers that bind randomly elsewhere in the genome (Liu & Whittier, 1995). This approach may require alternating between different annealing temperatures to favor specific versus non-specific priming, and is typically done in stages employing different transposon-specific primers in order to limit the amount of non-specific product. A method that can be readily employed in tandem with library preparation in modern deep sequencing approaches involves using one primer that binds to the transposon and another that anneals to an adaptor ligated to the fragmented end (Gallagher et al, 2011; Gawronski et al, 2009;

Goodman et al, 2009; Langridge et al, 2009; van Opijnen et al, 2009). Such approaches may also include affinity purification using biotinylated primers that anneal to transposon sequence (Gawronski et al, 2009) and the addition of recognition sites of the MmeI restriction enzyme, which cuts approximately 20 bp away from its recognition site, on the ends of the transposon (Goodman et al, 2009; van Opijnen et al, 2009), to achieve uniform fragment sizes. Schematics of these approaches are shown in **Figure 1.2**.

1.7 HOMOLOGOUS RECOMBINATION WITH SUICIDE PLASMIDS

Though the F factor discussed in **Section 1.4** was the first plasmid to be discovered, a number of further developments were needed before the true potential of the plasmid as a genetic tool could be realized. The first was the discovery of the proliferation of the so-called "R factors" in Japan (Falkow et al, 1966; Watanabe & Fukasawa, 1961). These were plasmids that provided antibiotic resistance and spread quickly due to the increased use of antibiotics in the post-war period. A great many of the plasmids commonly used in the lab today are derived from these R factors. Second was the identification, isolation, and application of enzymes such as restriction endonucleases (Kelly & Smith, 1970; Smith & Wilcox, 1970) and DNA ligases (Weiss & Richardson, 1967) from microbes and phage that could be used to site-specifically cut and re-ligate DNA. This led in the 1970s to the first useful methods for editing DNA in a predictable and reproducible manner (Cohen et al, 1973; Jackson et al, 1972; Maniatis et al, 1982), giving birth to the biotechnology industry. Molecular cloning techniques and plasmid biotechnology have loomed large over the fields of biology and medicine ever since.

A complete discussion of plasmid biotechnology is beyond the scope of this work, but following is a discussion of the ways in which plasmids have been employed to

modify genomic DNA. Besides serving as vectors for delivering the enzymes and nucleic acids used in the other approaches discussed herein, DNA carried on plasmids can be directly inserted into the genome via homologous recombination, in much the same way as for the F plasmid described in **Section 1.4**. The primary difference is that in this approach, the plasmid does not serve as a shuttle for DNA between genomes but as a means of delivery for artificially constructed DNA. Most commonly some method is employed to make replication of the plasmid conditional, allowing the cell to be forced to integrate the desired DNA sequence into the genome in order to continue to express the associated antibiotic resistance gene or other marker, which is also delivered to the genome. Such plasmids are thus referred to as "suicide plasmids."

The first suicide plasmids were actually used to deliver transposons and relied on the fact that the Mu prophage prevented replication of RP4 plasmids after transfer from *E. coli* to other strains such as *Agrobacterium tumefaciens* (van Vliet et al, 1978) and *Rhizobium leguminosarum* (Beringer et al, 1978). The first use of homologous recombination with a recombinant plasmid to deliver genes to the chromosome did not employ suicide plasmids *per se* but rather involved removal of the initial, gene-delivering plasmid via subsequent addition of a second plasmid whose presence was incompatible with the first (Ruvkun & Ausubel, 1981), but the use of suicide plasmids for the approach was adopted soon after (Meade et al, 1982).

Many additions and improvements to the method followed. For markerless modifications, a cassette containing a positive-selectable marker (such as antibiotic resistance) and a negative-selectable marker (such as the *sacB* gene, which is lethal to Gram-negative bacteria in the presence of sucrose) can be integrated first, followed by addition of DNA containing the desired modifications under selection against the negative (or "counter-selectable") marker (Ried & Collmer, 1987). Alternatively, the

entire suicide plasmid can be integrated, creating two copies of the target DNA sequence (the original chromosomal copy and the copy from the plasmid) in the genome, which then recombine back to a single copy upon counter selection, deleting the markers and other plasmid sequences in the process (Link et al, 1997). This second approach frequently results in resolution back to the wild-type, however, and screening is required to find cells harboring the desired mutations (see **Figure 1.3**). Simpler and more reliable methods for inducing "suicide" were also developed, including plasmids that require a specific gene, such as *pir* (Miller & Mekalanos, 1988) or *repA* (Leenhouts et al, 1996a) to reproduce, or plasmids that do not replicate at high temperatures (Biswas et al, 1993; Hamilton et al, 1989; Link et al, 1997; Luchansky et al, 1989).

The advantages of this approach are that it works in a wide variety of organisms and can be used to delete, replace, or insert DNA of several kilobases in a predictable manner. Disadvantages include low efficiency and potentially laborious cloning steps, and it is currently not preferred when other methods are available. For instance, the approaches described in **Section 1.10** have largely replaced suicide plasmids for targeted modifications in *E. coli*, except perhaps in certain special cases where complex, markerless gene replacements are desired, and the methods described in **Section 1.11** are becoming a standard solution for creating targeted gene knock-outs in other strains. However, approaches based on homologous recombination between chromosomes and plasmids have been used to achieve some impressive results, such as inserting an entire bacterial genome piecemeal into another (Itaya et al, 2005).

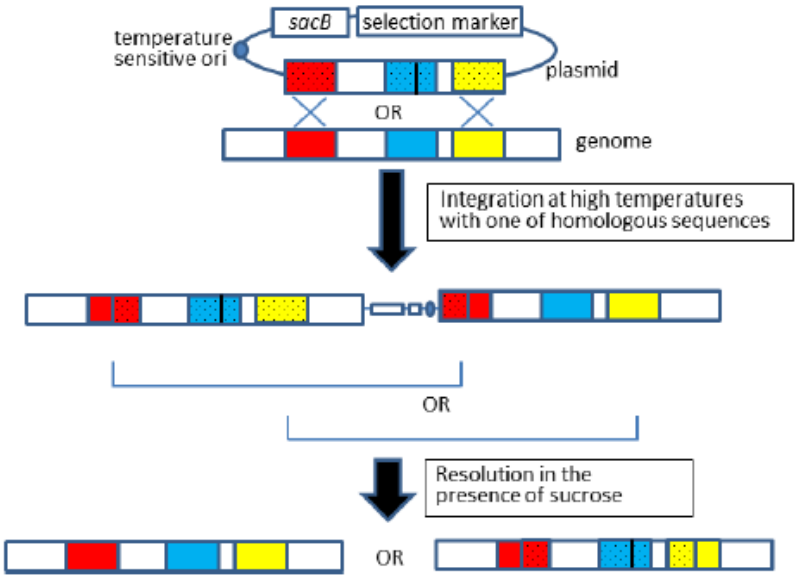

Figure 1.3. Schematic of suicide plasmid used for allelic replacement.

Colored boxes represent open reading frames, and the black bar represents the mutation to be introduced. The example shown depicts integration at the site of the red gene, but integration could occur anywhere along the region of homology. Figure from Nakashima & Miyazaki (2014), which is available under a Creative Commons Attribution license.

1.8 Site-Specific Recombinases

Site-specific recombinases are a class of enzymes that catalyze efficient recombination between specific sequences of DNA. This section will give particular attention to the Cre recombinase since it is arguably the best studied and most widely used (Anastassiadis et al, 2009) and is frequently employed in subsequent chapters of the present work.

The Cre protein (named because it *c*auses *re*combination) was first characterized in 1981 (Sternberg & Hamilton, 1981). Cre is a 38-kD enzyme that originates from the bacteriophage P1, which is commonly used for generalized transduction in *E. coli*. Cre seems to serve cyclize the linear genome of P1 and to resolve chromosome dimers

(Sternberg et al, 1981). Cre recognizes a sequence referred to as a "*lox*" site (for *lo*cus of crossing (*x*) over) and, in its basic implementation, will delete or invert the DNA between two *lox* sites depending on their relative orientation (see **Fig. 1.4**).

The *lox* site itself is 34 bases long and consists of 13-base-pair inverted repeats ("arms") flanking an 8-base-pair spacer region (see **Figure 1.5A**). The spacer is asymmetric and determines the relative oriention of the *lox* site. Dimers of Cre bind the inverted repeats of the *lox* site. Cre dimers bound to separate *lox* sites then bind together into tetramers, bringing the *lox* sites together such that the spacer regions are aligned. Recombination then takes place within the spacer region (Guo et al, 1997a; Hoess & Abremski, 1985).

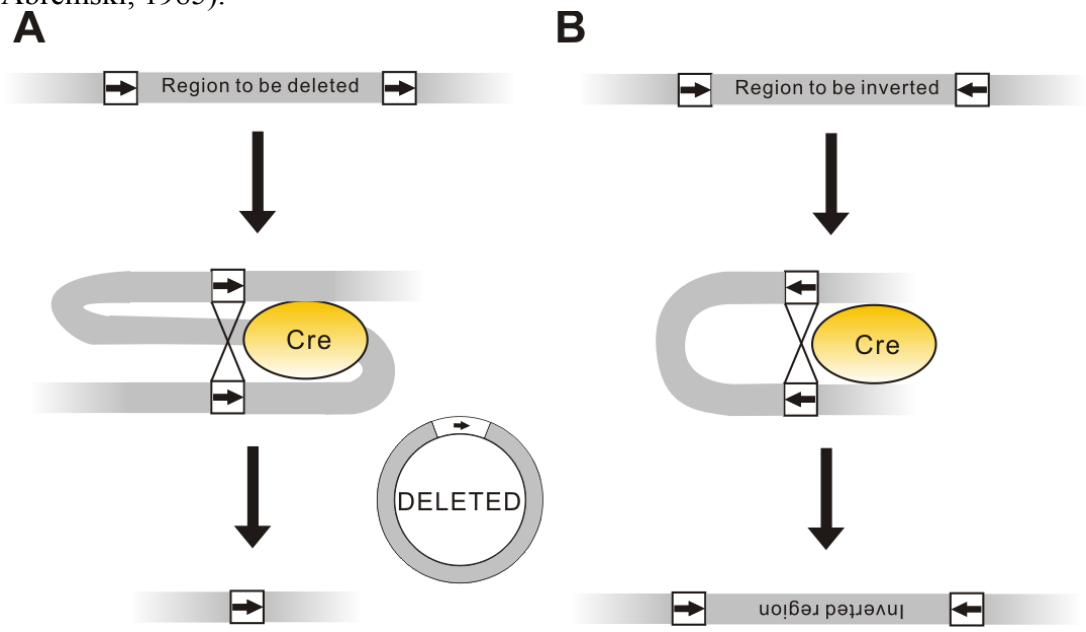

Figure 1.4. Cre-*lox* basics.

(A) shows the mechanism of deletion, and **(B)** shows the mechanism of inversion, where *lox* sites are represented as white boxes with black arrows indicating orientation.

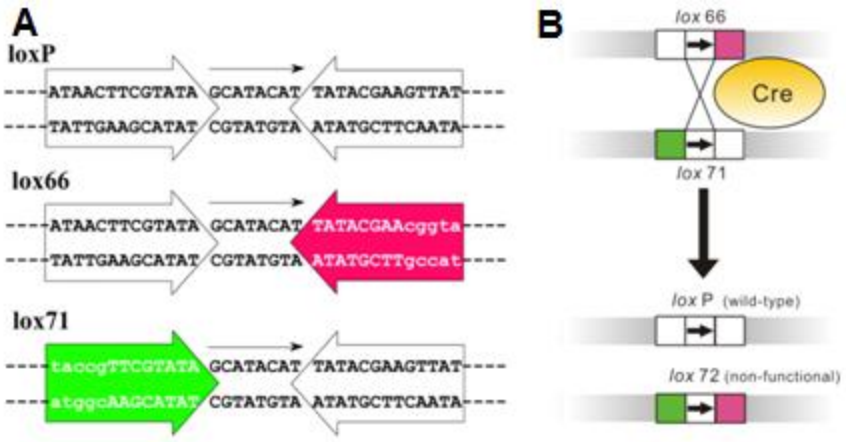

Figure 1.5. Sequences and functionality of wild-type (*loxP*) versus arm-mutant *lox* sites.

(A) Sequences of wild-type and arm-mutant *lox* sites. The arms are enclosed by large arrows, and the arrow above the spacer region indicates orientation. Bases mutated relative to wild-type are shown in lowercase. (Adapted from Langer et al. 2002) **(B)** Recombination between *lox66* and *lox71* generates a non-functional *lox72* site and a reconstituted *loxP* site.

Mutant *lox* sites provide some of the most useful characteristics of the system. The variant *lox* sites can be separated into two groups: mutations in the 13-base-pair inverted repeats (arm mutants), and mutations in the 8-base-pair spacer (spacer mutants). Arm mutants affect the ability of Cre to bind the *lox* site. The most widely used of these mutants are *lox66*, which contains mutations in the downstream arm, and *lox71*, which contains mutations in the upstream arm (Albert et al, 1995). These mutants are still recognized by Cre, but upon recombination form a *lox72* site in which both arms are mutated. The *lox72* site is no longer recognized by Cre, and thus the use of these arm mutants allows the direction of the recombination reaction to be controlled (see **Fig 1.5**). These *lox* sites have been used to for multiple marker removal (Arakawa et al, 2001; Lambert et al, 2007), directional cloning (Langer et al, 2002), and creating multiple large genomic deletions (Suzuki et al, 2005a).

The spacer determines the specificity of the *lox* site, and mutations here can create new *lox* sites capable of recombining with each other but incapable of recombining with *lox* sites containing a different spacer sequence (Hoess et al, 1986; Langer et al, 2002; Livet et al, 2007; Sauer, 1996; Siegel et al, 2001) (see **Tables 1.1** and **1.2**). Experimental evidence suggests that the second and seventh bases are most important for determining specificity, while a T is preferred at position 4 and a T or A is preferred at position 5 (Langer et al, 2002). Spacer mutants are typically used to force cassette exchange (in other words, insertion of new DNA) over deletion or inversion, though the *loxPsym* mutant was designed to remove the directionality of the spacer and create a completely symmetrical *lox* site (Hoess et al, 1986). The combination of arm and spacer mutants allows exquisite control over the type and direction of recombination catalyzed by Cre.

	12345678
loxP (wild-type)	ATAACTTCGTATAGCATACATTATACGAAGTTAT
lox511	ATAACTTCGTATAGTATACATTATACGAAGTTAT
lox511-I	ATAACTTCGTATAATGTATACTATACGAAGTTAT
loxFAS	ATAACTTCGTATATACCTTTCTATACGAAGTTAT
lox2272	ATAACTTCGTATAGGATACTTTATACGAAGTTAT
lox5171	ATAACTTCGTATAGTACACATTATACGAAGTTAT
loxm2	ATAACTTCGTATATGGTTTCTTATACGAAGTTAT
loxN	ATAACTTCGTATAGTATACCTTATACGAAGTTAT
loxPsym	ATAACTTCGTATAATGTACATTATACGAAGTTAT

Table 1.1. Sequences of selected *lox* spacer mutants.

(Sequences collected from Hoess et al. 1986; Siegel et al. 2001; Langer et al. 2002; and Livet et al. 2007)

	WT	511-I	511	FAS	2272	5171
WT	99.6 (± 0.7)					
511-I	1.4 (± 1.6)	99.2 (± 1.9)				
511	10.3 (± 1.4)	75.3 (± 9.3)	99.8 (± 0.3)			
FAS	0.2 (± 0.3)	5.7 (± 3.1)	0.0 (± 0.0)	99.4 (± 0.3)		
2272	0.5 (± 0.4)	0.3 (± 0.4)	1.6 (± 0.3)	1.7 (± 0.8)	99.7 (± 0.5)	
5171	5.7 (± 2.7)	77.7 (± 5.1)	99.9 (± 0.1)	4.8 (± 0.8)	0.1 (± 0.1)	99.5 (± 0.9)

Table 1.2. Recombination frequencies between selected *lox* spacer mutants.

(Data from Siegel et al. (2001))

Additionally, a great deal of work has been done in finding Cre homologs with altered specificities in order to further increase the range of operations that can be carried out simultaneously using this system. These include Dre (Anastassiadis et al, 2009; Sauer & McDermott, 2004), SCre and VCre (Minorikawa & Nakayama, 2011; Suzuki & Nakayama, 2011), and Vika (Karimova et al, 2013). In addition, Cre has been subjected to directed evolution to generate variants with modified substrate specificity (Buchholz & Stewart, 2001; Rufer & Sauer, 2002; Santoro & Schultz, 2002), including a variant called Tre that was designed to excise the HIV provirus from the mammalian genome (Sarkar et al, 2007).

Another widely used site-specific recombination system is FLP/FRT. FLP ("Flippase") was discovered in 1980 (Broach & Hicks, 1980) in the yeast 2μm plasmid, which is a selfish element that uses recombination to flip the orientation of its origin of replication in order to switch between phage-like rolling-circle amplification and bacteria-like theta replication in order to modulate its copy number (Volkert & Broach, 1986). The FLP recognition target (FRT) was determined in 1985 (Andrews et al, 1985; Jayaram, 1985; Senecoff et al, 1985), and, like Cre/*lox*, the minimal recognition sequence involves an 8-base-pair spacer flanked by two 13-base-pair inverted repeats. Though FLP is considered to be less active than Cre (Anastassiadis et al, 2009; Srivastava &

Gidoni, 2010), particularly at 37°C (Buchholz et al, 1996), more active variants of FLP, such as FLPe (Buchholz et al, 1998) and FLPo (Raymond & Soriano, 2007), have been engineered to narrow this gap. As with *lox* sites, FRT sites with mutant arms have been developed for directional recombination (Senecoff et al, 1988). It has been reported that these mutant arms may not work as well as might be desired, though the use of the FLPe variant may increase their utility (Huang et al, 1991; Nandy & Srivastava, 2011). Like Cre, work has also been done to generate variants of FLP with modified target-site specificity (Bolusani et al, 2006; Voziyanov et al, 2003).

Phage integrases can also perform similar functions, with the special characteristic of being naturally unidirectional. The Int (integrase) of λ phage is the canonical example of this class and catalyzes recombination between so-called attachment sites in the phage and bacterial genomes, referred to as *attP* and *attB*, respectively (Campbell, 1962; Hershey, 1971). In order for the backwards recombination to occur, both the Int and Xis proteins must be present (Guarneros & Echols, 1970). This is schematically shown in **Figure 1.6**.

Though the activity of phage integrases has been known since 1960s, it was not until the early 1990s that they were employed separately from the phage themselves to aid in genome engineering (Atlung et al, 1991; Diederich et al, 1992). In recent years a number of other phage integrases have been developed for biotechnological applications, including PhiC31 (Keravala et al, 2009; Thorpe & Smith, 1998) and PhiBT1 (Xu et al, 2008).

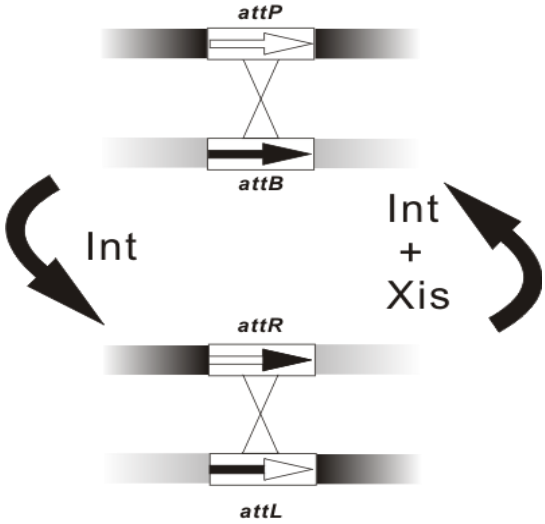

Figure 1.6. Activity of Int and Xis on their recognition targets.

Though site-specific recombinases are widely used in eukaryotes (Nagy, 2000; Siegal & Hartl, 2000; Srivastava & Gidoni, 2010; Turan et al, 2013; Wirth et al, 2007), they have remained relatively under-utilized in bacteria, being primarily used for removing the markers employed in other methods (some of the numerous examples include work reported in Abuin & Bradley (1996); Gilbertson (2003); Guldener et al. (1996); Lambert et al. (2007); Leibig et al. (2008); Pomerantsev et al. (2006); and Stockinger et al. (2005)). However, they have been used in a number of studies looking at the effects of large deletions and inversions on bacterial chromosomes in species including *Lactococcus lactis* (Campo et al, 2002; Campo et al, 2004), *E.coli* (Esnault et al, 2007; Fukiya et al, 2004; Valens et al, 2004), *Salmonella typhimurium* (Wilson et al, 2004) and *Corynebacterium glutamicum* (Suzuki et al, 2005a; Suzuki et al, 2005b). Additionally, Val and coworkers recently used phage integrases to fuse the two chromosomes of *Vibrio cholerae* into one (Val et al, 2012), and integrases were also used

to deliver large polyketide-synthase operons into the genomes of industrially optimized strains (Rodriguez et al, 2003).

1.9 GENOME SHUFFLING

Genome shuffling a method for rapid strain improvement that makes use of protoplast formation and subsequent fusion. A protoplast is a bacterium that has had most of its cell wall removed, and two such bacteria can then be fused into a single cell by the addition of an appropriate surfactant. This technique was first demonstrated in bacteria using *Bacillus subtilis* and *Bacillus megaterium* in 1976 (Fodor & Alfoldi, 1976; Schaeffer et al, 1976), and number of other bacteria soon followed (Peberdy, 1980). Placing multiple genomes from different strains in the same cell allows homologous recombination and trading of markers to occur, which was leveraged to map the genomes of species, such as *Staphylococcus aureus*, that had previously been difficult to analyze (Stahl & Pattee, 1983a; Stahl & Pattee, 1983b). Protoplasted bacteria also proved easy to transform (Chang & Cohen, 1979), which had previously been a problem for Gram-positive bacteria.

The field was reinvigorated in 2002 with the report of a technique called "genome shuffling," in which a large library of mutants is created and then repeatedly subjected to protoplast formation and fusion in between rounds of growth under a selective pressure or screening for a desired phenotype (Zhang et al, 2002). This allows beneficial mutations to spread rapidly through a population and allows strain improvement to proceed much more quickly than traditional methods of repeated rounds of simple mutagenesis and screening or selection. The initial example involved improving the production of tylosin by *Streptomyces fradiae*, and many other have followed, including improved acid tolerance and lactic acid production in Lactobacilli (John et al, 2008;

Patnaik et al, 2002; Wang et al, 2007), increased production of hydroxycitric acid in a *Streptomyces* species (Hida et al, 2007), and many others, most recently reviewed in Gong et al. (2009).

One drawback to these methods is that they do not function well in Gram-negative bacteria such as *E. coli*, which have a more complex membrane structure. Though protoplasting methods have been reported for Gram-negative bacteria (Dai et al, 2005), examples of protoplast fusion or genome shuffling of Gram-negative bacteria in the literature remain rare. While it is possible that the F plasmid could be exploited as an alternative (Quandt et al, 2013; Winkler & Kao, 2012), sharing fragments of genomes between cells is unlikely to be as efficient in facilitating evolution as combining multiple intact genomes, as can be achieved in the standard genome shuffling approaches in Gram-positive bacteria.

1.10 RECOMBINEERING[1]

In the 1960s, it was noticed that knocking out the homologous recombination system in *E. coli* did not stop homologous recombination from occurring between λ phage genomes (Takano, 1966; van de Putte et al, 1966), suggesting that λ possessed its own recombination system. Mutations that destroyed this ability were mapped to a locus that was dubbed "Red" (*re*combination *d*eficient) (Echols & Gingery, 1968; Signer & Weil, 1968). It turned out that there were two genes responsible, called *beta* and *exo* (Radding, 1970; Shulman et al, 1970) that are involved in λ recombination and are now called the Red system. Another neighboring gene, *gam*, is also involved and is frequently

[1] Portions of this section are adapted from Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* **9:** 685. This work is used with permission under a Creative Commons – Attribution license. The borrowed text was written by PJE with edits by ADE.

also considered as one of the Red functions. The product of *exo* is an 5'-3' exonuclease that acts on double stranded DNA and leaves 3' overhangs (Carter & Radding, 1971). The Beta protein binds single-stranded DNA and promotes annealing to complementary DNA (Kmiec & Holloman, 1981; Muniyappa & Radding, 1986). Together, then, Exo and Beta convert the ends of a linear DNA molecule to single-stranded DNA and promote annealing to complementary sequences. The Gam protein inhibits the native exonuclease machinery in E. coli that would otherwise degrade such DNA (Karu et al, 1975).

In the late 1990s it was found that over-expressing the *exo*, *beta*, and *gam* genes in *E. coli* greatly enhanced the frequency of homologous recombination in the cell (Murphy, 1998), allowing linear double-stranded DNA with short (approximately 40 base pairs) regions of homology on either end to be directly recombined into the genome (Datsenko & Wanner, 2000; Yu et al, 2000), entirely bypassing the use of plasmids as described in **Section 1.8**. This technique came to be called "recombineering" and has been widely used (for a review, see Murphy (2012)). It was, for instance, used by the Blattner lab to create their streamlined *E. coli* genome (Kolisnychenko et al, 2002; Posfai et al, 2006).

Soon after it was found that the *beta* gene by itself is sufficient to allow short, single-stranded DNA molecules to recombine into the genome (Ellis et al, 2001), the efficiency of which can be increased significantly in mutator strains (specifically, in strains lacking a functional *mutS* gene), which also confirmed long-standing suspicions that DNA replication played an important role in the process (Costantino & Court, 2003). In the latter case in particular, efficiencies can be on the order of 10% without the need for selectable markers (though the presence of markers can be exploited to push efficiency even higher (Carr et al, 2012)). This allows point mutations and other small

modifications to be made quickly and rapidly in *E. coli*. The method has also been automated (Wang et al, 2009), and it was this mechanized version, combined with use of the F plasmid, that was used to create an *E. coli* strain completely free of UAG stop codons, allowing that codon to be assigned to new, unnatural amino acids (Isaacs et al, 2011; Lajoie et al, 2013).

Though Red-based recombineering has become the default approach for genome engineering in *E. coli*, there are limitations. When using single-stranded DNA, mutator strains must be used to achieve reliable efficiency even in the single-digit range. Even then, the efficiency of inserting a sequence as large as a *lox* site is in the neighborhood of 1%, and the efficiency of deleting 10,000 bases of genomic sequence is approximately 0.1% (Wang et al, 2009). The use of the full complement of λ Red proteins allows larger pieces of double-stranded DNA to be inserted, but selectable markers are typically required and the size of possible insertions is limited to several thousands of bases. Additionally, use of this approach in other species has so far been limited and typically requires developing new recombineering functions for each system (Binder et al, 2013; Datta et al, 2008; Swingle et al, 2010; van Kessel & Hatfull, 2008).

1.11 TARGETRONS[2]

While introns are commonly associated with gene splicing in eukaryotic cells, introns are also present in bacteria, where they seem to function as selfish elements that insert into host genomes but splice themselves out of RNA transcripts in order minimize deleterious effects on the host. In recent years, the class referred to as "mobile group II

[2] Portions of this section are adapted from Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM (2014) Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mob DNA* **5:** 2. This work is used with permission under a Creative Commons – Attribution license. The borrowed text was written by PJE with edits by AML. The figures in this section were created by GM.

introns" has been the subject of much interest after it was found that they recognize new sites for insertion via base-pairing of the intron RNA with the target site DNA (Guo et al, 1997b), and that, furthermore, the site targeted can be changed by modifying the RNA sequence of the intron (Guo et al, 2000; Mohr et al, 2000). Such retargeted introns have come to be called "targetrons." Since these are frequently employed in later chapters of this work, they will be discussed in some detail here.

Mobile group II introns consist of a catalytically active intron RNA and an intron-encoded protein (IEP), which is a reverse transcriptase and also performs other functions (see **Fig. 1.7**) (Lambowitz & Zimmerly, 2011; Michel & Ferat, 1995; Qin & Pyle, 1998). Group II intron RNAs have a length of 400 to 800 nts, excluding the open reading frame encoding the IEP (Lambowitz & Zimmerly, 2011). They have little sequence similarity to each other, but fold into a conserved three-dimensional structure consisting of six interacting double helical domains (DI-DVI) (see **Figs. 1.7A** and **1.7B**) (Marcia et al, 2013; Michel & Ferat, 1995; Qin & Pyle, 1998; Toor et al, 2008).

The folded group II intron RNA contains an active site that uses specifically bound magnesium ions to catalyze RNA splicing via two sequential transesterification reactions that yield ligated exons and an excised intron lariat RNA, the same reaction mechanism used for the splicing of nuclear spliceosomal introns in eukaryotes (see **Fig. 1.7C**) (Pyle & Lambowitz, 2006). Because the transesterification reactions used for splicing are reversible, the intron RNA can also catalyze reverse splicing of the intron into RNA or DNA sites containing the ligated exon sequence, with reverse splicing into DNA playing a key role in intron mobility.

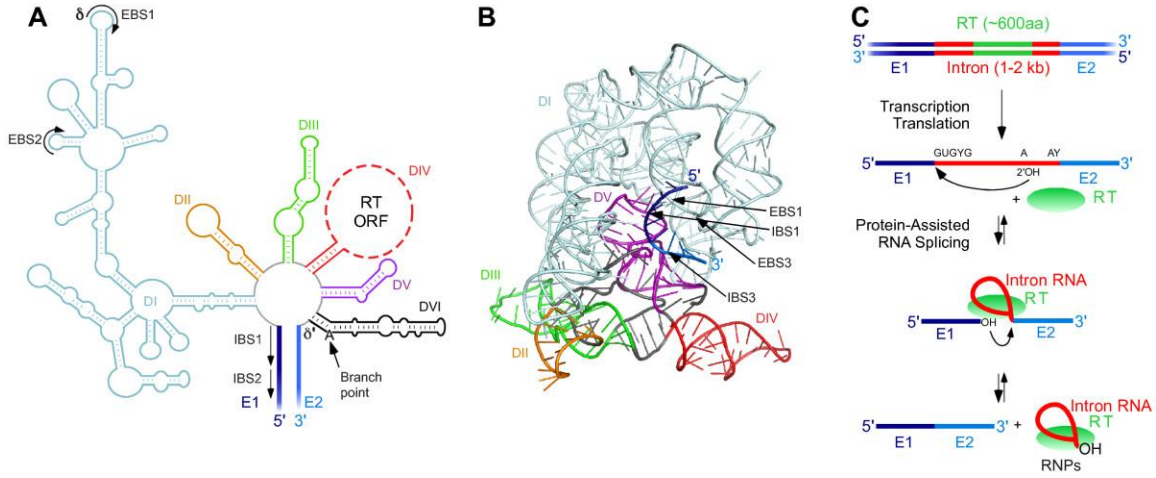

Figure 1.7. Group II intron RNA structure and splicing mechanism.

(A) Group II intron RNA secondary structure. The example shown is the *Lactococcus lactis* Ll.LtrB group IIA intron. Intron RNA domains are different colors, and the 5' and 3' exons (E1 and E2, respectively) are thicker dark and light blue lines, respectively. The large 'loop' region of DIV, which encodes the group II intron reverse transcriptase, is shown as a dashed line and not drawn to scale. **(B)** Crystal structure of the *Oceanobacillus iheyensis* group IIC intron. The ribbon diagram of the intron's structure was generated from Protein Data Bank file 3IGI (Toor et al, 2010) (http://www.pdb.org) with PyMol. Group II intron RNA domains are colored as in panel A. **(C)** Group II intron RNA splicing and reverse splicing. Double-stranded DNA is indicated by double lines and RNA as a single line. E1 and E2 are shown in dark and light blue, respectively; the intron and intron RNA are shown in red; and the intron-encoded RT is shown in green.

The process by which the intron inserts into a new DNA site is called "retrohoming," depicted in **Figure 1.8** and reviewed in Lambowitz & Zimmerly (2004) and Lambowitz & Zimmerly (2011). Retrohoming starts with the group II intron splicing out of a larger RNA molecule, typically a transcript of the gene in which the group II intron is inserted. Splicing is accomplished via folding of the intron RNA into a catalytic structure, with help of the IEP, which binds the intron RNA and stabilizes the active RNA tertiary structure. After splicing, the IEP remains tightly bound to the excised intron lariat RNA in a complex that initiates retrohoming by recognizing DNA target sequences by a combination of site-specific binding of the IEP and base pairing of

sequence motifs in the intron RNA, described in more detail below. The intron RNA then integrates directly into the DNA target site, while an endonuclease activity of the IEP cuts the opposite DNA strand slightly downstream of the insertion site, leaving an overhang with a cleaved 3' end that is used as a primer for synthesis of a cDNA copy of the inserted intron RNA by the IEP (Yang et al, 1996; Zimmerly et al, 1995a; Zimmerly et al, 1995b). The cDNA copy of the reverse-spliced intron RNA is integrated into the host genome by common cellular DNA recombination or repair mechanisms, a feature that contributes to the wide host range of group II introns. Host nucleases trim DNA overhangs, and ligases repair remaining nicks (Smith et al, 2005).

The key to using group II introns for gene targeting is their mode of DNA target site recognition. DNA target sequences are recognized by using both the IEP and base pairing of the intron RNA, with the latter contributing most of the DNA target specificity (Guo et al, 2000; Guo et al, 1997b). The major target site interactions for the introns that have been adapted for retargeting are shown in **Figure 1.8**.

The Ll.LtrB intron (often simply referred to as "LtrB") is a type IIA intron and as such contains three sequence motifs that recognize DNA target sites by base pairing. These are denoted EBS1, EBS2, and δ, and they base pair to complementary sequences in the DNA target site denoted IBS1, IBS2, and δ' (where EBS stands for "exon-binding site" and IBS stands for "intron-binding site"; these same interactions also occur upon splicing out of a larger RNA molecule). The Ll.LtrB IEP (commonly called "LtrA") recognizes nucleotides both upstream and downstream of the IBS/δ' sequences (colored purple and blue, respectively, in **Fig. 1.9**). Binding of the IEP promotes DNA melting (Singh & Lambowitz, 2001), enabling the intron RNA to base pair to the DNA target sequence, and DNA bending, which positions the target DNA properly for cleavage and priming of reverse transcription (Noah et al, 2006).

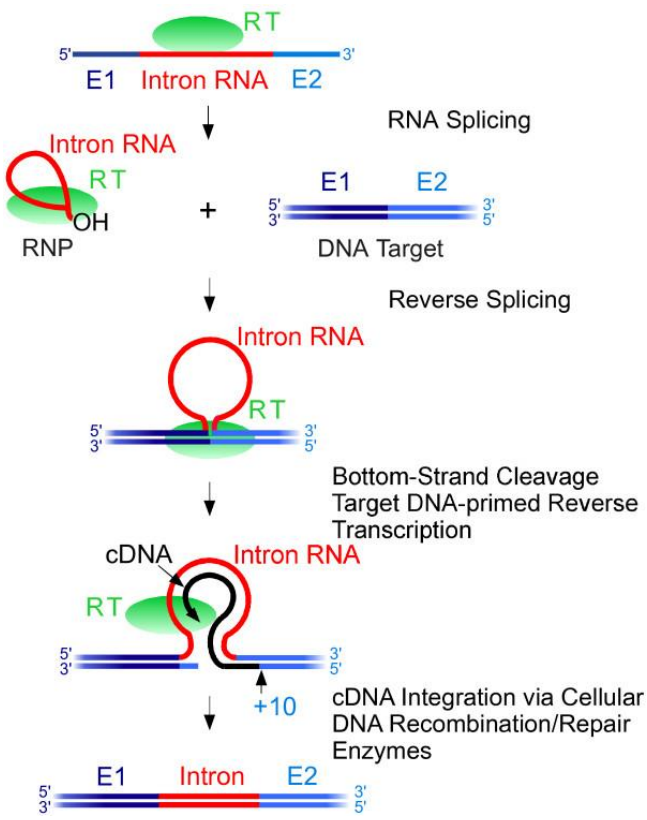

Figure 1.8. Group II intron retrohoming.

In the first step, the intron-encoded protein (IEP) binds to the intron in a larger initial transcript of a gene and promotes RNA splicing, resulting in a ribonucleoprotein complex that contains the excised intron lariat RNA and the tightly bound IEP. The complex recognizes DNA target sites by using both the IEP and base pairing of the intron RNA and then promote reverse splicing of the intron RNA into the top strand of the double-stranded DNA. After reverse splicing, the bottom DNA strand is cleaved by the endonuclease domain of the IEP, and the 3' end generated at the cleavage site is used as a primer for target DNA-primed reverse transcription of the inserted intron RNA. The resulting intron cDNA (black) is integrated into the host genome by cellular DNA recombination or repair mechanisms.

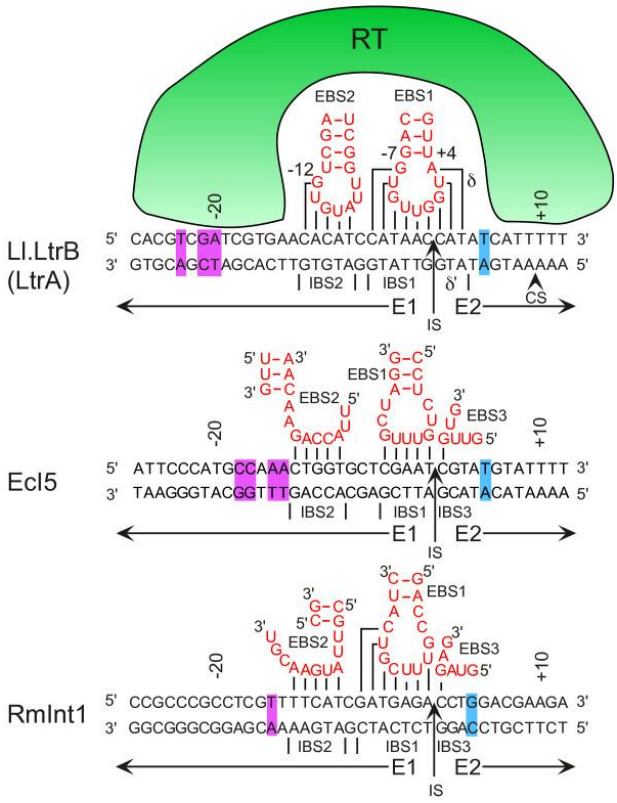

Figure 1.9. DNA target site recognition by mobile group II introns used as targetrons.

Portions of the intron RNA involved in the EBS1-IBS1, EBS2-IBS2, and δ − δ' or EBS3-IBS3 base-pairing interactions with the DNA target site are shown in red. Purple and blue highlights indicate base-pairs in the 5' and 3' exons (E1 and E2, respectively) that are important for DNA targeting and recognized by the IEP (denoted RT (reverse transcriptase) here). CS, bottom-strand cleavage site; IS, intron-insertion site.

EcI5 and RmInt1 are group IIB introns but also contain three sequence elements that recognize the DNA target site by base pairing. Specifically, EBS1, EBS2, and EBS3 base pair to corresponding IBS sequences in the target. The IEP again recognizes flanking sequences. In EcI5, a relatively well-studied example of this class (Zhuang et al, 2009), the IEP recognizes a similar number of residues as the IEP of Ll.LtrB, although the identities and locations of these residues differ. The RmInt1 IEP recognizes only two

critical nucleotide residues, but additional sequences may contribute (Jimenez-Zurdo et al, 2003).

These introns have high DNA-target specificity and integrate only rarely into off-target sites. For example, retrotransposition of the Ll.LtrB intron into off-target sites in the *E. coli* chromosome occurs at a frequency of 0.1-30 x 10^{-6} (Coros et al, 2005; Lambowitz & Zimmerly, 2011).

Although group II introns can and have been retargeted by finding the closest match to the native recognition site in a sequence to be targeted and modifying the base-pairing sequences of the intron to accommodate discrepancies, the rules by which introns recognize their target sites are actually more complex. For instance, the IEP recognizes different residues at the target site with different stringencies, and none of these recognition events are absolutely required for retrohoming to occur (Guo et al, 2000; Mohr et al, 2000; Zhong et al, 2003). If only the wild-type recognition sequence is used, then new targeting sites may be hard to come by, but knowing which bases can be varied and how is not a simple matter. The EBS/δ sequences may also differ in the stringency of required base-pairing interactions at different positions. Algorithms have therefore been developed for retargeting the Ll.LtrB (Perutka et al, 2004) and EcI5 (Zhuang et al, 2009) introns. These algorithms were developed by examining libraries of inserted mobile group II introns with randomized base-pairing motifs for the most frequently conserved residues and base-pairing interactions, and using these frequencies to generate weighting schemes for the various interactions. Potential target sites are then assessed using the weighted criteria and assigned a score. Although the algorithms have limitations and do not always correctly predict insertion frequency, typically a targetron efficient enough to be screened via colony PCR for insertions without selection can be found for any given stretch of 1,000 base pairs of DNA.

The actual retargeting process is carried out by using two sequential PCR reactions that modify the EBS/δ sequences within the intron to base pair to the DNA target site and modify the IBS sequences flanking the intron to base pair to the retargeted EBS sequences to allow the intron to splice out of the precursor RNA transcribed from the expression plasmid (Karberg et al, 2001; Perutka et al, 2004). The PCR product corresponding to a segment of the intron and upstream exon is then cloned into an expression vector. Alternatively, the entire region covering the IBS and EBS/δ sequences can be commercially synthesized in a single DNA molecule (for example, as a gBlock sold by IDT) that can be cloned directly into the vector (Enyeart et al, 2013). The δ' or IBS3 positions in the downstream exon are typically adjusted by cloning the PCR product into one of four parallel targetron vectors already containing the correct base for this interaction.

The typical configuration for the targetron cassette is one in which the open reading frame encoding the IEP is removed from the intron and expressed in tandem. This increases the efficiency of retrohoming and allows for disruptions of the targeted gene to be either conditional or non-conditional, depending on whether the intron is targeted to insert into the sense or antisense strand of the gene and whether or not the IEP remains present to aid in splicing of the intron from the mRNA (Frazier et al, 2003; Yao et al, 2006).

Targetrons have been used in a wide range of bacteria, including medically and commercially important species that had been recalcitrant to gene targeting by other methods (see **Table 1.3**). Because the initial reverse splicing and target-DNA-primed reverse transcription reactions are catalyzed by group II intron RNPs, and because the late steps of second-strand synthesis and cDNA integration are performed by common host factors (Eskes et al, 2000; Eskes et al, 1997; Smith et al, 2005; White & Lambowitz,

2012; Yao et al, 2013), there are in principle no limitations to the number of bacterial species in which targetrons might function. As mobile group II introns are present in the genomes of some archaea (Rest & Mindell, 2003), targetrons may prove useful in archaea, as well.

Genus	Primary references
Agrobacterium	(Yao & Lambowitz, 2007)
Azospirillum	(Malhotra & Srivastava, 2008)
Bacillus	(Akhtar & Khan, 2012)
Clostridium	(Chen et al, 2005; Heap et al, 2007)
Chlamydia	(Johnson & Fisher, 2013)
Ehrlichia	(Cheng et al, 2013)
Escherichia	(Karberg et al, 2001; Zhuang et al, 2009)
Francisella	(Rodriguez et al, 2008)
Lactococcus	(Frazier et al, 2003)
Listeria	(Alonzo et al, 2009)
Paenibacillus	(Zarschler et al, 2009)
Pasteurella	(Steen et al, 2010)
Proteus	(Pearson & Mobley, 2007)
Pseudomonas	(Yao & Lambowitz, 2007)
Ralstonia	(Park et al, 2010)
Salmonella	(Karberg et al, 2001)
Shewanella	(Enyeart et al, 2013)
Shigella	(Karberg et al, 2001)
Sinorhizobium	(Garcia-Rodriguez et al, 2011)
Sodalis	(Smith et al, 2013)
Staphylococcus	(Yao et al, 2006)
Vibrio	(Kumar et al, 2011)
Yersinia	(Palonen et al, 2011)

Table 1.3. Bacteria in which targetrons have been used successfully.

Targetrons have most frequently been used to generate knock-outs in bacteria. A great deal of work has been done using this method, with examples including identifying virulence factors (Alonzo et al, 2009; Buchan et al, 2009; Carter et al, 2011; Cheng et al,

2013; Francis et al, 2013; Sayeed et al, 2008; Steen et al, 2010) and potential drug targets (Zoraghi et al, 2010; Zoraghi et al, 2011), and examining the combinatorial effect of different genomic loci on protein expression (Rawsthorne et al, 2006).

Targetrons have been particularly widely used in strains of the genus *Clostridium* (reviewed in Enyeart et al. (2014)). Suicide plasmids were previously the only method of utility in these strains (Heap et al, 2007), but since clostridia typically have very low transformations frequencies (for instance, more than a milligram of plasmid is required to transform *Clostridium acetobutylicum* (Shao et al, 2007)), suicide plasmids are difficult to apply in these organisms, which include a number of medically and industrially important species.

Many bacteria of interest are difficult to transform due to restriction-modification systems. In *Staphylococcus aureus* (Corvaglia et al, 2010), *Clostridium acetobutylicum* (Dong et al, 2010), and *Clostridium cellulolyticum* (Cui et al, 2012), targetrons were used to knock out restriction enzymes, thereby opening clinical and environmental isolates to systematic mutational analysis. Targetrons have been developed for use in a number of pathogenic bacteria (see **Table 1.3**), opening up the possibility of using targetrons to develop vaccine strains of these organisms. Additionally, a thermostable targetron was recently developed (Mohr et al, 2013), allowing the technology to be applied to thermophilic bacteria, which include many industrially important strains.

Targetrons have also been used to deliver cargo genes, including genes for fluorescent proteins (Rawsthorne et al, 2006), phage resistance (Frazier et al, 2003), and antigens for release into a host's digestive system as a live vaccine (Chen et al, 2007). Sequences of less than 100 nt in length can usually be carried without impacting intron mobility (though see Chapter 2 for exceptions). Longer sequences may impair functionality, and sequences above 1,000 nt can drastically decrease efficiency. Domain

IV has been shown to be the best location to insert cargo genes for minimal impact on intron mobility (Plante & Cousineau, 2006).

The advantages of targetrons are their broad-host range specificity and their high efficiency. Disadvantages include the fact that retargeting and testing new introns is rather more laborious than the methods presented in **Section 1.10**, for example.

1.12 WHOLE-GENOME SYNTHESIS

In 2007, researchers at the Venter Institute announced that they had replaced the genome of one bacterium with that of another (Lartigue et al, 2007). This was followed by reports of complete chemical synthesis of a bacterial genome (Gibson et al, 2008), and then, by combining the two techniques, the creation of a bacterium whose genome was entirely synthetically constructed (Gibson et al, 2010). These techniques have since been refined by employing yeast to assemble, modify and safely store bacterial genomes (Benders et al, 2010; Karas et al, 2013; Lartigue et al, 2009). These advances raise the possibility of an entirely new paradigm for genome engineering: instead of modifying existing genomes, genomes having the desired properties could simply be designed, synthesized, and inserted into appropriate cells.

While these results are impressive, this technology is still in its infancy and is far from being routinely practicable for most practitioners. The genomes to which this technique have been applied are quite small, with the artificial genome that was expressed in a living cell being about 1.1 megabases (Gibson et al, 2010), and the largest genome cloned in yeast being about 1.8 megabases (Karas et al, 2013). These are quite small compared to *E. coli*, whose genome is about 4.6 megabases (Blattner et al, 1997). Whether and how these techniques can be scaled up to organisms with more complex genetics remains to be seen. Synthesizing entire genomes is also extremely resource

intensive (Enyeart & Ellington, 2011), and while DNA synthesis costs are falling, the rate of decrease has lagged behind that for DNA sequencing, for instance (Carr & Church, 2009). Finally, the modifications that were made to the wild-type genome sequence in the process of synthesizing new genomes have thus far been modest, not extending beyond watermarks and unintentional mutations (Gibson et al, 2010). It will be interesting to see if and when whole genome synthesis will be able implement new bacterial genome designs that could not be performed more easily by engineering naturally occurring chromosomes.

1.13 TARGETED NUCLEASES

Brief mention will be made of artificial targeted nucleases, such as zinc finger nucleases (Urnov et al, 2010), TAL effector nucleases (Sun & Zhao, 2013), and the CRISPR-Cas9 system (Cho et al, 2013; Cong et al, 2013; Mali et al, 2013), which were developed in recent years and have been widely employed in a variety of applications in eukaryotes, including human gene therapy (Gaj et al, 2013). These approaches rely on cutting the genome at a precisely defined point, which allows mutations to be introduced upon repair, including the insertion of new DNA at that site via homologous recombination. These approaches have not gained much traction in bacteria, however. For the CRISPR-Cas9 nuclease system in particular, targeting the system to genomic sites was found to be lethal (Bikard et al, 2012), but it has been found to function acceptably well in bacteria with exceptionally efficient homologous recombination systems, such as *Streptococcus* (Jiang et al, 2013). Similar phenomena are likely responsible for the fact that zinc finger nucleases and TAL effector nucleases are rarely used in bacterial systems.

1.14 SUMMARY AND PERSPECTIVE

Table 1.4 summarizes bacterial genome engineering methodologies. In considering conventional methods for modifying bacterial genomes, we note two issues. The first is that efficient, versatile, and broad-host range solutions for genome engineering are rare. The second is that many opportunities exist for combining these methods to achieve new functionalities, particularly as regards the utility of site-specific recombinases. The research presented in Chapters 2 through 4 herein was inspired by these observations, while Chapters 5 and 6 present some potential applications of that work.

	Primary uses	Advantages	Disadvantages
Serial dilution	Strain improvement	"Natural," universal application	Extremely slow, difficult to target and analyze
Mutagens	Creation of knock-outs and other mutant strains, strain improvement	Easy to use, random, universal application	Dangerous to handle, unpredictable effects, difficult to analyze, possibly difficult to maintain
F plasmid	Mapping, strain improvement	Useful for combining or dispersing mutations between strains	Narrow host range, somewhat labor intensive
Generalized transducing phage	Mapping, strain improvement	Useful for combining or dispersing mutations between strains (finer grained than F plasmid)	Narrow host range, somewhat labor intensive
Specialized transducing phage	Gene delivery	Reliable delivery of large DNA pieces	Narrow host range, rigid site specificity, somewhat labor intensive
Transposons	Creation of knock-out mutants, strain improvement	Efficient, random, mutations easy to maintain, relatively easy to analyze, broad host range	Not suitable for site-specific applications
Suicide plasmids	Knock-outs, gene replacements, deletions	Broad host range	Inefficient, somewhat labor intensive
Site-specific recombinases	Deletions, inversions, insertions	Extremely efficient, broad host range	Recognition sites must be positioned using other methods
Genome shuffling	Strain improvement	Fast	Does not work well in Gram negatives
Recombineering	Insertions, deletions, point mutations	Easy to use	Limited to small-size modifications, narrow host range
Targetrons	Knock-outs	Efficient, broad host range	Somewhat laborious
Whole-genome synthesis	Large-scale chromosomal remodeling	Virtually any type and any number of viable modifications can be made simultaneously	Extremely expensive, laborious, and currently limited to small genomes
Targeted nucleases	Insertions and short deletions	Site specificity	Lethal to most bacteria

Table 1.4. Summary of current bacterial genome engineering methods.

Chapter 2: Developing targetrons for delivery of functional genetic elements

2.1 INTRODUCTION

Targetrons are a useful, broad-host-range tool for integrating site-specifically into bacterial genomes, but to date they have primarily been employed for knock-outs. Targetrons can be used to carry cargo, however. In both the LtrB and EcI5 introns that have been developed into targetrons, the open reading frame of the intron encoded protein (IEP) has been removed and is expressed separately. The former site of the IEP in Domain IV is replaced by an MluI restriction site and serves as favorable location for inserting cargo (Plante & Cousineau, 2006). All of the work presented in this chapter involved cloning new elements into this MluI site.

We specifically looked at adding three different elements to targetrons for delivery to the genome. The first was *lox* sites, to enable a versatile and broad-host range solution to making large genomic rearrangements. The second was *Ter* sites, which are polar sequences normally found near the termination of replication in bacterial genomes, where they serve to orient and arrest the replication machinery. As part of a collaboration, we wanted to examine the effect of introducing *Ter* sites to unnatural genomic locations in order to better understand their function. Lastly, we put *lac* operators into targetrons to test the hypothesis that binding and dimerization of the *lac* repressor could be used to bring disparate regions of the chromosome together and increase the frequency of homologous recombination as a potential alternative to using site specific recombinases.

2.2 RESULTS

2.2.1 Engineering targetrons to carry *lox* sites[3]

We began by inserting *lox* sites into the MluI site of the *lacZ*-targeting introns LtrB.LacZ.635s and EcI5.LacZ.912s (see the **Appendix** for a complete list of targetrons employed in this work; the numbers 635 and 912 indicate the position in the *lacZ* gene at which the introns insert, and "s" (as opposed to "a") indicates that the introns insert into the sense strand of the gene). However, some of the initial *lox*-site constructs significantly reduced the integration efficiency of the introns. We hypothesized that the tight hairpins that were predicted to be formed by the symmetric *lox* sites (Zuker, 2003) might disrupt the tertiary structure of the intron, and thereby inhibit splicing or insertion. The *lox* site inserts were therefore redesigned to include the sequence "GAGAG" at both ends of the insert to increase the flexibility of the structures, as judged by the presence and size of non-base-pairing regions in the predicted structures, particularly at the base of the stem. This largely restored insertion efficiency (see **Fig. 2.1**). We hypothesize that this trend occurs because inserts having inflexible structures are more likely to interfere with proper folding of the catalytic structures of the intron than inserts having flexible structures, which can be moved out of the way of other formations.

[3] This section is adapted from Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* **9:** 685. This work is used with permission under a Creative Commons – Attribution license. All experiments described in this section were performed by PJE, and the borrowed text was written by PJE with edits by ADE.

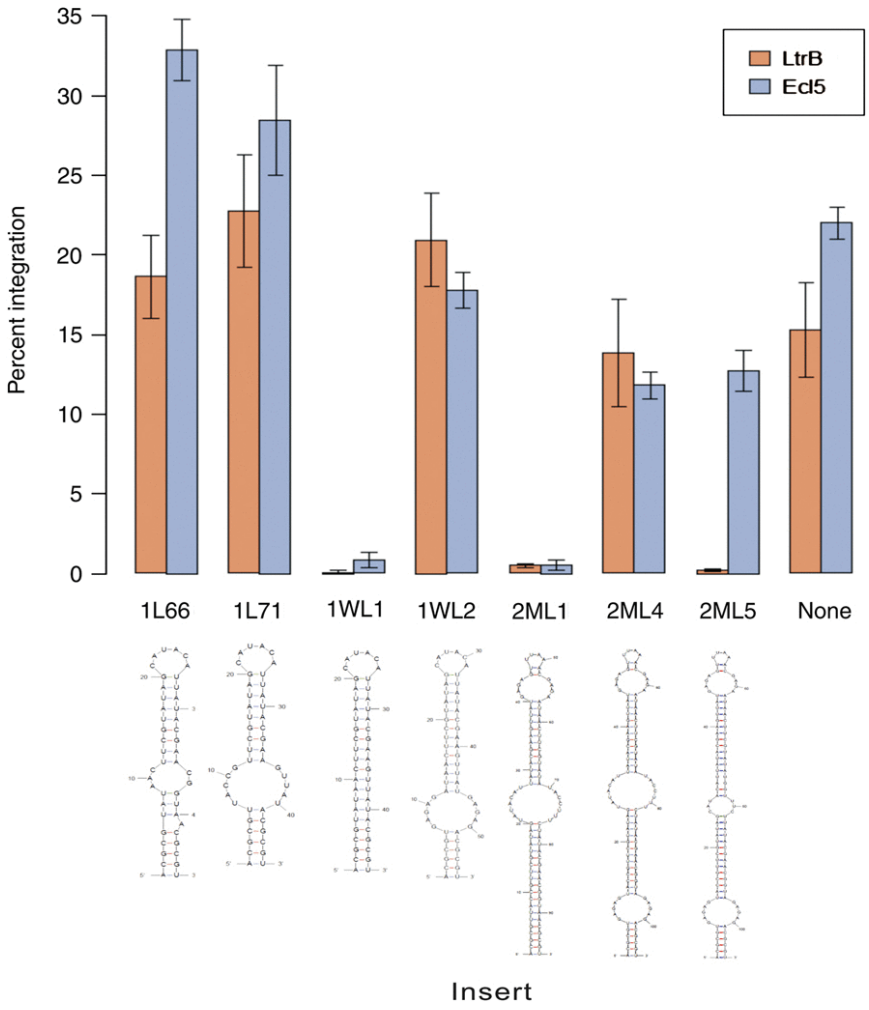

Figure 2.1. Effect of *lox* insert on intron efficiency.

Different *lox* sequences were inserted into the MluI site of the LtrB.LacZ.635s (Ll.LtrB) and EcI5.LacZ.912s (EcI5) introns (see **Appendix**), and efficiency was screened by counting the number of white colonies obtained. Error bars are standard error of three replicates, each representing a separate transformation of the intron plasmid into the recipient strain. The identities of the *lox* inserts are as follows, where all sequences are flanked by MluI sites: 1L66– *lox66*; 1L71– *lox71*; 1WL1– *lox*P (wild-type *lox* site); 1WL2– 1WL1 plus a flexible base; 2ML1– *lox511* with the *lox71* arm mutation (*lox511/71*) and *lox*FAS with the *lox*66 arm mutation (*loxFAS/66*), separated by a PmeI site and a short linker; 2ML4– 2ML1 plus a flexible base; and 2ML5– identical to 2ML4 except with *lox71* and *loxm2/66* instead of *lox511/71* and *loxFAS/66*. At the bottom are the RNA structures of the inserts as determined by Mfold (Zuker, 2003).

Statistical analyses (see **Section 2.4**) confirmed that the inserts fall into two classes: one of wild-type efficiency, and one of impaired efficiency. The 2ML5 insert

was the only insert that performed markedly differently in the Ll.LtrB intron versus the EcI5 intron. The poor performance of the 2ML5 insert in the Ll.LtrB intron may be due to its relative inflexibility at the central non-base-pairing region as compared to the 2ML4 insert, which is otherwise similar in structure.

2.2.2 Engineering targetrons to carry *Ter* sites

In bacteria with circular genomes, DNA replication starts at the origin of replication and proceeds bi-directionally around the chromosome until the two replication complexes meet and halt in the termination domain on the other side (Rocha, 2008). In *Escherichia coli*, the mechanism of this termination occurs via *Ter* sites, which are 21-basepair sequences of DNA found primarily in the termination domain and the left and right domains flanking it (see **Fig. 2.2**). The *Ter* sites have a specific orientation and, upon binding by a protein called Tus, will halt the DNA replication machinery when it approaches the *Ter*-Tus complex from one side but not the other (Hidaka et al, 1988; Hill et al, 1988). In particular, one set of five *Ter* sites all having the same orientation is found one side of the genome, and another set of five unidirectional *Ter* sites is found on the other. Both sets are oriented so as to arrest a DNA replication complex traveling from inside the termination domain to the outside, while allowing passage from outside in, with the end result being that replication is terminated in the termination domain (Duggin & Bell, 2009). The Tus protein has differing affinities for the different *Ter* sites depending on their sequence (Moreau & Schaeffer, 2012a; Moreau & Schaeffer, 2012b), which in turn correspond well to the frequency with which replication is arrested at a given *Ter* site (Duggin & Bell, 2009) (see **Fig. 2.2**).

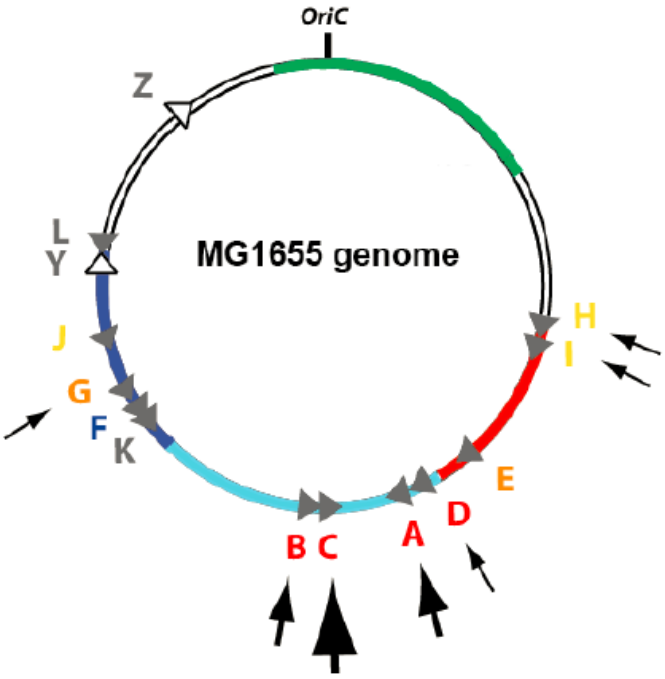

Figure 2.2. Locations of *Ter* sites in the *E. coli* genome.

Ter sites oriented to block a replication fork moving from the terminus to the origin are grey triangles, and those oriented to block a replication fork moving from the origin to the terminus are white triangles, where the direction the triangle points is the permissive direction. The letter designations of the *Ter* sites are encoded from blue to red according to increasing strength of Tus binding, as assayed in (Moreau & Schaeffer, 2012a) (the grey-lettered sites were not assessed and are farther from the consensus sequence than the others). The arrows represent the frequency of replication pausing at a given *Ter* site, as assessed in Duggin & Bell (2009). The coloring of the circle represents the different structural domains of the chromosome according to Valens et al. (2004) and Scolari et al. (2011), with the ori domain in green, the left domain in dark blue, the termination domain in light blue, the right domain in red, and the unstructured domains in white. Figure adapted from Moreau (2013). Used with permission.

In order to further study the biological activity of the *Ter* sites, we entered into a collaboration with Morgane Moreau and Patrick Schaeffer at James Cook University in Australia to use targetrons to deliver *Ter* sites ectopically to a site near the origin to assess the effect of placing *Ter* sites where they are not normally found. We chose one strong Tus-binding site (*TerB*) and two intermediate-strength Tus-binding sites (*TerH* and *TerJ*), the sequences of which are given in **Table 2.1**.

TerB (permissive orientation)	5′-ACTTTAGTTACAACATACTTATT-3′
TerB (non-permissive orientation)	5′-AATAAGTATGTTGTAACTAAAGT-3′
TerH (permissive orientation)	5′-GAGATAGTTACAACATACGATCG-3′
TerH (non-permissive orientation)	5′-CGATCGTATGTTGTAACTATCTC-3′
TerJ (permissive orientation)	5′-GCATTAGTTACAACTTACTGCGT-3′
TerJ (non-permissive orientation)	5′-ACGCAGTAAGTTGTAACTAATGC-3′

Table 2.1. Sequences of *Ter* sites delivered ectopically.

Given the issues with insert structure previously encountered, as described in **Section 2.2.1**, we examined the structures these sequences would form upon transcription into RNA, as shown in **Figure 2.3**. Though some of the sites do form hairpins, we judged that they were small enough and flexible enough that we did not need to add extra sequence to add flexibility.

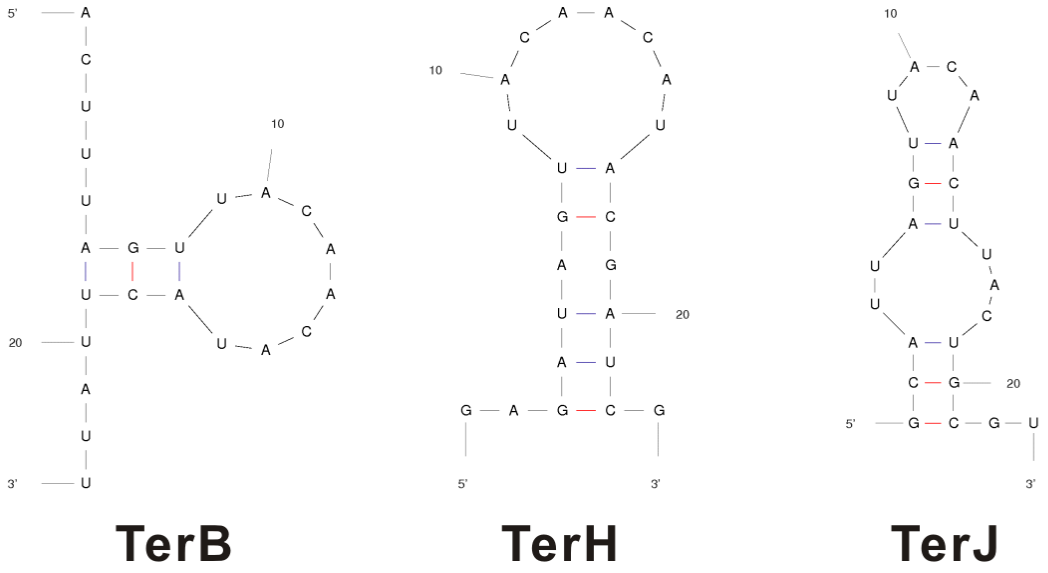

Figure 2.3. RNA structures of *Ter* inserts.

As for genomic integration sites, it was important to ensure as much as possible that the effects of delivering a *Ter* site to a given locus were a result of the *Ter* site itself and not from interrupting a gene at that site. We therefore sought to design targetrons to

insert into "safe insertion regions" (SIRs) that had been previously determined for the *E. coli* genome (Isaacs et al, 2011). We found suitable integration sites in SIR32.1 (about 1 kilobase away from the origin, on the left side) and in SIR5.6 (about 900 kilobases away from the origin, on the right side). We designed two introns, LtrB.SIR32.1 and EcI5.SIR2.3, to integrate into these respective loci (for sequences, see the **Appendix**), which retrohomed with efficiencies of approximately 10% and 80%, respectively.

We judged the integration site of LtrB.SIR32.1 to be too close to the origin to judge effects reliably and thus focused on EcI5.SIR2.3, into which we cloned each of the *Ter* sites in **Table 2.1** in both permissive and non-permissive orientations relative to the direction of DNA replication upon intron integration. We first assayed retrohoming efficiency in *E. coli* BL21(DE3). These results are shown in **Table 2.2**.

***Ter* site**	**Successful integrations**
TerB (permissive)	5/9
TerB (non-permissive)	0/28
TerH (permissive)	12/14
TerH (non-permissive)	12/14
TerJ (permissive)	12/14
TerJ (non-permissive)	12/14

Table 2.2. Integration efficiencies of EcI5.SIR5.6 introns carrying *Ter* sites.

The fact that no integrations were ever obtained with *TerB* in the non-permissive orientation suggests that this site is strong enough to cause replication arrest wherever it is found. On the other hand the *TerH* and *TerJ* sites do not seem to halt replication to an extent sufficient to result in unviable cells, regardless of their orientation.

We further proceeded to measure doubling times of strains harboring integrations of *Ter*-carriyng LtrB.SIR32.1 introns. These results are shown in **Table 2.3**.

Integrated *Ter* site	Doubling time (standard error)[4]
(None)	25.9 (± 0.5)
TerB (permissive)	25.5 (± 0.4)
TerH (permissive)	24.1 (± 0.1)
TerH (non-permissive)	23.7 (± 0.6)[5]
TerJ (permissive)	22.7 (± 0.3)
TerJ (non-permissive)	23.7 (± 0.3)

Table 2.3. Doubling times of *E. coli* BL21(DE3) strains harboring integrations of *Ter*-carrying EcI5.SIR5.6 introns.

No growth deficiencies were seen in any of the strains harboring ectopic *Ter* sites, suggesting that if such ectopic insertions can be obtained at all they will have little practical effect on DNA replication in the cell. In fact, the data suggests the strains grow better with *Ter* sites at this location, though it is unclear why this might be.

2.2.3 Engineering targetrons to carry *lac* operators

Homologous recombination can result in deletions or inversions, and the presence of two homologous targetrons in the same genome can serve as sites for such recombination to occur (see Chapter 3). This phenomenon has been applied as a means for generating genomic deletions in *Clostridium* (Jia et al, 2011). However, the process was slow (requiring about a week after placement of the introns) and inefficient (less than 0.5% for a two-gene operon). If the efficiency of homologous recombination could be site-specifically increased, homologous recombination between introns might serve as a more rapid alternative to creating genomic rearrangements, as compared to using introns to deliver *lox* sites, as described in **Section 2.2.1**, which requires subsequent addition of a Cre-expressing plasmid.

[4] Standard error is of three replicates.
[5] One outlier was removed. With outlier included, doubling time is 28 ± 4.

We hypothesized that the frequency of homologous recombination between two sites could potentially be increased by causing the two sites to spend more time in proximity to one another. One way to accomplish this using targetrons would be to include a sequence bound by a dimerizing protein, which would then bring the two targetron sequences together upon protein binding. One such well-studied sequence is the *lac* operator, which is bound by the *lac* repressor, which dimerizes with other repressors bound at other sites in order to loop out the intervening DNA (Lewis et al, 1996; Oehler et al, 1990). Thus we set out to examine whether including *lac* operators in targetrons could increase the efficiency of homologous recombination enough to be useful in biotechnological applications.

The structure of the *lac* operator when transcribed into DNA is shown in **Figure 2.4**. Since the *lac* operator forms a tight hairpin similar to the *lox* sites discussed in **Section 2.2.1**, we examined the differences in efficiency obtained from cloning a lac operator directly into *lacZ*-targeting L1.LtrB and EcI5 introns (specifically, LtrB.LacZ.635s and EcI5.LacZ.912s; see the **Appendix**) versus also including a non-base-pairing linker ("GAGAG") on both sides of the insert to improve flexibility. Counts of white colonies (presumably resulting from intron integrations into *lacZ*) from plates of individual intron electroporations are given in **Table 2.4**.

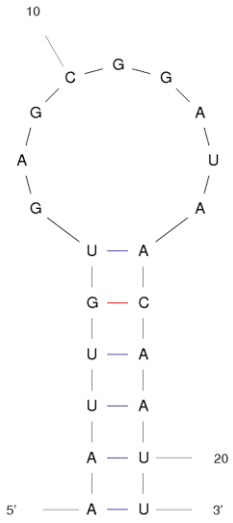

Figure 2.4. RNA structure of the *lac* operator, according to Mfold (Zuker, 2003).

LtrB.LacZ.635s		EcI5.LacZ.912s	
Insert type	White colony count	Insert type	White colony count
None	22/289 (7.6%)	None	154/544 (28.3%)
LacO	68/299 (22.7%)	LacO	69/338 (20.4%)
LacO + linker	55/231 (23.8%)	LacO + linker	75/318 (23.6%)

Table 2.4. Intron integration counts for *lacZ*-targeting LtrB and EcI5 introns carrying *lac* operators.

Though these data only represent single replicates, and no statistical analysis was performed, the results suggest that the deficit in retrohoming efficiency caused by the structural rigidity of the *lac* operator sequence is slight if it exists. The difference in effect as compared to the *lox* sites analyzed in **Figure 2.1** maybe be due to the fact that the *lac* operator sequence is approximately 50% shorter than the *lox*-site sequence, while the loop region at the end of the hairpin is approximately 50% longer in the *lac* operator, resulting in a hairpin that is both smaller and more flexible at the distal end.

To examine whether the presence of *lac* operators increases the frequency of homologous recombination, we first delivered one *lac* operator on the EcI5.LacZ.912s

intron. We then subsequently used the EcI5.LacZ.1806s intron to carry a second *lac* operator, and as a control, a *lox71* site into the genome of the strain carrying a *lac* operator at the EcI5.LacZ.912s integration point. The EcI5.LacZ.1806s intron is homologous to the EcI5.LacZ.912s intron and inserts in the same orientation approximately 900 bases away (see **Appendix**), meaning that homologous recombination between the two introns will result in a deletion the intervening bases. PCR was performed using primers outside the region of the *lacZ* gene containing both integration points in order to assess deletions between the two integration points. A gel displaying the results in shown in **Figure 2.5**.

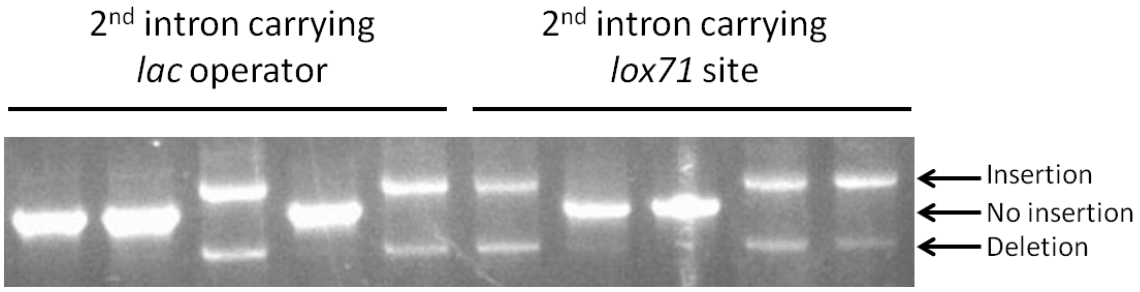

Figure 2.5. Gel of PCR amplicons for assessing formation of deletions between homologous introns carrying *lac* operators.

Figure 2.5 shows that deletions are seen upon integration of the EcI5.LacZ.1806s intron, as expected, but gives little reason to believe that the presence of *lac* operators in both integrated introns increased the rate of deletions as compared to delivering a *lox71* site, which will not be bound by the *lac* repressor, which presumably does bind the *lac* operator in the nearby EcI5.LacZ.912s intron. As a further examination, each of the five colonies found to harbor EcI5.LacZ.1806s intron insertions in **Figure 2.5** was restreaked, and the same PCR was repeated on five colonies from each of the five resulting plates. Both the insertion and deletion bands were once again seen in all cases, indicating that

the frequency of deletion within the colonies assayed in **Figure 2.5** is likely less than one in five without or without a *lac* operator in the second intron, and that isogenic isolates containing the desired deletion are not likely to be obtained without screening many colonies and/or repeated restreakings.

We thus concluded that if including *lac* operators in targetrons increases the frequency of homologous recombination, the magnitude of the increase is not enough to be useful in biotechnological applications as compared to using targetrons to deliver *lox* sites and then transforming a Cre-expressing plasmid.

2.3 DISCUSSION

The results in this chapter show that targetrons can be used to deliver a variety of genetic elements to genomic loci for applications in biotechnology and basic research. An in-depth treatment of the uses of the *lox*-site-carrying targetrons discussed in **Section 2.2.1** is found in Chapter 3. The *Ter*-site-carrying targetrons discussed in **Section 2.2.2** were used successfully to examine the effect of ectopic *Ter* sites on growth in *E. coli*. The introns targeting the safe insertion sites described in **Section 2.2.2** are also worthy of note for their potential general utility for delivering genetic cargo to the chromosome without interrupting other genetic elements in the genome. For example, these introns could be particularly useful for delivering *lox* sites for inserting large pieces of DNA via recombination-mediate cassette exchange, which is investigated in more detail in **Section 3.2.2**. Finally, while the results reported in **Section 2.2.3** did not support the hypothesis that the presence of *lac* operators increases the frequency of homologous recombination between introns, better results could potentially be obtained by including multiple *lac* operators in each intron.

Taken together, the data presented in this chapter also provides some initial guidelines for the design of targetrons carrying genetic elements. The structure of the RNA sequence inserted into the intron is important. If the structure is large and rigid, intron mobility may be significantly impaired. In particular, hairpins containing 14 or more base pairs with short loop regions on the distal end can cause major deficits, while hairpins of 6 or fewer base pairs may cause only minor impairments, or none at all. The addition of non-base-pairing loop regions, particularly at the base of the structure, can serve to mitigate deleterious effects, however.

2.4 MATERIALS AND METHODS[6]

2.4.1 Intron retargeting

Introns were designed as described elsewhere (Perutka et al, 2004; Zhuang et al, 2009). The algorithm is available at http://www.targetrons.com. Ll.LtrB-type introns were retargeted according to the Sigma Aldrich User Guide for the TargeTron Gene Knockout System (http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/General_Information/targetron-user-guide.Par.0001.File.tmp/targetron-user-guide.pdf), except that the primers were prepared differently to improve the yield of the PCR amplification. Specifically, 1 µL each of 20-µM solutions of the EBS2 and EBS2AS primers were diluted into 26 µL of water. 2 µL of this mixture and 1.4 µL each of 20-µM solutions of the IBS and EBS1 primers were used in the PCR amplification. The rest of the protocol was not substantially different from the Sigma-Aldrich protocol. Alternatively, the entire

[6] This section is adapted from Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* **9:** 685. This work is used with permission under a Creative Commons – Attribution license. The protocols here were developed by PJE, MND, and JP. The text was written by PJE with edits by other authors.

retargeted HindIII/BsrGI was ordered as a gBlock from IDT and cloned directly into the introns to be retargeted.

For the EcI5 introns, two different PCR amplifications were first executed using the IBS1/2S and EBS2AS primers in one, and the EBS1S and EBSR primers in the other. In these reactions, 2 μL each of 10-μM solutions of the two primers and at least 5 ng template (an EcI5 intron having the proper base at the +1 position) were used in 50 μL. The products were subjected to PCR clean-up, and then at least 5 ng of each were combined for use as the template of a second PCR amplification similar to the first except doubled to a total volume of 100 μL, with 8 μL of 10-μM EBSR and 2 μL of 10-μM IBS1/2S as the primers. The product was subjected to PCR clean-up, digested with AvaII and XbaI, and ligated into the EcI5 vector (having the proper base at the +1 position) cut with AvaII and XbaI.

2.4.2 Cloning of inserts into targetrons

To insert *lox*, *ter*, or *lacO* constructs into the introns, the intron plasmids were first cut with MluI in the presence of calf intestinal phosphatase. The inserts themselves were ordered as two complementary oligomers having 5'-phosphates and forming MluI sticky ends (i.e., having "CGCGT" on the 5' end and "A" on the 3' end) upon annealing. For sequences of oligomers used to make the *lox* inserts, see the online supplementary materials of Enyeart et al. (2013). The oligomers for the *Ter* inserts were designed by adding sequences for MluI sticky ends to the sequences listed in **Table 2.1**. For *lac* operator inserts, the sequences for MluI sticky ends were added to the sequence of the *lac* operator (AATTGTGAGCGGATAACAATT) plus the sequence "GAGAG" on both ends in cases where a flexible linker was added.

The oligomers for any given insert were annealed together by mixing 10 μL of 200-μM solutions of each of the oligomers with 80 μL of water, holding at 95°C for 20 minutes, and then allowing to cool (in some cases by ramping downward at 0.5°C/s until reaching 40°C, holding 20 minutes at 40°C, and then cooling at room temperature). The annealed oligomers were then ligated directly into the MluI-cut intron plasmids.

2.4.2 Intron induction

In *E. coli* strains, cells transformed with the intron expressing plasmid were grown overnight at 37°C in Luria-Bertani (LB) broth plus 34 μg/mL chloramphenicol, diluted to an OD_{600} of 0.05 in 5 mL of LB plus 34 μg/mL chloramphenicol, and then grown for one hour at 37°C. 250 μL of that culture was then inoculated into 5 mL of LB containing 200-μM IPTG (no antibiotic) and grown for 20 min (for Ll.LtrB-type introns) or 3 hours (for EcI5-type introns) at 37°C. (EcI5 introns can also be induced for the shorter time period, but efficiency is somewhat better using the longer period.) The cultures were then put on ice, and 50 μL of a ×100 dilution (for Ll.LtrB-type introns) or a ×1000 dilution (for EcI5-type introns) was then streaked on LB plates (non-selective) pre-warmed to 37°C. The plates were then incubated overnight, and intron integration was screened using colony PCR. A subset of positive colonies was then screened for loss of antibiotic resistance to indicate absence of the intron-expressing plasmid.

2.4.3 Doubling time measurements

Overnight cultures of the strains to be measured were diluted in LB to an OD_{600} of 0.001, and triplicates of 500 μL of that culture were placed in a 96-well plate (Nunc). All other wells (including all wells on edges) were filled with 500 μL of sterile media (LB). Growth was measured using a plate reader (Bio-Tek PowerWave 340), pre-heated to and

maintained at 37°C with a shaking intensity of 4 for 250 seconds at a time, with measurements taken every 300 seconds.

The results were plotted as $\log_2(OD_{600})$ versus time (min). In order to select the linear region of the curve, each point was assigned a correlation coefficient R^2 corresponding to the value of R^2 for the line consisting of that point and the three points before and after. Since variance was lower when the same time window was used for all three replicates, the resulting R^2 values were averaged for all three replicates at each time point. The longest stretch in which all these averaged R^2 values were equal to or greater than 0.99 was taken as the linear range. The slope of the least-squares linear fit of each replicate in that time range was then taken as the doubling time.

2.4.4 Statistical analyses

Statistical analyses were performed in R. For the data on the dependency of intron efficiency on insert type in **Figure 2.1**, analyses were performed on square-root-transformed data in order to obtain better homoscedasticity, which is a requirement for analysis of variance (ANOVA) comparisons (Rosner, 2011). Barlett's test for equality of variances (Rosner, 2011) gave 8.1×10^{-5} for the untransformed LtrB values, 0.30 for the transformed LtrB values, 0.10 for the untransformed EcI5 values, and 0.77 for the transformed EcI5 values. For normality tests, all data triplets gave a *P*-value of at least 0.01 in the Shapiro-Wilk normality test (Shapiro & Wilk, 1965), except the values for the 1WL1 insert in LtrB, which contained two values of zero. Replacing the LtrB.1WL1 data with values randomly selected from a normal distribution having the same mean as the actual data points (0.084%) and a standard deviation equal to 0.08% made no substantial difference in the results. Multiple pairwise comparisons for this and all other data were made using the Tukey method for correcting for multiple testing (Hsu, 1994).

An analysis of variance (ANOVA) of the results presented in **Figure 2.1** confirmed the dependence of intron efficiency on insert type (with *P*-values of 2.93×10^{-9} for LtrB.LacZ.635s and 3.04×10^{-11} for EcI5.LacZ.912s). The pairwise comparisons also confirmed that the inserts generally fall into two groups: one of approximately wild-type efficiency (with flexible structures) and one of markedly impaired efficiency (with relatively rigid structures). The *P*-values for the difference between the least efficient insert in the wild-type group and the most efficient insert in the impaired group were 6.8×10^{-5} for LtrB and 6.0×10^{-6} for EcI5.

Chapter 3: Generalized bacterial genome editing using targetrons and Cre/*lox*[7]

3.1 INTRODUCTION

Though synthetic biology has thus far been focused primarily on building circuits of small numbers of genes to perform tasks of interest (Kaern et al, 2003; Lu et al, 2009), in recent years, more interest is being taken in the genome as a whole as the unit of engineering (Dymond et al, 2011; Gibson et al, 2010; Isaacs et al, 2011). As interest in engineering bacterial genomes increases, so too will the need for efficient tools for manipulating these genomes. Though a variety of methods exist for engineering bacterial genomes (Hughes & Maloy, 2007; Miller, 1991), each has specific limitations in terms of site specificity, efficiency, versatility, and/or range of applicable bacterial species. Recombineering and related methods making use of phage recombinases have come into widespread use for small-scale modifications in *Escherichia coli* (Costantino & Court, 2003; Datsenko & Wanner, 2000; Wang et al, 2009; Yu et al, 2000), but use of this approach in other species has so far been limited and often requires developing new recombineering functions for each system (Datta et al, 2008; Swingle et al, 2010; van Kessel & Hatfull, 2008). On the other hand, site-specific recombinases such as the Cre-*lox* system are quite efficient and function in many organisms; indeed, the Cre-*lox* system has been claimed to function efficiently "in any cellular environment and on any kind of DNA" (Nagy, 2000). In bacteria the system has thus far been primarily used for selective

[7] This chapter is adapted from Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* **9:** 685. This work is used with permission under a Creative Commons – Attribution license. SMC and PJE performed the GFP and DEBS-TE insertions, which were planned by PJE, SMC, ATK-C, and ADE and analyzed by SMC and PJE. JP designed the introns and planned the *S. aureus* deletion. MND constructed most of the introns, and designed, performed, and analyzed the simultaneous triple deletion. JY, JTW, and AML designed and tested the *B. subtilis* introns, and EMQ designed and tested the *S. oneidensis* rDNA intron. PJE and ADE designed the rest of the experiments, and PJE performed and analyzed all other experiments. PJE wrote the text, and all authors participated in revising and editing the text.

marker removal, but it has, for example, been used to create large deletions in *E. coli* (Fukiya et al, 2004) and large inversions in *Lactococcus lactis* (Campo et al, 2004). However, positioning the recombination-recognition targets requires complementary genome-engineering approaches (typically with selectable markers), thus creating a chicken-and-egg problem.

Retargetable mobile group II introns are another tool that has been developed relatively recently. These so-called "targetrons" can be designed to insert into a given DNA site at efficiencies high enough that selectable markers need not be used. Mobile group II introns occur naturally in bacteria, eukaryotic organelles, and some archaea, and are thought to be precursors to the eukaryotic spliceosome (Lambowitz & Zimmerly, 2004). In these introns, the intron-encoded protein (IEP) aids in self-splicing and in the process of 'retrohoming,' in which the intron site-specifically reverse-splices into DNA.

While targetrons are conventionally used for gene knockouts, their efficiency, specificity, and broad applicability make them attractive for tandem use with other general-utility genome-engineering tools, such as site-specific recombinases. In this chapter, we used modified targetrons to efficiently carry *lox* sites to defined genomic loci and thereby developed a generalizable approach to genome editing that can be adapted with minimal modification to a wide variety of bacterial strains. We use this system, called GETR (Genome Editing via Targetrons and Recombinases), to generate large-scale chromosomal insertions, deletions, inversions, and one-step cut-and-pastes, and we demonstrate its use in the Gram-negative *Escherichia coli* and *Shewanella oneidensis* bacteria, as well as the Gram-positive *Staphylococcus aureus* and *Bacillus subtilis* bacteria.

3.2 RESULTS

3.2.1 Overview of genomic manipulations of *E. coli* chromosome

Introns were targeted to a variety insertion sites in *E. coli* (**Fig. 3.1**); these sites were chosen to flank genomic regions that had previously been shown to be non-essential and amenable to deletion (Fukiya et al, 2004; Kolisnychenko et al, 2002). A list of introns used in the present work is given in the **Appendix**. **Figures 3.2A** through **3.2D** show schematics for using this system to implement insertions, deletions, inversions, and cut-and-paste operations, respectively.

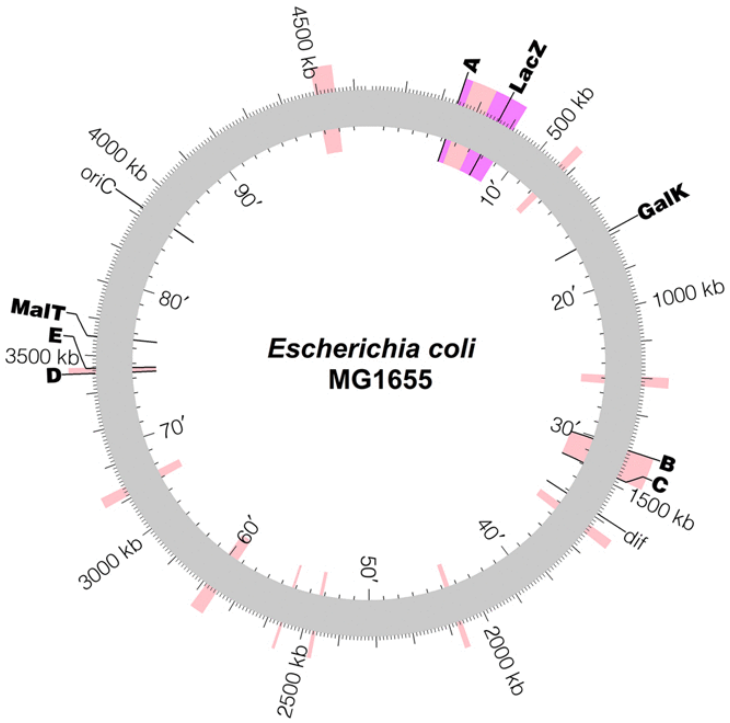

Figure 3.1. Genomic integration sites of the introns.

Insertion sites of introns used in the present work are labeled in bold type. Pink highlights are regions previously deleted by Kolisnychenko et al. (2002), and the purple highlight is a region previously deleted by Fukiya et al. (2004). The intron used for *lacZ* is Eci5.LacZ.1806s (see **Appendix**) unless otherwise noted. Image made using Circos (Krzywinski et al, 2009).

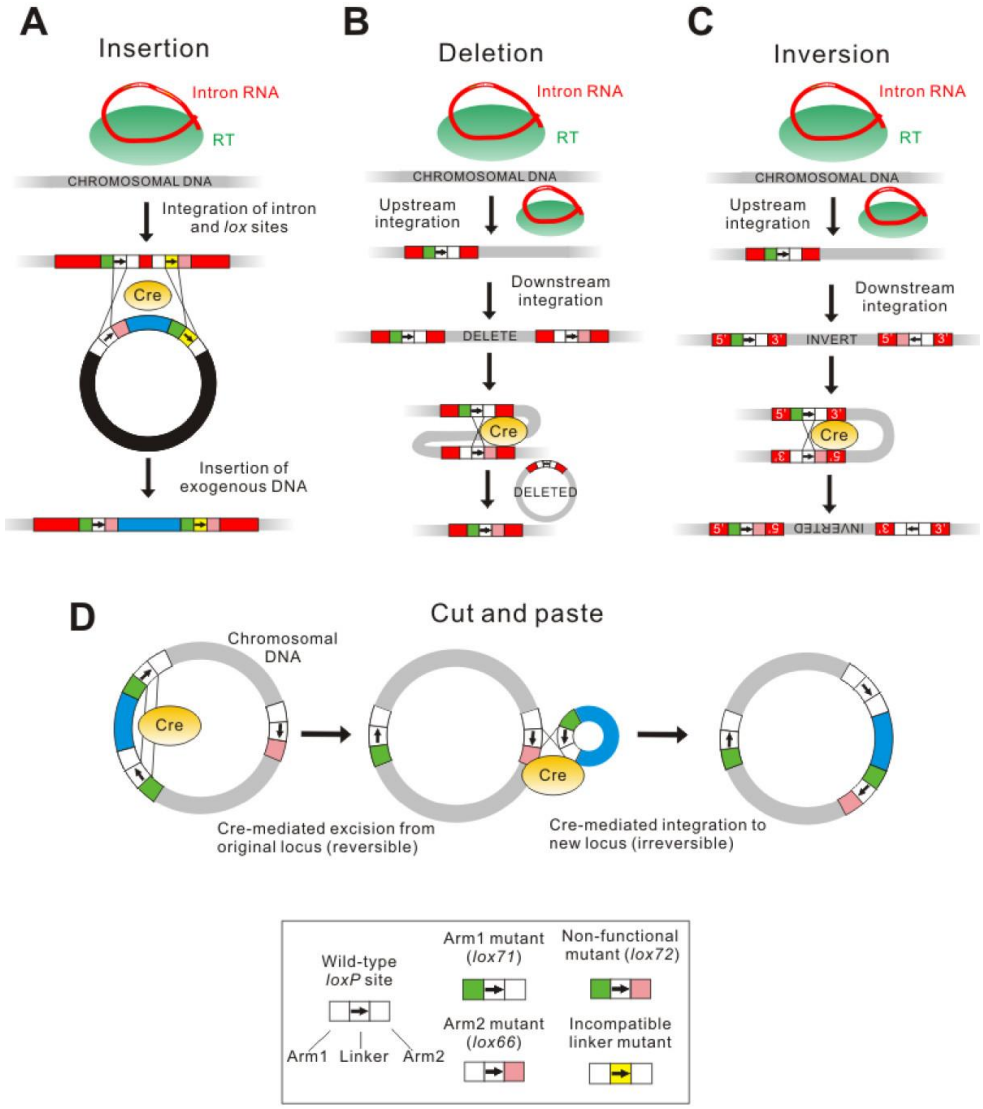

Figure 3.2. Genome edits performed.

(A) Inserting exogenous DNA (recombinase-mediated cassette exchange). Two *lox* sites having incompatible linker regions and differing arm mutations (for example, *lox71* and *lox66*) are delivered to the genome using an intron. The sequence to be inserted is then delivered between *lox* sites identical to those in the genome except having opposite arm mutations. The formation of non-functional *lox* sites (*lox72*) makes the process irreversible. **(B)** Procedure for deleting genomic sequences. After delivery of *lox* sites (*lox71* and *lox66*) on targetrons, Cre-mediated recombination then deletes the intervening region, leaving a non-functional *lox* site (*lox72*) behind. **(C)** Procedure for inverting genomic sequences. The procedure is the same as in panel **B**, except the *lox* sites have opposing orientations. **(D)** Procedure for one-step cut-and-paste after using introns to position *lox* sites (two *lox71* sites and one *lox66* site) as shown. The first (reversible) step is Cre-mediated deletion, followed by Cre-mediated reinsertion at the target site that is made irreversible by the formation of a non-functional *lox* site (*lox72*). Adapted from Enyeart et al. (2014), used with permission.

3.2.2 Insertions (recombination-mediated cassette exchange)

After targetron integration, genomic insertions were performed by recombination-mediated cassette exchange (RMCE), using the EcI5.LacZ.1806s intron to deliver an incompatible pair of *lox*P and *lox*m2 (Langer et al, 2002) sites to the genome (**Fig. 3.2A)**. The 1806s intron for inserting into *lacZ* was used for most subsequent modifications instead of the 912s intron due to its higher efficiency, approaching 97% (Zhuang et al, 2009). The use of incompatible linker mutations prevents inversion or deletion of the sequence between the *lox* sites, and the use of arm mutants makes the recombination reaction unidirectional and allows multiple insertions to be made without cross-reactivity. In order to examine the role of various experimental parameters (in particular, incubation time, copy number of the delivery plasmid, and strain background) on the efficiency of RMCE, we first delivered a T7 promoter to the genome along with the *lox* sites in the EcI5.LacZ.1806s intron, and separately provided both a promoterless GFPuv gene, flanked by *lox* sites on a pUC19 vector or a pACD vector (derived from pACYC184), and a Cre-expressing plasmid (pQL269[8] (Liu et al, 1998)). The pUC19 high-copy plasmid is present at about 500 copies per cell (Chambers et al, 1988), whereas pACYC is present at only about 20 copies per cell (Chang & Cohen, 1977). In co-transformed cells, GFP expression (via the endogenous T7 RNA polymerase) should only occur upon insertion into the genomic target site (see **Fig. 3.3A**). Two *E. coli* strains, HMS174(DE3) (a K-12 strain related to MG1655), and BL21(DE3) (a B strain), which contained intron-delivered *lox* sites were used and were plated at one, two, and three days after

[8] Concerning Cre plasmids, in initial work we made several attempts to put the *cre* gene on the same plasmid as the *lox* sites but found that the *cre* gene was quickly deleted. Sequencing plasmids harboring the deletion yielded results consistent with Cre-mediated recombination between one of the *lox* sites and the ribosome binding site of the *cre* gene. This occurred with the *cre* gene under control of the T7 promoter in a cell that did not express the T7 polymerase, and we thus abandoned further attempts to create a single-plasmid RMCE system.

transformation. Efficiency was gauged by manually counting colonies. The results are shown in **Figure 3.3B**, and the full statistical analysis of the results is presented in **Section 3.4.5**.

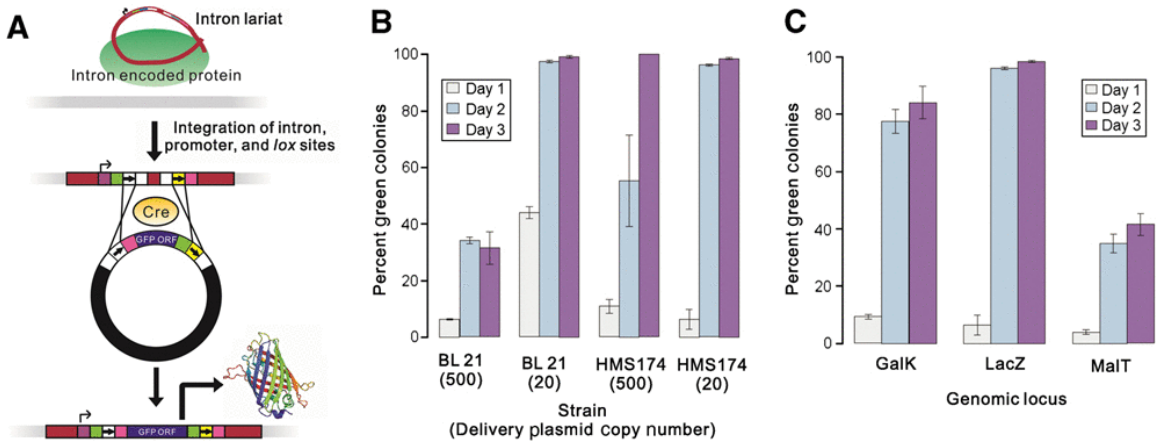

Figure 3.3. GFP reporter assay for Cre/*lox*-mediated gene insertion.

(**A**) Overview of the method. A T7 promoter is first delivered to the genome with an intron. A promoterless GFP ORF (with ribosome binding site) is then inserted via Cre/*lox*, such that GFP expression is only seen upon insertion. Color-coding as in **Figure 3.2**. (**B**) Results as a percentage of green colonies, by strain, delivery-plasmid copy number, and incubation time. Error bars are the standard error of three replicates. On day 3, the HMS174(DE3) (High) colonies were visually homogenous and were thus also assayed by PCR. (**C**) Results as a percentage of green colonies, by genomic location, in HMS174 using the lower-copy vector. The data for *lacZ* are identical to those for HMS174(DE3) (20) in (B). Error bars are the standard error of three replicates.

In interpreting these results, we first note the significant effect of increasing time on efficiency of insertion. This is likely because interaction between the delivery plasmid and the chromosome occurs at random during any given period, and thus the chance of an interaction occurring increases with time (though in general little is gained by waiting three days as opposed to two). The better performance of the lower-copy vector versus the high-copy vector is surprising at first but may be a result of the lower opportunity for Cre-mediated swapping of cassettes between plasmids in the lower-copy case. The effect of strain, which was not statistically significant except in interaction with other factors,

seems to be in modulating the influence of time and copy number. In particular, the effect of time was more pronounced in HMS174(DE3), and the effect of vector copy number was more pronounced in BL21(DE3).

We then further examined the effect of genomic location on RMCE insertion efficiency by repeating the experiment using the lower-copy pACD vector in HMS174(DE3) at two new loci, the *galK* gene and the *malT* gene. The results are shown in **Figure 3.3C**. While efficiency of integration into the *malT* locus was worse than at the other loci (see **Section 3.4.5** for a full statistical analysis), in all cases the efficiency was high enough by the second day that screening for insertions via colony polymerase chain reaction (PCR) could be easily performed.

To demonstrate not just the efficiency but the broad utility of this system, we then proceeded to insert the 12-kilobase DEBS1-TE polyketide synthase operon (Kodumal et al, 2004; Wiesmann et al, 1995) into the *lacZ* gene of *E. coli* K207-3 (Murli et al, 2003). The delivery-plasmid *lox* sites used in earlier experiments were inserted on either side of the operon in the pET26b-DEBS1-TE plasmid using conventional cloning methods. The pET vectors are built on the pBR322 backbone (Rosenberg et al, 1987), which is similar in copy number to the pACYC backbone used for the pACD plasmids (Green & Sambrook, 2012). Insertion of the entire operon into the *lacZ* gene was facile (as judged by PCR across an insertion junction) and also showed a trend of increased insertion efficiency as a function of incubation time; after three days of incubation, 25/25 screened colonies tested positive for the insertion. One of the features of this manipulation is that insertion also inactivated the *LacY* gene. Thus IPTG sensitivity is reduced due to a smaller amount being transported into the cell (Mehdi et al, 1971), and protein production can be more precisely modulated, resulting in a lower fitness load on the cell. When two of these colonies were screened more fully using overlapping PCR amplifications that

were subsequently sequenced, both were found to in fact contain the entire polyketide synthase operon without error (see **Fig. 3.4**). In principal, insertions of any size could be made at similar efficiency, limited only by the constraints of genome structure (Esnault et al, 2007).

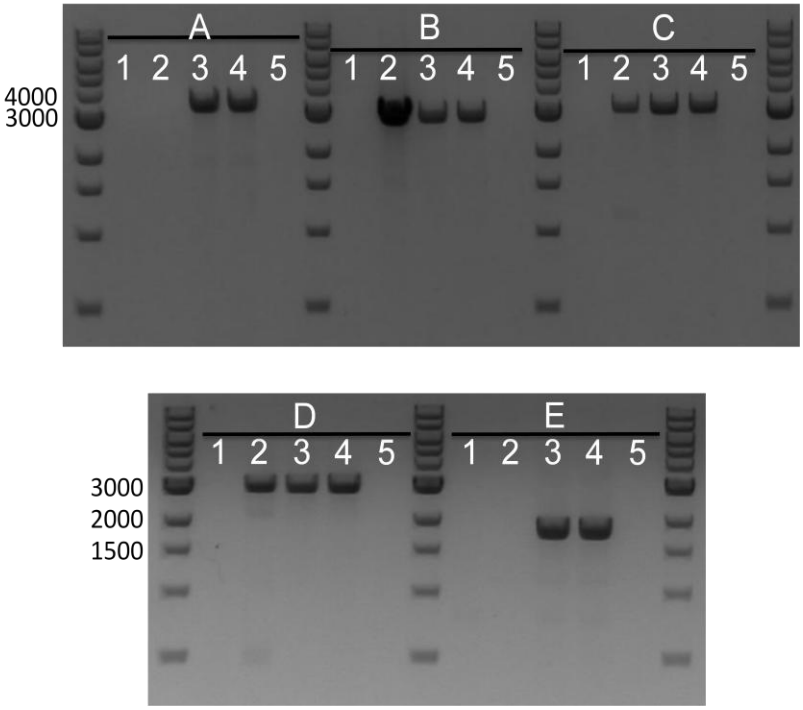

Figure 3.4. Verification of DEBS1-TE (polyketide synthase operon) genomic insertion.

A through E are overlapping PCRs covering the entire 12-kb operon, where the A and E PCRs in particular also amplify the flanking DNA intron sequence in the genome and should only be seen upon successful insertion. Lane 1: Unmodified *E. coli* K207-3; Lane 2: Plasmid pET26b-DEBS1TE-i (DEBS1-TE delivery plasmid); Lane 3: DEBS1-TE insertion clone 1 (*E. coli* K207-3 base strain); Lane 4: DEBS1TE insertion clone 2 (*E. coli* K207-3 base strain); Lane 5: Negative control (water). All bands are of the expected sizes and were further verified by sequencing. The primers used are listed in the supplementary material of Enyeart et al. (2013), available online.

We have devised a set of vectors to facilitate the use of RMCE insertion of cassettes into the genome: pX10, pX11, pX20, and pX21. All the vectors contain the *sacB* gene for counter-selection on sucrose, as well as two T7 terminators between the *sacB* gene and the *lox* sites to prevent read-through from the *sacB* promoter to the

delivery target between the *lox* sites. Vectors pX10 and pX11 contain the incompatible *lox* pair of *loxFAS/71* and *lox511/66* for use with the 2ML4 pair of *lox* sites (see **Fig. 2.1**) at the integration site, and vectors pX20 and pX21 contain the incompatible *lox* pair of *loxm2/71* and *lox66* for use with the 2ML5 pair of *lox* sites (see **Fig. 2.1**) at the integration site. For ease of cloning, vectors pX10 and pX20 contain a PmeI restriction site between the *lox* sites, while vectors pX11 and pX21 have the multiple cloning site from the pUC vectors between the *lox* sites.

3.2.3 Deletions

We used the same methods to demonstrate larger scale manipulations of the cellular genome of *E. coli* MG1655(DE3). First, we attempted large-scale deletions (**Fig. 3.2B**). The *A*-*lacZ*, *D*-*E*, and *B*-*C* regions (see **Fig. 3.1**) were deleted both sequentially (in the order given) and simultaneously. A set of three mutually incompatible *lox* sites (*loxP*, *lox2272*, and *loxN* (Livet et al, 2007)) was used for the simultaneous triple deletion. The use of arm mutants once again allowed multiple deletions to be made without cross-reactivity.

Screening for the deletions was performed via colony PCR as depicted in **Figure 3.5A**. Three PCR amplifications can be used to characterize each deletion: two amplicons that bridge the genomic sites at which *lox*-carrying targetrons are inserted, and one amplicon that bridges the expected deletion between those sites. The first two PCRs testing for insertion are expected to give relatively small bands (several hundred base pairs) when performed on the wild-type strain and larger bands (about 1 kilobase larger) when performed on a strain harboring targetrons at the expected sites. Upon successful deletion of the intervening region, all of these bands should no longer be generated. Instead, the PCR that bridges the two sites should yield a new band of a predicted size.

Doing all three PCR amplifications on all three strains (wild-type, wild-type harboring insertions, and recombined (induced with Cre)) provided clear diagnostic signatures of the recombination events. When artifact bands near the sizes of expected bands were observed, these were further analyzed by sequencing to ensure that they did not represent an off-target rearrangement. The predicted sizes of all of the expected amplicons of the present work (as shown in the gels of **Figures 3.5** through **3.10**) are listed in **Table 3.1**.

The gels used for verifying the occurrence of these deletions are shown in **Figure 3.5**. In particular, **Figure 3.5B** shows the three PCR amplifications performed on the three successive, engineered strains that ultimately resulted in a deletion of 121 kilobases between the *A* and *lacZ* loci. The deletion-bridging amplicon (Au/Ld-I in **Fig. 3.5B**) from the strain that was finally exposed to the Cre protein (*E. coli* MG1655 E1) was sequenced and found to conform to expectations.

As deletions are added, verification becomes more complex but is performed in exactly the same manner. **Figure 3.5C** shows the PCR amplifications for verifying the sequential double-deletion strain (*E. coli* MG1655 E6) that harbors deletions of both the *A-lacZ* and *D-E* regions. The Eu/Dd-I band for verifying the deletion of the *D-E* region is small because the recombination results in an inverted repeat of intron sequences that is subsequently removed by homologous recombination. Sequencing results confirmed this interpretation. There were some unexpected PCR-amplified bands. These bands, in particular the Eu/Dd-I, Au/Ad-U, and Au/Ad-I bands, were sequenced and found to match genomic sequences from unrelated regions, and not to off-target rearrangements.

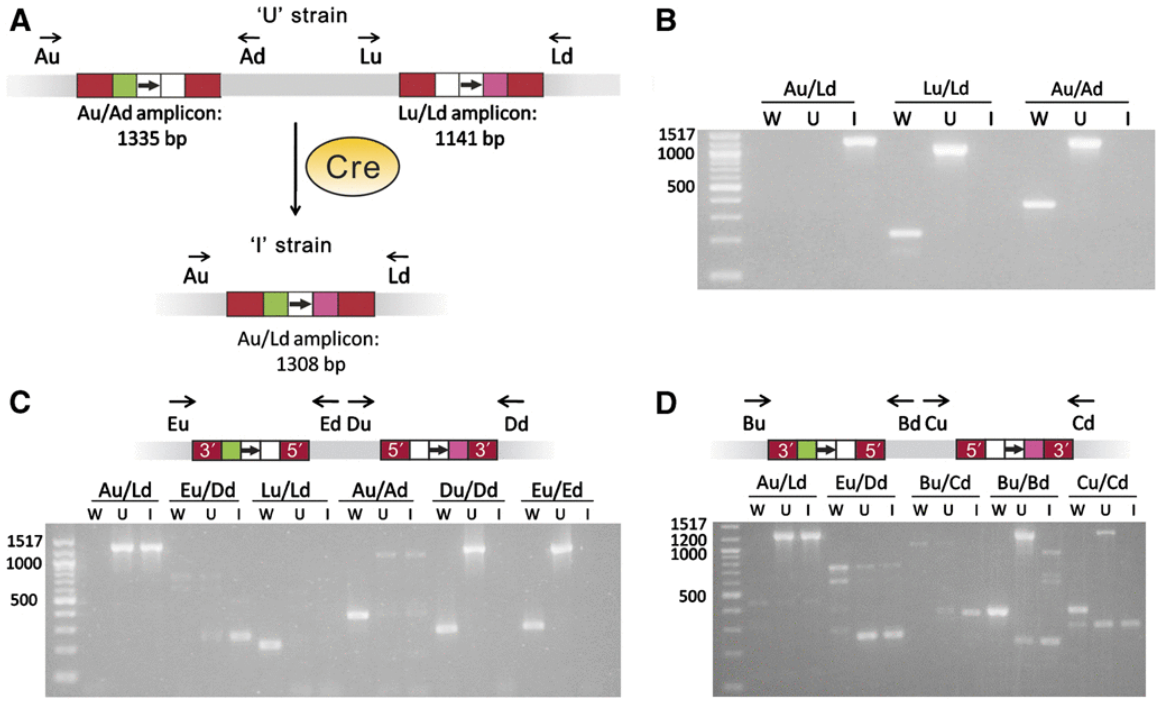

Figure 3.5. Verification of genomic deletions.

In the figure, 'W' refers to the wild-type *E. coli* strain MG1655(DE3); 'U' refers to the relevant uninduced strain, in which introns and *lox* sites have been placed but Cre has not been added; and 'I' refers to the induced strain, which results from Cre-mediated recombination of the 'U' strain. For primers, the first letter indicates the genomic location the primer amplifies (where 'L' refers to the *lacZ* locus), and the subsequent 'u' or 'd' designates the primer as 'up' or 'down.' PCR products are designated by the two primer names separated by a slash. '5′' or '3′' refers to the sense strand of the intron. (**A**) Methodology, using the deletion of the *A-lacZ* region as an example. (**B**) Verification of the strain (*E. coli* MG1655 E1) containing a deletion of the *A-lacZ* region, as shown in (**A**). (**C**) Verification of the sequential double-deletion strain (*E. coli* MG1655 E6), with schematic corresponding to the 'U' strain. 'U' here is *E. coli* MG1655 E1 with introns inserted to delete the *D-E* region. The Eu/Dd PCR amplifies the *D-E* deletion site (the *D-E* deletion leaves an inverted repeat behind). (**D**) Verification of the sequential triple-deletion strain (*E. coli* MG1655 E10), with schematic corresponding to the 'U' strain. 'U' here is *E. coli* MG1655 E6 with intron insertions for the deletion of the *B-C* region. Bu/Cd amplifies the *B-C* deletion site (the *B-C* deletion leaves behind an inverted repeat).

Primers	Event detected	Expected size (bp)	Inverted repeat generated[9]	Relevant figures
Au/Ad	LtrB insertion	379 (before) 1335 (after)	No	3.5B, 3.5C, 3.6B, 3.7A, 3.7B
Bu/Bd	EcI5 insertion	367 (before) 1286 (after)	No	3.5D
Cu/Cd	EcI5 insertion	388 (before) 1307 (after)	No	3.5D
Du/Dd	EcI5 insertion	285 (before) 1204 (after)	No	3.5C, 3.6C
Eu/Ed	EcI5 insertion	284 (before) 1203 (after)	No	3.5C, 3.6C, 3.6D, 3.7A, 3.7B
Lu/Ld	EcI5 insertion	222 (before) 1141 (after)	No	3.5B, 3.5C, 3.6B, 3.6D, 3.7A, 3.7B
Lu_0/Ld_0	LtrB insertion	246 (before) 1226 (after)	No	3.6D
Iu/Id	LtrB insertion	357 bp (before) 1312 (after)	No	3.8B
SDu/SDd	LtrB insertion	514 (before) 1469 (after)	No	3.8B
Ad/Ld	Inversion	1747	No	3.6B
Au/Ld	Deletion	1308	No	3.5B, 3.5C, 3.5D, 3.7A, 3.7B
Au/Lu	Inversion	729	No	3.6B
Bu/Cd	Deletion	644	Yes	3.5D
Du/Eu	Inversion	1268	No	3.6C
Ed/Ad	Cut-and-paste	1785	No	3.7B
Ed/Dd	Inversion	1139	No	3.6C
Eu/Ad	Cut-and-paste	1192	No	3.7A
Eu/Dd	Deletion	546	Yes	3.5C, 3.5D
Eu/Ld	Inversion	1165	No	3.6D
Eu/Ld_0	Inversion	1093	No	3.6D
Eu/Lu	Cut-and-paste	586	Yes	3.7B
Lu/Ed	Inversion/cut-and-paste	1179	No	3.6D, 3.7A
Lu_0/Ed	Inversion	1336	No	3.6D
Iu/SDd	Deletion	1432	No	3.8B
SBd/Yd	Inversion	783	Yes	3.9B

Table 3.1. Expected sizes of amplicons for verifying intron insertions and Cre/*lox* recombinations.

[9] If an inverted repeat is generated, the actual size will usually be significantly smaller than the expected size due to loss of most of the repeated sequences via homologous recombination.

Primers	Event detected	Expected size (bp)	Inverted repeat generated	Relevant figures
sorrsAu/ltrbint30r	LtrB insertion	1354	No	3.10B, C
sorrsBu/ltrbint30r	LtrB insertion	1317	No	3.10B, C
sorrsCu/ltrbint30r	LtrB insertion	1328	No	3.10B, C
sorrsDu/ltrbint30r	LtrB insertion	1166	No	3.10B, C
sorrsEu/ltrbint30r	LtrB insertion	1345	No	3.10B, C
sorrsFu/ltrbint30r	LtrB insertion	1329	No	3.10B, C
sorrsGu/ltrbint30r	LtrB insertion	1592	No	3.10B, C
sorrsHu/ltrbint30r	LtrB insertion	1624	No	3.10B, C
sorrsIu/ltrbint30r	LtrB insertion	1340	No	3.10B, C

Table 3.1, cont.

Similarly, **Figure 3.5D** shows the PCR amplifications for verifying the sequential triple-deletion strain (*E. coli* MG1655 E10) that contains a deletion of the *B-C* region in addition to the *A-lacZ* and *D-E* regions. The unexpected band in the Bu/Cd-U lane was sequenced and found to be from an unrelated genomic region. The Bu/Cd-I amplicon that confirms the deletion also represents the formation and removal of an inverted repeat, which was confirmed by sequencing. The simultaneous triple deletion strain (E9) and a strain containing a single deletion of the *D-E* region (E11) were verified in the same manner.

As was the case with insertions, the efficiency of deletions approached 100%, with the expected deletion being found in every colony tested. Off-target recombination was rare; in strains designed for deletion, only 1/60 was found to have an inversion when screened after Cre induction; recombination between *lox*72 sites was not detected in any of the modifications reported herein. The removal of inverted repeats upon formation of the the *D-E* and *B-C* deletions is interesting in that the *lox* sites are removed entirely and the size of the scar is reduced from hundreds to tens of base pairs.

3.2.4 Inversions

GETR proved to be robust for other types of recombination that are not easily achieved by other methods. The same methods used to detect deletions can be used to detect inversions, except that four different PCRs are used to verify an inversion: two amplicons bridging the insertions sites and two amplicons bridging the new ends of the inversion (see **Fig. 3.6A**). Several inversions (see **Fig. 3.2C**) were executed: namely, between the *A*-*lacZ*, *B*-*lacZ*, *E*-*lacZ*, and *D*-*E* loci. All colonies screened by PCR soon after addition of Cre tested positive for the expected inversion. Some inversions were only detected immediately after adding Cre and were not detected at later time points. This is in line with previous studies of inversions in the *E. coli* genome, some of which are not well tolerated (Esnault et al, 2007). Inversions into the *lacZ* locus were transient when an inverted repeat was formed and subsequently deleted but were stable when non-homologous introns were used, suggesting that the intron sequences may function as a buffer against otherwise deleterious rearrangements at this site.

Recombination back to the original state via homologous recombination of the introns could be detected in some cases but was not seen when non-homologous introns were used. In other words, inversions between *lox* sites in homologous introns may be reversible, but inversions between *lox* sites in non-homologous introns are irreversible. We also tested for the presence of uninverted chromosomes soon after induction of an irreversible inversion between the *E* and *lacZ* loci. All ten colonies assayed tested positive for both inverted and uninverted chromosomes, though uninverted chromosomes were not found after restreaking.

Gels for verifying the stable inversions are found in **Figure 3.6**. **Figure 3.6B** shows the PCRs used to verify an inversion between the *A* and *lacZ* loci, as depicted in **Figure 3.6A**. **Figure 3.6C** shows an analogous set of PCRs for verifying an inversion

between the *D* and *E* loci. Since these introns are present in opposite orientations upon integration, inversions can occur via homologous recombination, and this inversion is in fact detected at low levels prior to the addition of Cre. Similarly, the uninverted (reinverted) state can still be detected after the addition of Cre. These bands were confirmed by sequencing. In those strains where inversion had occurred in the absence of Cre, unrecombined *lox* sites were found, while in those strains where inversions back to the wild-type state had apparently occurred after the induction of Cre, recombined *lox* sites were found. These results are consistent with homologous recombination between the introns rather than catalyzed recombination between the *lox* sites. No artifacts were seen in these instances, consistent with the presence of a template that could be amplified by the primers.

Figure 3.6D demonstrates the difference between using homologous introns (where both introns are of the EcI5 type) versus non-homologous introns (one EcI5, one Ll.LtrB) for delivering *lox* sites to create substantially identical inversions. When non-homologous introns are used, as in the E5 strain, the inversion is only detected (via the Lu_0/Ed-I and Eu/Ld_0-I bands, which are of the expected size and were confirmed by sequencing) after adding Cre, and reversions back to the original state via homologous recombination were not detected. However, when homologous introns were used, as in the E2 strain, PCR products that should only have been seen upon inversion were also seen in the absence of Cre, and furthermore the uninverted state could still be detected after the addition of Cre, consistent with homologous recombination between introns. The bands of the sizes expected for recombination events, whether due to Cre-*lox* or homologous recombination, were confirmed by sequencing.

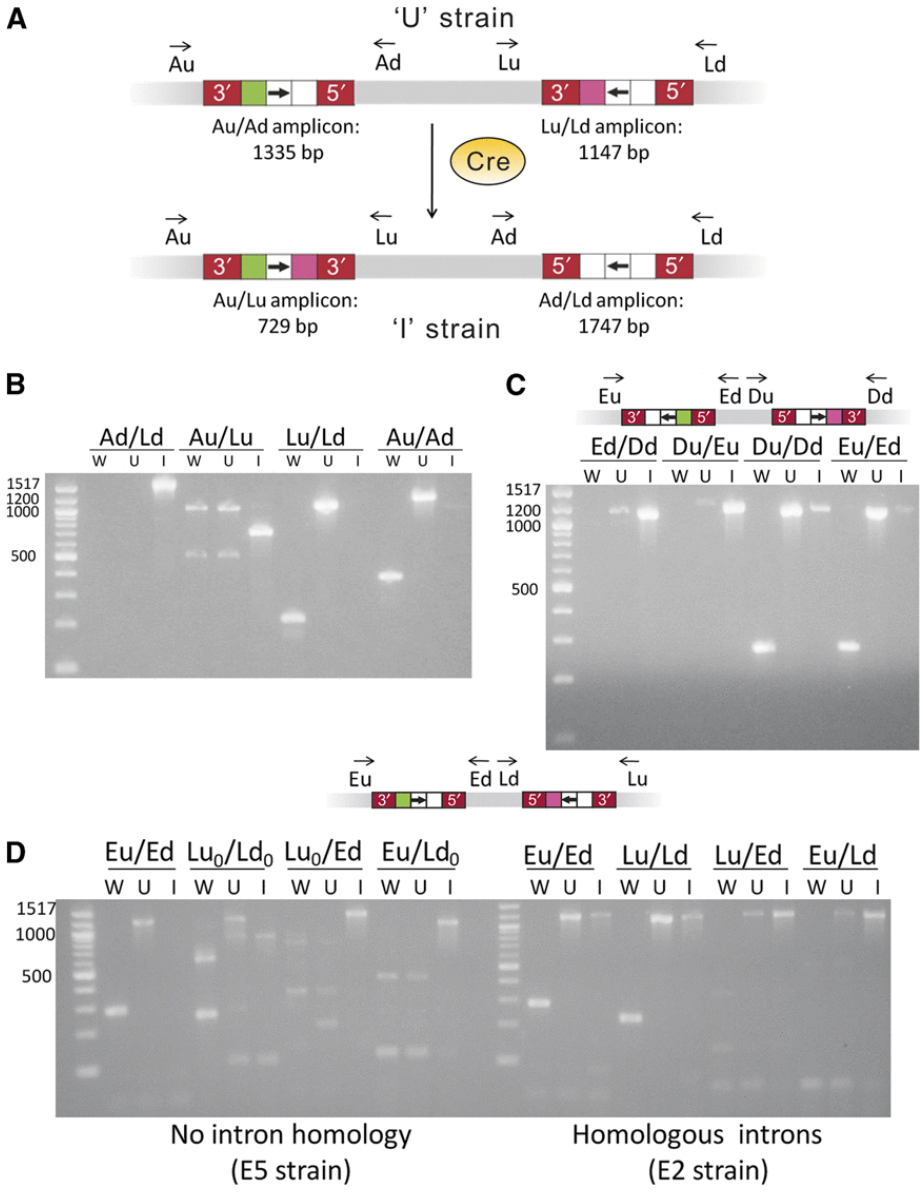

Figure 3.6. Verification of genomic inversions.

Letter designations are as in **Figure 3.5**. (**A**) Methodology, using the inversion of the *A-lacZ* region as an example. '5′' or '3′' refers to the sense strand of the intron. (**B**) Verification of the strain containing an inversion of the *A-lacZ* region (*E. coli* MG1655 E3), as shown in (**A**). (**C**) Verification of the strain containing an inversion of the *D-E* region (*E. coli* MG1655 E4), with schematic corresponding to the 'U' strain. (**D**) Comparison of a strain containing an inversion of the *E-lacZ* region using homologous introns (*E. coli* MG1655 E2) and a strain containing the same inversion using non-homologous introns (*E. coli* MG1655 E5), with schematic corresponding to the 'U' strains. The subscript in Lu_0 and Ld_0 signifies that these primers amplify the insertion site of the LtrB.LacZ.635s intron rather than the EcI5.LacZ.1806s intron used elsewhere.

3.2.5 One-step cut-and-paste

Finally, we used combinations of three *lox* sites to effect unique one-step cut-and-paste reactions (**Fig. 3.2D)**. We use the term "one-step" because the designated region moves directly from one part of the genome to another upon adding Cre, without the requirement for a stable intermediate, such as a plasmid, to act as a shuttle. In particular, we transferred the *D-E* region to the *B* locus, and the *A-lacZ* region to the *E* locus. Six different PCR amplifications were used to validate a stable cut-and-paste reaction, as shown in **Figure 3.7**: three amplicons bridging the three intron integration sites (the left set of three triplets in **Figures 3.7A** and **B**), one amplicon bridging the site of the "cut" (the fourth set of triplets in **Figs. 3.7A** and **B**), and two amplicons bridging the boundaries of the "paste" region (the fifth and sixth set of triplets in **Figs. 3.7A** and **B**). The *D-E* to *B* transfer was detected (via the "cut" and "paste" bridging amplicons) in every colony soon after exposure to Cre but was not stable upon restreaking.

The *A-lacZ* region was stably translocated to the *E* locus in both possible orientations, with 3/5 and 4/5 colonies positive for the translocations after overnight growth in liquid culture after transformation. Gels for verifying these recombinations can be found in **Figure 3.7**. In the case shown in **Figure 3.7A**, intron homology allows inversions to occur back and forth between the *lacZ* and *E* loci, but the complete cut-and-paste was only seen upon addition of the Cre protein. **Figure 3.7B** shows a similar case where the orientation of the *lox* site at the target (*E*) locus is reversed with respect to the case shown in **Figure 3.7A**. The expected rearrangement was obtained, but since the insertion at *E* is in the opposite orientation, inversions and reversions resulting from intron homology were avoided. The key bands for confirming the cut-and-pastes (Au/Ld-I, Lu/Ed-I, Eu/Ad-I, Eu/Lu-I, and Ed/Ad-I) as well as the bands expected to

result from homologous recombination (Lu/Ld-U and Lu/Ed-U in **Fig 3.7A**) were confirmed by sequencing.

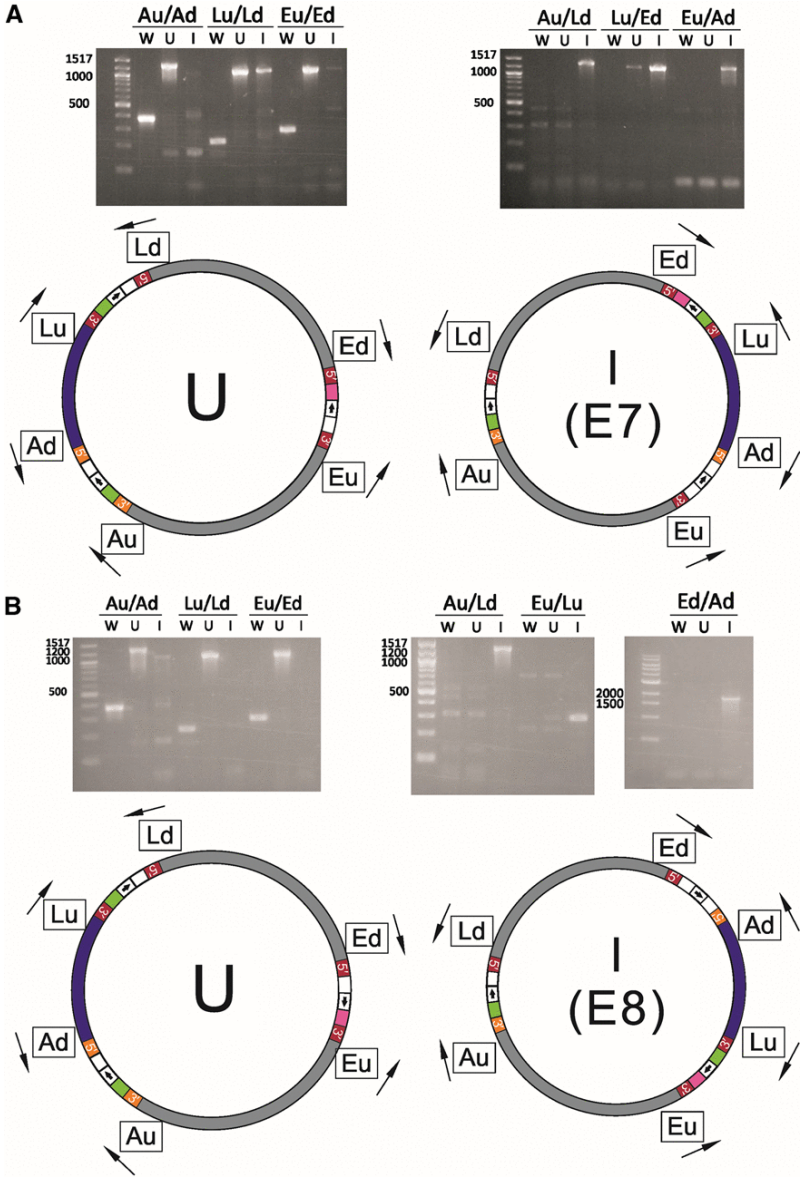

Figure 3.7. Verification of one-step cut-and-pastes.

Orange is Ll.LtrB intron sequence, which is non-homologous with respect to EcI5 intron sequence shown in red. Letter and number designations are as in **Figure 3.5**. (**A**) Verification of the strain (*E. coli* MG1655 E7) containing a cut-and-paste (translocation) of the *A-lacZ* region to the *E* locus in the reverse orientation. (**B**) Verification of the strain (*E. coli* MG1655 E8) containing a cut-and-paste (translocation) of the *A-lacZ* region to the *E* locus in the forward orientation.

Loci recombined[10]	Largest distance between recombined loci	Type of modification	Inverted repeat generated	Stability[11]	Recombination in absence of Cre[12]
A-lacZ	121 kb	Deletion	No	Stable	Yes
A-lacZ	121 kb	Deletion	No	Stable	No
B-C	81 kb	Deletion	Yes	Stable	No
D-E	14 kb	Deletion	Yes	Stable	No
A-lacZ	121 kb	Inversion	Yes	Unstable	No
A-lacZ	121 kb	Inversion	No	Stable	No
B-lacZ	1 Mb	Inversion	Yes	Unstable	No
E-lacZ	1.5 Mb	Inversion	No	Quasi-stable	Yes
E-lacZ	1.5 Mb	Inversion	No	Stable	No
D-E	14 kb	Inversion	No	Quasi-stable	Yes
A-lacZ to *E*	1.5 Mb	Cut-and-paste	Yes	Stable	Some
D-E to *B*	2.1 Mb	Cut-and-paste	Yes	Unstable	Some

Table 3.2. Summary of intra-genomic rearrangements in *E. coli*.

3.2.6 Growth of *E. coli* strains with chromosomal rearrangements

A summary of the genomic rearrangements generated in *E. coli* is given in **Table 3.2**, and a list of *E. coli* strains containing these rearrangements is given in **Table 3.3**, along with doubling times measured for these strains as a proxy for fitness. A statistical analysis of the doubling times indicated that the strains fall broadly into two groups, one group having approximately wild-type growth, and the other group having impaired growth (see **Section 3.4.5**). All strains that either lacked the *A-lacZ* region or had an

[10] The LtrB.LacZ.635s intron was used in the first A-lacZ deletion (with recombination in the absence of Cre) and in the stable A-lacZ and E-lacZ inversions; the EcI5.LacZ.1806s intron was used in all other cases (see **Appendix**).

[11] "Stable" means the recombination remained present unchanged through multiple rounds of regrowth. "Unstable" means the recombination was detected initially but was not detected after multiple rounds of regrowth. "Quasi-stable" means the recombination was still detected after multiple rounds of regrowth, but back-recombination due to homologous recombination was also detected.

[12] Recombination was only seen in the absence of Cre when (1) the introns were homologous and (2) the introns were oriented so as to allow the homologous recombination to occur. "Some" recombination in the absence of Cre for the cut-and-pastes refers to the fact that inversions caused by homologous recombination were detected, but the complete cut-and-paste did not occur without Cre.

inversion between the *E* and *lacZ* regions showed significantly impaired growth. Interestingly, the strains containing a cut-and-paste of the *A-lacZ* region to the *E* site displayed wild-type growth rates.

Strain	Doubling Time	Standard Error[13]	Description
MG1655	24.94	0.1	
MG1655(DE3)	24.44	0.2	base strain for E1-E11
E1	28.33	0.5	A-lacZ deletion
E2	33.25	0.6	E-lacZ inversion (reversible)
E3	22.03	0.1	A-lacZ inversion (irreversible)
E4	24.65	0.4	D-E inversion (reversible)
E5	30.5	1.2	E-lacZ inversion (irreversible)
E6	33.94	0.8	lacZ-A, D-E deletion
E7	25.24	0.5	lacZ-A region to E, reverse orientation
E8	23.65	0.4	lacZ-A region to E, forward orientation
E9	29.27	1.0	lacZ-A, D-E, B-C simultaneous deletion
E10	31.39	0.2	lacZ-A, D-E, B-C sequential deletion
E11	24.45	0.5	D-E del

Table 3.3. Doubling times of *E. coli* strains with intra-genomic rearrangements.

3.2.7 Genome engineering in diverse bacteria

While GETR is obviously broadly useful for creating virtually any type of rearrangement, the real utility of the method appears when moving beyond *E. coli* as a model system. We therefore applied the method to making genomic modifications in three additional phylogenetically diverse species: *Staphylococcus aureus*, *Bacillus subtilis*, and *Shewanella oneidensis*.

The Gram-positive bacterium *Staphylococcus aureus* is an intensely studied human pathogen, and the rise of drug-resistant strains in recent years has given new urgency to the development of prophylactic and therapeutic approaches to treatment (Otto, 2012). We therefore attempted to delete the 15-kilobase *Staphylococcus aureus*

[13] Error is from three replicates.

pathogenicity island 1 (SaPI-1) from *Staphylococcus aureus* RN10628 (Ubeda et al, 2009), in order to create a strain that might serve as a live vaccine. Highly efficient Ll.LtrB-type introns were generated that could integrate in the first gene (*int*) of SaPI-1 and also downstream of the pathogenicity island. After transformation of the Cre-expressing plasmid pRAB1 (Leibig et al, 2008), 40/40 colonies tested contained cells harboring the expected deletion, and 19/40 colonies tested still harbored SaPI-1. The deletion was detected and verified via PCR and sequencing (see **Fig. 3.8**). No chromosomes containing SaPI-1 were detected in restreaked colonies. The deletion was stably maintained.

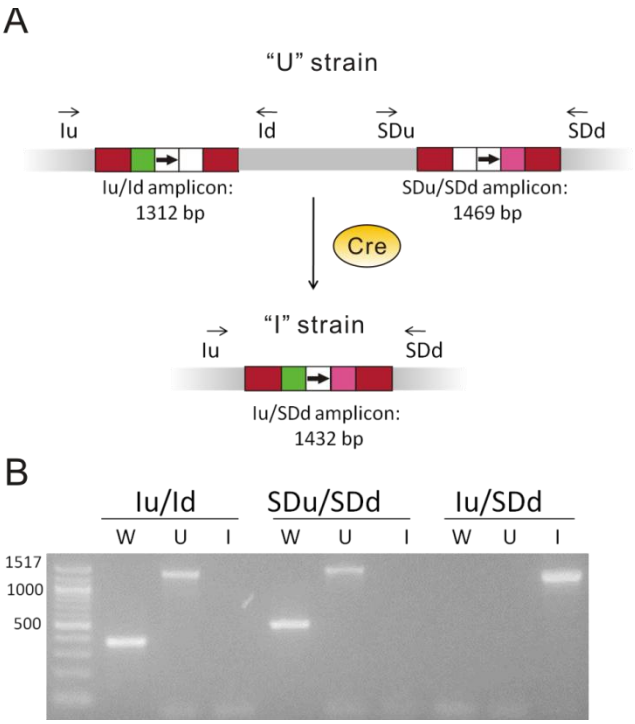

Figure 3.8. Deletion in *Staphylococcus aureus*.

Letter designations are as described in **Figure 3.5**. **(A)** Methodology, showing schematics of the PCRs used to verify the deletions, where Iu and Id primers amplify the *int* insertion site, and the SDu and SDd primers amplify the *SAPI-B* insertion site. **(B)** Verification of the strain (*S.aureus* RN10628 E1) containing a deletion of the SaPI (*int*/*SAPI-B*) region, as shown in (A). The Iu/SDd-I band was further verified by sequencing.

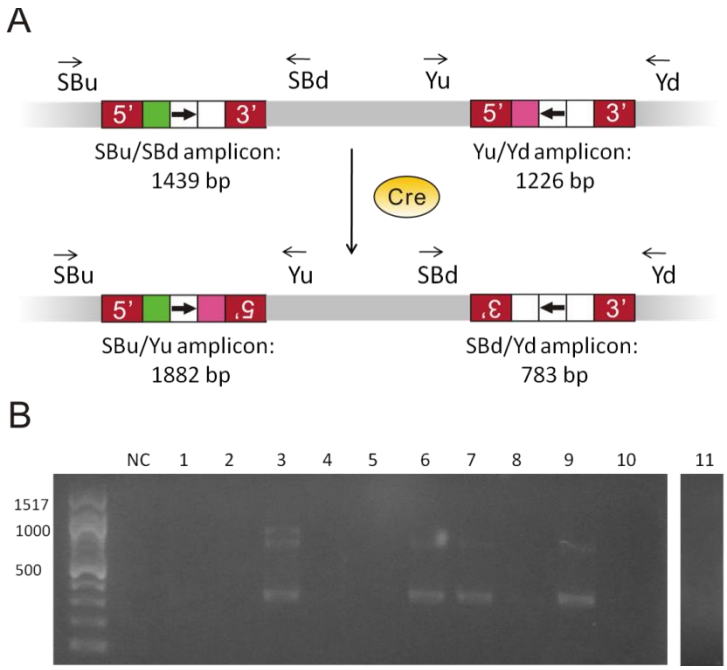

Figure 3.9. Inversion in *Bacillus subtilis*.

Letter designations are as described in **Figure 3.5**. **(A)** Methodology, where the SBu and SBd primers amplify the *sacB* intron insertion site, and the Yu and Yd primers amplify the *yhcS* insertion site. **(B)** Screening for inversions via PCR using the SBd/Yd primer pair on *B. subtilis* colonies containing intron insertions as depicted at the top in (A), after the addition of the Cre-expressing plasmid. The negative control (NC) was the same strain, except without the addition of Cre. The smaller, brighter bands are consistent with deletion of the inverted repeat formed by the inversion, but dimmer bands corresponding to the expected amplicon size are seen in all four lanes that gave bands. (The source of the uppermost band in lane 3, but it is assumed to be an artifact.) All four PCR products were sequenced, and the results confirmed the occurrence of the expected inversion between the *sacB* and *yhcS* loci followed by removal of the intron and *lox* sequences by homologous recombination. These bands were not found in individual colonies upon restreaking, and thus the inversion was judged to be unstable.

Bacillus subtilis is a model system for the study of Gram-positive bacteria, including studies on sporulation (Earl et al, 2008; Higgins & Dworkin, 2012). We designed and built two Ll.LtrB-type introns that inserted into the sense strands of the *sacB* and *yhcS* (*srtA*) genes of *B. subtilis* at efficiencies of 98% and 91%, respectively (Whitt, 2011; Yao, 2008). We used these introns to deliver a *lox*71 site to the *sacB* locus and a *lox*66 site to the *yhcS* locus in *B. subtilis* 168, positioning the intervening region for inversion. Upon addition of the Cre-expressing plasmid pCrePA (Pomerantsev et al,

2006), 4/11 screened colonies tested positive for the inversion via colony PCR (**Fig. 3.9**). Sequencing of the PCR products gave a sequence consistent with the expected inversion. The inversion covers about 1.5 Mb of the 4.2 Mb genome and was not seen upon restreaking, indicating that it was not well tolerated.

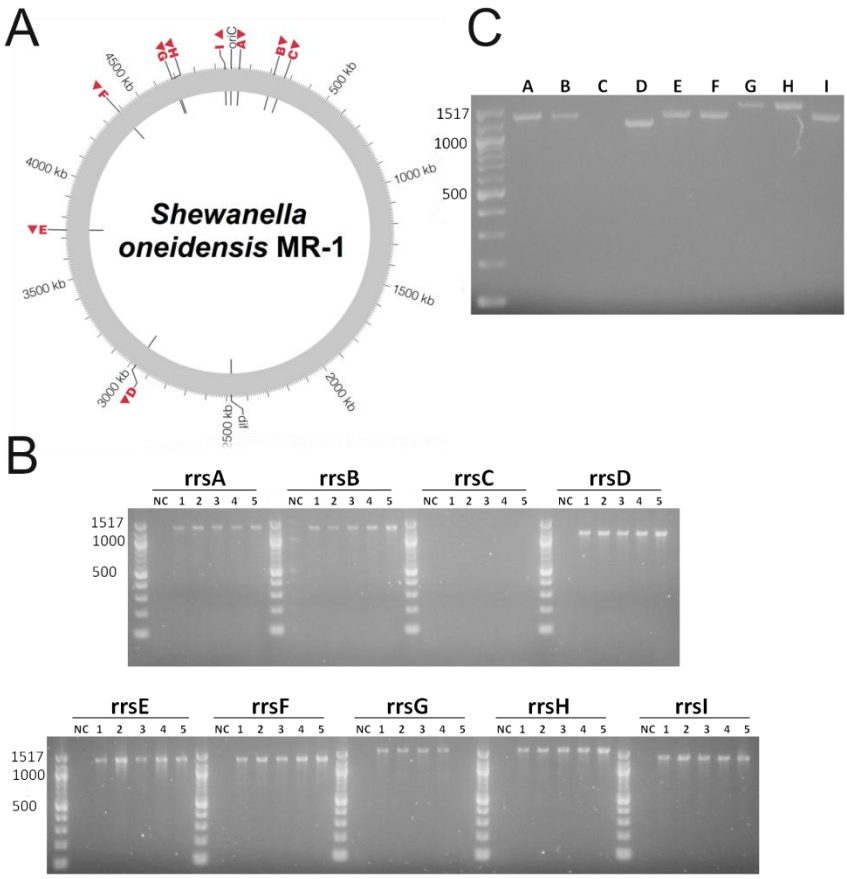

Figure 3.10. Modifications in *Shewanella oneidensis*.

(A) Schematic of the *S. oneidensis* genome, showing locations and orientations of the *rrs* genes. **(B)** The results of PCR amplifications to determine intron insertions into each *rrs* gene in five isolates of *S. oneidensis* transformed with RP4.T5.rDNA.798s.1WL2R. One primer binds to intron sequence, and the other binds to a unique genomic region outside the *rrs* gene. The "NC" lanes were performed on untransformed cells. **(C)** A repetition of the PCRs in (B) performed on a single colony grown by inoculating isolate 5 into liquid culture and streaking the overnight culture on a plate.

Shewanella oneidensis is a Gram-negative bacterium that is a model system for extracellular electron transfer, with potential applications in bioremediation and energy (Fredrickson et al, 2008). We initially designed EcI5 introns to insert into the *nrfA* gene, which has been shown to increase current production when inactivated (Bretschger et al, 2007), but were unable to find evidence of integration. We then borrowed an Ll.LtrB targetron designed by Erik Quandt to insert into the ribosomal *rrs* genes (*rrsA* through *rrsI*) in *S. oneidensis* (see **Fig. 3.10A**). A *loxP* site was inserted into the MluI site of the intron to facilitate subsequent genomic rearrangements. The intron, which was cloned onto the broad host-range plasmid RP4, was introduced into *S. oneidensis* via conjugation. We screened single colonies for insertions at each site by PCR using one primer complementary to the intron and another primer complementary to a unique chromosomal sequence near the insertion site. We initially found insertions in all copies of the *rrs* gene except *rrsC* (**Fig. 3.10B**); subsequent PCRs using other primers to more specifically detect insertions in *rrsC* did yield bands. We found the same pattern of PCR bands after growing one of the colonies overnight in liquid culture, freezing at -80°C, restreaking, and repeating the PCR amplifications on one of the resultant colonies (**Fig. 3.10C**).

Several attempts were made to generate rearrangements by introducing Cre-expressing plasmids into the *S. oneidensis* strain harboring *lox* sites in all nine copies of the *rrs* gene. We initially cloned the *cre* gene onto the broad-host range plasmid pBAV1K (Bryksin & Matsumura, 2010) (along with a gene encoding chloramphenicol acetyltransferase, to provide a selectable marker not found on the RP4 plasmid) to generate pBAV1KC.Cre. Despite several attempts at transforming this plasmid into the nine-*lox S. oneidensis* strain, the only chloramphenicol resistant strains that were obtained tested negative for the presence of the *cre* gene upon PCR screening. However,

the plasmid was successfully transformed into wild-type *S. oneidensis* MR-1. Reasoning that perhaps the high copy number of this plasmid caused problems, we next looked into putting the *cre* gene onto the pWV01 vector (Kok et al, 1984), which is the low-copy vector from which pBAV1K was derived. The resulting plasmid, which we called pCre.LAC, was not successfully transformed into either the wild-type or the *lox*-harboring *S. onedensis*. Finally, we examined placing the *cre* gene on the pBT-2 and pBTML-4 broad-host range plasmids (Lynch & Gill, 2006) containing the pBBR1 replicon, which has previously been used in *S. oneidensis* (Johnson et al, 2010). Thus we built plasmids pBT.Cre and pBTML.Cre, which differ from each other in that the latter contains an origin of transfer. Attempts to transform pBT.Cre into the nine-*lox S. oneidensis* strain were also unsuccessful. We thus concluded that inducing recombination between *lox* sites in this strain was likely lethal. However, the suite of broad host-range Cre-expressing plasmids generated herein should be useful for future work in applying these methods to new species.

3.3 DISCUSSION

As synthetic biology continues to advance, there will be an increasing emphasis on the genome as the unit of engineering, which allows much larger swaths of DNA to be manipulated than is possible with plasmid-based methods and enhances our ability to study the structure and function of genomes. This is already evidenced by the synthesis and transplantation of whole genomes by the Venter Institute (Gibson et al, 2008; Gibson et al, 2010; Lartigue et al, 2007; Lartigue et al, 2009) and the development of technologies such as MAGE that can site-specifically perturb multiple sites in a genome (Isaacs et al, 2011; Wang et al, 2009). However, both technologies are still time- and

resource-intensive and are currently limited to a relatively small number of organisms (Enyeart & Ellington, 2011).

We have therefore combined the well-known Cre-*lox* recombinase system and the adaptable targetron technologies to create a method we dub Genome Editing via Targetrons and Recombinases (GETR). GETR presents several advantages in comparison to recombineering and related methods that make use of the λ Red functions. One of these advantages is the fact that GETR is very efficient, and, while the use of selectable markers is required for temporary plasmid maintenance, markers need not be used for selecting or maintaining genomic modifications. Recombineering using single-stranded DNA is simple to execute and useful for making small changes such as point mutations, but percent efficiencies are typically in the single digit range in mutator strains (specifically, *mutS* mutants) and are much lower in wild-type strains. Even then, the efficiency of inserting a sequence as large as a *lox* site is in the neighborhood of 1%, and the efficiency of deleting 10,000 bases of genomic sequence is approximately 0.1% (Wang et al, 2009). The use of the full complement of λ Red proteins allows larger pieces of double-stranded DNA to be inserted (Datsenko & Wanner, 2000), but selectable markers are typically required and the size of possible insertions is limited to several thousands of bases. Manipulations such as inversions and cut-and-paste operations are also impossible using these methods alone.

Another advantage is that targetrons function at high efficiency in many bacterial strains and thus provide an appealing alternative to recombineering functions in many contexts. While the λ Red system has been used outside of *E. coli*, it typically does not function as well in other organisms and in such cases generally requires 500 nucleotides of target-site homology on either side of the integration cassette to obtain reliable results (Beloin et al, 2003; Derbise et al, 2003; Jia et al, 2010; Lesic & Rahme, 2008; Rossi et al,

2003). This requires significantly more labor than the 30-40 nucleotides of homologous sequence required in *E. coli*. In *Pantoea ananatis*, the system only worked well after selection of mutants resistant to the toxic effects of the λ Red proteins (Katashkina et al, 2009). In some organisms, such as *Pseudomonas syringae* (Swingle et al, 2010) and *Mycobacterium tuberculosis* (van Kessel & Hatfull, 2007; van Kessel & Hatfull, 2008), alternative recombineering functions have been discovered, but these do not exceed 0.1% efficiency without selection and also typically require at least 500 nucleotides of homology on either side for reliable results with selectable markers. Recombineering using single-stranded oligonucleotides for making point mutations has been reported in *Lactobacillus* species, but the electroporation of 100 μg of DNA (1,000 times the optimal amount in *E. coli*) was required for efficient mutagenesis (van Pijkeren & Britton, 2012). Wang and coworkers were also able to demonstrate gene disruption in *B. subtilis* using single-stranded DNA, but the method required the use of selectable markers and the generation of single-stranded DNA long enough to encode those markers (Wang et al, 2012). After a search for a suitable recombinase, Binder and coworkers were able to obtain over 4% efficiency using 1 μg of single-stranded DNA in *C. glutamicum*, however, which is a relatively promising result (Binder et al, 2013). Datta and coworkers have identified a number of other possible recombineering proteins from a variety of species (Datta et al, 2008), but to our knowledge none of these have yet been demonstrated as recombineering tools in their natural hosts.

Another common method of genome engineering is the use the use suicide plasmids. For instance, temperature-sensitive integrable plasmids have been developed for all the systems described here other than *S. oneidensis* (Biswas et al, 1993; Hamilton et al, 1989; Link et al, 1997; Luchansky et al, 1989), and systems based on plasmids requiring expression of the *pir* (Kolter et al, 1978; Miller & Mekalanos, 1988) or *repA*

(Leenhouts et al, 1996b) genes to replicate have also been frequently employed. These systems are most useful for gene replacements. For the types of modifications discussed in the present work, suicide plasmids present many of the same limitations as recombineering, such as requirements for selectable markers and large regions of homology, and are limited by poor efficiency and relatively high background. The profusion of research into alternative recombineering systems in recent years, described above, is symptomatic of broader dissatisfaction with suicide plasmids as genetic tools, and the present system represents a favorable alternative to suicide plasmids for large-scale genomic modifications.

A more recent addition to the set of tools available for genome engineering is the CRISPR/Cas9 system, which adapts the site-specific RNA-mediated restriction system of bacteria toward making targeted double-strand breaks in genomic DNA (Cho et al, 2013; Cong et al, 2013; Hwang et al, 2013; Mali et al, 2013). Methods of genome engineering relying solely on the creation of double-strand breaks have not traditionally gained much traction in bacterial systems. Besides the requirement for selectable markers, the efficiency of double-strand break repair tends to be poor in bacteria, since most prokaryotes are only capable of repairing breaks via homologous recombination, and those that can carry out non-homologous end joining have only a rudimentary system for doing so (Aravind & Koonin, 2001; Hefferin & Tomkinson, 2005). CRISPR-Cas9-mediated cutting of genomic DNA has been shown to be lethal to bacteria (Bikard et al, 2012), but Jiang and coworkers have recently reported that this method can be used to select for the integration of mutated DNA homologous to the cut site (Jiang et al, 2013).

However, the CRISPR/Cas9 system alone is only of functional efficiency in bacteria that have very active recombination systems, such as *Streptococcus pneumoniae*, and in those systems the CRISPR-Cas9 expression construct must also be integrated into

the genome along with a selectable marker and then subsequently removed. In *E. coli* the CRISPR-Cas9 system has been shown to increase the efficiency of recombineering by cleaving the genome at unmodified sites (and thereby selecting for modified strains), but this method also has the inherent limitations of recombineering; i.e., the requirement of a mutator strain for high efficiency, limitation to relatively small changes, and generally poor efficiency in systems other than *E. coli*. That said, it is possible that a more general application of CRISPR-Cas9 could be to increase the efficiency targetron-mediated mutagenesis. Finally, recent work by Fu and coworkers (Fu et al, 2013) demonstrates extensive off-target mutagenesis by CRISPR-Cas9, often at efficiencies comparable to the degree of on-target mutagenesis.

We have demonstrated the utility of targetron-delivered *lox* sites by deleting up to 120 kilobases of the *E. coli* genome and 15 kilobases of the *S. aureus* genome, inverting up to 1.5 megabases (one third) of both the *E. coli* and *B. subtilis* genomes, and stably translocating 121 kilobases of the *E. coli* genome to another locus 1.5 megabases away. Efficiencies of the Cre-mediated recombinations are typically near 100%. This method compares favorably with another recently reported method for using targetrons to make genomic deletions (Jia et al, 2011) that relied on homologous recombination between introns and reported an efficiency of 2/648 for the deletion of a two-gene operon, requiring seven rounds of growth and transfer to new media.

While large-scale inversions were presented here primarily as a demonstration of the lack of size limits for generating rearrangements using our method, artificial inversions have traditionally been used for studying genome structure and its constraints (Campo et al, 2004; Esnault et al, 2007; Garcia-Russell et al, 2004; Guijo et al, 2001; Hill & Gray, 1988; Rebollo et al, 1988; Segall et al, 1988; Valens et al, 2004), and the

approaches presented herein allow such studies to be more easily performed in many more systems.

The one-step cut-and-paste method we present is of particular interest given that it allows one piece of a genome to be inserted within another site, without the accumulation of intervening intermediates, an operation that is not possible with any other technique. The cut-and-paste method could also be applied to more nuanced studies of genome structure constraints. For instance, the effect of moving different structural domains or of swapping two domains, such as the Ori and Ter domains, could be examined. Additionally, expression levels tend to be dependent on genomic location, with, for instance, genes nearer the origin tending to be more highly expressed (Cooper & Helmstetter, 1968; Rocha, 2008), and thus cut-and-pastes could be used as a simple means for modulating the overall expression levels of super-operons (Lathe et al, 2000; Rogozin et al, 2002) or other large genetic units. The ability to move DNA between species without regard for inherent similarities or phylogenetic relationships opens up the possibility of using genomic editing for rapidly adapting bacterial genomes.

The use of the Cre/*lox* system also allows large pieces of foreign DNA to be integrated into genomes at high efficiency. An initial recombination occurs between a *lox* site on the plasmid and a *lox* site in the genome, serving to integrate the entire plasmid into the genome, and a second recombination event then occurs between the other two *lox* sites and removes the plasmid sequence. We found no evidence for a difference in efficiency between inserting 1 kilobase and 13 kilobases into the *E. coli* genome via RMCE. Given the high efficiency observed during the construction of large deletions and inversions, the limiting factor in RMCE would thus seem to be the initial encounter between the plasmid and the genome, and not the size of the insertion. The

speed and efficiency of the second recombination event is presumed to be rapid and essentially 100% efficient, similar to the other intragenomic recombinations we report.

Targetron genomic engineering technology can be readily practiced by almost any lab. The algorithm for retargeting the targetrons is available online. The targeting sites in the intron can be changed via restriction cloning of a short fragment of DNA that can be created via two PCR reactions or synthesized in its entirety (see **Section 2.4.1**), followed by the typical time required for ligation, transformation, and sequence validation. Retargeting and the addition of *lox* sites can be performed for multiple introns in parallel. Following electroporation into the target strain, intron induction requires only one day, and plated induction colonies grow after one day. The method is similar in complexity to *lox*-site placements with λ Red, but is an improvement on recombineering in that no selectable markers are required and it can be used in strains where λ Red performs poorly. Similarly, Cre-mediated recombination requires one day for electroporation of the Cre-expressing plasmid (and, for RMCE insertions, the delivery plasmid), and one to two days for the cells to grow and for recombination to occur. Though we used plasmids (one plasmid carrying the targetron, one plasmid carrying the Cre gene, and, as necessary, a plasmid or other vector carrying DNA to be integrated, delivered by electroporation or conjugation) to deliver targetrons in the present study, phage, direct electroporation or other methods could potentially be used, as well.

The scars left by the GETR method and the possibility of unplanned homologous recombination between introns are potential drawbacks, but we have shown that these can be avoided by careful planning. If intron and *lox* site orientations are designed so that inverted repeats form upon Cre-mediated recombination, the repeats will be deleted by the cell, removing most of the intron scar. However, the fact that certain inversions into the *lacZ* locus were viable when a scar was present but not when the scar was

removed indicates that such scars may serve as a buffer against deleterious genomic rearrangements. Unwanted homologous recombination between introns can be prevented by the use of non-homologous introns (EcI5 and Ll.LtrB), or by targetron-mediated disruption of the *recA* gene.

Removing the genome-modifying plasmids was also simple. Except in the case of *S. oneidensis*, which required the continued presence of the intron-encoded protein to allow the intron to splice out from the rRNA genes, a significant fraction (at least 1/3) of colonies were found to have lost the intron-expressing plasmid after the induction process. The Cre-expressing plasmids employed all contained temperature-sensitive origins of replication, and the delivery plasmids for RMCE had the *sacB* gene for counter-selection on sucrose, allowing these plasmids to be easily removed, as well.

In summary, GETR is a new method for genome engineering that can be adapted for use in a variety of bacteria with minimal modifications and without significant loss of functionality. Large, specific, and varied changes can be made with high efficiency. This approach presents certain advantages over recombineering, particularly when working in strains not closely related to *E. coli*, or when the use of selectable markers is impractical or undesirable. In the case of *Staphylococcus aureus* in particular, recent work has made it possible to transform clinical strains (Corvaglia et al, 2010), opening the way to genome editing of otherwise drug-resistant bacteria to create vaccine strains. As concerns about increasing drug resistance of pathogenic bacteria continue to mount, such strains may prove to be a viable alternative to antibiotics. We also expect the system to be of general utility to synthetic biologists looking to engineer entire genomes, particularly those looking to work in systems other than *E. coli*.

3.4 MATERIALS AND METHODS

Intron retargeting, *lox* site cloning, and intron induction in *E. coli* were performed as described in **Section 2.4**. Doubling time measurements were also performed as described in **Section 2.4.3**, except measurements were made every 540 seconds, with shaking for 500 seconds. Introns used in this work are listed in the **Appendix**. A complete list of plasmids, strains, and oligomers not reported in **Section 3.4.1**, as well as detailed cloning procedures for the delivery plasmids described in **Section 3.2.2** and the *rrs*-targeting intron used in *S. oneidensis*, are available in the online supplementary materials of Enyeart et al. (2013).

3.4.1 Construction of broad host-range Cre-expressing plasmids

Oligomers used in the construction of the broad host-range Cre-expressing plasmids pBAV1KC.Cre, pCre.LAC, pBT.Cre, and pBMTL.Cre are given **Table 3.4**.

Primer name	Sequence
bavliu	TCCCAGTCACGACGTTGTAAAAC
bavliu_creu	GTTTTACAACGTCGTGACTGGGACCCAGGCTTTACACTTTATGCTTC
bavlid	AATCCAGAGGCATCAAATAAAACGA
bavlid_catdown	TCGTTTTATTTGATGCCTCTGGATTCGGGTCGAATTTGCTTTC
catdown	CGGGTCGAATTTGCTTTC
cdpar	TGGCAGAAATTCGAAAGCAAATTCGACCCGCGCTTAGTGGGAATTTGTACC
cd-pbt22f	TGGCAGAAATTCGAAAGCAAATTCGACCCGATTCAGGACGAGCCTCAGACTC
cred	CTTTACCGCTGATTCGTGGAACAG
cred_catup2	CTGTTCCACGAATCAGCGGTAAAGCGTTGATCGGCACGTAAGA
lid2	GACGGAGCCGATTTTGAA
lidpaf2	CATAATTGTGGTTTCAAAATCGGCTCCGTCAATAGGATGAATCCGAACCTCATTA
lid-pbt1559r	CATAATTGTGGTTTCAAAATCGGCTCCGTCGGTGAAAGTTGGAACCTCTTACG

Table 3.4. Oligomers (primers) used to construct broad host-range Cre-expressing plasmids.

To construct pBAV1KC.Cre, first the chloramphenicol-acetyltransferase gene was amplified from pX20 using the cred_catup2 and bavlid_catdown primers, the *cre*

gene (including the *lac* promoter) was amplified from pQL269 using the bavliu_creu and cred primers, and the pBAV1K backbone was amplified from the plasmid pBAV1K.lacI.t5.798s (contains the *lacI* gene, plus *ltrA* and the *rrs*-targeting intron under a t5 promoter; provided by E. Quandt) so as to include the *lacI* gene (but not the genes for the intron or the intron-encoded protein) using the bavliu and bavlid primers. These three amplicons were then assembled into pBAV1KC.Cre using the method of Gibson (Gibson, 2011; Gibson et al, 2009). Thus, in addition to the *cre* gene, pBAV1KC.Cre also carries the *lacI* gene for regulating the *lac* promoter that controls Cre expression, as well as genes for resistance to both kanamycin and chloramphenicol.

For pCre.LAC, an amplicon containing the *lacI*, *cre*, and *cat* (chloramphenicol acetyltransferase) genes was amplified from pBAV1KC.Cre using the catdown and lid2 primers, while the temperature-sensitive variant of the pWV01 backbone was amplified from pCrePA using the cdpar and lidpaf2 primers. These two amplicons were then subjected to Gibson assembly to yield pCre.LAC, which also harbors *lacI* and *cre* under the *lac* promoter on a low-copy, temperature-sensitive vector granting resistance to chloramphenicol and erythromycin.

For pBT.Cre and pBMTL.Cre, the vector backbones of pBT-2 and pBMTL-4, respectively, were amplified using the cd-pbt22f and lid-pbt1559r primers. These amplicons were each subjected to Gibson assembly with the catdown/lid2 amplicon above to yield pBT.Cre and pBMTL.Cre. Both of these plasmids have *lacI* and *cre* under the *lac* promoter and provide chloramphenicol resistance. pBT.Cre also grants kanamycin resistance, while pBMTL.Cre grants tetracycline resistance. pBMTL.Cre also contains the *mob* gene and can be transferred between cells when RP4 transfer functions are provided *in trans* (Lynch & Gill, 2006).

3.4.2 Intron induction in non-*E. coli* strains

Intron induction in *S. aureus* RN10628 and B. subtilis 168 was performed as described elsewhere (Yao et al, 2006). Tryptic soy broth (TSB) was used as the growth medium for *S. aureus*, and LB broth was used for *B. subtilis* (with 5 μg/mL erythromycin) and *S. oneidensis* (with 50 μg/mL kanamycin). The T5.rDNA.798s.1WL2R intron was not formally induced.

3.4.3 Induction of Cre-mediated recombination

For intra-molecular recombinations in *E. coli*, the plasmid pQL269 (Liu et al, 1998) was electroporated into cells that were then plated on LB plus 100 μg/mL spectinomycin and grown at 30°C until colonies appeared. Occurrence of recombination was screened using colony PCR, and a subset of positive colonies were restreaked on LB (non-selective) and grown overnight at 42°C to cure the plasmid. Freezer stocks were made from these cells, and the analyses shown in **Figures 3.5** through **3.8** were performed on cells streaked from these stocks. The procedure was essentially the same in *S. aureus* RN10628, except that the cells were electroporated with pRAB1 (Leibig et al, 2008) and grown initially on tryptic soy agar (TSA) plus 10 μg/mL chloramphenicol. *B. subtilis* 168 was electroporated with pCrePA and grown on LB plus 5 μg/mL erythromycin.

3.4.4 Cre-mediated genomic insertion (recombination-mediated cassette exchange)

To assay insertion efficiency, delivery plasmids pACDX3S-GFP and pUC19X3S-GFP were used. These plasmids contain the GFP open reading frame (ORF) flanked by T7 terminators and the *lox*71 and *lox*m2/66 sites. Each of the GFP delivery plasmids was transformed into *E. coli* BL21(DE3) Gold and *E. coli* HMS174(DE3) having a T7 promoter, as well as *lox*66 and *lox*m2/71 sites complementary to those in the delivery

plasmids, integrated at the *lacZ*, *galK*, or *malT* locus. The Cre-expressing plasmid pQL269 was transformed into the strains, which were then grown at 30°C in liquid culture. At days one, two, and three after transformation of pQL269, aliquots from each culture were spread on LB plates. The plates were grown overnight at 37°C, and then imaged using a UV backlight and a B&W 061 dark-green filter. Identical strains lacking pQL269 were used as negative controls at each time point. Green colonies were counted manually to determine insertion efficiency. The entire three-day procedure was performed three separate times.

The insertion of DEBS1-TE was performed similarly, using pET26b-DEBS1TE-i as the delivery plasmid. Insertion was assayed by colony PCR using primers flanking the 5' end of the insertion three days after transformation of pQL269. Selected positive clones were then further assayed by overlapping PCRs covering the entire operon after removal of the delivery plasmid.

3.4.5 Statistical analyses

The analyses of the data on Cre-mediated insertion (RMCE) efficiency were also performed on square-root-transformed values. The *P*-values of Bartlett's test for transformed and untransformed data were 0.11 and 0.99, respectively, for the data shown in **Figure 3.3B**, and 4.4×10^{-5} and 0.37, respectively, for the data shown in **Figure 3.3C**. Even for the data shown in **Figure 3.3B**, the *P*-values resulting from an ANOVA performed on the square-root-transformed data were more stringent and were in better agreement with the results of pairwise comparisons, and thus the square-root-transformed data was used for the analysis. All triplets had *P*-values of at least 0.04 in the Shapiro-Wilk normality test, except for HMS174 containing a high-copy vector on day three, where all three values were 100%. Replacing these values with values randomly drawn

from a normal distribution having a mean of 100% and a standard deviation of 1% (similar to the other triplets near 100%) made no substantial difference in the analysis results.

A multifactorial analysis of variance performed on the results in **Figure 3.3B** indicated that time (*P*-value = 7.837×10^{-14}) and delivery-plasmid copy number (*P*-value = 2.363×10^{-8}) were significant factors, but strain (*P*-value = 0.2708) was not. However, the interactions between strain and time (*P*-value = 1.588×10^{-5}) and between strain and copy number (*P*-value = 7.243×10^{-7}) were significant. The interaction between copy number and time (*P*-value = 0.04234) and the three-way interaction between all factors (*P*-value = 0.02650) were significant at the 0.05 level but not the 0.01 level. Subsequent comparisons between days (corrected for multiple comparisons) showed a significant difference between day one and day two (*P*-value = 1.1×10^{-5}) and between day one and day three (*P*-value = 1.0×10^{-6}), but not between day two and day three (*P*-value = 0.698). Multiple pairwise comparisons for this and all other data were made using the Tukey method for correcting for multiple testing (Hsu, 1994).

A multifactorial analysis of variance performed on the results in **Figure 3.3C** indicated that location significantly affects insertion efficiency (*P*-value = 7.2×10^{-7}). Time also proved once again to be a significant factor, with a *P*-value of 2.6×10^{-13}. Subsequent pairwise comparisons, corrected for multiple testing, showed that the *malT* locus differed significantly from the *lacZ* and *galK* loci on days two and three (maximum *P*-value = 4.1×10^{-3}) but not on day one (minimum *P*-value = 0.61). The *lacZ* and *galK* loci were not found to differ significantly (*P*-value = 0.72).

For the doubling-time data in **Table 3.3**, all strains had a Shapiro-Wilk *P*-value of at least 0.01. Bartlett's test gave a *P*-value of 0.13 for the original data and 0.19 for square-root-transformed data. Using square-root-transformed data versus untransformed

data made little difference in the quantitative results (*P*-values) and no difference in qualitative results (determinations of statistically significant differences), and thus the *P*-values presented below for doubling times are those for untransformed data.

An analysis of variance performed on the results in **Table 3.3** showed that doubling time is highly dependent on the type of rearrangement present (*P*-value = 1.85×10^{-14}). Subsequent pairwise comparisons corrected for multiple testing showed that the strains fall broadly into groups: a high-growth group having approximately wild-type doubling times and a low-growth group having impaired doubling times, where the *P*-value for the difference between the slowest member of the high-growth group and the fastest member of the low-growth group was 0.028.

Chapter 4: *Lox*-carrying transposons for generating libraries of genomic rearrangements

4.1 INTRODUCTION

In Chapter 3, targetron technology was adapted to deliver *lox* sites to specific genomic loci and thereby execute planned rearrangements. This is obviously useful, but, as discussed in Chapter 1, random approaches to genome engineering are also powerful, for two related reasons. The first, which could be considered the basic-science rationale, is that random approaches (mutant screening or selection) are the most consistently reliable option for learning what is not already known, at least in the field of genetics. The second reason is that random approaches allow the experimenter to harness the power of evolution via natural selection, which allows the discovery of optimization solutions without any *a priori* knowledge about what those solutions might be. This might be called the biotechnological rationale. We therefore adapted transposons to carry *lox* sites to random genomic loci in order to create libraries of strains containing genomic rearrangements upon expression of the Cre protein, which are then analyzed by deep sequencing.

In the context of the basic-science rationale, the reason for creating such libraries is to probe and press the limits of bacterial genome structure. This question has previously been explored primarily via the generation of planned deletions (Fukiya et al, 2004; Kolisnychenko et al, 2002; Posfai et al, 2006; Suzuki et al, 2005a; Suzuki et al, 2005b; Wilson et al, 2004) or inversions (Campo et al, 2002; Campo et al, 2004; Esnault et al, 2007; Garcia-Russell et al, 2004; Guijo et al, 2001; Hill & Gray, 1988; Rebollo et al, 1988; Segall et al, 1988; Valens et al, 2004) or by the analysis of naturally occurring inversions (Eisen et al, 2000; Hill & Gray, 1988; Liu et al, 2006; Liu & Sanderson, 1996; Louarn et al, 1985).

Studies such as these, combined with the fact that genome structure is highly conserved across bacterial lineages, have led to the general assumption that many fundamental aspects of bacterial genome organization are immutable (Rocha, 2008). However, very little of the potential search space for alternative genomic rearrangements has been explored. Among those works presenting artificial rearrangements, none examine more than a few dozen different modifications, while the study of naturally occurring recombinations is limited to those that occur via homologous recombination between repeated genomic elements, which in bacteria is primarily the ribosomal genes. This represents a very small fraction of the ways in which the structure of the genome could in theory be rearranged. While some characteristics such as the requirement for replichore balance are likely not subject to much modification (Esnault et al, 2007; Itaya et al, 2005), a question remains as to whether the evolutionarily conserved elements of bacterial genome structure are the best possible arrangement or simply one of many possible arrangements alternate arrangements that might have been chosen. In other words, does the current bacterial genome structure really represent the highest peak of the fitness landscape, or is it simply a peak from which other peaks are difficult to reach? A more thorough search through the space of potential structures could shed more light on this issue.

In the biotechnological context, random methods used for strain improvement have focused primarily on generating and combining knock-out mutants (see Chapter 1). Expanding this focus and implementing large-scale genomic rearrangements may be useful in finding strains better suited to a specific task or environment.

There have been a few attempts to apply random methods to the generation of genomic rearrangements. For instance, Goryshin and coworkers utilized a nested transposon structure, in which the entire genome is essentially made into a transposon

that then inserts randomly into itself, to create randomized rearrangements in *Escherichia coli* (Goryshin et al, 2003). However, likely due to the paucity of high-throughput genomic analysis approaches available at the time, only a small number of such rearrangements were generated and analyzed. The authors reported that the method could be applied recursively and that they were in the process of using it to build a minimal genome, but to our knowledge these results were never published. Yu and coworkers have also previously used transposons to randomly deliver *lox* sites to the *E. coli* genome, but they only placed one *lox* site per genome, and rearrangements were generated by rationally selecting *lox* insertions in different strains and combining them via P1 transduction before introducing the Cre protein to catalyze recombination (Yu et al, 2002).

Thus a truly random approach to creating genomic rearrangements that achieves library sizes and analysis throughput sufficient to begin providing serious answers to the questions raised above does not yet exist. The work in this chapter aims to provide such an approach, as well as initial data relating to the questions at hand. Specifically, we randomly delivered two *lox* sites per genome on modified *mariner* transposons (mobilized by the hyperactive mutant of the *Himar1* transposase (Lampe et al, 1999)) to create two separate libraries (biological replicates) of genomic rearrangements, each consisting of several million variants. We then subjected them to successive rounds of growth and dilution for approximately 200 generations in both LB and M9 media, and performed targeted high-throughput sequencing on the libraries at generations 0, 25, 95, and 195 in order to examine the genetic dynamics over the course of the selections. The results are presented below.

4.2 RESULTS

4.2.1 Summary of initial efforts

A primary obstacle facing the implementation of such a method is the question of how to analyze the results. Though numerous methods exist for determining the site of transposon integrations (see **Section 1.6** and **Fig. 1.2**), these methods are usually used under the assumption that the genomic sequence on one side of the insertion site is contiguous with the genomic sequence on the other side with reference to genomic arrangement of the unmodified strain. Thus many of these approaches only give information about the genomic sequence on one side of the insertion site and cannot be applied to the present case, where the insertion site is also a recombination site, and information about the genomic context on both sides of the insertion is required in order to assess the nature of the recombination.

We initially attempted an approach analogous to the restriction/ligation approach shown on the left side of **Figure 1.2**, where sequencing primer-binding sites were included within the transposon, and the DNA was sheared and ligated. PCR amplification using the sequencing primers failed to yield trustworthy results, however. This approach could potentially be improved upon by employing end-repair or enzymatic digestion, but given the poor results of the initial implementation and the fact that even under ideal circumstances this approach could yield false rearrangements resulting from inter- rather than intra-molecular ligation, we decided to abandon it.

Next we tried skipping the enrichment step entirely and simply sequencing everything, sorting out reads containing both sides of a transposon integration computationally after the fact. However, the fraction of usable reads was very small, ranging from 0.004 to 0.006% of all reads, which was inadequate to obtain the depth of coverage we required without incurring great expense. We next attempted to increase the

proportion of usable reads via pulldown with a biotinylated primer that annealed to the transposon sequence, but this yielded only about five-fold enhancement, which was still insufficient. The reason the pull-down performed poorly is likely due to the fact that, after the marker is removed by Cre-*lox* recombination, the insertion site consists of a *lox* site flanked by the inverted repeats of the transposon. Since the majority of the *lox* sequence also consists of an inverted repeat, the insertion site as a whole is largely palindromic. This means that when the DNA is melted to allow primer binding, the insertion site folds up into a tight hairpin, which likely impedes primer binding. In support of this hypothesis, it was found that insertion sites containing *lox71* or *lox66* sites, which, as shown in **Figure 2.1**, form imperfect hairpins, were enriched in the pulled-down sequences as opposed to *loxP* or *lox72* sites, which form much tighter hairpins. This also meant that the pull-down was biased against insertion sites where a recombination event had occurred.

Next we tried adding extra sequence into the transposon to serve as a binding site for primer pull-down, but for unknown reasons the addition of this sequence significantly impaired transposon efficiency. This issue could potentially be remedied by changing the sequence included for primer binding, but we decided to switch focus to the approach described in **Section 4.2.2** instead. However, if the issue of reduced transposon efficiency could be resolved, this approach could potentially be used in tandem with that used in **Section 4.2.2** to further improve efficiency.

4.2.2 Overview of methodology

The approach used to generate and analyze the libraries of genomic rearrangements is shown schematically in **Figure 4.1**.

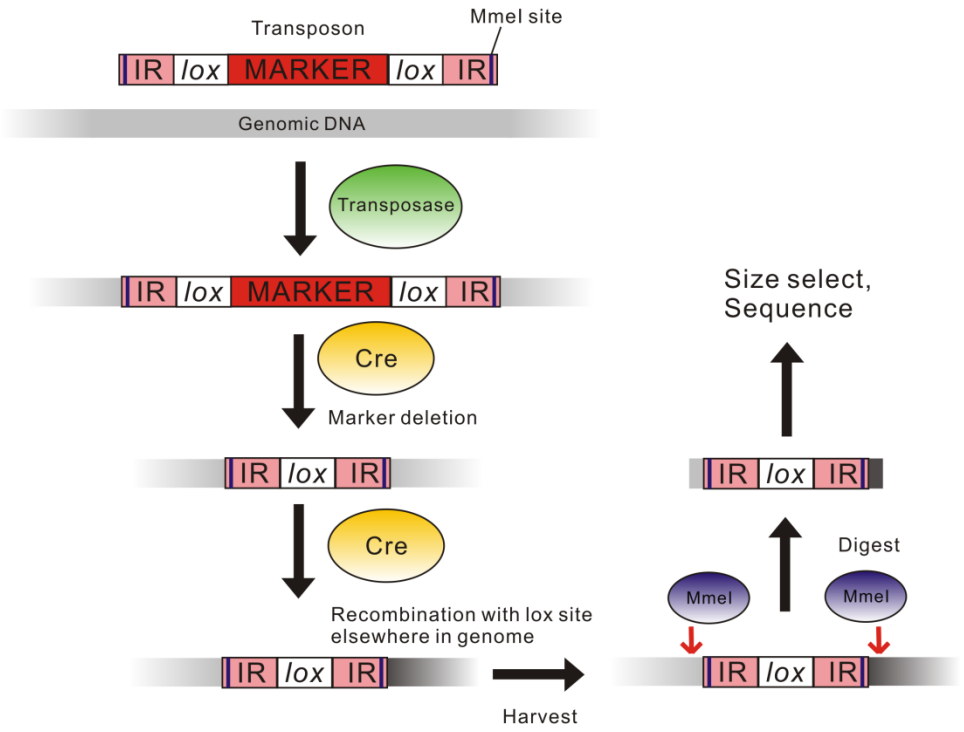

Figure 4.1. Methodology for delivering *lox* sites and screening genomic libraries.

The "IR" box is the inverted repeat of the transposon; black is used to denote DNA originally from a different region of the genome than the grey genomic DNA.

Transposons are constructed by PCR amplification of a selectable marker using primers that add a *lox* site and the transposon inverted repeat on both sides. The transposase is expressed from a plasmid (pQLH1) in the cell, and the transposon PCRs are electroporated separately. This avoids the issues with background that are seen using conventional methods based on suicide plasmids. For instance, in another study using this transposon expressed in *E. coli* from a conditionally replicating plasmid that also expressed the transposase, 2% of resistant target colonies were also found to harbor antibiotic resistance found only on the plasmid (Chiang & Rubin, 2002). This is acceptable for single insertions, but given the transposon efficiency, measured in the same work, of about 3×10^{-3}, it can become a progressively larger problem in multiple

rounds of mutagenesis. Not keeping the transposon on a plasmid with the transposase also means that transposition will not occur in cloning or donor strains harboring the plasmid. While transposome technology, in which a Tn5 transposon and transposase are assembled together *in vitro* and then directly electroporated (Goryshin et al, 2000), is another potential option for this application, we chose the current approach because the Tn5 inverted repeats cannot be readily adapted to carry MmeI recognition sites (see below), and because the approach used here also does not require protein purification or the purchase of a commercial kit, except as required for standard PCR reactions.

The efficiency of transposition of the amplicons varied depending on the nature of the insert and the sequence used for the inverted repeats but was typically in the range of 10^{-4} to 10^{-6} integration events per microgram of PCR product, as measured by number of resistant colonies. No viable resistant colonies were seen upon transforming transposons harboring kanamycin or chloramphenicol resistance into strains lacking the transposase (tiny colonies were occasionally seen with kanamycin, but these did not grow upon inoculation into LB containing kanamycin).

Once the *lox* sites have been placed, Cre is introduced on the plasmid pQL269A, which has the same origin as the transposase-expressing plasmid but a different selectable marker, and the cells are grown overnight. Cre-mediated recombination between *lox* sites removes the markers and then causes rearrangements between *lox* sites at different sites in the genome. Both plasmids pQLH1 and pQL269A are temperature sensitive and can be removed by growth at 42°C.

The sequencing and analysis methodology relies on the fact that the inverted repeat of the *Himar1 mariner*-type transposon can be modified to contain a recognition site for the restriction enzyme MmeI, as discussed in **Section 1.6**. Upon digestion of genomic DNA with MmeI, this results in all fragments containing an insertion being of a

uniform size containing approximately 16 bases of genomic sequence on either side of the transposon DNA. These fragments can be enriched via a size-selection step and then sequenced. In the samples, analyzed here, the percentage of mate-paired sequencing reads that were found to have the transposon inverted repeat in at least one read ranged from 0.6 to 3.5%, and the percentage that had the inverted repeat in both reads ranged from 0.3 to 3%.

4.2.3 Library creation and selection

We started by making two separate libraries (biological replicates) in which each strain contained two *lox* sites randomly delivered to the genome. *E. coli* MG1655 was used as the base strain. The first set of *lox* sites were *lox71* sites delivered with a kanamycin resistance gene. The efficiency of transposition was estimated to be $6.0 \pm 0.4 \times 10^4$ insertions μg^{-1} (calculated from the averages for each replicate, which in turn were calculated from three different platings), and the estimated library sizes were $1.6 \pm 0.2 \times 10^5$ for replicate 1 and $1.4 \pm 0.2 \times 10^5$ for replicate 2, which correspond to approximately 30× coverage of one insertion in every kilobase of the genome.

The second set of *lox* sites were *lox66* sites delivered with a chloramphenicol resistance gene (*cat*). The efficiency of transposition was estimated to be $4.94 \pm 0.02 \times 10^5$ insertions μg^{-1}, and the estimated library sizes were $5.41 \pm 0.9 \times 10^6$ for replicate 1 and $5.4 \pm 0.1 \times 10^6$ for replicate 2, which correspond to approximately 1× coverage of every possible pairwise combination of insertions in every two kilobases of the genome. The efficiency of insertion of a transposon that was identical except had the wild-type inverted-repeat sequence (lacking the MmeI recognition site) was $1.5 \pm 0.3 \times 10^6$ insertions μg^{-1} when electroporated under similar conditions.

The Cre-expressing plasmid pQL269A was then transformed. The total numbers of transformants of pQL269A were estimated to be $2.7 \pm 0.1 \times 10^6$ and $1.16 \pm 0.04 \times 10^6$, respectively, for replicates 1 and 2 (transformation efficiency under these conditions: $1.3 \pm 0.2 \times 10^5$ μg^{-1}). Since a large number of the recombinations executed by Cre were likely lethal, it is difficult to put an exact number on the diversity of the resulting libraries, but taking that fact into account, the rearrangement libraries seem to be large enough to represent good coverage of the two-transposon libraries they were generated from.

The pQL269A-transformed cells were incubated for approximately one day in LB at 30°C, and the end of this incubation was considered as "generation 0" for purposes of the subsequent selection. Next each replicate was diluted by a factor of 100 into LB and M9 minimal media (0.4% glucose) and grown overnight with shaking at 42°C to remove the temperature-sensitive pQL269A plasmid and any residual pQLH1 plasmid. The cultures were then diluted ×1000 into fresh media and regrown to saturation at 37°C nineteen more times. Glycerol stocks were made and genomic DNA was isolated after growth of the 3rd, 10th, and 20th dilutions after generation 0, which correspond to approximate generations 25, 95, and 195.

As an initial assessment of whether the selections were successful in producing strains better adapted to their respective environments, growth curves were measured for each library under each condition at each for generations 0, 25, 95, and 195. These results are shown in **Figure 4.2**. The growth curves are consistent with significant improvement in fitness compared to wild-type, with some improvement seen starting even at generation 0, with fitness increasing over time and then leveling off. These results appear to be consistent with a successful selection. The M9 libraries seem to have

a greater fitness relative to wild-type than the LB libraries, suggesting that wild-type *E. coli* MG1655 is closer to optimal for growth in LB as opposed to M9.

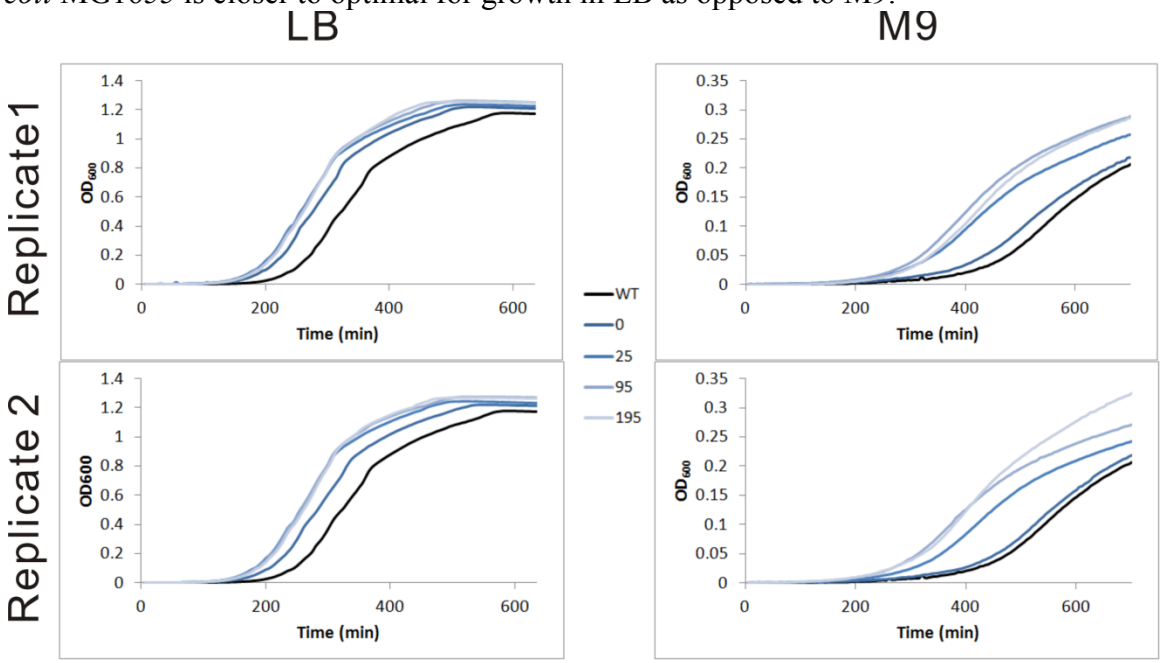

Figure 4.2. Growth curves of wild-type and genome-rearrangement libraries at various time points.

Each curve is an average of three replicates (except wild-type in M9, which is an average of two).

Additionally, the generation-195 LB libraries were grown in M9, and the generation-195 M9 libraries were grown LB in order to assess whether the growth benefits attained were general or environment-specific. These results are shown in **Figure 4.3**. The "winners" from both conditions outgrew wild-type in the opposite condition, consistent with some general benefit to growth *in vitro* from the selection procedure in all cases. It is interesting to note that the M9 winners in particular outperform wild-type in LB primarily in the later stages of growth, when nutrients are presumably more scarce, akin to the situation in M9 media. Neither set of winners seem

to grow as well in the opposite environment as the winners selected in that environment, which indicates some degree of specialization and is not unexpected.

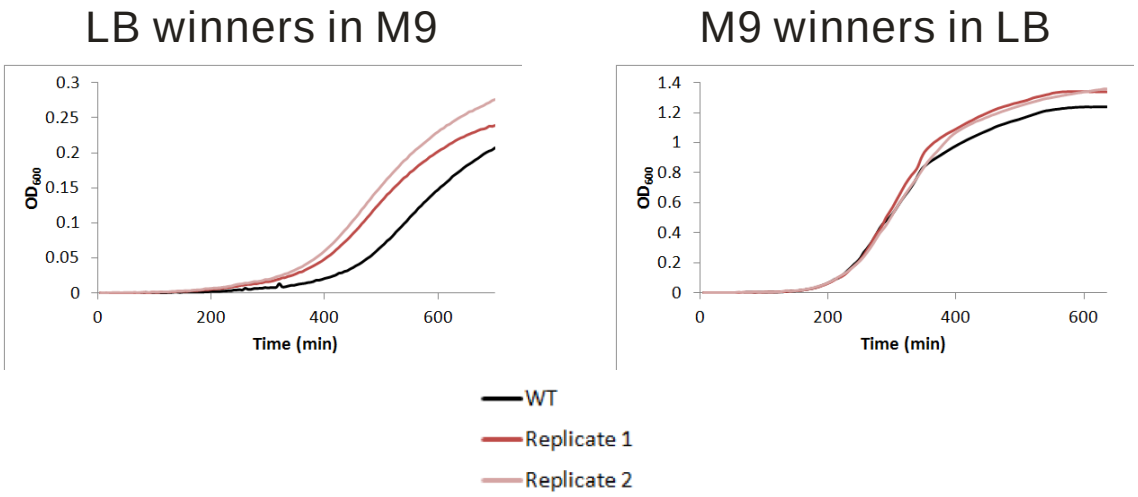

Figure 4.3. Growth curves testing specialization of end-point genomic rearrangement libraries.

Each curve is an average of three replicates (except wild-type in M9, which is an average of two).

4.2.4 Analysis of detected rearrangements

The key issue in the sequencing analysis is determining what type of rearrangement, if any, is represented by a given mate-pair read covering an insertion. A schematic showing the approach used to identify rearrangement types is found in **Figure 4.4**. In short, each mate-pair can be represented four numbers: the wild-type genomic positions corresponding to the MmeI cut-site on one end, the insertion junction on the same end, the insertion junction on the other end, and the MmeI cut-site on the other end. These four positions can further be simplified to three binary values representing the direction that must be traversed to move from one position to the next in the shortest distance on the wild-type chromosome.

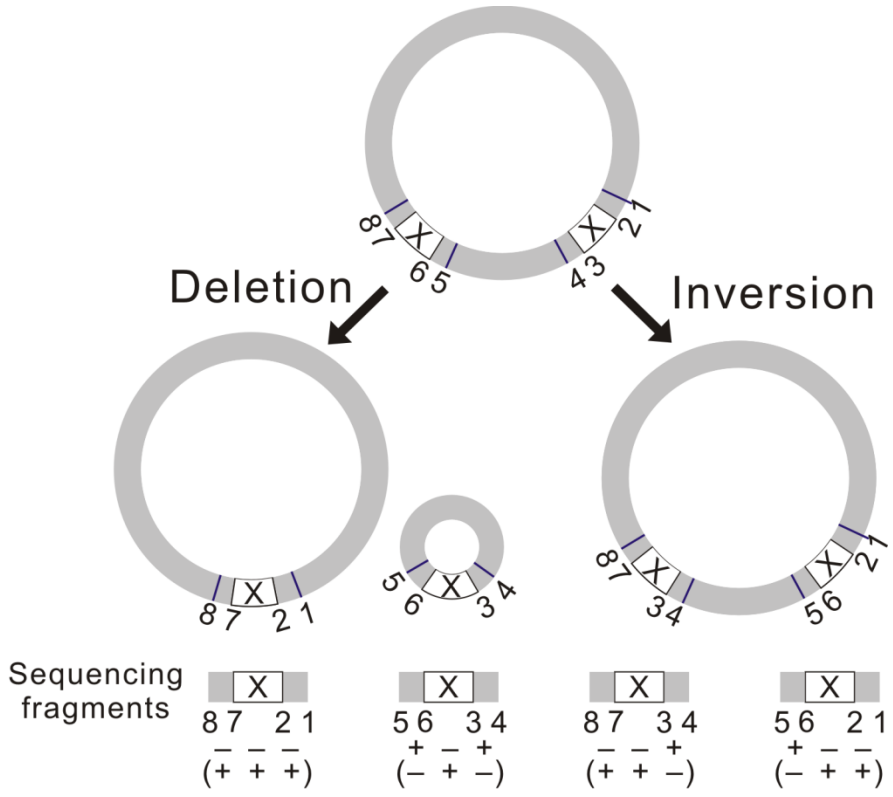

Figure 4.4. Outline of method for identifying rearrangements from sequencing data.

The "X" boxes represent *lox*-containing transposon insertion sites. The numbers are arbitrary and meant simply to represent relative locations on the wild-type chromosome. Numbers at box edges represent wild-type genomic positions at the insert junctions, and numbers at the lines outside represent genomic positions at the MmeI cut site. "+" or "–" represents the direction that must be traversed to move from one numbered position to the next adjacent numbered position in the shortest distance; the pattern of these directions can be used to identify the rearrangement.

If the cut-site/junction direction is different on one side than on the other, then an inversion is present. If the cut-site/junction directions are the same on both sides, and the distance between the two junction sites is small (the exact cutoff is somewhat arbitrary; ten bases was used here), then no rearrangement has occurred and the reads represent a simple insertion. If the distance between the two junction sites is large, a deletion has occurred. Since a recombination-mediated deletion in a circular chromosome actually involves resolving the chromosome into two separate circles, it is important to determine which of those two deletions is represented by a given deletion read, without simply

assuming one or the other. If the direction representing the shortest distance between the two junction points is the same direction as the cut-site/junction directions, then the DNA represented by the reads was on the larger genomic fragment, and the apparent deletion is the smaller fragment. If the direction between junctions is different, then the DNA was on the smaller fragment, and the apparent deletion was the larger fragment.

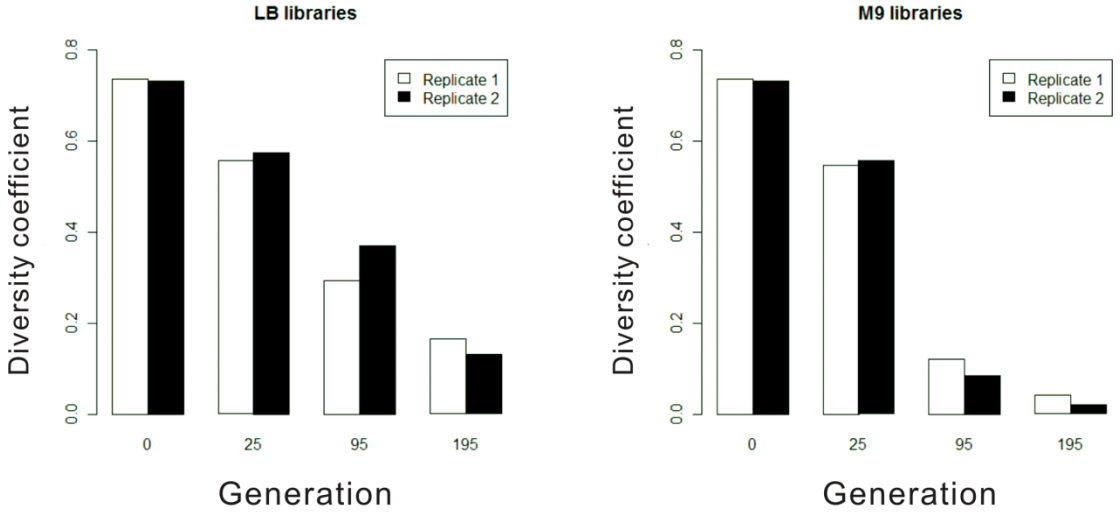

Figure 4.5. Library diversity over the course of selection.

The diversity coefficient is the number of unique reads (specifically, the number of reads remaining after removing all duplicates) divided by the total number of reads. Note that the generation 0 bars are identical in both plots.

Bar plots showing the change in diversity of insertion reads over the course of the selection are shown in **Figure 4.5**. Diversity declines significantly over time, again as would be expected from a successful selection. Diversity becomes particularly low in the M9 libraries, which may mean that a small number of changes are most important for adapting to M9, while a larger number of less important changes can be used to adapt to LB. This may also support the hypothesis that wild-type MG1655 is better adapted to growth in LB than in M9, since any extremely important adaptations to growth in LB seem to have already been made.

Figure 4.6 shows the changes in the distribution of types of rearrangements over the course of the selections. Besides deletions, inversions, and no rearrangement, two other types were assigned. In some cases, the genomic sequence on one or both sides of the insertion cannot be assigned to a single genomic locus. Insertions into ribosomal genes are one example. Since it is difficult to make a determination of the type of rearrangement without knowing exactly where the insertion is placed, such reads are labeled "undetermined." The "alignment failure" type results when the alignment program fails to find a match in the MG1655 genome for the genomic sequence on one or both sides of the insertion. Particularly in generation zero, these may represent plasmid DNA. The integration of other exogenous DNA or simple sequencing error may also contribute to the presence of such reads.

In generation 0, approximately equal numbers of deletions, inversions, and simple insertions are seen. It is initially surprising to see so many reads representing non-rearrangements given the apparent high efficiency of Cre-mediated recombination seen in Chapter 3. These insertions would not show up in the analysis if Cre had not deleted the marker gene in the transposon, however, so Cre clearly functioned in these strains. One possible explanation is that the finer detail available in this data allows us to see that Cre is actually less efficient than is apparent from the bulk information presented in Chapter 3. Another explanation, which is not mutually exclusive with the first, is that the existence of unrecombined insertions is selected for in cells that would die if recombination between distant *lox* sites was successful.

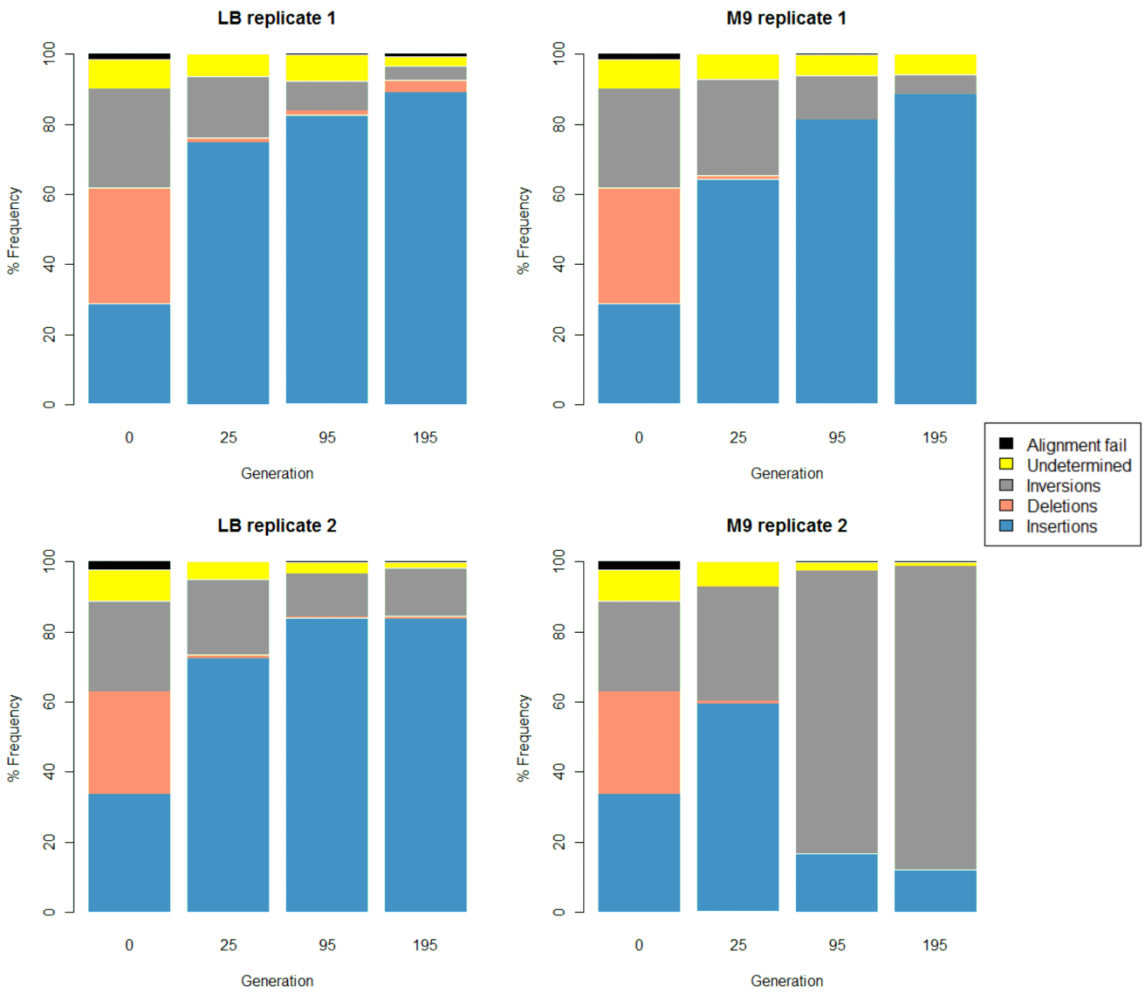

Figure 4.6. Frequencies of rearrangement types over the course of the selections.

In most cases genomic rearrangements decrease in frequency as the selection progresses, supporting the hypothesis that the wild-type genome structure is best. Deletions are particularly disfavored, especially in M9, which is expected since the chance of deleting an important gene for any random deletion is quite high (generation 0 can be expected to include a large number of functionally dead cells). Replicate 2 in M9 represents a counter-example to the trend of the other libraries in disfavoring rearrangements, however, as inversions (in particular, two specific inversions) come to

dominate the population by the end of the selection. Even in replicate 1, inversions seem to be better favored in M9 than in LB, which is somewhat surprising but may also be related to the hypothesis that more drastic measures are required to adapt *E. coli* MG1655 to M9 than to LB.

The box plots of the distribution of sizes of inversions and deletions in the libraries over the course of the selections are shown in **Figures 4.7** and **4.8**, respectively.

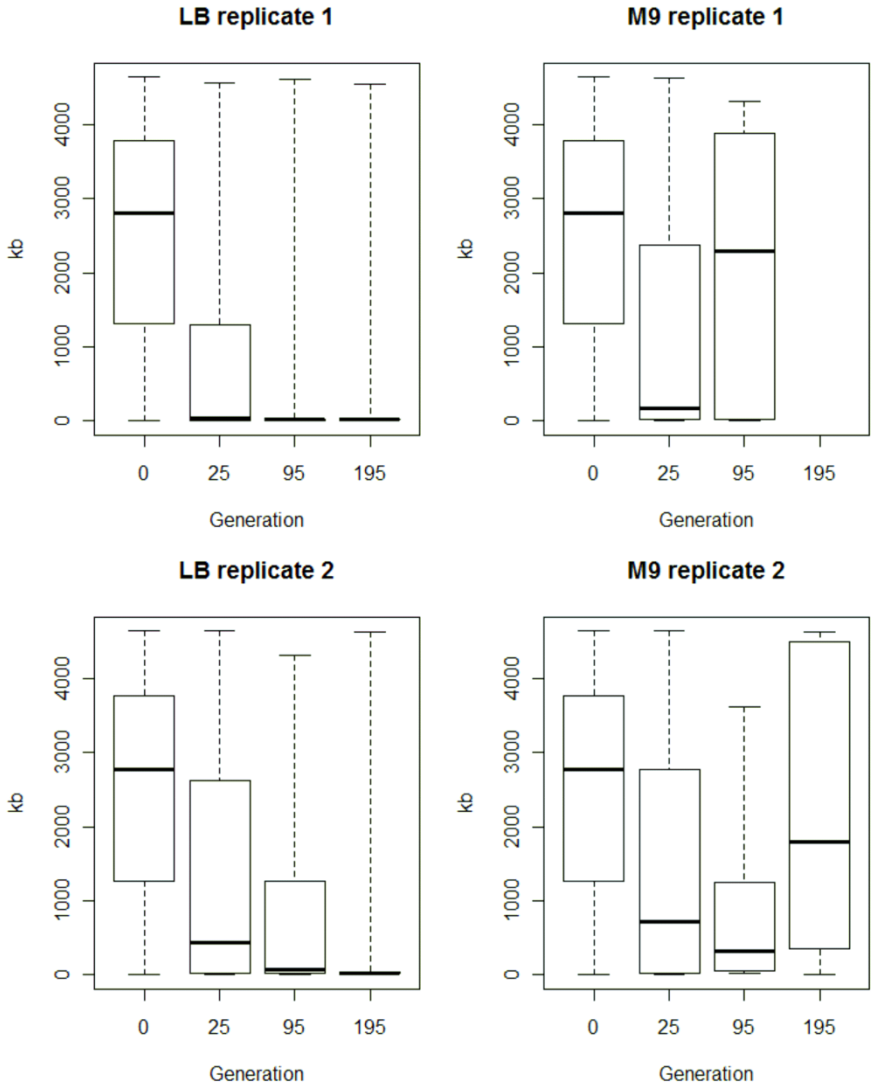

Figure 4.7. Box plots of distribution of deletion sizes over the course of the selection.

The general trend is that large deletions are disfavored, which is as expected. However, a small number of apparently extremely large deletions seem to persist, and, in M9, perhaps even be relatively favored. (The number of deletions detected in the later stages of the M9 selection was very small however, and thus the opportunity for sampling error is great.) It is unlikely that these reads truly represent strains lacking most of the genome. Since many such deletions were detected in significant numbers, however, it is also difficult to discount them as sequencing error. The explanation may have to do with the fact that during exponential growth *E. coli* harbors numerous active replication forks and therefore multiple copies of its genome within each cell (Cooper & Helmstetter, 1968; Rocha, 2008). Thus, if a large deletion were to occur in one chromosomal copy, part or all of the deleted fragment could potentially homologously recombine into another copy. Though much of the insertion would likely be removed by further recombination, parts including the deletion site might remain. Such rearrangements could be selected for if they result in adaptive gene duplications. Alternatively, the presence of large regions of homology in the sets of ribosomal genes could allow for an intra-chromosomal cut-and-paste reaction to occur between ribosomal genes on the deleted fragment and those in the remainder of the genome. Further work will be required to assess the true nature of these apparent deletions.

The distributions of inversion sizes as shown in **Figure 4.8** also show a general trend of decreasing size, though not as pronounced as the trend for deletions, which is to be expected. Additionally, it does seem that certain large and adaptive inversions exist, particularly in the M9 libraries.

Finally, graphical visualizations of the modifications most commonly detected are shown in **Figure 4.9** for replicate 1 and **Figure 4.10** for replicate 2. Not all detected modifications are depicted; there are too many to show for generations 0 and 25, and

modifications found at extremely low frequencies are not shown for generations 95 and 195. The wide variety of deletions and inversions present at generation 0 is clear, and the quick reduction in the diversity of those rearrangements is also apparent. The increased tolerance of the M9 libraries for inversions can also be seen by comparing the LB and M9 libraries at generation 25; the apparent reduction in inversion diversity in later M9 generations is a result of a small number of inversions dominating the population. Genetic interpretations of the most common rearrangements are given in **Section 4.2.5**.

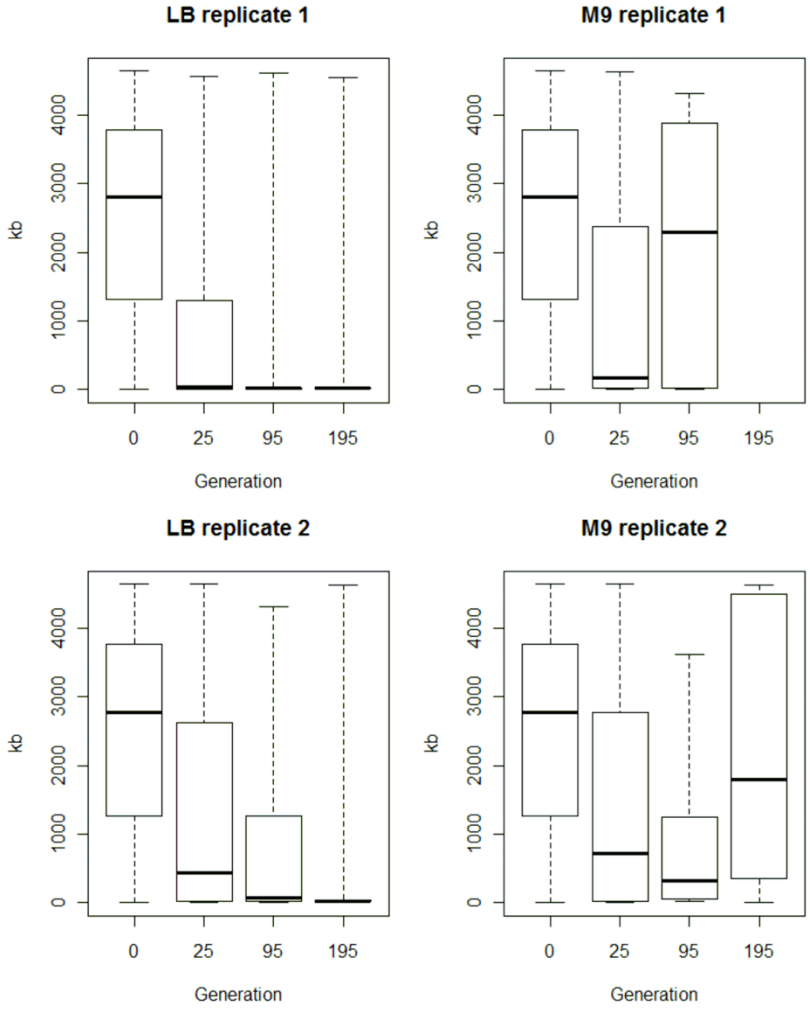

Figure 4.8. Box plots of distribution of inversion sizes over the course of the selection.

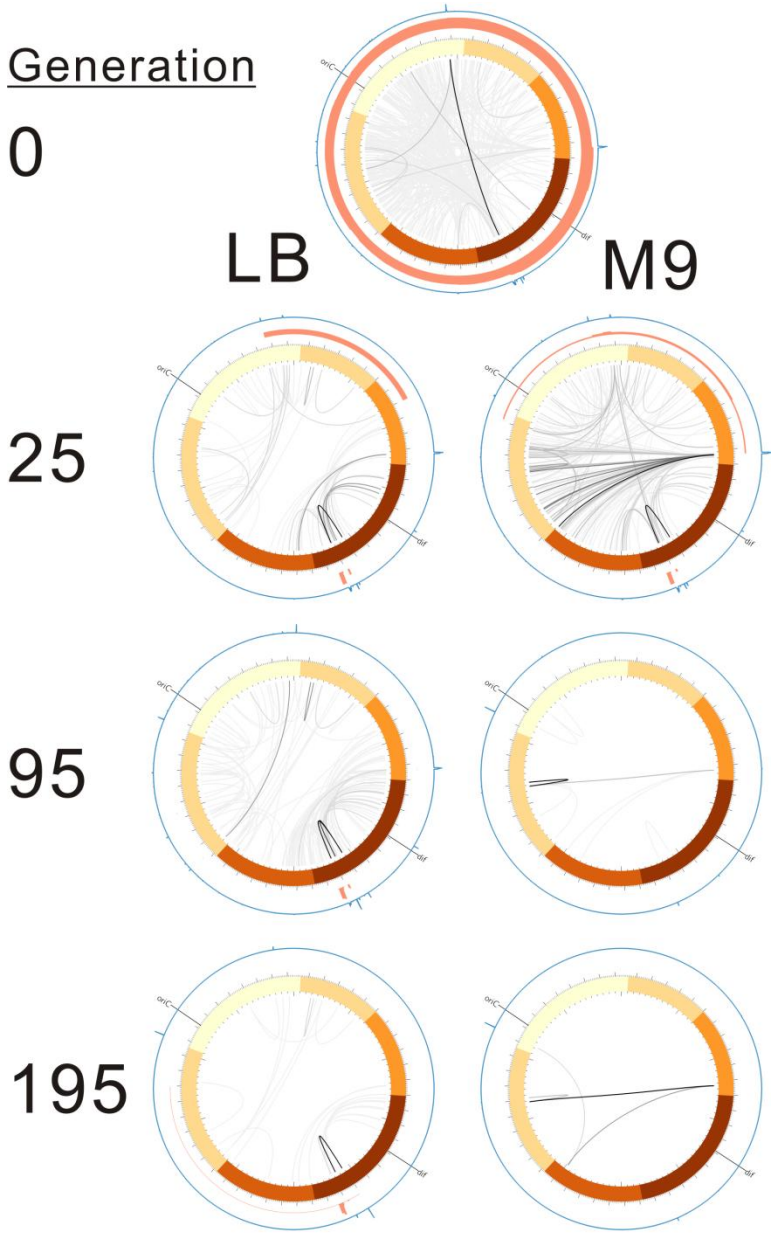

Figure 4.9. Graphical depictions of common genomic modifications in replicate 1.

From the outside, the opposing black lines denote locations of the *ori* and *dif* sites; the blue ring represents the frequency of unrecombined insertions; the red ring represents the frequency of deletions; the thick ring denotes locations in the MG1655 genome, colored by structural domain (Valens et al, 2004); and the links in the interior represent inversions, colored according to frequency, with darker links being more common. The most frequent 15% of rearrangements are shown for generation 0, the top 30% for generation 25, the top 80% for generation 95, and the top 90% for generation 195. Made using Circos (Krzywinski et al, 2009).

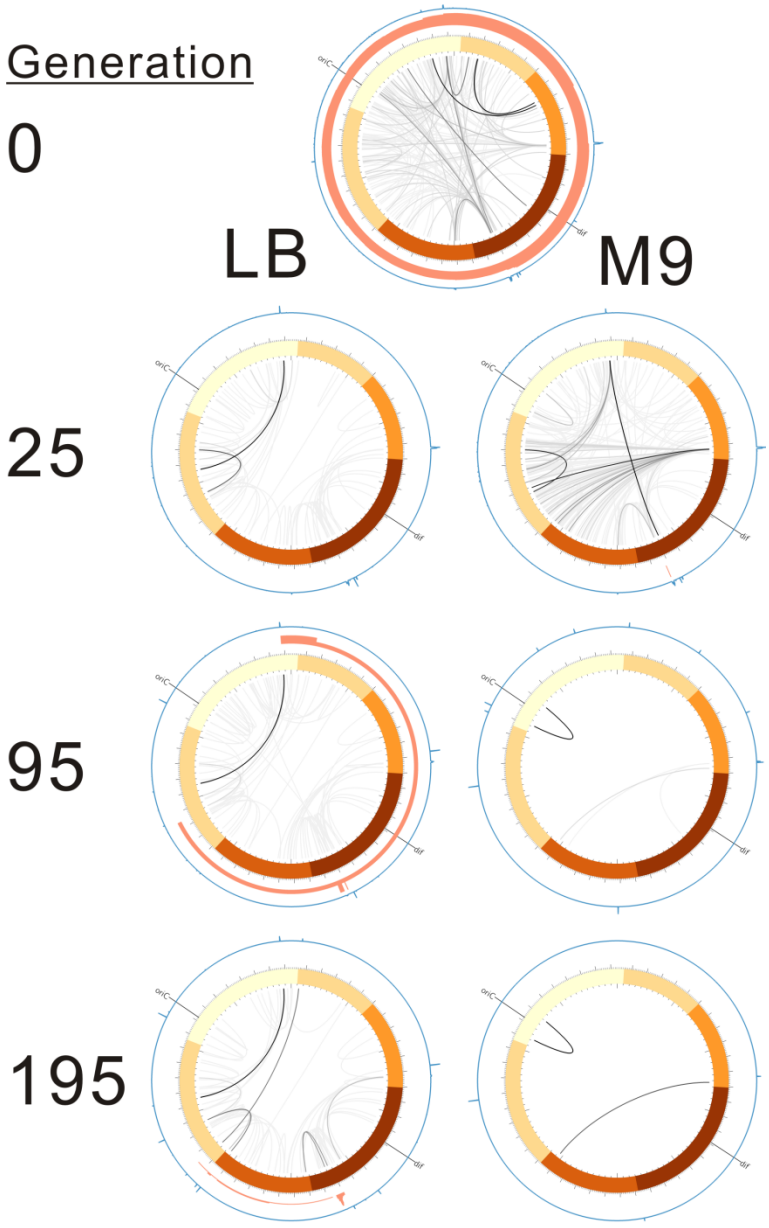

Figure 4.10. Graphical depictions of common genomic modifications in replicate 2.

From the outside, the opposing black lines denote locations of the *ori* and *dif* sites; the blue ring represents the frequency of unrecombined insertions; the red ring represents the frequency of deletions; the thick ring denotes locations in the MG1655 genome, colored by structural domain (Valens et al, 2004); and the links in the interior represent inversions, colored according to frequency, with darker links being more common. The most frequent 15% of rearrangements are shown for generation 0, the top 30% for generation 25, the top 80% for generation 95, and the top 90% for generation 195. Made using Circos (Krzywinski et al, 2009).

4.2.5 Genetic interpretation of commonly detected modification

One interesting point in the data representations shown in **Figures 4.9** and **4.10** is the fact that certain unrecombined insertions are already very common in generation 0. These are likely insertions that gave a selective advantage during the initial stages of library construction. They do not necessarily remain common over the course of the selections, however, because some are more adaptive than others over longer periods, and because different selection pressures presumably come to bear in the M9 libraries. The set of these insertions is also very similar between replicates 1 and 2, indicating that both have good coverage of the space of simple insertional mutations.

A number of these insertions cluster in a region to left of the *dif* site, from positions 1.96 to 2.02 Mb on the genome. This region is also commonly deleted. These insertions are all in genes for flagellar synthesis. The peak in that region, which also typically remains the longest in the later stages of selection, is in the *flhCD* operon, which is one of the primary regulators of flagellar synthesis (Liu & Matsumura, 1994). Another set of extremely common insertions are those in the region from 1.13 to 1.14 Mb, which is the *flg* operon that also consists of genes necessary for flagellar synthesis and export. These observations can likely be explained by the fact that *E. coli* MG1655 is not only motile but hyper-motile compared to other strains; it contains an insertional mutation in the regulatory region of the *flhCD* operon that upregulates the expression of the operon (Barker et al, 2004). Since building flagella is both clearly energy intensive and dispensable for growth with shaking *in vitro*, shutting down this system is an obvious choice for improving growth under such conditions. Knocking out the *flhCD* operon of *E. coli* MG1655 has in fact previously been shown to be adaptive even in the mouse intestine (Gauger et al, 2007).

The other set of common insertions in the early-stage libraries are in the *hsdR* and *hsdM* genes at about 4.58 Mb (near the top of the circles in **Figures 4.9** and **4.10**). These encode the components of the native restriction system, EcoKI. Inactivation of these genes may have been selected for in the library construction stage in order to facilitate plasmid maintenance, though these mutants also often persist throughout the selection, particularly in the LB libraries.

Next, the most successful modifications in the various libraries will be discussed. **Table 4.1** shows the ten most common modifications in the LB-replicate 1 library.

Frequency	Position 1	Position 2	Rearrangement type	Rearrangement size	Position 1 Gene	Position 2 Gene
25.66%	1907558	1907558	None		*yobF*	*yobF*
22.53%	3771847	3771848	None		[intergenic]	[intergenic]
6.79%	4534941	4534942	None		[intergenic]	[intergenic]
6.14%	1978167	1978166	None		*flhD*	*flhD*
1.52%	1915600	1977300	Inversion	61700	[intergenic]	*flhC*
1.34%	2022100	2044700	Deletion	22600	*fliP*	[intergenic]
1.09%	3260360	3260361	None		*tdcE*	*tdcE*
0.74%	1978426	1978427	None		[intergenic]	[intergenic]
0.71%	1977368	1977369	None		*flhC*	*flhC*
0.68%	2021200	2044700	Deletion	23500	*fliN*	[intergenic]

Table 4.1. Most common genomic modifications at generation 195 in LB-replicate 1 library

Knock-outs and deletions of flagellar genes are well represented among the top ten, and the only inversion in the top ten also has an anchor site in *flhC*. The gene apparently knocked out in the dominant variant, *yobF*, is not very well studied, but is known to be expressed in stationary phase (Hemm et al, 2008) and to increase sensitivity to cell-envelope stress and acid (Hobbs et al, 2010).

The second most common variant is in the intergenic space between the *yibIH* operon and the *mtlADR* operon. Little is known about the former except that it seems to be involved in nitrate/nitrite redox reactions (Constantinidou et al, 2006), whereas the latter encodes machinery for mannitol import and phosphorylation (Postma et al, 1993). The upstream regulatory elements of both operons are found in this intergenic space, which also hosts one of the highest concentrations of cAMP-receptor protein (CRP) binding sites in the genome (Shimada et al, 2011). CRP binding sites are apparently frequently found in intergenic spaces between divergently transcribed genes. The transposon insertion site is just outside the CRP binding sites, nearer to the *yibIH* operon, and may thus affect expression of both operons.

Frequency	Position 1	Position 2	Rearrangement type	Rearrangement size	Position 1 Gene	Position 2 Gene
10.55%	1078829	1078828	None		*putA*	*putA*
8.70%	3828680	3828679	None		[intergenic]	[intergenic]
6.11%	3159829	3159828	None		*ygiQ*	*ygiQ*
6.01%	2843087	2843086	None		*hycH*	*hycH*
4.44%	4581382	4581381	None		*hsdS*	*hsdS*
4.23%	1824229	1824228	None		*cho*	*cho*
3.17%	2866790	2866789	None		*rpoS*	*rpoS*
2.61%	4581300	3348100	Inversion	1233200	*hsdS*	*npr*
2.31%	59717	59716	None		*rluA*	*rluA*
2.13%	4582929	4582928	None		*hsdM*	*hsdM*

Table 4.2. Most common genomic modifications at generation 195 in LB-replicate 2 library

The top performing modifications for replicate 2 in LB are shown in **Table 4.2**. As compared to replicate 1, replicate 2 has a less clear winner and seems to be more enriched for simple insertional mutants. The sum of frequencies of insertions in *hsdS* or *hsdM* in the top ten would put them in second place overall. The top performing variant is an insertion in *putA*. PutA acts in proline degradation and in the presence of proline

also represses expression of the *put* operon, which also encodes a proline transporter, PutP (Larson et al, 2006; Zhou et al, 2008). Knocking out this gene may thus increase proline uptake.

The second place intergenic insertion is also between two divergently transcribed genes, *gltS* and *xanP*, and thus could potentially affect regulation of both. GltS is one of four glutamate transporters in the cell (Schellenberg & Furlong, 1977), while XanP is a xanthine transporter (Karatza & Frillingos, 2005). The insertion site is fifteen bases upstream of the start codon of *gltS* and thus very likely has an effect on its expression. Knocking out *gltS* has been shown to be beneficial both in butanol production (Smith & Liao, 2011) and in surviving ionizing radiation (Harris et al, 2009), though it is not clear why this should be.

Frequency	Position 1	Position 2	Rearrangement type	Rearrangement size	Position 1 Gene	Position 2 Gene
49.12%	3815794	3815795	None		[intergenic]	[intergenic]
12.84%	2016285	2016284	None		*fliH*	*fliH*
2.91%	3820697	4169232 (+6 more)	Undetermined		*ligB*	[intergenic]
2.18%	3411475	3411476	None		*fis*	*fis*
2.06%	3820696	226349 (+6 more)	Undetermined		*ligB*	*fliH*
1.92%	1134100	3379500	Inversion	2245400	*flgF*	*yhcM*
1.49%	757756	757755	None		*sdhB*	*sdhB*
1.32%	1756861	1756862	None		*pykF*	*pykF*
1.29%	755760	755761	None		*sdhD*	*sdhD*
1.22%	1965984	1965984	None		*flhB*	*flhB*

Table 4.3. Most common genomic modifications at generation 195 in M9-replicate 1 library

The top ten most common modifications at the end of the selection for replicate 1 in M9 are shown in **Table 4.3**. In this case we have a very clear winner, which is an insertion in the intergenic space between the *rph* and *pyrE* genes. The *pyrE* gene encodes orotate phosphoribosyltransferase, which takes part in pyrimidine biosynthesis and is required for growth in minimal media (Patrick et al, 2007). The *rph* and *pyrE* genes are actually part of an operon, and there is an attenuator in this intergenic space that serves to negatively regulate *pyrE* (Andersen et al, 1992). The insertion site is directly adjacent to the attenuator sequence on the *pyrE* side and thus seems likely to function in removing attenuation.

FliH is another protein involved in flagellar biosynthesis and is required for targeting the flagellar ATP synthase to the export machinery (McMurry et al, 2006). Though it of course makes sense to down-regulate the flagellar machinery, as discussed above, it is interesting that this particular mutation would prove more adaptive here than, for instance, the *flhCD* knock-outs discussed above.

The two "undetermined" modifications in M9-replicate 1 almost certainly represent the two end points of an inversion. Doing a BLAST search on the multiply matched sequence turns up matches the ribosomal genes *rrnA*, *B*, *C*, *D*, *E*, *G*, and *H*. The nature of the inversion is such to potentially allow the regions downstream of the *ligB* gene to be controlled by the strong ribosomal promoter. The first operon that is downstream of *ligB* and oriented so as to be able to be activated by the ribosomal promoter is *rph-pyrE*, the same operon modified by the insertion in the top-performing variant.

Another interesting high-frequency hit in the M9-replicate 1 reads is *fis*. Fis plays a wide variety of roles (Finkel & Johnson, 1992; Travers et al, 2001) and is one of the most abundant DNA-binding proteins in the cell during log phase (Ali Azam et al, 1999).

Thus it is surprising to see this gene on the list of competitive insertional mutants. However, Fis is not required for growth, even in minimal media (Baba et al, 2006; Joyce et al, 2006). A close look at the insertion shows that it serves to introduce an in-frame stop codon in 3' end of the gene. Mapping the downstream codons to a crystal structure of Fis bound to DNA (Stella et al, 2010), as shown in **Figure 4.11**, shows that the insertion is perfectly placed to remove the DNA binding domain while leaving the rest of the protein intact. Thus it seems this insertion results in a variant of Fis that is still able to dimerize but is not able to bind DNA.

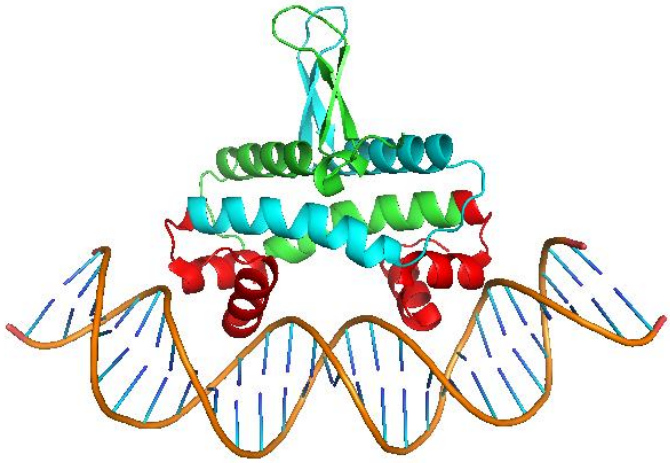

Figure 4.11. Structure of the detected Fis truncation.

Fis binds as a dimer, and the monomers are shown in blue and green. The residues that are removed by the insertion detected in the M9-replicate 1 library are shown in red. Structure generated from Protein Data Bank (http://www.pdb.org) file 3IV5 (Stella et al, 2010) using PyMol.

The ten most common rearrangements found at the end of the M9-replicate 2 selection are given in **Table 4.4**. The winner here is unique among the samples studied here not only in being an inversion but also in representing the highest fraction of the total of any of the winners. The *cyaA* gene at one end of the inversion encodes adenylate cyclase, which is likely dispensable in this context since cAMP signaling is not used when glucose is the only carbon source present, as is the case here. The other side of the

inversion is four bases upstream of the *pyrE* gene. Taken as whole, the inversion is situated to remove the attenuator upstream of *pyrE* and place *pyrE* under control of the *cyaA* promoter.

Frequency	Position 1	Position 2	Rearrangement type	Rearrangement size	Position 1 Gene	Position 2 Gene
51.33%	3993200	3815700	Inversion	177500	cyaA	[intergenic]
34.12%	2810900	1173900	Inversion	1637000	mprA	[intergenic]
2.09%	3379282	3379281	None		yhcM	yhcM
1.19%	1137239	1137238	None		flgI	flgI
1.17%	3379173	3379173	None		yhcM	yhcM
0.80%	2176115	2176114	None		gatZ	gatZ
0.62%	1133817	1133818	None		flgF	flgF
0.53%	3379774	3379775	None		yhcM	yhcM
0.52%	2573800	1135400	Inversion	1438400	eutT	flgG
0.40%	2315014	2315014	None		rcsD	rcsD

Table 4.4. Most common genomic modifications at generation 195 in M9-replicate 2 library

The second-place inversion is more difficult to interpret, but also intriguing. At generation 95 the *cyaA-pyrE* inversion actually represented 63% of the detected modifications, with the second-place inversion at 11%, but by generation 195 the second inversion seems to be closing the gap. The *mprA* gene at one end of this inversion encodes a multidrug resistance regulator that regulates the operon of which it is a part, which includes multidrug resistance pumps, and that activates transcription in response to antimicrobial agents (Lomovskaya et al, 1995). Thus under present conditions, this gene would seem to be dispensable and its promoter weakly activated. An RNA polymerase moving from a promoter on the other side of the inversion would transcribe the genes for the multidrug resistance pumps in the operon downstream of *mprA*, which is not obviously adaptive in the present circumstances. In considering whether the inversion is

placed to put another gene under the control of the *mprA* promoter, the first gene that would be in position to be transcribed in this way is *mfd*. The Mfd protein removes stalled RNA polymerases from DNA lesions (Selby & Sancar, 1993), but it is not clear what benefit would be gained from expressing this protein under the *mprA* promoter. Furthermore, this inversion serves to severely disrupt the replichore balance of the chromosome, adding to the mystery of what benefit it provides the cells in which it is found.

4.3 DISCUSSION

The results presented above represent a large initial dataset for exploring the limits of genome plasticity and the utility of artificially induced genomic rearrangements for directed evolution of bacteria. Perhaps the most striking aspect of this data is the differences that arise when the same libraries are grown in LB versus M9 (0.4% glucose) minimal media. In comparison to LB, the libraries selected in M9 appear to achieve greater fitness relative to wild-type, become pared down to lower diversity, and yet seem more tolerant to major genomic rearrangements. Indeed, the most successful single variant in any of the libraries, as judged by the proportion of the population occupied (51%), was an inversion around the origin in replicate 2 grown in M9. The other inversion in M9-replicate 2 occupied 34% of the population, which places it third among proportional rankings, after the winning variant in M9-replicate 1. Taken together, these observations suggest that the wild-type *E. coli* strain MG1655 is better adapted to growth in LB than in M9, and that more drastic changes to genome structure are better tolerated when the progenitor organism is poorly adapted to its environment.

The low diversity in the end-stage M9 libraries as well as a reproducible trend in the top performing variants points to one major issue preventing *E. coli* from being as

well adapted to growth in M9 as it might otherwise be: the attenuation of the *pyrE* gene, which is an essential gene in minimal media and codes for orotate phosphoribosyltransferase (OPRTase). This enzyme synthesizes orotidine 5'-monophosphate from orotic acid and phosphoribosyl pyrophosphate (Bhatia et al, 1990; Scapin et al, 1994; Scapin et al, 1995). Orotidine 5'-monophosphate in turn is an intermediate in the synthesis of uridine monophosphate (UMP), which is required for RNA synthesis.

The method of attenuation works in a fashion somewhat analogous to the well-known *trp* attenuator, where low levels of tryptophan cause the ribosome to stall at a string of tryptophan codons, in turn preventing formation of a terminating hairpin, which allows RNA polymerase to transcribe the entire operon (Platt, 1981). The *pyrE* gene is activated in a similar fashion by low levels of UMP. The site where transcription is terminated is enriched in uridine residues, and the mechanism in this case seems to be that the RNA polymerase slows down in this region when UMP is low, allowing a pursuing ribosome to catch up and prevent formation of the termination hairpin (Andersen et al, 1992; Bonekamp et al, 1984). As it turns out, *E. coli* strains MG1655 and W3110 harbor a two-base deletion near the end of the *rph* gene (Jensen, 1993). This results in a premature stop codon, which in turn means the ribosome leaves the RNA before ever reaching the attenuator region, leaving the terminator hairpin intact.

The importance of fixing this problem for growth in minimal media is evident both in the number of different ways it was solved, and in the large growth advantages that clearly incurred to those individuals that solved it. The dominant strain in M9-replicate 1 harbors an insertion six bases downstream of the attenuated transcriptional stop site and presumably serves to break the attenuation. The most common inversion in M9-replicate 1 seems to put the entire operon under the control of a ribosomal promoter,

which likely compensates for reduced read-through in the attenuator with increased transcription frequency, and may also incur some benefit in coupling UMP regulation to ribosomal regulation. Finally, the dominant variant in M9-replicate 2 is an inversion that releases *pyrE* completely from control of the *rph* promoter and attenuator and instead places it under control of the promoter for the adenylate cyclase gene, which is negatively regulated by cAMP-CRP (Aiba, 1985) and is therefore presumably freed from repression due to the interruption of the gene responsible for producing cAMP. These solutions also highlight the potential utility of artificial inversions for rewiring gene expression to better match the demands of the local environment. An interesting experiment might be to reconstruct the ancestral sequence of the *rph-pyrE* operon (specifically, by reversing the two-base deletion) in order to examine how the solution these strains have found to the problem of *pyrE* regulation compares to that which was had previously evolved.

Though explanations are evident for many of the mutations and rearrangements found here, others are harder to explain. The truncated Fis protein generated in M9-replicate 1 is very curious, for instance, and expression studies or protein binding studies such as two-hybrid assays might be done to examine the effect of this mutation. Additionally, the second-most common modification in M9-replicate 2 is a large inversion that seems to severely unbalance the chromosome without resulting in an obviously beneficial change in gene expression. Since this strain represents nearly one third of the final library, it should be easy to isolate. A full-genome sequence of the strain may yield more clues as to what has happened. Alternatively, an attempt could be made to recapitulate this inversion using λ Red or the methods described in Chapter 3 in order to verify that it is real. Finally, the presence and even persistence of apparently very large deletions in the libraries is intriguing. The fact that these large deletions are relatively persistent in the M9 libraries relative to smaller deletions, despite the fact that

deletions as a whole are more disfavored in the M9 libraries compared to the LB libraries, would seem to indicate that the regions are not actually deleted but rather displaced. Thus an attempt to artificially recreate one or more of these deletions, such as the large deletion that can be seen at generation 25 of M9-replicate 1 in **Figure 4.9**, followed by sequencing of survivors in order to see what happened, could be instructive. An attempt could also be made to generate the deletion in a *recA*- strain in order to test the hypothesis that homologous recombination is responsible for the maintenance of these deletions.

The data presented here suggest that genomic rearrangements are more likely to be adaptive when the cell is poorly matched to its environment. However, another goal in examining the limits of genome plasticity would be to explore the space of neutral rearrangements. How many different genome rearrangements could be made that are equivalent to the current arrangement? This is difficult to answer in the present case since neutral mutations are quickly drowned out by adaptive ones. Perhaps performing experiments such as these in a truly wild-type bacterium in the environment it evolved in could yield more data on neutral rearrangements. Alternatively, the data presented here constitutes a starting point for constructing strains ultra-adapted to growth in culture. For instance, the cells in these contexts clearly benefit from removing genes for flagellar proteins and restriction enzymes, and of course the problem with *pyrE* attenuation must be resolved for optimal growth in M9. F plasmid (Quandt et al, 2013; Winkler & Kao, 2012) or P1 transduction (Donath et al, 2011) could potentially be employed to combinatorially find the best set of modifications.

This approach could of course also be of potential utility in adapting strains to industrial conditions, such as high temperatures or the presence of organic solvents. Redoing the selections presented here under such conditions, such as at 42°C or the in the

presence of ethanol, would be an obvious next step, both for further testing the hypothesis that major rearrangements are more likely to be selected for when the strain is poorly adapted to its environment, and to test the power of this approach in yielding industrially useful results.

Additional work in this area could also include building libraries containing more *lox* sites per genome. For instance, three *lox* sites per genome would allow cut-and-paste reactions, as described in Chapter 3, and six *lox* sites per genome would allow entire regions to swapped. If it is the case that there are other viable but drastically different genome configurations that are nonetheless difficult to reach from the current position on the fitness landscape, such more extreme measures might be more likely to find them.

Alternatively, a modification of this approach could be used to analyze double knock-out libraries. In method used here, double knock-outs are created, then recombined, and then selected. However, if the recombinations were instituted after the selection, they would serve to bring together the two selected insertion sites at one locus for easy identification in subsequent sequencing analysis. Cre would likely have to be tightly and inducibly expressed from the genome for such methods, and steps would have to be taken to assess the potential biasing effect of distance and lethal rearrangements on the final makeup of the sequencing pool, but the end result could yield a powerful new tool for genetic analysis.

Finally, the fact that both *mariner* transposons and the Cre/*lox* system function efficiently across the domains of life means that this approach can be widely adapted to different species. These results thus open a gate to large number of potential follow-up studies with significance for both basic science and industrial biotechnology.

4.4 MATERIALS AND METHODS

4.4.1 Plasmid construction

To construct the temperature-sensitive hyperactive Himar1-expressing plasmid pQLH1, the *himar1* gene was amplified via polymerase chain reaction from plasmid pSC189 (Chiang & Rubin, 2002) using primers himar1u5p (ATGGAAAAAAAGGAATTTCGTGTTT) and himar1d5p (TTATTCAACATAGTTCCCTTCAAGA) (though these primers were ordered with 5' phosphates, this was not necessary for the method in which the plasmid was finally constructed), and the backbone of the pQL269 plasmid was amplified without the *cre* gene using primers pqATGr_hi1uOL (AAACACGAAATTCCTTTTTTTCCATTTAACACTCAGCGGCCGCCTAG) and pqTAGf_h1dOL (TCTTGAAGGGAACTATGTTGAATAGCCCGGGAAGCCGAATTCG). The two amplicons were then assembled into the pQLH1 plasmid using the Gibson method (Gibson, 2011). This plasmid has the rep101ts origin, is temperature sensitive, and harbors spectinomycin resistance.

To construct pQL269A, the pQL269 backbone was amplified without the *aadR* gene that imparts spectinomycin resistance using primers pq6111f5p (TGTCTAACAATTCGTTCAAGCCGAC) and pq4945r5p (TCTCCACGCATCGTCAGG), and the ampicillin resistance gene from pUC19 was amplified using primers pUCampru (GTCAGGTGGCACTTTTCGG) and pUCamprd (AACTCACGTTAAGGGATTTTGGTCA). (All these primers had 5' phosphates, though technically only one pair requires them.) These two amplicons were ligated together with T4 ligase, transformed, and plated on LB with ampicillin to yield colonies harboring pQL269A, which is essentially identical to pQL269A except that it expresses

ampicillin resistance instead of spectinomycin resistance. An ampicillin-resistant variant of pQLH1, pQLH1A, was also constructed in the same manner but was not used in the present work.

4.4.2 Transposon construction and electroporation

To generate an amplicon consisting of *Himar1* inverted repeats (mutated to carry MmeI recognition sites) flanking *lox66* sites flanking a kanamycin resistance gene, primers kantuIS (GAGAGAGATACAGGTTGGATGATAAGTCCCCGGTCTATAACTTCGTATAGCATACATTATACGAACGGTAGGAAAGCCACGTTGTGTCTC) and kantdIS (GAGAGAGATACAGGTTGGATGATAAGTCCCCGGTCTTACCGTTCGTATAATGTATGCTATACGAAGTTATTGAGGTCTGCCTCGTGAAG) were used to amplify the kanamycin resistance gene from a variant of the pACD.EcI5.1806s plasmid carrying kanamycin resistance instead of chloramphenicol resistance (the kanamycin resistance gene originated from pUC4K). Analogously, primers cattuIS (GAGAGAGATACAGGTTGGATGATAAGTCCCCGGTCTTACCGTTCGTATAGCATACATTATACGAAGTTATCGTTGATCGGCACGTAAGA) and cattdIS (GAGAGAGATACAGGTTGGATGATAAGTCCCCGGTCTATAACTTCGTATAATGTATGCTATACGAACGGTACGGGTCGAATTTGCTTTC) were used to amplify the chloramphenical acetyltransferase gene from the pX20 plasmid so as to flank the *cat* gene with the modified inverted repeats and *lox71* sites.

Initial tests indicated transposition efficiency was better with extra sequence on the ends of the transposons and after ligating the transposons into circles. So the kantuIS/kantdIS and cattuIS/cattdIS PCRs were digested with DpnI to remove residual plasmid and then used as a template for a second PCR reaction using bp2_udM

(5Phos/GGAAGTGCTGGAGTTATGCGGAGAGAGATACAGGTTGGATGATAA) as both the forward and reverse primer. Phusion was used as the polymerase.

400 to 800 µL of these PCR reactions were subjected to PCR clean-up and recovered in 150 to 300 µL, which were concentrated in a vacuum concentrator, resuspended in 20 µL, desalted using cellulose filters having a pore size of 0.025 µm (Millipore catalog no. VSWP01300), divided in two, and ligated overnight at room temperature with 3 µL of T4 DNA ligase (NEB, 20 U/µL) in 30 µL. These ligations were then pooled, concentrated in a vacuum concentrator, resuspended in 15 µL, and desalted. (The pQL269A plasmid was also concentrated and desalted in the same way before electroporation.)

Electroporations were performed using a Bio-Rad Micropulser and chilled 0.1-cm cuvettes at 1.8 kV, with 1 to 1.5 µL of the ligated PCR products in 40 to 60 µL freshly made electrocompetent cells (*E. coli* MG1655 harboring the pQLH1 plasmid) per electroporation.

Growth curves were measured as described in **Section 2.4.3**.

4.4.3 Preparation of sequencing libraries

The protocol used to prepare the sequencing libraries was a modification of the ddRAD protocol (Peterson et al, 2012) available at http://www.bit.ly/ddRAD. Genomic DNA was prepared using a kit from Sigma (GenElute, catalog no. NA2110-1KT). 1 to 4 µg of genomic DNA were digested using 2 U of MmeI per microgram of genomic DNA and 50 µM S-adenosyl methionine in the NEB CutSmart buffer in 50 to 100 µL total reaction volumes. Digestions were performed for five hours at 37°C (though likely little was gained from digesting five hours versus one), and subjected to column clean-up without heat inactivation.

For adaptors, P5.1M (ACACTCTTTCCCTACACGACGCTCTTCCGATCTNN) and P5.2M (5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT) were annealed together, and P7.1M (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNN) and P7.2M (5Phos/AGATCGGAAGAGCGAGAACAA) were annealed together by combining 15 µL of 100 µM solutions of each oligomer plus 0.3 µL of 5 M NaCl, keeping at 95°C for five minutes, then ramping down to 14°C at 0.1°C per second. 5 µL aliquots were stored at –20°C.

After inputting the results from BioAnalyzer analyses of the restriction digests into the ddRAD ligation molarity calculator, it was decided to use 2 µL of 20× dilutions of each of the annealed adaptor solutions per µg of digested DNA (0.5 to 2 µg total) in ligations with 2 µL 20 U/µL T4 DNA ligase plus buffer and water to 40 µL. The ligations were put at 16°C for one hour, 65°C for 10 minutes, and then ramped back down to 16°C at 0.1°C per second. The ligations were then subjected to a bead-based cleanup (Agencourt Ampure XP) using a ratio of 1.5 beads:ligation mixture.

Size selection to 195±10 basepairs was performed using a Pippin Prep system. The samples were recovered in 50 µL. The size-selected ligations were then PCR amplified using Phusion with ddRAD primers, with 25 µL of sample in 100 µL reactions. The thermocycle was as follows:

1. 98°C for 30s
2. 98°C for 10s
3. 65°C for 30s
4. 72°C for 30s
5. repeat 2-4 11 times
6. 72°C for 5 min

7. hold at 4°C

The PCRs were cleaned up using beads as above. If subsequent Bioanalyzer results indicated the presence of amplified adaptor-adaptor ligations, these were removed via bead-based size selection (which could potentially be performed as the final bead clean-up, instead).

Sequencing was performed on an Illumina MiSeq V2 with 2×100 mate-pair reads. Generations 0 and 25 were sequenced on one lane, and generations 95 and 195 were sequenced on another. The indices of the ddRAD primers were used to differentiate samples, though barcodes could also be added to the P5M adaptor sequence above just before the MmeI overhangs.

4.4.4 Analysis of sequencing results

The presence of the 5'-most fifteen bases of the modified transposon inverted repeat on both reads of the mate-pair was taken as required to accept a mate-pair read as covering an insertion site. A reasonably good alignment to the inverted repeat sequence could be potentially be used instead, but for present purposes we decided to err on the side of making correct determination. Once a transposon-containing mate-pair read was identified, non-genomic sequence was stripped out, an identification of the type of *lox* site present was added to the tag, and both reads were added to a single FASTQ file containing all the usable reads for a given sample. These were then aligned to the *E. coli* MG1655 genome using BWA (samse) with the default parameters. Analysis of recombinations was then performed on the aligned sequences, as discussed in **Section 4.2.4** and diagrammed in **Figure 4.4**. Any rearrangements occurring within the same 100 bases of genome sequence were considered the same for analysis purposes, though this number could potentially be modified up or down as desired. Other than BWA and

Circos, all the analyses were performed using scripts written in Python by the author of this work. These scripts are available upon request.

Chapter 5: Toward implementing and improving extracellular electron transfer in *Escherichia coli*

5.1 INTRODUCTION

Certain bacteria such as *Geobacter sulfurreducens* and *Shewanella oneidensis* are capable of the remarkable feat of exporting electrons from the cell for purposes of anaerobic respiration (Lovley et al, 2004; Nealson et al, 2002). In short, in the absence of oxygen, such bacteria search out oxidants (often metals) in their environment to act as terminal electron acceptors and transfer electrons across the outer membrane to these oxidants. This process has a number of potential applications, including electricity production via microbial fuel cells (Logan, 2010), bioremediation via the reduction of soluble uranium (VI) to insoluble uranium (IV) (Wall & Krumholz, 2006), and interfacing manmade electrical systems with microbial biochemical systems. It has also been shown that the same pathway can be used to drive electrons into the cell (Ross et al, 2011), raising the possibility of using electricity directly as a reducing equivalent for improving the efficiency of biosynthesis. Thus there is also great interest in being able to transfer the machinery for extracellular electron transfer to heterologous hosts in order to combine them with the advantageous properties of other strains, such as the variety of genetic tools and well-understood metabolic networks of *Escherichia coli*, for example.

The pathway for extracellular electron transport is probably best understood for *S. oneidensis*. A schematic of this pathway as currently understood is shown in **Figure 5.1**. The multi-heme cytochrome CymA can be considered the start of pathway. CymA is localized to the periplasm, anchored to the inner membrane, and can take electrons from the quinol molecules that serve as electron shuttles in the electron transport chain (Shi et al, 2012). CymA seems to transfer electrons directly to another multi-heme cytochrome, MtrA (Firer-Sherwood et al, 2011; Schuetz et al, 2009), which is expressed from the

mtrCAB operon that forms the core of this pathway. MtrA localizes to the periplasm (Beliaev et al, 2001) and forms a complex with the other members of the *mtr* operon, the outer-membrane proteins MtrB and MtrC (Ross et al, 2007). MtrB is not a cytochrome but has been hypothesized to aid in the transfer of electrons from MtrA to MtrC across the outer membrane (Hartshorne et al, 2009). MtrB is also required for proper localization of MtrC and OmcA (Myers & Myers, 2002). MtrC also binds OmcA (Shi et al, 2006), and these two cytochromes have been implicated as the terminal reductases in the chain (Borloo et al, 2007). All five of these proteins cause large drops in current density when absent (Bretschger et al, 2007).

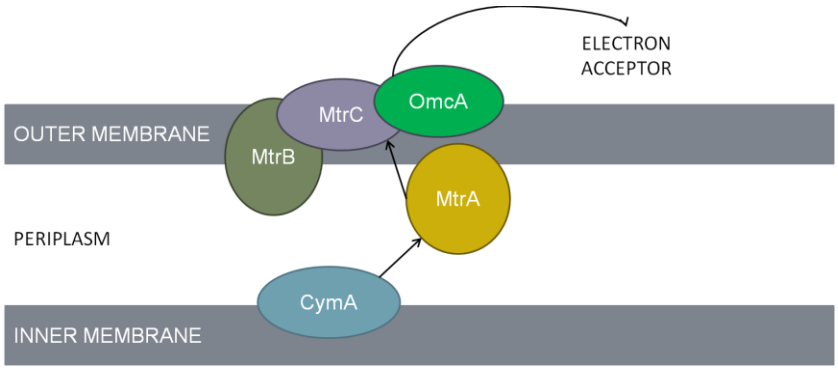

Figure 5.1. Extracellular electron transport pathway in *Shewanella oneidensis*.

Given that CymA alone (Gescher et al, 2008) or MtrA alone (Pitts et al, 2003) have been shown to allow *E. coli* to reduce solubilized iron(III), and that OmcA has been shown to properly localize to the outer membrane in *E. coli* (Donald et al, 2008), we hypothesized that the entire pathway could be reconstituted in *E. coli*, and set out to do so.

5.2 RESULTS

5.2.1 Expression of the *mtrCAB* operon in *Escherichia coli*

Since MtrA can interface with the rest of the *E. coli* metabolic machinery, and the MtrCAB complex forms a membrane-spanning electron conduit, as described above, we reasoned that the expression of the *mtrCAB* operon alone might be sufficient to observe extracellular electron transfer to solid substrates in *E. coli*. We therefore amplified both the entire *mtrCAB* operon and, separately, the *omcA* gene, directly from the *S. oneidensis* genome and cloned them onto a modified pBR322 vector under control of the *tacI* promoter. We transformed these plamids (pRS.M expressing the MtrCAB operon and pRS.O expressing OmcA) into *E. coli* BL21 along with a plasmid (pEC86 (Arslan et al, 1998)) constitutively expressing cytochrome maturation genes, and subjected them to a visually screenable iron reduction assay (Cho & Ellington, 2007) by placing them in anaerobic conditions with a non-fermentable compound (glycerol) as the only carbon source and amorphous iron(III) as the only electron acceptor, thus making iron reduction a requirement for survival. In different samples, pRS.M was induced using 0, 1, 10, 50, 200 μM IPTG, and pRS.O was induced using 10 and 200 μM IPTG. *S. oneidensis* was used as a positive control (with lactate instead of glycerol), and *E. coli* BL21, *E. coli* BL21 plus the pEC86 plasmid, *E. coli* BL21 plus the pRS.M plasmid, and sterile media were used as negative controls. The results after 10 days are shown in **Figure 5.2**.

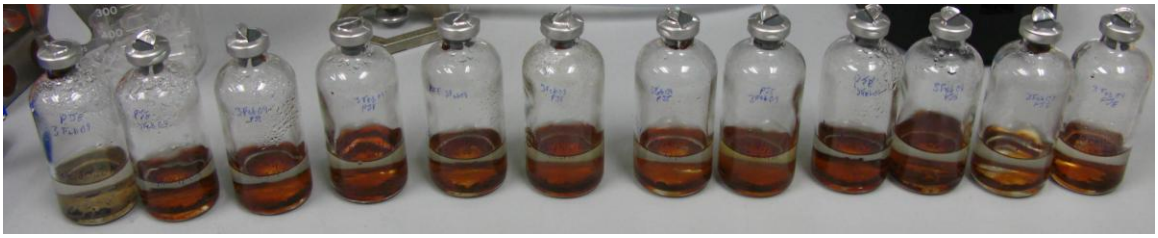

Figure 5.2. Insoluble iron(III) reduction assay for testing functionality of *Shewanella cytochromes* in *E. coli*.

The left-most bottle is a positive control harboring wild-type *Shewanella oneidensis* MR-1. The rest are samples containing cytochrome-expressing plasmids at various levels of induction in *E. coli* BL21, as well as negative controls.

Visually, no difference could be detected between negative controls and those samples containing cytochrome-expressing *E. coli* strains. After one month, the bottles were shaken, and 50 μL from each of the *E. coli*-containing samples was inoculated into LB plus the antibiotics pertaining to the expected plasmids, and grown overnight aerobically at 37°C. Growth was seen only in the cultures corresponding to wild-type *E. coli*, *E. coli* plus pRS.M, and *E. coli* plus pRS.M and pEC86 induced at 1 μM IPTG. Freezer stocks were made of the latter two cultures.

Not long after, Jensen and coworkers reported functional expression of the *mtrCAB* operon in *E. coli* under a T7 promoter using a much more sensitive assay for insoluble iron reduction (Jensen et al, 2010), and we suspended further experiments in this vein.

5.2.2 Survey of phylogenetic variants of the *mtrCAB* operon

Though the MtrCAB proteins of *Shewanella oneidensis* MR-1 have served as the model system for studying extracellular electron transfer, a great many alternatives exist, and it is possible that another variant or combination of variants might yield better performance in *E. coli*. Even within *S. oneidensis* there is another operon, *mtrDEF*, which is situated very near *mtrCAB* in the genome and encodes a similar set of

membrane-spanning redox active proteins. However, these proteins have not been much studied until recently (Bucking et al, 2010; Clarke et al, 2011; Coursolle & Gralnick, 2010), and functional expression in a heterologous host remains to be attempted. The *mtrCAB* operon itself still holds some mysteries; the intergenic region between *mtrC* and *mtrA*, which is highly conserved among orthologues of the operon (J. Swaminathan, personal communication), contains a putative hairpin structure that may be involved in regulating expression levels (Beliaev et al, 2001), but this hypothesis remains to be tested.

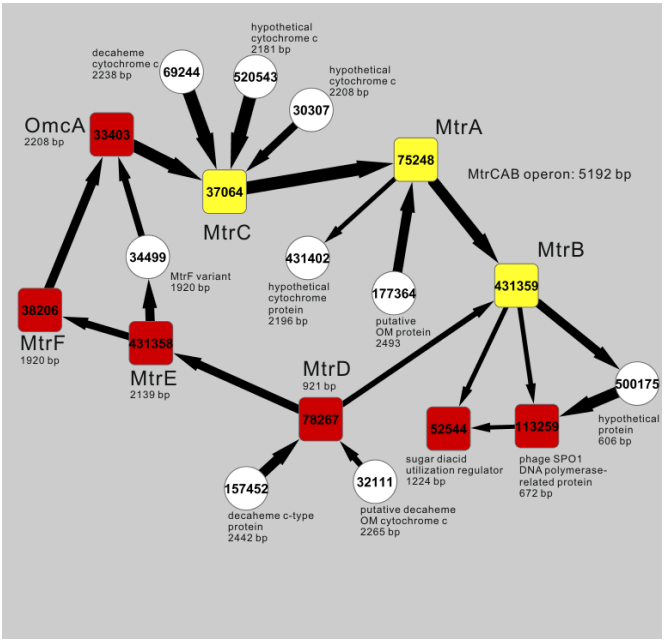

Figure 5.2. Bioinformatic analysis of conserved gene orders of orthologues of *mtrCAB*, *mtrDEF*, and *omcA*.

Each number represents a set of orthologues from different species. Squares represent orthologues found in *S. oneidensis*, with yellow squares for the MtrCAB operon and red squares for other *S. oneidensis* genes. Circles represent genes not found in *S. oneidensis*. Arrows represent a conserved gene ordering, with the thickness of the arrow representing the frequency of the ordering. Data generated by J. Swaminathan.

With help from Jagannath Swaminathan, we therefore performed an initial bioinformatic analysis of gene order and identity in the region of *mtrCAB* operon variants in various bacteria. A summary of the results is shown in **Figure 5.2**.

A list of the locus tags of orthologues of the MtrC, MtrA, and MtrB proteins (corresponding to group numbers 37064, 177364, 75248, 431359, and 431402 in **Figure 5.2**) is given in **Table 5.1**.

MtrC	MtrA	MtrB
(Group 37064)	(Group 75248)	(Group 431359)
SO_1778	SO_1777	SO_1776
Shewmr7_2578	Shewmr7_2579	Shewmr7_2580
Sfri_2637	Sfri_2638	Sfri_2639
Shew_2525	Shew_2526	Shew_2527
Shewana3_2676	Shewana3_2677	Shewana3_2678
Sputw3181_2623	Sputw3181_2624	Sputw3181_2625
Shewmr4_2510	Shewmr4_2511	Shewmr4_2512
Sama_1208	Sama_1207	Sama_1206
Sbal_1589	Sbal_1588	Sbal_1587
Shew185_1578	Shew185_1577	Shew185_1576
Sputcn32_1478	Sputcn32_1477	Sputcn32_1476
Sbal195_1612	Sbal195_1611	Sbal195_1610
Spea_2698	Spea_2699	Spea_2700
Ssed_1525	Ssed_1524	Ssed_1523
Shal_2784	Shal_2785	Shal_2786
Swoo_3125	Swoo_3126	Swoo_3127
Sbal223_2765	Sbal223_2766	Sbal223_2767
swp_3278	swp_3279	swp_3280
(Group 177364)	Acid_7896	VVA0644
Gmet_0910	Rfer_4082	VP1218
Acid_7897	Gura_3626	VV2_0135
Rfer_4083	ACP_0479	AHA_2766
	GM21_0397	(Group 431402)
		Gura_3627
		GM21_0398
		RPC_2959
		amb3018

Table 5.1. List of phylogenetic variants of genes in the *mtrCAB* operons.

The first locus tag in each column is from Shewanella oneidensis MR-1. Group numbers refer to **Figure 5.2**.

5.3 DISCUSSION

Further work in this area could involve synthesizing numerous alternatives to the *S. oneidensis* MtrCAB operon and screening these alternatives for activity in *E. coli*. Beyond testing rational designs such as the *mtrDEF* operon or an *mtrCAB* operon lacking the intergenic region between *mtrC* and *mtrA*, unnatural *mtrCAB* operons could be randomly assembled from the various orthologues listed in **Table 5.1**. An example of how this might be done using Gibson assembly (Gibson, 2011; Gibson et al, 2009) is shown in **Figure 5.3**. Additionally, designs including genes commonly associated with the members of the *mtrCAB* and *mtrDEF* operons depicted in **Figure 5.2** could also be synthesized and tested.

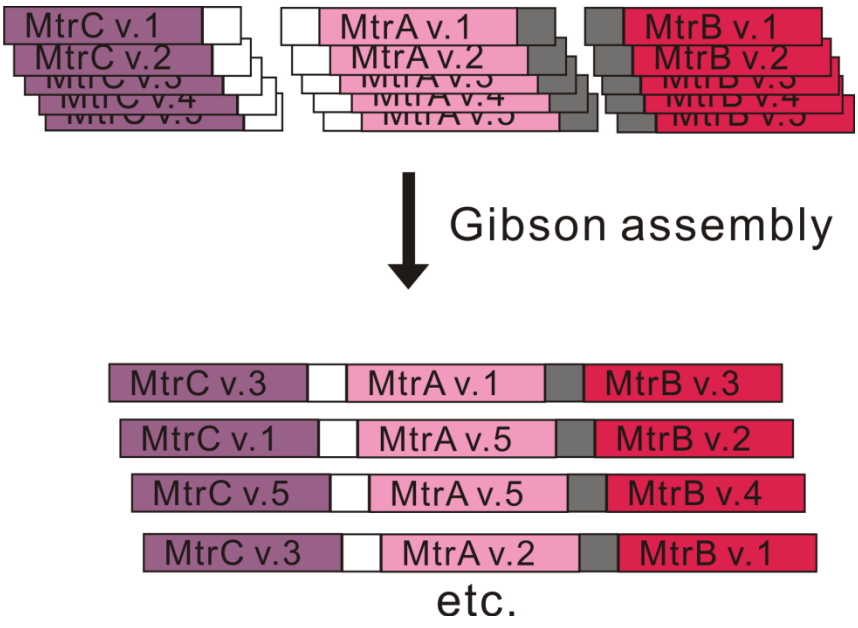

Figure 5.3. Schematic of randomized operon construction.

The white and grey boxes represent regions of homologous sequence; "v. 1," "v. 2," etc. represent different orthologues of the gene in question.

Such designs could yield alternatives to MtrCAB that are more suitable for expression in heterologous hosts such as *E. coli*, but even if that is not the case a great

deal could be learned about the biology of these operons from such work. Furthermore, though synthetic biology to date has generally focused on achieving the goal of rationally designed "plug-and-play" genetic architectures, this would represent a more biologically inspired design approach of performing selections on randomized pools and extracting evolutionarily tested designs from phylogenetic data.

Research in this field can also benefit substantially from the methods developed in other parts of this work. For instance, a complete set of genes expressing CymA, MtrCAB, and OmcA is more than 8000 base pairs, which begins to push the limits of the amounts of DNA advisable to put on plasmids. Furthermore, maximal functional proteins levels of the cytochromes and the cytochrome maturation genes require relatively low expression rates (Goldbeck et al, 2013), which could be achieved by using the methods of **Section 3.2.2** to place these genes on the chromosome instead of on plasmids. Indeed, CymA was functionally expressed from a genomic locus in *E. coli* due to problems with toxicity when expressed on a plasmid (Gescher et al, 2008). Additionally, transposon selection methods such as those described in **Chapter 4** could be used in tandem with the survival assay of **Section 2.2.1** to optimize strains for higher efficiency output.

5.4 MATERIALS AND METHODS

5.4.1 Plasmid construction

Amplicons containing the *omcA* gene and the *mtrCAB* operon were amplified from *Shewanella oneidensis* MR-1 genomic DNA (ATCC #700550D) using the Roche Expand Long Range enzyme mix with primers having the sequences "CCTTTGCTAATGTGTGACTTGG" and "CAGGTGGGATCTAATTCC" (for *omcA*)

or "CTTTAGAGGCTGATCTCATTCA" and "CATAGCGGTTAAGCAATGC" (for *mtrCAB*). These amplicons were cloned into TOPO vectors and verified by sequencing.

Next, the *mtrCAB* operon was amplified from the TOPO vector using the primers "GAGAGAGAGAGAGAGAGAGATCTAGAATGATGAACGCACAAAAATCA" and "GAGAGAGAGAGAGAGACTCCTCGAGTTAGAGTTTGTAACTCATGCTCA," and the *omcA* gene was similarly amplified using primers "GAGAGAGAGAGAGAGAGGAGTCTAGAATGATGAAACGGTTCAATTTC" and "GAGAGAGAGAGAGAGAGAGGCTCGAGTTAGTTACCGTGTGCTTC." These amplicons were digested with XhoI and XbaI and ligated into similarly digested pRS.1 plasmid (containing ampicillin resistance and the lacIq gene on the pBR322 vector backbone; courtesy of Randall Hughes) to place the genes under the control of the *tacI* promoter, resulting in plasmids pRS.M and pRS.O, respectively.

5.4.2 Insoluble iron(III) reduction assay

The amorphous iron(III) gel was made by suspending 27 g of $FeCl_3 \bullet 6H_2O$ in 150 mL water, adjusting the pH to 7 using 10 N NaOH (using unadjusted $FeCl_3 \bullet 6H_2O$ to lower the pH upon overshoot). The suspension was stirred for 30 minutes, and the pH adjusted again. The suspension was then washed six times using water, spinning at 5000 RPM in a Beckman-Coulter Avanti J-20 XP centrigue for 20 minutes each time, discarding the supernatant.

The media used for the iron(III) reduction/survival assays consisted of 50 mM phosphate buffer (pH 7.4, prepared at 2X concentration beforehand), 0.06 M NaCl, 1% v/v ATCC mineral mix, 1% v/v ATCC vitamin mix, and 0.1 M carbon source (glycerol for *E. coli*, lactate (pH 6.4) for *S. oneidensis*). Each bottle contained 20 mL of this media (which included 1 mL of overnight culture added as the source of the cells) and

approximately 0.8 mL of the iron(III) gel. The bottles were sealed and degassed with pure nitrogen delivered via needles inserted through the stoppers for 30 minutes to create an anaerobic environment. Bottles containing *E. coli* were incubated at 37°C, while those containing *S. oneidensis* were incubated at 30°C.

Chapter 6: Mathematical models of synthetic microbial implementations of comparative advantage

6.1 Introduction

One of the justifications for developing methods such as those described in Chapter 3 (and in **Section 3.2.2** in particular) is that as the circuits employed in synthetic biology grow larger and more complex, it becomes more difficult logistically and more detrimental to the host to express such machinery from plasmids. Keeping part or all of this machinery on the chromosome then becomes more appealing, and if the sequences of DNA involved are at all large, then few methods other than those described in **Section 3.2.2** will facilitate that task. Furthermore, once these circuits are in place, methods such as those described in Chapter 4 can potentially be used to optimize their function. In this chapter we present mathematical models for such complex circuits as a proof of principle in advance of actually building them. In particular, these circuits are designed to replicate the characteristics of comparative advantage when expressed in bacteria.

Comparative advantage is a mathematical concept in economics and is thought to underlie many trade interactions. The theory is usually credited to David Ricardo (Ricardo, 1817), but Richard Torrens is also recognized as having made key insights (Torrens, 1815). In simple terms, comparative advantage demonstrates that as long as two groups have differing efficiencies in producing two or more goods, it is typically to the advantage of both to engage in trade, even if one group produces all of the relevant goods with higher efficiency than the other. Though it is tempting to think of economics as a zero-sum game and assume that if one group is gaining another group must be losing, this is not necessarily the case, and comparative advantage provides a mathematical proof of this assertion. Comparative advantage has of course traditionally been applied to the study of human interactions, but the mathematical universality of the

concept implies that comparative advantage could come into play anytime two self-interested entities with different resource bases come into peaceful contact with each other. For instance, two bacteria that produce and export useful metabolites at different efficiencies could find profit in trading with each other. The question then arises as to whether comparative advantage could be implemented in microbial systems.

In order to design and test such models, we must first specify what the conditions and expected results of comparative advantage are. Fortunately, comparative advantage involves specific requirements for the interacting parties and makes specific predictions about their subsequent behavior. The example that Ricardo used to illustrate the concept involves the production of wine and cloth by England and Portugal, where each has a set amount of man-hours that can be allocated toward producing either wine or cloth. Both countries are capable of producing both products, but Portugal is better at producing both than England. Specifically, Portugal can produce a greater amount of each product for the same amount of man-hours than England can. Intuition might then suggest that it is not in Portugal's interest to trade with England for either product, but this is not necessarily the case. If Portugal produces the wine it needs more efficiently than it produces cloth, and England produces the cloth it needs more efficiently than it produces wine, then cloth is more valuable to Portugal than wine, and wine is more valuable to England than cloth. It can be shown that both sides can profit by shifting resources into making the more efficient product and then trading for the other. This allows both countries individually to consume more wine and cloth through trade than either country could produce on its own. In practice, any two entities of sufficient complexity should be able to find a trading scheme that is to the advantage of both.

The requirements for demonstrating comparative advantage are therefore (1) that both countries have differing efficiencies for producing the two products, where the two

have opposite relative efficiencies, and one has better absolute efficiency in producing both, and (2) that neither country can produce more of one product without producing less of the other. The specific predictions that comparative advantage makes for situations in which two such entities enter into trade are: (1) to reach maximum levels of consumption and production, both countries will specialize in manufacturing the product they make most efficiently; (2) under such conditions the amount of product available for consumption is greater for both countries than if they had not entered into trade; and (3) the country that is less efficient overall will specialize more than the more efficient country in order to balance out the higher production of the other. With this information we can begin to design microbes that might be capable of engaging in comparative advantage-like trading and formulate testable hypotheses about how they will behave.

As for how such bacteria might be designed, the tools of the burgeoning field of synthetic biology can be used to engineer bacteria to demonstrate desired behaviors. In particular, the practice of using synthetic biology to model social or ecological interactions has come to be called "synthetic ecology" (Dunham, 2007). Building on the artificial genetic oscillators (Elowitz & Leibler, 2000) and switches (Gardner et al, 2000) that formed the foundation of synthetic biology, the sub-field of synthetic ecology has thus far successfully modeled a number of social systems in microbes, including mutualist interactions between two strains trading essential nutrients (Biliouris et al, 2012; Hu et al, 2010; Kerner et al, 2012; Kubo et al, 2013; Shendure et al, 2005; Shou et al, 2007; Wintermute & Silver, 2010), predator prey relationships (Balagadde et al, 2008), and communally synchronized cyclic behavior (Danino et al, 2010; Mondragon-Palomino et al, 2011). Additionally, a great deal of work in recent years has concerned competition between cooperative producers and selfish consumers in designed microbial populations (Celiker & Gore, 2013; Chuang et al, 2009; Chuang et al, 2010; Craig

Maclean & Brandon, 2008; Diggle et al, 2007; Gore et al, 2009; Greig & Travisano, 2004; Nahum et al, 2011; Rainey & Rainey, 2003; Sanchez & Gore, 2013; Tanouchi et al, 2012; Waite & Shou, 2012).

Such engineered biological models occupy a useful intellectual territory between, on the one hand, mathematical and computational models, which can be criticized for being too simple to accurately represent reality or for experimenter bias in selecting parameters or other model characteristics, and, on the other, natural biological systems, which tend to be extremely complex and frequently involve confounding variables that interact in unpredictable ways with the phenomenon of interest. Synthetic biological models, on the other hand, provide the experimenter with a significant measure of control over the system's behavior but still ultimately play out in the context of actual living organisms with all their inherent complexity and unpredictability. In the present case, using comparative advantage as a model provides a framework for implementing more nuanced models of cooperation than the synthetic ecology systems that have been implemented thus far, which typically involve trade between two strains with mutually exclusive capabilities, or cases where the strains can be simply partitioned into "producers" and "cheaters."

The microbial context also provides an interesting test of the generality of comparative advantage. In particular, the systems employed by bacteria to sense and respond to their environment rely on non-linear feedback mechanisms, and direct measurements and calculations of the sort humans might employ when engaged in trade cannot be used. In the biological context, cooperation can be seen as a problem when considered in the light of basic Darwinian ideas about organisms' struggle to maximize their fitness relative to others (Hamilton, 1963). Kin selection, by which an individual helps another individual with similar genes and thereby increases the likelihood of those

genes being passed on, is a commonly cited solution (Frank, 1998; Smith, 1964). Comparative advantage can be considered as an example of division of labor, which is one type of mutually beneficial cooperation. Division of labor is most well known among eusocial insect species (Page & Erber, 2002) but occurs among bacteria, as well (Crespi, 2001; Shapiro, 1998), with a well-studied bacterial example being the division into nitrogen-fixing and photosynthetic cells in certain cyanobacteria (Muro-Pastor & Hess, 2012). Such cases also frequently involve complete specialization on the part of the participants, whereas comparative advantage need only deal in shifts in relative specialization. Additionally, division of labor frequently involves a kin selection component, but comparative advantage has no such requirement. Thus the question of whether comparative advantage is generalizable enough to serve as a solution to the evolutionary problems faced by microbes is an interesting one.

Below we present mathematical models of both experimenter-controlled and self-regulating microbial systems designed to demonstrate comparative advantage in a bacterial system, as well as analyses of how these models perform. We find that the principles of comparative advantage do extend to these systems, and further that external stress increases the benefit gained from cooperative trading.

6.2 MATHEMATICAL MODELS

6.2.1 Basic model

We chose to employ as a model system an extension of the one-component system used by Chuang and coworkers (Chuang et al, 2009; Chuang et al, 2010) to study the interactions between "producer" bacteria, which produce and distribute necessary metabolites to the entire community, and "non-producer" bacteria, which make use of the metabolites provided by the producers but contribute nothing in return. Specifically, both

bacterial strains are made to grow in the presence of antibiotics and must produce an antibiotic-resistance protein in order to reproduce. However, the gene for this protein is only expressed when a chemical signal referred to as a "quorum-sensing molecule" or "autoinducer" is present in the culture medium. This signal is manufactured by the producer strain at a certain cost to growth so that the community as a whole may grow, and the relative success of the community is assessed by measuring the growth rate (the individual strains are also tagged with different fluorescent proteins in order to allow the relative success of the two types to be measured).

To modify this system to replicate comparative advantage, we propose adding a second antibiotic along with a second antibiotic resistance gene under the control of a second autoinducer. (However, the genes activated by the autoinducers do not necessarily need to code for antibiotic resistance proteins, but could also code for essential amino acids or other essential molecules.) In such a system, the two autoinducers would be the "products" traded between the two groups, and growth would be the measured output variable.

To develop a model for how such bacteria might grow, we start with the Monod equation (Monod, 1942; Monod, 1949) for modeling microbial growth, which is equivalent to the Michaelis-Menten equation used in enzyme kinetics (Lehninger et al, 2013):

$$\frac{dC}{dt} = C\left[\frac{VS}{(K + S)}\right] \qquad (1)$$

Here C is the density of the bacteria, S is the concentration of a substance required for growth, V is the maximum rate at which S can be converted into growth, and K is the value of S at which this rate is one half of V.

Since here we wish the bacteria to be dependent on two different products for growth, which we will call I_1 (the concentration of autoinducer 1) and I_2 (the concentration of autoinducer 2), we accordingly replace S in equation (1) with the arithmetical product of I_1I_2, which results in sigmoidal growth dynamics and reduces the growth rate to zero in the absence of either product:

$$\frac{dC}{dt} = C\left[\frac{VI_1I_2}{(K + I_1I_2)}\right] \qquad (2)$$

Next we add a quantity to force the system to adhere to logistic growth, with Z as the carrying capacity. While this term is not strictly necessary since we are interested not in the final density of the cells so much as the growth rate at which that density is reached, without this term the system grows to infinity, and it is difficult to devise a consistent rule for determining the range over which the growth rate should be measured.

$$\frac{dC}{dt} = C\left[\frac{VI_1I_2}{(K + I_1I_2)}\right]\left[1 - \frac{C}{Z}\right] \qquad (3)$$

Finally, we add a penalty term to represent the growth deficit that results from producing the autoinducers, which requires separate equations for the two strains, which we will call "A" and "B":

$$\frac{dC_A}{dt} = C_A\left[\frac{VI_1I_2}{\left(1 + \frac{I_{A1} + I_{A2}}{C_AP}\right)(K + I_1I_2)}\right]\left[1 - \frac{C_A + C_B}{Z}\right] \qquad (4)$$

$$\frac{dC_B}{dt} = C_B\left[\frac{VI_1I_2}{\left(1 + \frac{I_{B1} + I_{B2}}{C_BP}\right)(K + I_1I_2)}\right]\left[1 - \frac{C_A + C_B}{Z}\right] \qquad (5)$$

I_{A1} and I_{B1} here are the concentrations of autoinducer 1 produced by strain A and strain B, respectively, and I_{A2} and I_{B2} analogously represent the amounts of autoinducer 2 produced by the two strains, where $I_1 = I_{A1} + I_{B1}$ and $I_2 = I_{A2} + I_{B2}$. Since both strains should be essentially identical except in their differing production rates of the two

autoinducers, we can safely assume that the parameters V, K, P, and Z are the same for both strains.

In the penalty term $\left(1 + \frac{I_{N1}+I_{N2}}{C_N P}\right)$, P is analogous to the inhibition coefficient (K_I) in Michaelis-Menton kinetics (Lehninger et al, 2013) and determines the severity of the penalty, with smaller values of P leading to larger penalties. This formulation was chosen with reference to the kinetics of enzyme inhibition to penalize the strains for making more of the autoinducers, without allowing for the possibility of negative growth (we assume bacteriostatic rather than bacteriocidal antibiotics). Since the concentration of autoinducer should be a linear function of the density of cells at any given time, for determining the growth penalty this concentration should be divided by the density (concentration) of cells in order to avoid penalizing the cells for the presence of other cells in addition to penalizing them for production. (For instance, if, as a control, strains A and B are made to be identical, then the results should be the same for starting at (C_A, C_B) = (n, n) and at (C_A, C_B) = (2n, 0). If C_N is omitted from the denominator in the penalty term, this will not be the case.)

The penalty term could also be multiplied by K alone or by I_1I_2 alone, depending on the nature of the inhibition. However, if the inhibition affects the apparent K, then the growth inhibition will be lessened at higher values of the I_1I_2 product, which is not the behavior we would expect from such a system. Thus the growth inhibition due to producing the autoinducers should only affect the apparent V, as in equations (4) and (5) above. In other words, the situation corresponds to non-competitive inhibition in enzyme kinetics.

We note that this model is similar to one previously developed for the one-component model (Chuang et al, 2010). We increase the complexity of that model by taking into account two autoinducers instead of one and by including a more

sophisticated penalty function, and we simplify the model by not assuming a minimum growth rate.

Equations (4) and (5) can be considered output functions for converting production rates of the autoinducers into measurable variables, but the key design aspects of the system come from the question of how those production rates are determined. We have devised two methods by which this might be done. In the first (Conception 1), the production rates are determined according to the concentrations of exogenous inducers added by the experimenter. In the second (Conception 2), the bacteria control the production rates themselves through feedback regulation. Conception 1 has the advantage of allowing greater control over the system and greater freedom in exploring the parameter space, while Conception 2 is more intellectually pleasing as a self-regulating system.

6.2.2 Model for Conception 1

In this conception, the experimenters manually control the amounts of the autoinducers by changing the concentration of four exogenous inducers (γ_{A1}, γ_{A2}, γ_{B1}, γ_{B2}) that modulate the promoters that control expression of the genes for producing the autoinducers I_1 and I_2 in strains A and B:

$$I_{A1} = C_A \, k_{A1} \, \gamma_{A1} \qquad (6)$$

$$I_{A2} = C_A \, k_{A2} \, \gamma_{A2} \qquad (7)$$

$$I_{B1} = C_B \, k_{B1} \, \gamma_{B1} \qquad (8)$$

$$I_{B2} = C_B \, k_{B2} \, \gamma_{B2} \qquad (9)$$

The k_{Ni} coefficients represent the strength of the promoters regulated by the exogenous inducers and determine how effectively these inducers stimulate production of the autoinducers. Therefore, setting these coefficients to proper values allows

implementation of the necessary differences in efficiency for satisfying the requirement of comparative advantage. Equations (6) through (9) could also be represented using Michaelis-Menten kinetics, but this simpler formulation can be used if functions for converting inducer concentrations into gene expression are experimentally determined so as to yield a set of inducer values that result in a linear response in autoinducer concentration.

To force the strains to make a trade-off between producing one autoinducer or the other, we can require that $\gamma_{A1} + \gamma_{A2} = \gamma_A$ and $\gamma_{B1} + \gamma_{B2} = \gamma_B$, where γ_A and γ_B are constant for each strain and represent a resource base that the respective strains have exclusive access to. In other words, we give each strain a set amount of resources (the exogenous inducers) that can be allocated to producing one or the other autoinducer. (For simplicity we here assume that the exogenous inducers are active over the same concentration ranges, but normalizing coefficients could be added as necessary for specific inducers.) Further, we can set "rheostat" values R_A and R_B to represent how that allocation has been made (specifically, the extent of specialization in making autoinducer 1), where:

$$R_A = \frac{\gamma_{A1}}{\gamma_{A1} + \gamma_{A2}} = \frac{\gamma_{A1}}{\gamma_A} = 1 - \frac{\gamma_{A2}}{\gamma_A} \qquad (10)$$

$$R_B = \frac{\gamma_{B1}}{\gamma_{B1} + \gamma_{B2}} = \frac{\gamma_{B1}}{\gamma_B} = 1 - \frac{\gamma_{B2}}{\gamma_B} \qquad (11)$$

Equations (6) through (9) can then be converted to:

$$I_{A1} = C_A k_{A1} \gamma_A R_A \qquad (12)$$

$$I_{A2} = C_A k_{A2} \gamma_A (1 - R_A) \qquad (13)$$

$$I_{B1} = C_B k_{B1} \gamma_B R_B \qquad (14)$$

$$I_{B2} = C_B k_{B2} \gamma_B (1 - R_B) \qquad (15)$$

By combining constants such that $\kappa_{A1} = k_{A1}\,\gamma_A\,R_A$, $\kappa_{A2} = k_{A2}\,\gamma_A(1 - R_A)$, $\kappa_{B1} = k_{B1}\,\gamma_B\,R_B$, and $\kappa_{B2} = k_{B2}\,\gamma_B(1 - R_B)$, and then substituting into equations (1) and (2), we obtain:

$$\frac{dC_A}{dt} = C_A\left[\frac{V(C_A\kappa_{A1} + C_B\kappa_{B1})(C_A\kappa_{A2} + C_B\kappa_{B2})}{\left(1 + \frac{\kappa_{A1} + \kappa_{A2}}{P}\right)(K + (C_A\kappa_{A1} + C_B\kappa_{B1})(C_A\kappa_{A2} + C_B\kappa_{B2}))}\right]\left[1 - \frac{C_A + C_B}{Z}\right] \qquad (16)$$

$$\frac{dC_B}{dt} = C_B\left[\frac{V(C_A\kappa_{A1} + C_B\kappa_{B1})(C_A\kappa_{A2} + C_B\kappa_{B2})}{\left(1 + \frac{\kappa_{B1} + \kappa_{B2}}{P}\right)(K + (C_A\kappa_{A1} + C_B\kappa_{B1})(C_A\kappa_{A2} + C_B\kappa_{B2}))}\right]\left[1 - \frac{C_A + C_B}{Z}\right] \qquad (17)$$

These equations can be non-dimensionalized to:

$$\frac{dx}{d\tau} = x\frac{(1 - x - y)(\alpha x + \beta y)(\gamma x + \delta y)}{(1 + \alpha + \gamma)(\epsilon + (\alpha x + \beta y)(\gamma x + \delta y))} \qquad (18)$$

$$\frac{dy}{d\tau} = y\frac{(1 - x - y)(\alpha x + \beta y)(\gamma x + \delta y)}{(1 + \beta + \delta)(\epsilon + (\alpha x + \beta y)(\gamma x + \delta y))} \qquad (19)$$

Where

$$x = \frac{C_A}{Z}\,;\quad y = \frac{C_B}{Z}\,;\quad \tau = Vt\,;\quad \alpha = \frac{\kappa_{A1}}{P}\,;\quad \beta = \frac{\kappa_{B1}}{P}\,;\quad \gamma = \frac{\kappa_{A2}}{P}\,;\quad \delta = \frac{\kappa_{B2}}{P}\,;\quad \epsilon = \frac{K}{Z^2P^2}$$

In the non-dimensionalized equations, α, β, γ, and δ are the four parameters that determine the relative efficiencies of the two strains, while ε performs the role of K in the original equations (the other parameters fall out during the course of non-dimensionalization). Though the rheostat values R_A and R_B are not explicitly included in these equations, since they are dimensionless they could be separated out and included by replacing α, β, γ, and δ with αR_A, βR_B, $\gamma(1 - R_A)$, and $\delta(1 - R_B)$.

6.2.3 Model for Conception 2A

For Conception 2, we designed gene circuits that will allow the bacteria to make their own decisions about how to modulate autoinducer production via feedback regulation, a very common approach in molecular systems (Lehninger et al, 2013). In the simplest scheme, each autoinducer inhibits its own synthesis by inducing expression of a repressor that represses the gene of that autoinducer. In order to implement a trade-off, each autoinducer should also induce expression of the other autoinducer. We call this implementation Conception 2A, a diagram of which is shown by the black arrows in **Figure 6.1**.

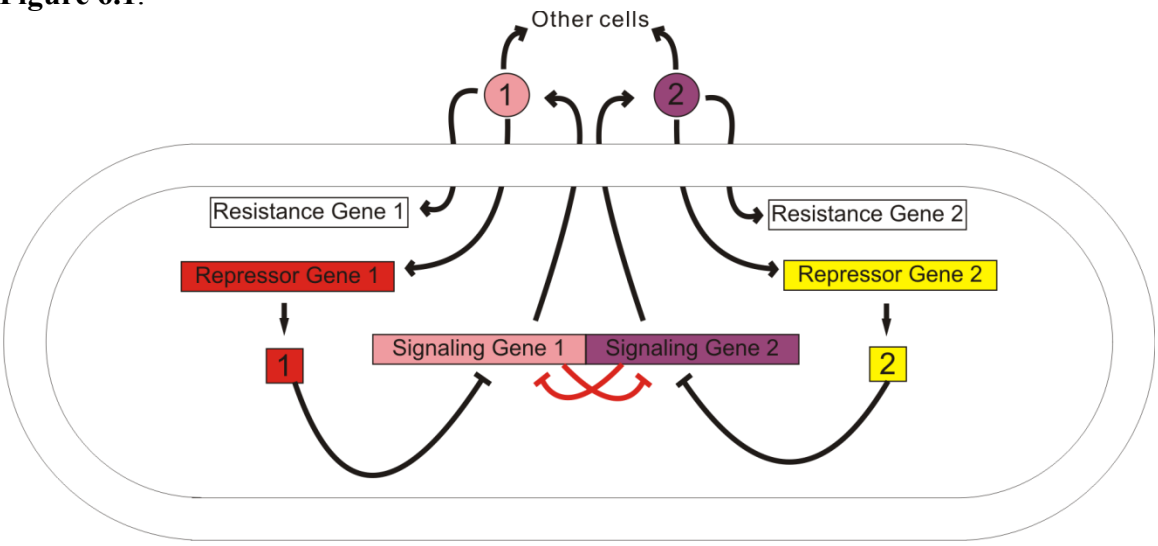

Figure 6.1. Gene circuits of Conception 2.

Circles represent signaling molecules (products), and squares represent repressors. Pointed arrows indicate activation, and flat arrows represent inhibition. The black arrows represent the interactions in Conception 2.1, and the red arrows are interactions added in Conception 2.2. Note that the signaling molecules are the only components that can leave to influence other cells.

In a simplified mathematical model of this circuit, we can imagine each autoinducer is produced according to Michaelis-Menten kinetics, where the autoinducer acts as its own inhibitor, and the other autoinducer acts as the substrate. We can therefore represent Conception 2A with the six coupled equations (20) through (25).

$$\frac{dI_{A1}}{dt} = C_A \frac{V_{A1}(I_{A2} + I_{B2})}{k_2\left(1 + \frac{I_{A1} + I_{B1}}{K_I}\right) + (I_{A2} + I_{B2})} \quad (20)$$

$$\frac{dI_{A2}}{dt} = C_A \frac{V_{A2}(I_{A1} + I_{B1})}{k_1\left(1 + \frac{I_{A2} + I_{B2}}{K_I}\right) + (I_{A1} + I_{B1})} \quad (21)$$

$$\frac{dI_{B1}}{dt} = C_B \frac{V_{B1}(I_{A2} + I_{B2})}{k_2\left(1 + \frac{I_{A1} + I_{B1}}{K_I}\right) + (I_{A2} + I_{B2})} \quad (22)$$

$$\frac{dI_{B2}}{dt} = C_B \frac{V_{B2}(I_{A1} + I_{B1})}{k_1\left(1 + \frac{I_{A2} + I_{B2}}{K_I}\right) + (I_{A1} + I_{B1})} \quad (23)$$

$$\frac{dC_A}{dt} = C_A \left[\frac{V(I_{A1} + I_{B1})(I_{A2} + I_{B2})}{\left(1 + \frac{I_{A1} + I_{A2}}{C_A P}\right)\left(K + (I_{A1} + I_{B1})(I_{A2} + I_{B2})\right)}\right]\left[1 - \frac{C_A + C_B}{Z}\right] \quad (24)$$

$$\frac{dC_B}{dt} = C_B \left[\frac{V(I_{A1} + I_{B1})(I_{A2} + I_{B2})}{\left(1 + \frac{I_{B1} + I_{B2}}{C_B P}\right)\left(K + (I_{A1} + I_{B1})(I_{A2} + I_{B2})\right)}\right]\left[1 - \frac{C_A + C_B}{Z}\right] \quad (25)$$

Here V_{A1}, V_{A2}, V_{B1}, and V_{B2} represent the maximum expression rates of the genes that code for the autoinducers and are the quantities that will be varied in order to create the conditions necessary to demonstrate comparative advantage. In an actual gene circuit, these parameters could be varied by changing the strength of the ribosome binding sites. The autoinducers are not produced directly from the genes, of course, but are produced by enzymes that are translated from RNA molecules that are transcribed from the genes. However, this simplified model should be sufficient as an initial test of the feasibility of the system. The constants k_1 and k_2 are the concentrations of I_1 and I_2, respectively, at which expression reaches half-maximum. K_I is the inhibition constant and represents the affinity of the repressor for the operator that it binds. The inhibition term in equations (20) through (23) affects the apparent k_i in this case and not the V_{Ni}, because we expect large amounts of the substrate autoinducer to overcome the inhibition and allow maximal expression. In terms of enzyme kinetics, this is competitive

inhibition, which makes sense because the repressor and the RNA polymerase should be competing for access to the same stretch of DNA. In the actual gene circuit we would likely need at least two different k_i and two different K_I, but for theoretical purposes we assume they are all the same.

6.2.4 Model for Conception 2B

A potential failing of Conception 2A is that the strains do not differentiate between products made by themselves and products made by others. Specifically, in order to properly allocate resources according to the comparative advantage model, each cell must decrease the production of one autoinducer in response to increased production by itself of the other, yet simultaneously increase production of the first autoinducer in response to increased production by other strains of the other autoinducer. Intracellular RNA-based inhibition mechanisms (Isaacs et al, 2004; Lucks et al, 2011; Na et al, 2013; Saito & Inoue, 2009) could be used to allow a cell to respond separately to the amount of autoinducer it produces as opposed to the total amount of autoinducer present. We call this modification Conception 2B, which is shown schematically by both the black and red arrows in **Figure 6.1**, and is implemented mathematically by replacing equations (20) through (23) with equations (26) through (29):

$$\frac{dI_{A1}}{dt} = C_A \frac{V_{A1}(I_{A2} + I_{B2})}{k_2\left(1 + \frac{I_{A1} + I_{B1}}{K_I}\right)\left(1 + \frac{I_{A2}}{K_{Iint}}\right) + (I_{A2} + I_{B2})} \quad (26)$$

$$\frac{dI_{A2}}{dt} = C_A \frac{V_{A2}(I_{A1} + I_{B1})}{k_1\left(1 + \frac{I_{A2} + I_{B2}}{K_I}\right)\left(1 + \frac{I_{A1}}{K_{Iint}}\right) + (I_{A1} + I_{B1})} \quad (27)$$

$$\frac{dI_{B1}}{dt} = C_B \frac{V_{B1}(I_{A2} + I_{B2})}{k_2\left(1 + \frac{I_{A1} + I_{B1}}{K_I}\right)\left(1 + \frac{I_{B2}}{K_{Iint}}\right) + (I_{A2} + I_{B2})} \quad (28)$$

$$\frac{dI_{B2}}{dt} = C_B \frac{V_{B2}(I_{A1} + I_{B1})}{k_1\left(1 + \frac{I_{A2} + I_{B2}}{K_I}\right)\left(1 + \frac{I_{B1}}{K_{Iint}}\right) + (I_{A1} + I_{B1})} \quad (29)$$

Here we add an extra inhibition term, $\left(1 + \frac{I_{Ni}}{K_{Iint}}\right)$, as a simplified model of RNA-based inhibition that serves to reduce production of one autoinducer in response to increased production of the other in the same strain, where K_{Iint} serves as the inhibition coefficient. In actual bacteria, RNA levels could be measured by reverse transcription and quantitative PCR.

6.2.5 Implementation

For analysis, the equations derived above were integrated in Matlab. Integration was started from (C_A, C_B) = (1, 1) (alternatively, (0, 1) or (1, 0) for monoculture controls) and, in Conception 2, $I_{A1} = I_{A2} = I_{B1} = I_{B2} = 1$. Integration was then continued for 5 arbitrary time units for each parameter set in Conception 1, and for 100 arbitrary time units for each parameter set in Conceptions 2A and 2B (Conception 1 being much more computationally intensive than Conception 2). Parameters whose values are not specified elsewhere were set to one.

Growth rate was defined as the fraction of increase per unit time from the start of growth until reaching half the carrying capacity defined by the Z parameter (i.e., if the start point is C_0, half carrying capacity is $C_{0.5}$, and the time between is t, the growth rate equals $(C_{0.5}/C_0)^{1/t} - 1$). Since values at exactly half the carrying capacity ($C_{0.5}$) could not be directly obtained without using extremely resource-intensive integration parameters, we instead used the values for the growth of strains A and B (C_A and C_B) at the time points before and after the moment when $C_A + C_B$ reached the threshold value $C_{0.5}$ to estimate the time $t_{0.5}$ required to reach $C_{0.5}$, and then used $t_{0.5}$ to estimate the values of C_A and C_B (and, in Conception 2, I_{A1}, I_{A2}, I_{B1}, and I_{B2}) when $C_{0.5}$ was reached. R^2 values for both linear and $\log_2$-linear fits of the data points in the vicinity of $C_{0.5}$ were >0.99, indicating that either could be used to provide accurate estimations of values at the point

of reaching $C_{0.5}$. The results of simple linear estimation proved to be more robust to variations in the stiffness of integration, however, and so we employed the linear method. Specifically, using the subscript α to denote integrated values obtained at the point just before reaching $C_{0.5}$ and using the subscript β to denote values obtained just after reaching $C_{0.5}$, $t_{0.5} = t_\alpha + (C_{0.5} - C_\alpha)(t_\beta - t_\alpha)/(C_\beta - C_\alpha)$, and the value of any other variable x at $t_{0.5}$ is then calculated as $x_{0.5} = x_\alpha + (t_{0.5} - t_\alpha)(x_\beta - x_\alpha)/(t_\beta - t_\alpha)$.

6.3 RESULTS

6.3.1.1 Conception 1: Analysis of non-dimensionalized equations

We analyzed the non-dimensionalized equations (18) and (19) in order make initial checks as to whether the model is behaving as expected. Isoclines (lines along which one of the variables is fixed) occur at $x = 0$ and $y = 0$. The intersection between the two isoclines at (0, 0) is a fixed point, where neither variable changes. Another region where neither variable changes is along the fixed line defined by $y = 1 - x$. The presence of a fixed line makes sense because, if we consider the fixed-line steady-state as equivalent to the system at carrying capacity (where $x + y = 1$), we do not expect the system to reach any particular value (x, y), but we do expect the ratio between x and y at the steady state to depend on the ratio between x and y at the starting point.

A flow field showing the direction and magnitude of flux at various points in the x-y plane is plotted in **Figure 6.2A**, in which the (0, 0) fixed point appears to be an unstable node, and the fixed line appears to be stable. This is consistent with any non-zero concentration of cells growing toward carrying capacity. A phase portrait showing trajectories over time from various starting points is plotted in **Figure 6.2B**, which again is consistent with all non-zero, non-negative starting points moving eventually onto the

fixed line (reaching carrying capacity) and then stopping. Thus in these respects the model is behaving as expected.

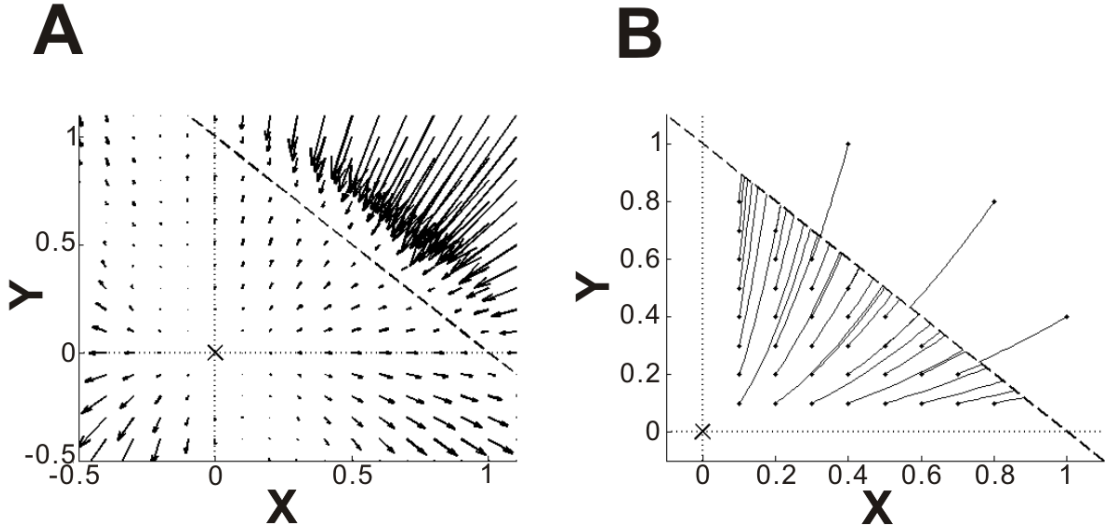

Figure 6.2. Graphical analysis of the non-dimensionalized equations of Conception 1.

Isoclines (which in this case also correspond to the x and y axes) are solid lines, the fixed line is a dotted line, and the fixed point is given by an X. The triangular region enclosed by the lines is the biologically relevant region of the parameter space. The parameter values used were $\alpha = 2$, $\beta = 0.5$, $\gamma = 1.5$, $\delta = 1$, and $\varepsilon = 1$, which represent values that could potentially recreate comparative-advantage-like dynamics. **(A)** Phase plane analysis of the system, where each arrow is a vector representing the direction and magnitude of flow in the system at the point at which the arrow starts. **(B)** Phase portrait of the system, where the black lines plot trajectories in the system from various starting points denoted by black dots.

6.3.1.2 Conception 1: Example growth curves

We performed further investigations of Conception 1 using equations (16) and (17), since the variables and constants therein are more easily interpretable in biological terms. **Figure 6.3A** shows an example of growth curves for strains A and B under conditions potentially compatible with comparative advantage (specifically, $k_{A1} = 2$, $k_{A2} = 1.5$, $k_{B1} = 0.5$, $k_{B2} = 1$, meaning that strain A makes product 1 more efficiently that product 2, strain A makes product 2 more efficiently than product 1, and strain A makes both products more efficiently that strain B). The expected sigmoidal growth curve is

seen for both strains. Strain B grows faster than strain A, which is also in line with expectations since strain B produces less and thus receives a lesser growth penalty. Both rheostat values R_A and R_B are set to 0.5 in this example, which means both strains specialize equally in both products (autoinducers), and differences in production result entirely from the differences in the values for k_{A1}, k_{A2}, k_{B1}, and k_{B2}. (R_A and R_B vary between zero and one and represent the extent to which strain A and strain B, respectively, have specialized in making product (autoinducer) 1 as opposed to product 2.) **Figures 6.3B** and **6.3C**, which show the amounts of products 1 and 2, respectively, being produced at each time point show that, as expected, the highest level of production under these circumstances is strain A's production of product 1, followed by strain A's production of product 2, then strain B's production of product 2, and finally the lowest production is for strain B and product 1.

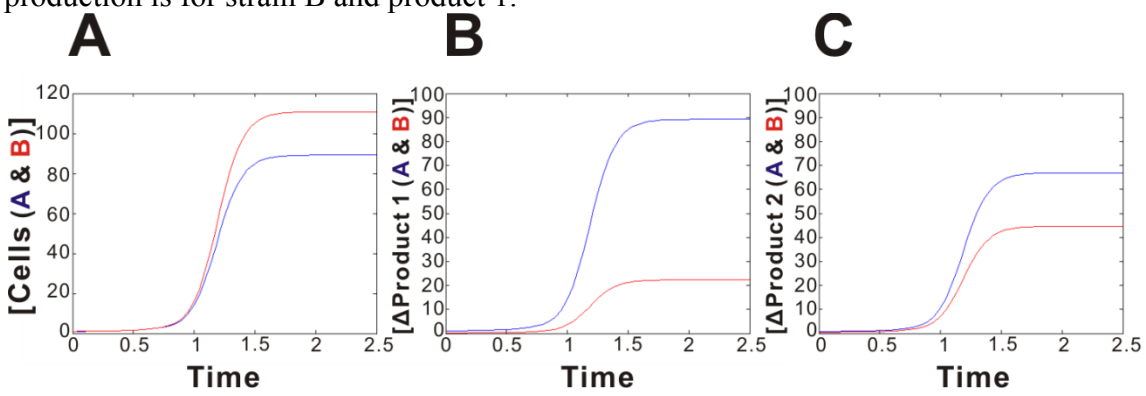

Figure 6.3. Example growth curves for Conception 1.

Parameter values are K = 20, P =20, V = 10, Z = 200, k_{A1} = 2, k_{A2} = 1.5, k_{B1} = 0.5, k_{B2} = 1, and (R_A, R_B) = (0.5, 0.5). Values for strain A are shown in blue, and values for strain B are shown in red. Units are arbitrary. **(A)** Cumulative cell concentration over time starting from a value of 1 for both strains. **(B)** The amount of product 1 (I_1) being produced by each strain at each time point. **(C)** The amount of product 2 (I_2) being produced by each strain at each time point.

6.3.1.3 Conception 1: Investigations into the (R_A,R_B) parameter space

In order to examine the effects of different concentrations of inducers, which correspond to different degrees of specialization between the two strains, we created heat maps examining the growth rate of the system at every combination of the rheostat values R_A and R_B in increments of 0.005. Examples of such heat maps are shown in **Figure 6.4**, and relevant output values are given in **Table 6.1**.

We first examined the dynamics of two strains with identical efficiencies allowed to specialize separately, as shown in **Figures 6.4A** through **6.4C**. There are two equivalent optimal (R_A,R_B) points. Strain A grows better when specialized to produce less than Strain B, and vice versa. Thus it is in the interest of both strains to produce less and have the other produce more. This diametrical opposition is graphically illustrated by the difference between **Figure 6.4B** and **Figure 6.4C**, which show the specialization preferences for maximal growth of each individual strain. In ecological terms, we could say that both strains are attempting to occupy the same niche, which means that at least one has to take up a suboptimal position.

Next we looked at strains designed to replicate comparative advantage. In this and all subsequent cases, strain A is the high-level producer, and strain B is the low-level producer. **Figures 6.4D** through **6.4F** show heat maps of these strains at three different levels of the parameter K. We chose to examine K because it represents the amount of difficulty the strains face in converting the products to growth and as such can likely be modulated by changing environmental factors such as the concentration of antibiotics. It is also the only parameter of out of K, P, V, and Z that still remained in some form after non-dimensionalization (as ε in equations (18) and (19)), supporting the idea that it is in some sense the most important of these parameters. Increasing K from 5 to 20 (compare **Figs. 6.4D** and **6.4E**) narrows the parameter space in which high growth can occur. The

high-growth parameter space expands again somewhat at K = 80 (**Fig. 6.4F**), though we note that the area representing the lowest growth rates (the area of the darkest color) is greater, as well. The values in **Table 6.1** also show that increasing K reduces growth rates, as expected.

We also note that the degrees of specialization occurring at the optimal (R_A,R_B) are in line with expectations: both strains are specializing in the product they make more efficiently, and strain B is completely or almost completely specialized, as can be seen from the optimal (R_A,R_B) values given for the comparative advantage cases in **Table 6.1**. We also note that changing K does not seem to significantly change the optimal (R_A,R_B). Only the combined-growth heat maps are shown for these cases because the heat maps of the individual preferences did not look much different, which would seem to indicate that the interests of the two strains are now aligned. In ecological terms, they can now occupy different niches, and cooperation now becomes more advantageous.

In **Figures 6.4G** through **6.4I** we examine the effects of changing the efficiencies between the two strains. **Figure 6.4G** and **Figure 6.4H** represent cases where the efficiencies of strain A or B are reversed with respect to the case in **Figure 6.4E**. This results in a state of absolute advantage, where strain A still makes both products more efficiently than strain B, but now both have similar relative efficiencies. The growth rates are lower than for the comparative advantage case shown in **Figure 6.4E** (see **Table 6.1**), as expected, and furthermore, the optimal (R_A,R_B) involves strain A specializing in the product it makes less efficiently, most likely because strain B is unable to take up the slack if strain A specializes in its preferred product. A case in which the efficiencies of both strain A and strain B are reversed with respect to **Figure 6.4E** is shown in **Figure 6.4I**, which as expected successfully recapitulates the specialization and improved growth characteristics of comparative advantage.

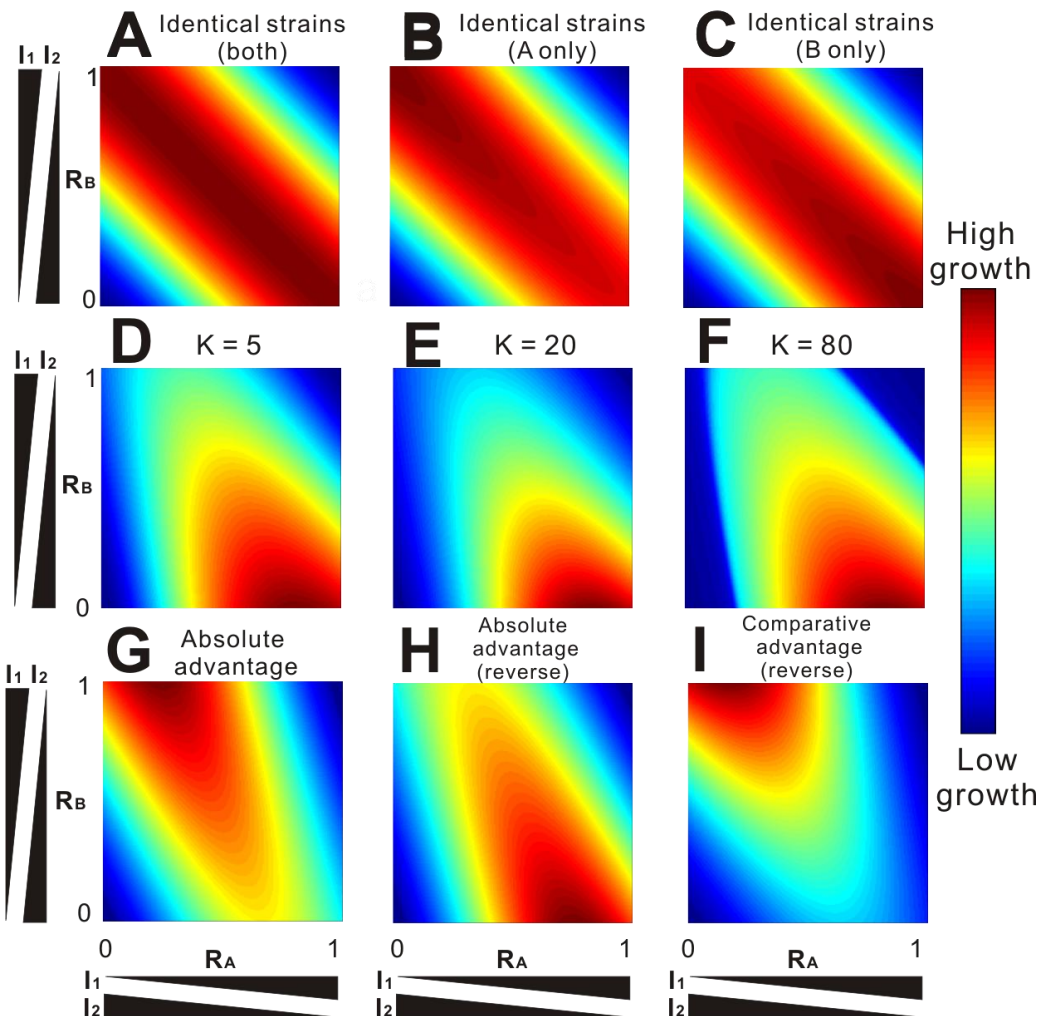

Figure 6.4. Heat maps for representative parameter sets in Conception 1.

The x-axis is R_A (the specialization of strain A in making product 1 versus product 2), the y-axis is R_B (the specialization of strain B in making product 1 versus product 2), and color represents relative growth rate. Default parameter values are K = 20, P =20, V = 10, Z = 200, k_{A1} = 2, k_{A2} = 1.5, k_{B1} = 0.5, k_{B2} = 1, which are conditions that should result in behavior consistent with comparative advantage. **(A)** through **(C)** represent a control case where two identical strains are permitted to specialize differently, such that $k_{A1} = k_{B1} = 2$ and $k_{A2} = k_{B2} = 1.5$. (A) depicts overall growth rates, (B) depicts the growth rates of strain A, and (C) depicts the growth rates of strain B. The optimal (R_A,R_B) for (B) and the optimal (R_A,R_B) for (C) are equivalent and together represent the optimal (R_A,R_B) for (A). **(D)** through **(F)** represent different values for K in the comparative advantage context. Specifically, K = 5 in (D), 20 in (E), and 80 in (F). **(G)** through **(I)** show variations on the efficiencies of production between strains. Specifically, k_{A1} = 2, k_{A2} = 1.5, k_{B1} = 1, k_{B2} = 0.5 (absolute advantage control) in (G), k_{A1} = 1.5, k_{A2} = 2, k_{B1} = 0.5, k_{B2} = 1 (absolute advantage control with reversed specializations) in (H), and k_{A1} = 1.5, $k_{A1} = k_{A2}$ = 2, k_{B1} = 1, k_{B2} = 0.5 (comparative advantage positive control with reversed specializations) in (I).

Relevant figure	Optimal (R_A,R_B)	Overall growth rate at optimal (R_A,R_B)	Growth rate of strain A at optimal (R_A,R_B)	Growth rate of strain B at optimal (R_A,R_B)
4A through 4C (Identical strains)	(0, 0.995) and (0.995, 0)	89.1	93.8 and 84.5	93.8 and 84.5
4D (K = 5)	(0.82, 0)	524	456	595
4E (K = 20)	(0.835, 0)	45.5	41.7	49.3
4F (K = 80)	(0.835, 0)	3.51	3.36	3.65
4G (Absolute advantage k_{B1} = 1, k_{B2} = 0.5)	(0.24, 1)	30.4	28.8	31.9
4H (Absolute advantage k_{A1} = 1.5, k_{A2} = 2)	(0.76, 0)	30.4	28.8	31.9
4I (Comparative advantage – reverse)	(0.165, 1)	45.5	41.7	49.3

Table 6.1. Output values for the heat maps in Figure 6.4.

We further note that strain B consistently grows better than strain A, which is a result of its lower production and therefore lower growth penalty. Thus strain A and strain B can be considered partly analogous to the producer and non-producer strains examined by Chuang and coworkers (Chuang et al, 2009; Chuang et al, 2010); alternatively, the work of Chuang and coworkers can be considered as a limit case for the system presented here.

6.3.1.4 Conception 1: Effects of varying K, P, and V

Next we examined in more detail the effects on the system of varying K, P, and V. As mentioned above, K represents the degree of difficulty with which the strains convert products into growth and can likely be externally modulated by changing the antibiotic concentration. The parameter P represents the growth penalty for generating product, with small values of P corresponding to a greater penalty. The P parameter is likely intrinsic to the system and not easily modified. The parameter V represents the

maximum rate with which the products can be converted into growth. The V parameter could likely be modified by changing the strength of the promoters regulating the antibiotic resistance genes activated by the products.

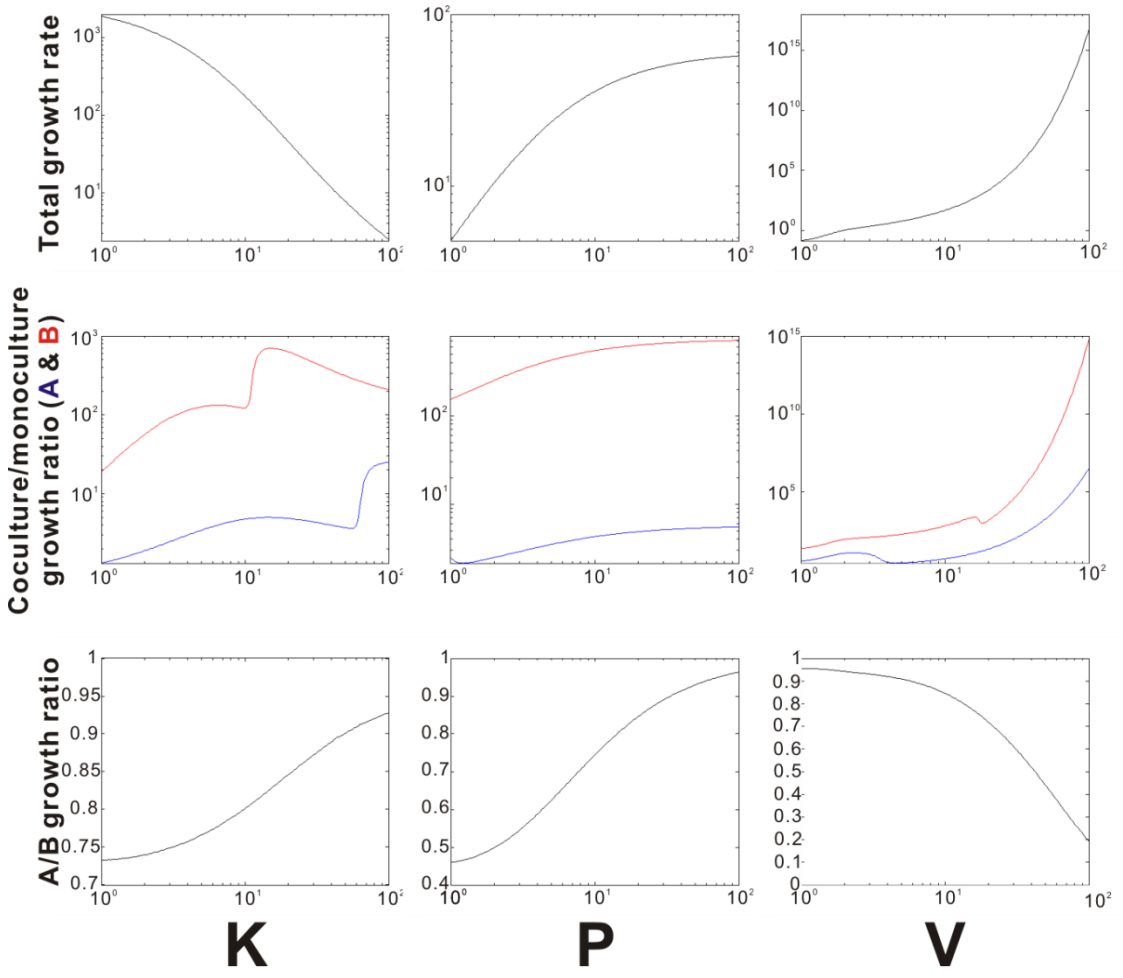

Figure 6.5. Effect of K, P, and V parameters on growth characteristics at the optimal (R_A,R_B) in Conception 1.

In all cases the x-axis represents the parameter value, and the y-axis is the resultant growth (absolute or relative). K = 20, P =20, V = 10, Z = 200, k_{A1} = 2, k_{A2} = 1.5, k_{B1} = 0.5, k_{B2} = 1, except when a specified parameter is being varied. The first column of graphs shows the effects of varying K, the second column shows the effects of varying P, and the third column shows the effects of varying V. The first row shows the effect of parameter changes on the combined growth rate of strains A and B. The second row shows the effect of parameter changes on the ratio of the coculture and monoculture growth rates of the two strains, with values specific to strain A in blue, and values specific to strain B in red. The last row shows the effect of parameter changes on the ratio of the coculture growth rates of strain A and strain B.

The effects on growth of varying these three parameters are explored in **Figure 6.5**. Increasing K decreases overall growth rates (first row, first column of **Fig. 6.5**), as expected, and also generally increases the benefit to cooperation as judged by the ratios of coculture versus monoculture growth rates for the two strains (second row, first column of **Fig. 6.5**), where apparent thresholds exist which when crossed significantly increase the benefit to coculture, which then levels off. Increasing K also narrows the growth advantage of strain B (third row, first column of **Fig. 6.5**) perhaps because as growth becomes harder to achieve, the penalty for producing more products carries less weight compared to the benefit for successfully achieving growth.

Increasing P (second column in **Fig. 6.5**) improves growth (both absolute growth and growth in coculture relative to monoculture) up to a point and then stops, which is in line with expectations, as the penalty term in equations (1) and (2) goes to one as P goes to infinity, removing its effect. Increasing P also narrows the gap between the strains, for the same reasons.

Increasing V increases growth rate (first and second rows, third column in **Fig. 6.5**), of course, and significantly widens the gap in growth rates between strain A and B (third row, third column in **Fig. 6.5**), likely because higher V makes strain B less dependent on strain A for growth. Note also that the benefit to cooperation is consistently higher for strain B than for strain A (second row in **Fig. 6.5**). Threshold effects as seen for the K parameter also seem to occur with changing V (second row, third column in **Fig. 6.5**).

The effects of varying K, P, and V on specialization are shown in **Movies 1**, **2**, and **3**, respectively, which show how the heat map changes as these parameters are varied over the same ranges as in **Figure 6.5**. The movies are available online at http://bit.ly/1cwG52S (Movie 1), http://bit.ly/1hdbgQy (Movie 2), and

http://bit.ly/1c6pb7K (Movie 3). Increasing K generally either narrows the range of the high-growth parameter space or expands the range of the low-growth parameter space, which is not an unsurprising result of more difficult growth conditions. Increasing P initially somewhat narrows the high-growth parameter space but ceases to have much effect at higher values. As V increases the high-growth parameter space narrows between V = 1 and V = 1.5, increases from approximately V = 1.5 to V = 3, and thereafter decreases. In a somewhat similar fashion to P, changing V seems to have the most effect on the specialization landscape when V is small; when V is big the options narrow, perhaps as a result of accentuating the advantage gained by finding the optimal level of specialization. The optimal (R_A,R_B) (the reddest area in the heat maps) experiences little change in location as these parameters change, indicating the robustness of the specialization effect. (Also note that heat map intensities are normalized by frame, and the reddest area with one parameter set does not necessarily correspond in absolute growth rate to the reddest area with another parameter set.)

Conception 1 thus demonstrates comparative advantage. Specifically, the optimal growth conditions are those under which both strains specialize in the product they make more efficiently, with the weaker strain specializing more. Under these conditions, both strains grow better together than alone, and also grow better than under comparable absolute advantage conditions.

6.3.2.1 Conception 2: Example growth curves

Figure 6.6A shows examples of growth curves for strains A and B in both Conceptions 2A and 2B under conditions potentially compatible with comparative advantage. (Conceptions 2A and 2B both involve self-regulating gene circuits, where Conception 2B contains an additional "rheostat" mechanism for forcing an intracellular

trade-off in production between the two products, as shown in **Fig. 6.1**.) The curves are similar to those seen for Conception 1 in **Figure 6.3A**, with Conception 2B resulting slower growth than Conception 2A, with a narrower gap between strains A and B. As seen in **Figures 6.6B** and **6.6C**, which show production levels of products 1 and 2, respectively, in a fashion analogous to **Figures 6.3B** and **6.3C**, product levels are lower yet better balanced in Conception 2B than in Conception 2A.

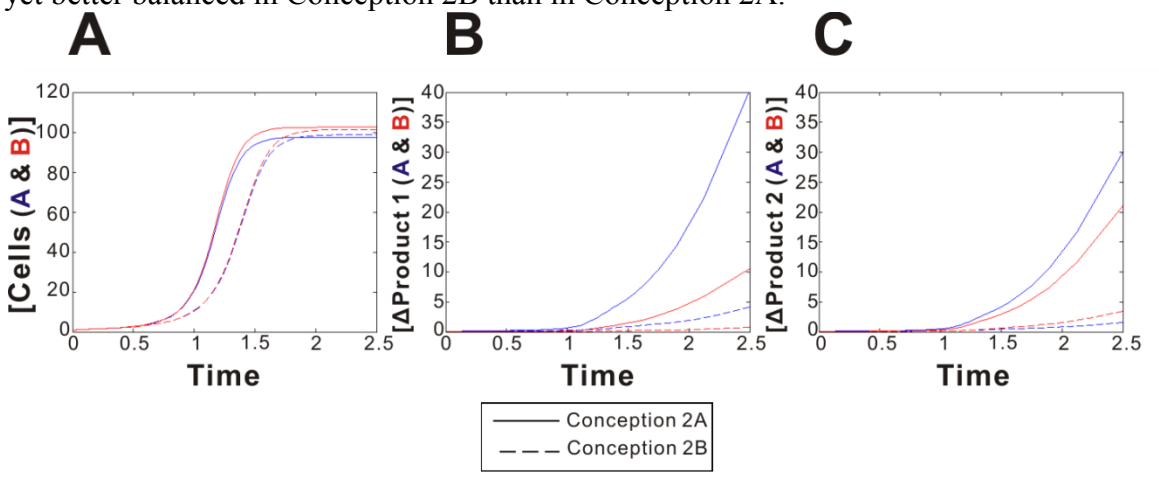

Figure 6.6. Example growth curves for Conception 2.

Parameter values are K = 20, P =20, V = 10, Z = 200, V_{A1} = 2, V_{A2} = 1.5, V_{B1} = 0.5, V_{B2} = 1, which are conditions that should be compatible with comparative advantage. Otherwise the format is identical to **Figure 6.3**, with **(A)** showing cumulative cell concentration, **(B)** showing the change in production of product 1, and **(C)** showing the change in production of product 2.

6.3.2.2 Conception 2: Effects of varying K, P, and V

We next repeated the analyses from **Figure 6.5** using the equations for Conceptions 2A and 2B. These results are shown in **Figure 6.7**. In most cases the effects are similar, but some differences exist. Perhaps most important is the fact that, as opposed to Conception 1, coculture is not more advantageous than monoculture over the entire parameter space, particularly for strain A. In other words, cooperation between the two strains is only beneficial to both under certain conditions.

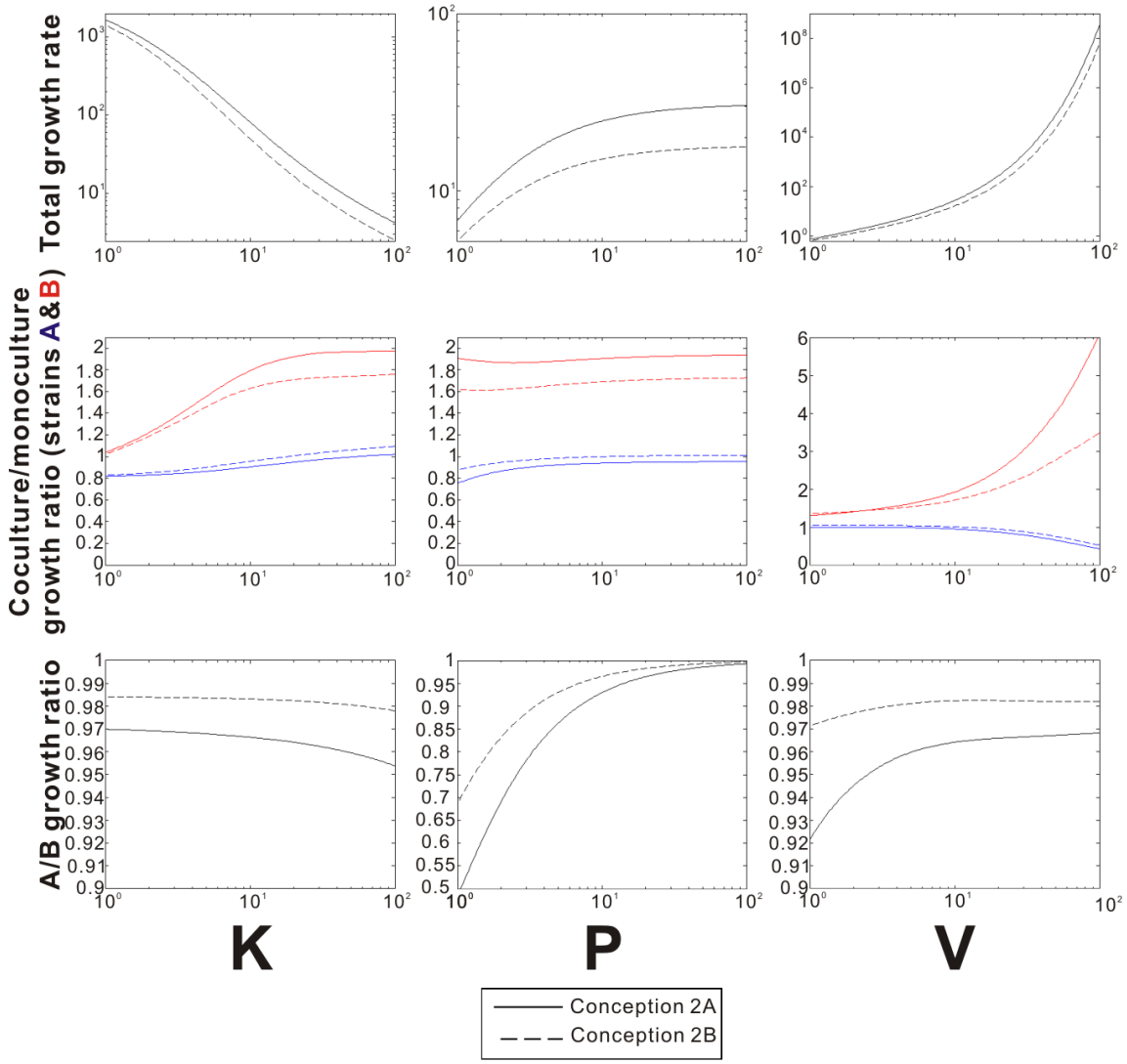

Figure 6.7. Effect of K, P, and V parameters on growth characteristics in Conception 2.

Non-varied parameter values are as in **Figure 6.6**. Data for Conception 2A is always a solid line, while data from Conception 2B is always a dotted line. Otherwise the format is identical to **Figure 6.5**.

Specifically, coculture is advantageous for Strain A when K is above 55 for Conception 2A or above 18 for Conception 2B; when P is above 10 for Conception 2B (and never for Conception 2A when P is between 1 and 100); and when V is less than 11

for Conception 2B (and never for Conception 2A when V is between 1 and 100). (When not being varied, the parameters here are K = 20, P = 20, and V = 10.)

Thus besides being interesting as a self-regulating system, Conception 2 is also interesting in providing more stringent conditions under which comparative advantage can work. In light of the discussion above, these conditions would seem to be a relatively light growth penalty for production (high P) combined with otherwise difficult growth conditions (high K and low V). In other words, capable individuals in trying circumstances benefit from cooperation, but in comfortable environments may be better off alone (see **Figs. 6.7C** through **6.7F**).

The V parameter affecting the maximum growth rate is interesting in that increasing it is beneficial for strain B but has the reverse effect on strain A, perhaps because the penalty term makes higher overall production rates more advantageous to lower-producing strain. On the other hand, Conception 2B, which adds an internal rheostat, results in increased benefit to cooperation for strain A but not for strain B, as well as lower absolute growth rates, perhaps because the internal rheostat makes it more difficult for strain B to allocate production to its less efficient (and less growth-penalizing) product, thus preventing it from gaining benefits from strain A without providing benefits in return.

Another interesting difference between Conceptions 1 and 2 is that in Conception 2, increasing K slightly reduces the ratio of the growth rates between strains A and B (third row, first column in **Figure 6.7**), perhaps because of a combination of the less precise method for finding the optimal specialization values in Conception 2 and the fact that as K increases, the fraction of total production assumed by strain A increases slightly, which also increases its growth penalty. Also note that the A/B growth ratio is

more robust to changing parameters in Conception 2B than in 2A (third row in **Figure 6.7**).

6.3.2.3 Conception 2: Investigating specialization

Next we examined the extent of specialization in the strains of Conception 2 by comparing the production ratios of the two products for each strain when grown together and separately, across the same parameter variations as were studied in **Figure 6.7**. These results are shown in **Figure 6.8**.

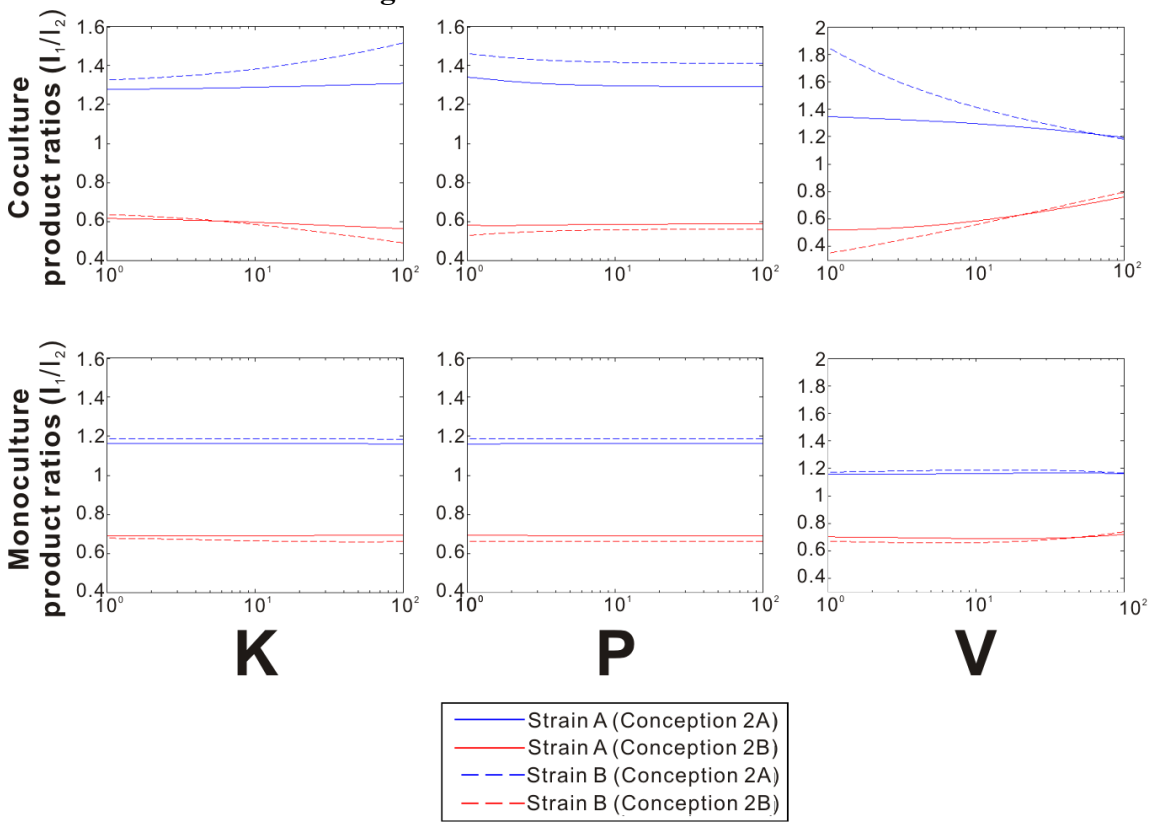

Figure 6.8. Specialization in Conception 2.

In all cases the x-axis represents the parameter value, and the y-axis is the production ratio of product 1 to product 2. Blue represents strain A, and red represents strain B. Results from Conception 2A are solid lines, and results from Conception 2B are dotted lines. Non-varied parameter values are as in **Figure 6.6**. The first column of graphs shows the effects of varying K, the second column shows the effects of varying P, and the third column shows the effects of varying V. The first row displays results for coculture of strains A and B, and the second row displays results for monoculture.

Figure 6.8 shows that in nearly all cases specialization is greater (i.e., the production ratios are farther from one) in coculture than in monoculture, and also that specialization is usually greater in Conception 2B than in Conception 2A, consistent with the better relative performance of Conception 2B in coculture versus monoculture. Strain B also usually specializes more than Strain A, as expected in comparative advantage, though the difference is not always large. Also interesting is that changing the parameters K, P, and V has little effect on specialization in monoculture but much more noticeable effects on specialization in coculture. Specialization is particularly responsive to changing parameter values in Conception 2B. Increasing K serves to increase specialization in the higher efficiency and higher penalty product, consistent with adversity requiring greater effort, while increasing P or V serves to decrease specialization, consistent with permissive conditions allowing laxity.

6.3.2.4 Conception 2: Further investigation of the K-P-V parameter space

To examine more fully the effects of changing the parameters K, P, and V, we made heat maps depicting changes in the benefit to coculture and the degree of specialization for both strains A and B across the K-P, K-V, and P-V planes in Conception 2B. We chose to focus on Conception 2B since, as discussed previously, it better adhered to the expectations of comparative advantage over a larger parameter space and was more responsive to changing parameters. The results are shown in **Figure 6.9**. The results are consistent with those shown in **Figures 6.7 and 6.8** in that coculture is favored at high K and high P, with strain A benefitting more from coculture at low V and strain B benefitting more at high V. Specialization, on the other hand, is highest at high K, low P, and low V for both strains, as seen previously.

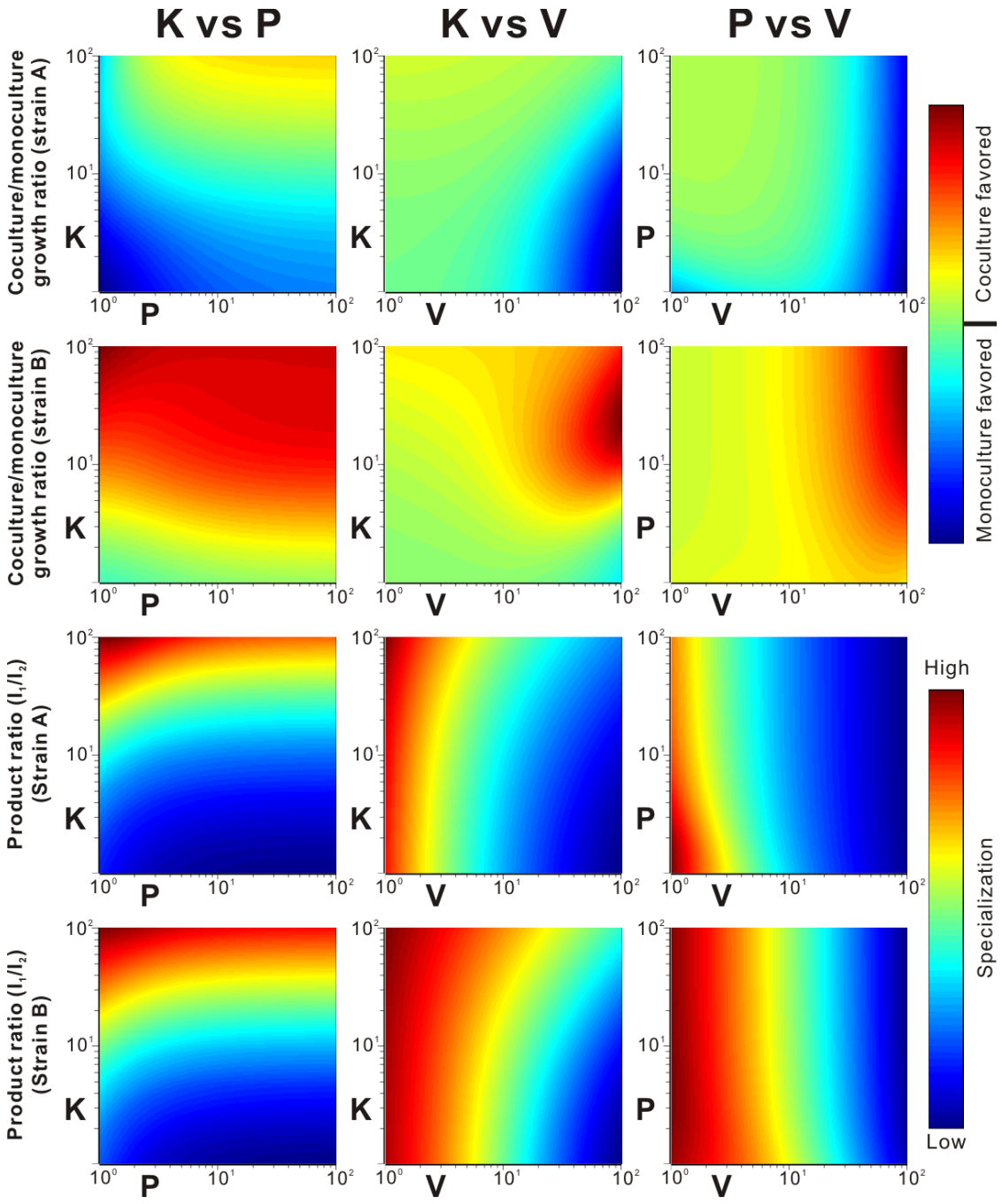

Figure 6.9. Exploration of the K-P-V parameter space in Conception 2B.

The first and second rows display the relative coculture/monoculture growth ratios for strains A and B, respectively, normalized such that a ratio of 1 is the median value on the color scale. The third and fourth rows display relative I_1/I_2 production ratios for strains A and B, respectively, where the color represents the absolute value of $I_1/I_2 - 1$. Non-varied parameter values are as in **Figure 6.6**.

The relative influence of the different parameters can also be assessed from such heat maps, in that when a banding pattern is seen across a particular axis, the value represented by that axis can be considered to have a stronger effect on the characteristic depicted than the value on the other axis. By this reasoning, we can judge that K and V have a similar degree of influence over the growth benefit from coculture, and both have much more influence than P. In specialization, V has the strongest effect, followed by K, and then P. These trends are also consistent with the data presented in **Figures 6.7** and **6.8**.

6.3.2.5 Conception 2: Alternate efficiency regimes

Next we wanted to compare the previous comparative advantage cases to a case of absolute advantage differing only in that the efficiencies of strain B are swapped. These results are shown over a range of values for K in the first two rows in **Figure 6.10**. (All the data in **Figure 6.10** is from Conception 2B, but equivalent results were obtained for Conception 2A, with trends in coculture growth benefit and specialization relative to Conception 2B similar to those in **Figs. 6.7** and **6.8**.) Specifically, the first row in **Figure 6.10** shows that the strains in the comparative advantage case gain a consistently higher benefit from cooperation than those in the absolute advantage case, and the second row in **Figure 6.10** shows that strain A specializes less in the absolute advantage case, while strain B specializes somewhat more and in the opposite direction from the comparative advantage case. Interestingly, this is the reverse of the absolute advantage effect seen in Conception 1 (see **Figs. 6.4G** and **6.4H**), where strain A specialized in its less efficient product. The difference in effect in Conception 2 may result from strain A's ability to influence the production of strain B through its own production.

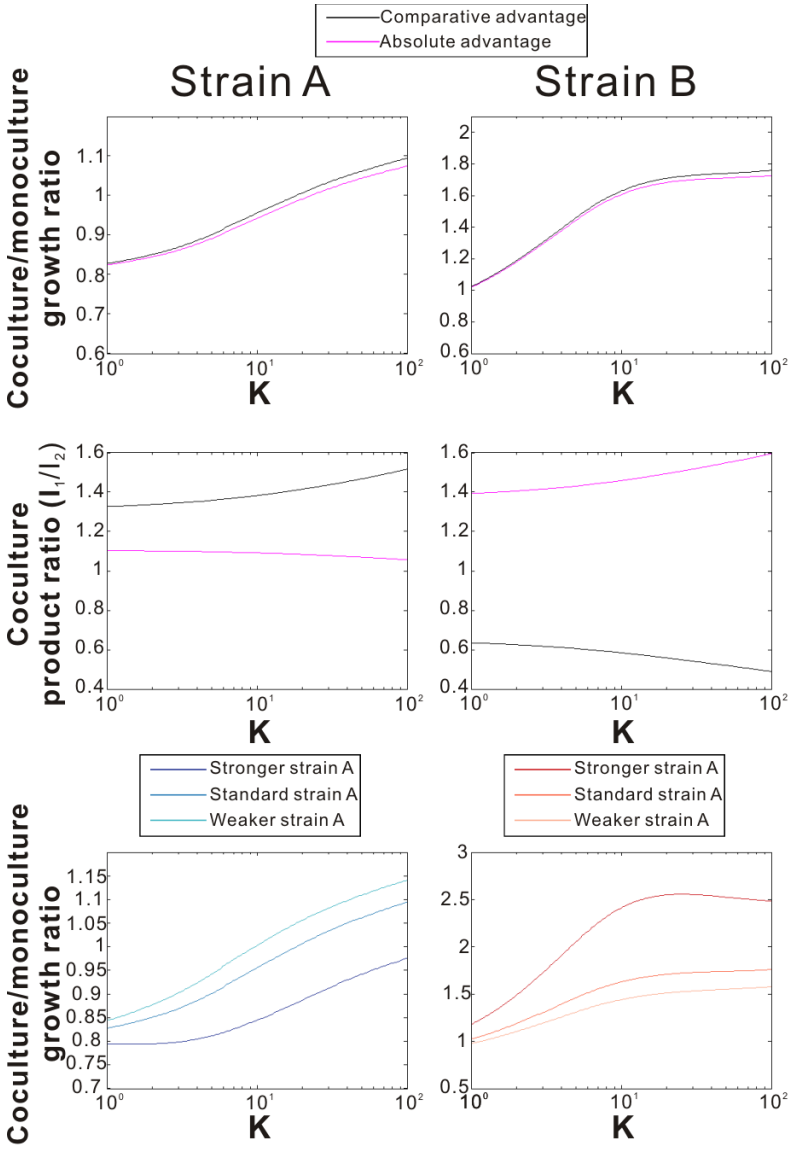

Figure 6.10. Alternate efficiency regimes in Conception 2B.

In all cases the x-axis represents the value of K, and the y-axis is either coculture/monoculture growth ratio (growth benefit) or the product 1/product 2 production ratio (specialization). Unvaried parameter values are as in **Figure 6.6** unless otherwise noted. The first and second rows show the difference in effect on growth benefit and specialization, respectively, between the comparative advantage case shown in **Figures 6.7** and **6.8** (black lines), and an absolute advantage case that is identical except that the efficiencies for strain B have been switched such that $V_{B1} = 1$ and $V_{B2} = 0.5$ (magenta lines). The third row shows the effect on growth benefit resulting from changing the efficiency of production of strain A, where the darkest colors represent a case where $V_{A1} = 4$ and $V_{A2} = 3$ (stronger strain A), the lightest colors represent a case where $V_{A1} = 1.5$ and $V_{A2} = 1.2$ (weaker strain A), and the intermediate colors represent the case where $V_{A1} = 2$ and $V_{A2} = 1.5$ (standard strain A), as in **Figure 6.7**.

Finally, in order to examine more closely the factors that determine whether cooperation is beneficial, particularly for strain A, we looked at varying the production efficiency of strain A. Specifically, we examined the differences in coculture benefit for strain A and strain B for three different value sets: $V_{A1} = 1.5$ and $V_{A2} = 1.2$ (weaker strain A), $V_{A1} = 2$ and $V_{A2} = 1.5$ (standard strain A), and $V_{A1} = 4$ and $V_{A2} = 3$ (stronger strain A) over a range of K values. The results are shown in the third row of **Figure 6.10**, which shows that the benefit to strain A of coculture decreases as its production strength increases, while the opposite trend is seen for strain B, which gains more from cooperation when strain A has higher production strength. Thus it seems that the stronger strain A is in comparison to strain B, the more of the burden of production it takes on.

Thus we have shown that a self-regulating microbial genetic circuit can in theory demonstrate the principles of comparative advantage, particularly when a strict trade-off in production is enforced, in adverse conditions, when the penalty for production is not too large, and when the difference in advantage between the two systems is not overly great.

6.4. DISCUSSION

We have shown that comparative advantage can in principle be implemented using simple signaling and feedback systems similar to those found in bacteria, which significantly broadens the demonstrated reach of the theory. Specifically, under the right conditions in both Conception 1, where the production levels are controlled by the experimenter, and Conception 2, which employs self-regulating genetic circuits, both strains grew better together than separately and specialized in the product they could

generate more efficiently, with the weaker strain specializing more completely than the stronger strain.

One interesting difference that arose in this model versus traditional models of comparative advantage is the variability in when and to what extent cooperation is beneficial to both parties in the comparative advantage context. Traditional models of comparative advantage assume that the total amount of effort expended by a given party remains constant, regardless of the amount produced. This is difficult to implement in a microbial system using current tools, and so we designed the system in terms analogous to having a relatively constant amount of raw materials, where greater or lesser effort can then be expended to produce greater or lesser amounts of products from those raw materials. This complicates the system in that the benefit to trade must be weighed against the disadvantages of increasing production. This difference results from including the penalty term and is responsible for much of the surprising behavior of these models. Situations such as this, where differential resource access is relatively constant but the effort that may be required to make use of those resources is variable, also have relevance to human economic systems, and thus these results may also have applicability beyond microbial genetics. For instance, if two countries of similar size both have arable land of comparable productivity per unit area, but one country has a greater amount of arable land than the other, the country with more farmland can produce more food, but only upon investing more effort.

The most interesting property emerging from the use of the penalty term is a general trend of more stringent conditions increasing the benefit to cooperation. In particular, increasing the difficulty with which the products can be converted into growth (the K parameter), which can be considered at least partly a result of environmental effects, is generally associated with increased benefit to cooperation. In the current

system this effect could likely be modulated by changing the antibiotic concentration. Hu and coworkers have in fact demonstrated that increased antibiotic concentration (below lethal limits) increases the benefit to cooperation in a simpler version of the system analyzed here, where each strain produces a single signaling molecule that the other needs to survive (Hu et al, 2010).

Similar findings have also resulted from a number of other studies of cooperation in both synthetic and natural microbial systems. For instance, in studies of yeast that secrete sugar-degrading enzymes in a cooperative fashion, the presence of environmental stressors such as low population density of common-good producers (Greig & Travisano, 2004; Sanchez & Gore, 2013; Waite & Shou, 2012), low nutrient concentration (Gore et al, 2009; Waite & Shou, 2012), and competition from other species (Celiker & Gore, 2013), has been shown to increase the benefit to cooperation. Similar trends have also been noted in *Pseudomonas* systems (Brockhurst, 2007; Brockhurst et al, 2007), which use signaling molecules of the type we have modeled here. In wild-type organisms these signaling molecules are frequently used to coordinate the formation of biofilms, which are known to form in response to adverse conditions, including the presence of antibiotics (Li & Tian, 2012; Mah, 2012). Additionally, a number of microbial systems exist that are normally unicellular but in stringent environments enter into cooperative relationships in which certain individuals sacrifice themselves and/or their ability to reproduce in order to benefit others. Such systems include the yeast *Saccharomyces cerevisiae*, in which older cells undergo programmed cell death and bequeath their nutrients to the younger generation (Fabrizio et al, 2004), as well as the bacterium *Myxococcus xanthus* (Fiegna et al, 2006) and the amoeba *Dictyostelium purpureum* (Mehdiabadi et al, 2006), which form multicellular fruiting bodies in which some cells reproduce and others are relegated exclusively to structural roles.

One interesting point to consider is the idea that antibiotics themselves could in some cases act as signals to induce cooperation. Cornforth and Foster have recently hypothesized that microbes may detect competitors via competition-induced stress, which is supported by the fact that production of antibiotics and toxins is frequently up-regulated by stressors such as nutrient limitation, cell damage, and oxidative stress, which could result from the presence of competitors, but is rarely enhanced by strictly abiotic stressors such as heat-shock and osmotic stress (Cornforth & Foster, 2013). Additionally, a number of investigators have shown that many antibiotics induce specific transcriptional changes in bacteria at concentrations too low to affect growth, leading some to hypothesize that antibiotics act as signaling molecules in natural environments (Davies, 2006; Fajardo & Martinez, 2008). The types of genes up-regulated depend on the strain and the compound but in a number of cases include genes for biofilm formation and cell adhesion (Davies et al, 2006; Linares et al, 2006), functions obviously associated with intercellular cooperation. Thus it may be the case that, at least in some contexts, microbial antibiotic production plays a dual role of attacking competitors while signaling friendly cells to band together for mutual benefit.

In contrast to the trend of environmental stress encouraging cooperation, making it intrinsically harder for the strains to produce the antibiotic resistance genes typically reduces the benefit to cooperation. Specifically, the V and P parameters, which represent respectively the maximum rate at which product can be turned into growth and the growth penalty for making the products, are likely intrinsic to the system and unlikely to be significantly modulated by changing external conditions such as antibiotic concentration. Changing either of these parameters can be considered as changing the cost of cooperation, and previous work in designed bacteria (Chuang et al, 2010) and yeast (Gore et al, 2009) has shown that increasing the cost of cooperation decreases the

benefit gained. The one quasi-exception we noted was in Conception 2, where the higher-producing strain (strain A) benefits more from cooperation at lower V, which is likely because a lower intrinsic production rate minimizes the opportunity for the lower-producing strain (strain B) to profit from the higher production of strain A without contributing production itself.

A next step would be testing these models *in vivo*, and the results presented here provide useful guidelines and testable hypotheses for designing such experiments. The first step would be to build a base strain that expresses two antibiotic resistance genes in response to two respective signaling molecules. For the signaling molecules, the rhl and lux systems have been reported to be orthogonal (Hu et al, 2010), and methods exist for reducing cross-talk should it occur (Brenner et al, 2007; Smith et al, 2008). Alternatively one or the other of the rhl and lux systems could be used together with an orthogonal peptide signal (Marchand & Collins, 2013). Once the base strains are built, the accuracy of equations (4) and (5) could then be verified by exogenously adding the signaling molecules at different concentrations to examine the effect on growth. Values for V and K could then be estimated. These parameters could also be checked for their dependence on antibiotic concentration.

For Conception 1, the genes for producing the signaling molecules would need to be put under the control of four different exogenous inducers, and then mapping functions for generating a linear response in growth rate from inducer concentrations would need to be experimentally determined. A value for P could also be obtained from these experiments. Equivalents to the heat maps presented here could then be generated by growing the cells in 96-well plates along gradients of inducer concentrations and measuring growth rates directly.

For Conception 2 in particular, the results presented here give a great deal of guidance for how the gene circuits should be designed and tested to maximize the chances of successfully replicating the effects of comparative advantage. Specifically, better performance should be seen under Conception 2B, with relatively weak promoters for the antibiotic resistance genes (lower V), with relatively narrow differences in efficiencies of production between the two strains, and at higher antibiotic concentrations (higher K). Although Conception 2A grew faster than 2B, this may be because Conception 2B simply adds negative feedback to 2A. If an additional internal positive feedback loop was added, these growth deficits might disappear. Additionally, more detailed simulations of transcription, translation, enzyme catalysis, and export could also be developed as necessary to guide the design of both Conceptions 1 and 2.

Besides expanding the demonstrated reach of the comparative advantage theory and proving that bacteria are capable of implementing and benefitting from such trading relationships, the implementation of such systems *in vivo* would also expand our toolkit for engineering bacterial consortia to efficiently execute human-directed tasks (Shong et al, 2012). As a hypothetical example, in a consortium of two microbes that can both produce a desired product but have differing efficiencies in performing different steps of the relevant metabolic pathway, if the microbes are designed to allow trading of intermediates, responsive signaling networks such as those modeled in Conception 2 herein could be used to maximize overall production and to continuously adapt to changes in environmental variables, such as the concentrations of intermediates, products, and other cells. The results presented herein represent a theoretical proof of this principle.

Finally, it is possible and perhaps likely that wild-type microbes enter into comparative-advantage-like social interactions in natural settings. In order to find such

interactions, a series of pair-wise coculture growth experiments could be performed, and the results compared to the monoculture case, as was reported, for instance, by (Foster & Bell, 2012). Those strains that grew better together than separately are candidates for natural examples of microbial comparative advantage, and this could potentially be confirmed by expression studies comparing gene regulation in monoculture and coculture, and/or mutagenesis studies to determine the genes required for mutualism. It is also possible that comparative advantage might be more likely to come into play in tandem with kin selection. Comparative advantage could potentially arise even between genetically identical cells if they are subjected to different environments, such as the center versus the edge of a colony or biofilm. Expression studies examining the differences between cells in different locations in a biofilm (Lenz et al, 2008) could potentially turn up such examples. It may be that comparative advantage is universally applied in all domains of life.

Appendix: Targetrons used in the present work

This appendix lists the names, sources, integration sites, insertion efficiencies, and retargeting primers for all the targetrons used in this work. For LtrB introns, the constant retargeting primer (used in tandem with the IBS primer) is EBS2AS (AATTAGAAACTTGCGTTCAGTAAACACAACTTATAC), and for EcI5 introns, the constant retargeting primer (used in tandem with the IBS1/2S primer is EBSR (TATCCGGTCCATTACAGACTGGCATTC). Note that the primers for the *lacZ*-targeting introns were inferred from the sequences of the intron-expressing plasmids and were not used to construct the introns used in this work.

Notes on sources for efficiency data: Intron efficiency for LtrB.LacZ.635s is the average from the three plates used in calculating the efficiency of the unmodified intron in **Figure 2.1**. Intron efficiencies for EcI5.LacZ.912s and EcI5.LacZ.1806s are as reported in Zhuang et al, 2009. Intron efficiencies for the LtrB.SIR32.1 and EcI5.SIR5.6 introns are from all introns carrying *Ter* sites that did not destroy integration efficiency. Intron efficiencies for LtrB.SacB.1221s are as reported in Yao, 2008. Introns efficiencies for LtrB.YhcS.187s are as reported in Whitt, 2001. Intron efficiencies for all other introns are for all uses of the intron during the course of studies in Chapters 2 and 3 and include tests in various strains and with any *lox* inserts that form hairpins with flexible bases (in addition to tests using the unmodified intron). Partial insertions (bands corresponding to both uninserted and inserted states seen upon colony PCR) were counted as insertions.

LtrB.LacZ.635s (*E. coli*; insertion site sequence from strain MG1655)
Source: Perutka & Lambowitz, unpublished results.

```
LtrB IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
364894|364895(-)  T A T G T G G C G G A T G A G C G G C A T T T T C C G T G A C G T C T C G T T G C T G C A
```

Inserts in sense strand of *lacZ*.
Insertion efficiency: 15.3%

Retargeting primers:

```
635s-IBS       AAAAAAGCTTCGTCGATCGTGAACATTTTCCGTGAGTGCGCCCAGATAGGGTG
635s-EBS1      CAGATTGTACAAATGTGGTGATAACAGATAAGTCCCGTGACGTAACTTACCTTTCTTTGT
635s-EBS2      TGAACGCAAGTTTCTAATTTCGGTTAAATGTCGATAGAGGAAAGTGTCT
```

EcI5.LacZ.912s (*E. coli*; insertion site sequence from strain MG1655)
Source: Zhuang et al., 2009

```
EcI5 IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
364617|364618(-)   A A C G T C G A A A A C C C G A A A C T G T G G A G C G C C G A A A T C C C G A A T C T C
```

Inserts in sense strand of *lacZ*.
Insertion efficiency: 68±7.1%

Retargeting primers:

```
912s-IBS1/2S   CCCCTCTAGAAGAATTCCCATGCCAAAACTGTGGAGCGCCGTGCGACATGAAGTCG
912s-EBS1S     CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTGGCGCCTAGGGGTACACGGAC
912s-EBS2AS    TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTACTGTTAACGACGCTTCAGC
```

EcI5.LacZ.1806s (*E. coli*; insertion site sequence from strain MG1655)
Source: Zhuang et al., 2009

```
EcI5 IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
363723|363724(-)   T T T G G C G A T A C G C C G A A C G A T C G C C A G T T C T G T A T G A A C G G T C T G
```

Inserts in sense strand of *lacZ*.
Insertion efficiency: 97±0.4%

Retargeting primers:

```
1806s-IBS1/2S CCCCTCTAGAAGAATTCCCATGCCAAACGATCGCCAGTTCGTGCGACATGAAGTCG
1806s-EBS1S   CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTGAACTCTAGGGGTACACGGAC
1806s-EBS2AS  TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTCGATCTAACGACGCTTCAGC
```

LtrB.A (*E. coli*; insertion site sequence from strain MG1655)
Source: This work

```
LtrB IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
243410|243411(-)  T G A A G T G C G G A T A A A A A C A G C A A C A A T G T G A G C T T T G T T G T A A T T
```

Insertion efficiency: 6/239 (2.5%)

Retargeting primers (constant primer listed in **Supplementary Table 5**):

```
A-IBS          AAAAAAGCTTATAATTATCCTTAAGCAACAATGTGGTGCGCCCAGATAGGGTG
A-EBS1         CAGATTGTACAAATGTGGTGATAACAGATAAGTCAATGTGAGTAACTTACCTTTCTTTGT
A-EBS2         TGAACGCAAGTTTCTAATTTCGATTTTGCTTCGATAGAGGAAAGTGTCT
```

LtrB.SIR32.1 (*E. coli*; insertion site sequence from strain MG1655)
Source: This work

```
LtrB IS       -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
3922698|3922698(+) A A C A C T G G T G A T A A A G C G T G C T T C A G A T C A C A T A T T G C G C A T G T T
```

Insertion efficiency: 9/27 (33%)

Retargeting primers:

```
SIR32.1-IBS    AAAAAAGCTTATAATTATCCTTATGCTTCAGATCAGTGCGCCCAGATAGGGTG
SIR32.1-EBS1   CAGATTGTACAAATGTGGTGATAACAGATAAGTCAGATCACATAACTTACCTTTCTTTGT
SIR32.1-EBS2   TGAACGCAAGTTTCTAATTTCGATTAAGCATCGATAGAGGAAAGTGTCT
```

EcI5.B (*E. coli*; insertion site sequence from strain MG1655)
Source: This work

```
EcI5 IS           -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
1399016|1399017(-)  T G C A G A C A T T G A C C G A A A G T C A G C G T T T T G G T T T A C G C A T A G C A G
```

Insertion efficiency: 26/94 (27.7%)

Retargeting primers (constant primer listed in **Supplementary Table 5**):

```
B-IBS1/2S      CCCCTCTAGAAGAATTCCCATGCCAAAAGTCAGCGTTTTGGTGCGACATGAAGTCG
B-EBS1S        CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTCAAAACTAGGGGTACACGGAC
B-EBS2AS       TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTAGTCATAACGACGCTTCAGC
```

EcI5.C (CMP) (*E. coli*; insertion site sequence from strain MG1655)
Source: This work

```
EcI5 IS              -30       -25        -20       -15       -10        -5       -1+1      +5
+10       +15
1479630|1479631(+)   T G T A T T G A T G G A G C T A A T G A T G A T G A T T C A G T T A T G G T G G T C A G T
```

Insertion efficiency: 11/31 (35.5%)

Retargeting primers:

```
C-IBS1/2S      CCCCTCTAGAAGAATTCCCATGCCAAATGATGATGATTCAGTGCGACATGAAGTCG
C-EBS1S        CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTTGAATCTAGGGGTACACGGAC
C-EBS2AS       TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTTGATGTAACGACGCTTCAGC
```

EcI5.D (*E. coli*; insertion site sequence from strain MG1655)
Source: This work

```
EcI5 IS           -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
3452370|3452371(-)  T G G A T C G C A T C G C T T A A A G T C G G G G A C A A A A A A T T G C C T G T T G T G
```

Insertion efficiency: 8/67 (11.9%)

Retargeting primers:

```
D-IBS1/2S      CCCCTCTAGAAGAATTCCCATGCCAAAAGTCGGGGACAAAGTGCGACATGAAGTCG
D-EBS1S        CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTTTTGTCTAGGGGTACACGGAC
D-EBS2AS       TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTAGTCGTAACGACGCTTCAGC
```

EcI5.E (*E. coli*; insertion site sequence from strain MG1655)
Source: This work

```
EcI5 IS           -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
3466270|3466271(+)  C A C A C C G T T A A A G C G A A T C A G C G T A T C G C T G G C A T A A A G C G T T T C
```

Insertion efficiency: 53/93 (57.0%)

Retargeting primers:

```
E-IBS1/2S      CCCCTCTAGAAGAATTCCCATGCCAAATCAGCGTATCGCTGTGCGACATGAAGTCG
E-EBS1S        CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTAGCGACTAGGGGTACACGGAC
E-EBS2AS       TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTTCAGCTAACGACGCTTCAGC
```

EcI5.SIR5.6 (*E. coli*, insertion site sequence from strain MG1655)
Source: This work

```
EcI5 IS        -30      -25      -20      -15      -10      -5     -1+1     +5       +10      +15
182213|182214(-) A T T G T G C A A A T G C C T A A A G G A T G A T G A A G A T G T A T G G A G T T G T G G
```

Insertion efficiency: 53/65 (81.5%)

Retargeting primers:

```
SIR5.6-IBS1/2S     CCCCTCTAGAAGAATTCCCATGCCAAAAGGATGATGAAGAGTGCGACATGAAGTCG
SIR5.6-EBS1S       CAGGCTTGAACCAAAAGGTATGTGGTTGGTTACTCCTCTTCTTCCTAGGGGTACACGGAC
SIR5.6-EBS2AS      TACCTTTTGGTTCAAGCCTGTCAGCATCTTTGGCTTGTTAGGATTAACGACGCTTCAGC
```

LtrB.SAPI-int (*S. aureus*, insertion site sequence from strain NCTC 8325)
Source: This work

```
LtrB IS          -30      -25      -20      -15      -10      -5     -1+1     +5       +10      +15
953084|953085(-)  G A T G A A A T G G A T A G T A A A G G T T A T G T A T A T C A A G T T A A C A A A G A T
```

Inserts in sense strand of the *int* gene.
Insertion efficiency: 27/28 (96.4%)

Retargeting primers:

```
SAPI-int-IBS       AAAAAAGCTTATAATTATCCTTAGGTTACGTATATGTGCGCCCAGATAGGGTG
SAPI-int-EBS1      CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTATATCATAACTTACCTTTCTTTGT
SaPI-int-EBS2      TGAACGCAAGTTTCTAATTTCGATTTAACCTCGATAGAGGAAAGTGTCT
```

LtrB.SAPI-B (*S. aureus*, insertion site sequence from strain NCTC 8325)
Source: This work

```
LtrB IS          -30      -25      -20      -15      -10      -5     -1+1     +5       +10      +15
968024|968025(-)  T C T T T G G T G G A T T A A T C A T T G G T A T C G T T C C A T A T T T A T T G A A A A
```

Insertion efficiency: 23/23 (100.0%)

Retargeting primers:

```
SAPI-B-IBS       AAAAAAGCTTATAATTATCCTTATTGGTCTCGTTCGTGCGCCCAGATAGGGTG
SAPI-B-EBS1      CAGATTGTACAAATGTGGTGATAACAGATAAGTCTCGTTCCATAACTTACCTTTCTTTGT
SAPI-B-EBS2      TGAACGCAAGTTTCTAATTTCGGTTACCAATCGATAGAGGAAAGTGTCT
```

LtrB.SacB.1221s (*B. subtilis*, insertion site sequence from strain 168)
Source: Yao 2008

```
LtrB IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
3537232|3537233(+)T T A A A A A T G G A T C T T G A T C C T A A C G A T G T A A C C T T T A C T T A C T C A
```

Inserts into the sense strand of the *sacB* gene.
Insertion efficiency: 47/48 (97.9%)

Retargeting primers:

```
1221s-IBS        AAAAAAGCTTATAATTATCCTTACCTAACGATGTAGTGCGCCCAGATAGGGTG
1221s-EBS1       CAGATTGTACAAATGTGGTGATAACAGATAAGTCGATGTAACTAACTTACCTTTCTTTGT
1221s-EBS2       TGAACGCAAGTTTCTAATTTCGGTTTTAGGTCGATAGAGGAAAGTGTCT
```

LtrB.YhcS.168s (*B. subtilis*, insertion site sequence from strain 168)
Source: Whitt 2011

```
LtrB IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
995184|995184(+)  A A T A G C A C A G A T C A A G C A A A G A A C A A A G C A T C A T T T A A G C C T G A G
```

Inserts into the sense strand of the *yhcS* (*srtA*) gene.
Insertion efficiency: 91 ± 5%

Retargeting primers:

```
186s-IBS         AAAAAAGCTTATAATTATCCTTAAAGAACAAAGCAGTGCGCCCAGATAGGGTG
186s-EBS1        CAGATTGTACAAATGTGGTGATAACAGATAAGTCAAAGCATCTAACTTACCTTTCTTTGT
186s-EBS2        TGAACGCAAGTTTCTAATTTCGGTTTTCTTCCGATAGAGGAAAGTGTCT
```

LtrB.rDNA.798s *(S. oneidensis*, insertion site sequence from strain MR-1 *rrsA* gene)
Source: Erik Quandt (Enyeart et al., 2013)

```
LtrB IS          -30       -25       -20       -15       -10       -5      -1+1      +5        +10       +15
787|788s          A A G C G T G G G G A G C A A A C A G G A T T A G A T A C C C T G G T A G T C C A C G C C
```

Inserts into the sense strand of the *rrs* genes in *S. oneidensis*.
Insertion efficiency: Not directly applicable, but all tested colonies contained the insertion in most copies of the *rrs* gene.

Retargeting primers:

```
798s-IBS         AAAAAAGCTTATAATTATCCTTAGGATTCGATACCGTGCGCCCAGATAGGGTG
798s-EBS1        ACAAAGAAAGGTAAGTTAAGGGTATCGACTTATCTGTTATCACCACATTTGTACAATCTG
798s-EBS2        TGAACGCAAGTTTCTAATTTCGATTAATCCTCGATAGAGGAAAGTGTCT
```

References

Abuin A, Bradley A (1996) Recycling selectable markers in mouse embryonic stem cells. *Mol Cell Biol* **16:** 1851-1856

Adams MR (1985) Vinegar. In *Microbiology of Fermented Foods*, Wood BJB (ed). London ; New York: Elsevier Applied Science Publishers

Aiba H (1985) Transcription of the Escherichia coli adenylate cyclase gene is negatively regulated by cAMP-cAMP receptor protein. *J Biol Chem* **260:** 3063-3070

Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ (1998) Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc Natl Acad Sci U S A* **95:** 8927-8932

Akhtar P, Khan SA (2012) Two independent replicons can support replication of the anthrax toxin-encoding plasmid pXO1 of Bacillus anthracis. *Plasmid* **67:** 111-117

Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J* **7:** 649-659

Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol* **181:** 6361-6370

Alonzo F, 3rd, Port GC, Cao M, Freitag NE (2009) The posttranslocation chaperone PrsA2 contributes to multiple facets of Listeria monocytogenes pathogenesis. *Infect Immun* **77:** 2612-2623

Alper H, Miyaoku K, Stephanopoulos G (2005) Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* **23:** 612-616

Ames BN, Durston WE, Yamasaki E, Lee FD (1973a) Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci U S A* **70:** 2281-2285

Ames BN, Lee FD, Durston WE (1973b) An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc Natl Acad Sci U S A* **70:** 782-786

Anastassiadis K, Fu J, Patsch C, Hu S, Weidlich S, Duerschke K, Buchholz F, Edenhofer F, Stewart AF (2009) Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. *Dis Model Mech* **2:** 508-515

Andersen JT, Poulsen P, Jensen KF (1992) Attenuation in the rph-pyrE operon of Escherichia coli and processing of the dicistronic mRNA. *Eur J Biochem* **206:** 381-390

Andrews BJ, Proteau GA, Beatty LG, Sadowski PD (1985) The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. *Cell* **40:** 795-803

Arakawa H, Lodygin D, Buerstedde JM (2001) Mutant loxP vectors for selectable marker recycle and conditional knock-outs. *BMC Biotechnol* **1:** 7

Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11:** 1365-1374

Arora R (2012) *Microbial biotechnology : energy and environment*, Wallingford, Oxfordshire: Cambridge CAB International.

Arslan E, Schulz H, Zufferey R, Kunzler P, Thony-Meyer L (1998) Overproduction of the Bradyrhizobium japonicum c-type cytochrome subunits of the cbb3 oxidase in Escherichia coli. *Biochem Biophys Res Commun* **251:** 744-747

Atlung T, Nielsen A, Rasmussen LJ, Nellemann LJ, Holm F (1991) A versatile method for integration of genes and gene fusions into the lambda attachment site of Escherichia coli. *Gene* **107:** 11-17

Auerbach C, Robson JM, Carr JG (1947) The Chemical Production of Mutations. *Science* **105:** 243-247

Azuma Y, Hosoyama A, Matsutani M, Furuya N, Horikawa H, Harada T, Hirakawa H, Kuhara S, Matsushita K, Fujita N, Shirai M (2009) Whole-genome analyses reveal genetic instability of Acetobacter pasteurianus. *Nucleic Acids Res* **37:** 5768-5783

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 2006 0008

Badarinarayana V, Estep PW, 3rd, Shendure J, Edwards J, Tavazoie S, Lam F, Church GM (2001) Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat Biotechnol* **19:** 1060-1065

Balagadde FK, Song H, Ozaki J, Collins CH, Barnet M, Arnold FH, Quake SR, You L (2008) A synthetic Escherichia coli predator-prey ecosystem. *Mol Syst Biol* **4:** 187

Barker CS, Pruss BM, Matsumura P (2004) Increased motility of Escherichia coli by insertion sequence element integration into the regulatory region of the flhD operon. *J Bacteriol* **186:** 7529-7537

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461:** 1243-U1274

Belhocine K, Mak AB, Cousineau B (2008) Trans-splicing versatility of the Ll.LtrB group II intron. *RNA* **14:** 1782-1790

Beliaev AS, Saffarini DA (1998) Shewanella putrefaciens mtrB encodes an outer membrane protein required for Fe(III) and Mn(IV) reduction. *J Bacteriol* **180:** 6292-6297

Beliaev AS, Saffarini DA, McLaughlin JL, Hunnicutt D (2001) MtrC, an outer membrane decahaem c cytochrome required for metal reduction in Shewanella putrefaciens MR-1. *Mol Microbiol* **39:** 722-730

Beloin C, Deighan P, Doyle M, Dorman CJ (2003) Shigella flexneri 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein. *Mol Genet Genomics* **270:** 66-77

Bender J, Kleckner N (1986) Genetic evidence that Tn10 transposes by a nonreplicative mechanism. *Cell* **45:** 801-815

Benders GA, Noskov VN, Denisova EA, Lartigue C, Gibson DG, Assad-Garcia N, Chuang RY, Carrera W, Moodie M, Algire MA, Phan Q, Alperovich N, Vashee S, Merryman C, Venter JC, Smith HO, Glass JI, Hutchison CA, 3rd (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res* **38:** 2558-2569

Beringer JE, Beynon JL, Buchanan-Wollaston AV, Johnston AWB (1978) Transfer of the drug-resistance transposon Tn5 to Rhizobium. *Nature* **276:** 633-634

Bhatia MB, Vinitsky A, Grubmeyer C (1990) Kinetic mechanism of orotate phosphoribosyltransferase from Salmonella typhimurium. *Biochemistry* **29:** 10480-10487

Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA (2012) CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12:** 177-186

Biliouris K, Babson D, Schmidt-Dannert C, Kaznessis YN (2012) Stochastic simulations of a synthetic bacteria-yeast ecosystem. *BMC Syst Biol* **6:** 58

Binder S, Siedler S, Marienhagen J, Bott M, Eggeling L (2013) Recombineering in Corynebacterium glutamicum combined with optical nanosensors: a general strategy for fast producer strain generation. *Nucleic Acids Res* **41:** 6360-6369

Biswas I, Gruss A, Ehrlich SD, Maguin E (1993) High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol* **175:** 3628-3635

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of Escherichia coli K-12. *Science* **277:** 1453-1462

Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489:** 513-+

Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *P Natl Acad Sci USA* **105:** 7899-7906

Bolusani S, Ma CH, Paek A, Konieczka JH, Jayaram M, Voziyanov Y (2006) Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites. *Nucleic Acids Res* **34:** 5259-5269

Bonekamp F, Clemmesen K, Karlstrom O, Jensen KF (1984) Mechanism of UTP-modulated attenuation at the pyrE gene of Escherichia coli: an example of operon polarity control through the coupling of translation to transcription. *EMBO J* **3:** 2857-2861

Bordi C, Butcher BG, Shi Q, Hachmann AB, Peters JE, Helmann JD (2008) In vitro mutagenesis of Bacillus subtilis by using a modified Tn7 transposon with an outward-facing inducible promoter. *Appl Environ Microbiol* **74:** 3419-3425

Borloo J, Vergauwen B, De Smet L, Brige A, Motte B, Devreese B, Van Beeumen J (2007) A kinetic approach to the dependence of dissimilatory metal reduction by Shewanella oneidensis MR-1 on the outer membrane cytochromes c OmcA and OmcB. *FEBS J* **274:** 3728-3738

Bouhenni R, Gehrke A, Saffarini D (2005) Identification of genes involved in cytochrome c biogenesis in Shewanella oneidensis, using a modified mariner transposon. *Appl Environ Microbiol* **71:** 4935-4937

Brenner K, Karig DK, Weiss R, Arnold FH (2007) Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc Natl Acad Sci U S A* **104:** 17300-17304

Bretschger O, Obraztsova A, Sturm CA, Chang IS, Gorby YA, Reed SB, Culley DE, Reardon CL, Barua S, Romine MF, Zhou J, Beliaev AS, Bouhenni R, Saffarini D, Mansfeld F, Kim BH, Fredrickson JK, Nealson KH (2007) Current production and metal oxide reduction by Shewanella oneidensis MR-1 wild type and mutants. *Appl Environ Microbiol* **73:** 7003-7012

Broach JR, Hicks JB (1980) Replication and recombination functions associated with the yeast plasmid, 2 mu circle. *Cell* **21:** 501-508

Brockhurst MA (2007) Population bottlenecks promote cooperation in bacterial biofilms. *PLoS One* **2:** e634

Brockhurst MA, Buckling A, Gardner A (2007) Cooperation peaks at intermediate disturbance. *Curr Biol* **17:** 761-765

Bryksin AV, Matsumura I (2010) Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PLoS One* **5:** e13244

Buchan BW, McCaffrey RL, Lindemann SR, Allen LA, Jones BD (2009) Identification of migR, a regulatory element of the Francisella tularensis live vaccine strain iglABCD virulence operon required for normal replication and trafficking in macrophages. *Infect Immun* **77:** 2517-2529

Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* **16:** 657-662

Buchholz F, Ringrose L, Angrand PO, Rossi F, Stewart AF (1996) Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nucleic Acids Res* **24:** 4256-4262

Buchholz F, Stewart AF (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol* **19:** 1047-1052

Bucking C, Popp F, Kerzenmacher S, Gescher J (2010) Involvement and specificity of Shewanella oneidensis outer membrane cytochromes in the reduction of soluble and solid-phase terminal electron acceptors. *FEMS Microbiol Lett* **306:** 144-151

Calos MP, Miller JH (1980) Transposable elements. *Cell* **20:** 579-595

Campbell AM (1962) Episomes. *Adv Genet* **11:** 101-145

Campo N, Daveran-Mingot ML, Leenhouts K, Ritzenthaler P, Le Bourgeois P (2002) Cre-loxP recombination system for large genome rearrangements in Lactococcus lactis. *Appl Environ Microbiol* **68:** 2359-2367

Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P, Le Bourgeois P (2004) Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol Microbiol* **51:** 511-522

Carl PL (1970) Escherichia Coli Mutants with Temperature Sensitive Synthesis of DNA. *Mol Gen Genet* **109:** 107-&

Carr PA, Church GM (2009) Genome engineering. *Nat Biotechnol* **27:** 1151-1162

Carr PA, Wang HH, Sterling B, Isaacs FJ, Lajoie MJ, Xu G, Church GM, Jacobson JM (2012) Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* **40:** e132

Carter DM, Radding CM (1971) The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. *J Biol Chem* **246:** 2502-2512

Carter GP, Awad MM, Hao Y, Thelen T, Bergin IL, Howarth PM, Seemann T, Rood JI, Aronoff DM, Lyras D (2011) TcsL is an essential virulence factor in Clostridium sordellii ATCC 9714. *Infect Immun* **79:** 1025-1032

Casadaban MJ, Cohen SN (1979) Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: in vivo probe for transcriptional control sequences. *Proc Natl Acad Sci U S A* **76:** 4530-4533

Cavalli LL, Lederberg J, Lederberg EM (1953) An infective factor controlling sex compatibility in Bacterium coli. *J Gen Microbiol* **8:** 89-103

Celiker H, Gore J (2013) Cellular cooperation: insights from microbes. *Trends Cell Biol* **23:** 9-15

Chambers SP, Prior SE, Barstow DA, Minton NP (1988) The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68:** 139-149

Chang S, Cohen SN (1977) In vivo site-specific genetic recombination promoted by the EcoRI restriction endonuclease. *Proc Natl Acad Sci U S A* **74:** 4811-4815

Chang S, Cohen SN (1979) High frequency transformation of Bacillus subtilis protoplasts by plasmid DNA. *Mol Gen Genet* **168:** 111-115

Chen Y, Caruso L, McClane B, Fisher D, Gupta P (2007) Disruption of a toxin gene by introduction of a foreign gene into the chromosome of Clostridium perfringens using targetron-induced mutagenesis. *Plasmid* **58:** 182-189

Chen Y, McClane BA, Fisher DJ, Rood JI, Gupta P (2005) Construction of an alpha toxin gene knockout mutant of Clostridium perfringens type A by use of a mobile group II intron. *Appl Environ Microbiol* **71:** 7542-7547

Cheng C, Nair AD, Indukuri VV, Gong S, Felsheim RF, Jaworski D, Munderloh UG, Ganta RR (2013) Targeted and Random Mutagenesis of Ehrlichia chaffeensis for the Identification of Genes Required for In vivo Infection. *PLoS Pathog* **9:** e1003171

Chiang SL, Rubin EJ (2002) Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene* **296:** 179-185

Cho EJ, Ellington AD (2007) Optimization of the biological component of a bioelectrochemical cell. *Bioelectrochemistry* **70:** 165-172

Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31:** 230-232

Chuang JS, Rivoire O, Leibler S (2009) Simpson's paradox in a synthetic microbial system. *Science* **323:** 272-275

Chuang JS, Rivoire O, Leibler S (2010) Cooperation and Hamilton's rule in a simple synthetic microbial system. *Mol Syst Biol* **6:** 398

Clarke TA, Edwards MJ, Gates AJ, Hall A, White GF, Bradley J, Reardon CL, Shi L, Beliaev AS, Marshall MJ, Wang Z, Watmough NJ, Fredrickson JK, Zachara JM, Butt JN, Richardson DJ (2011) Structure of a bacterial cell surface decaheme electron conduit. *Proc Natl Acad Sci U S A* **108:** 9384-9389

Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A* **70:** 3240-3244

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819-823

Constantinidou C, Hobman JL, Griffiths L, Patel MD, Penn CW, Cole JA, Overton TW (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as Escherichia coli K12 adapts from aerobic to anaerobic growth. *J Biol Chem* **281:** 4802-4815

Cooper S, Helmstetter CE (1968) Chromosome replication and the division cycle of Escherichia coli B/r. *J Mol Biol* **31:** 519-540

Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in qene expression after 20,000 generations of evolution in Escherichia coli. *P Natl Acad Sci USA* **100:** 1072-1077

Cornforth DM, Foster KR (2013) Competition sensing: the social side of bacterial stress responses. *Nat Rev Microbiol* **11:** 285-293

Coros CJ, Landthaler M, Piazza CL, Beauregard A, Esposito D, Perutka J, Lambowitz AM, Belfort M (2005) Retrotransposition strategies of the Lactococcus lactis Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol* **56:** 509-524

Corvaglia AR, Francois P, Hernandez D, Perron K, Linder P, Schrenzel J (2010) A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains. *Proc Natl Acad Sci U S A* **107:** 11954-11958

Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100:** 15748-15753

Coursolle D, Gralnick JA (2010) Modularity of the Mtr respiratory pathway of Shewanella oneidensis strain MR-1. *Mol Microbiol*

Craig Maclean R, Brandon C (2008) Stable public goods cooperation and dynamic social interactions in yeast. *J Evol Biol* **21:** 1836-1843

Crespi BJ (2001) The evolution of social behavior in microorganisms. *Trends Ecol Evol* **16:** 178-183

Cui GZ, Hong W, Zhang J, Li WL, Feng Y, Liu YJ, Cui Q (2012) Targeted gene engineering in Clostridium cellulolyticum H10 without methylation. *J Microbiol Methods* **89:** 201-208

d'Herelle F (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. *Comptes Rendus Acad Sciences* **165:** 373-375

Dai M, Ziesman S, Ratcliffe T, Gill RT, Copley SD (2005) Visualization of protoplast fusion and quantitation of recombination in fused protoplasts of auxotrophic strains of Escherichia coli. *Metab Eng* **7:** 45-52

Danino T, Mondragon-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. *Nature* **463:** 326-330

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97:** 6640-6645

Datta S, Costantino N, Zhou X, Court DL (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A* **105:** 1626-1631

Davies J (2006) Are antibiotics naturally antibiotics? *J Ind Microbiol Biotechnol* **33:** 496-499

Davies J, Spiegelman GB, Yim G (2006) The world of subinhibitory antibiotic concentrations. *Curr Opin Microbiol* **9:** 445-453

Derbise A, Lesic B, Dacheux D, Ghigo JM, Carniel E (2003) A rapid and simple method for inactivating chromosomal genes in Yersinia. *FEMS Immunol Med Microbiol* **38:** 113-116

Diederich L, Rasmussen LJ, Messer W (1992) New cloning vectors for integration in the lambda attachment site attB of the Escherichia coli chromosome. *Plasmid* **28:** 14-24

Diggle SP, Griffin AS, Campbell GS, West SA (2007) Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450:** 411-414

Donald JW, Hicks MG, Richardson DJ, Palmer T (2008) The c-type cytochrome OmcA localizes to the outer membrane upon heterologous expression in Escherichia coli. *J Bacteriol* **190:** 5127-5131

Donath MJ, 2nd, Dominguez MA, Withers ST, 3rd (2011) Development of an automated platform for high-throughput P1-phage transduction of Escherichia coli. *J Lab Autom* **16:** 141-147

Dong H, Zhang Y, Dai Z, Li Y (2010) Engineering clostridium strain to accept unmethylated DNA. *PLoS One* **5:** e9038

Duckworth DH (1976) "Who discovered bacteriophage?". *Bacteriol Rev* **40:** 793-802

Duggin IG, Bell SD (2009) Termination structures in the Escherichia coli chromosome replication fork trap. *J Mol Biol* **387:** 532-539

Dunham MJ (2007) Synthetic ecology: a model system for cooperation. *Proc Natl Acad Sci U S A* **104:** 1741-1742

Dutta D, Gachhui R (2007) Nitrogen-fixing and cellulose-producing Gluconacetobacter kombuchae sp nov., isolated from Kombucha tea. *Int J Syst Evol Micr* **57:** 353-357

Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL, Boeke AC, Gottschling DE, Chandrasegaran S, Bader JS, Boeke JD (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477:** 471-476

Earl AM, Losick R, Kolter R (2008) Ecology and genomics of Bacillus subtilis. *Trends Microbiol* **16:** 269-275

Echols H, Gingery R (1968) Mutants of bacteriophage lambda defective in vegetative genetic recombination. *J Mol Biol* **34:** 239-249

Eisen JA, Heidelberg JF, White O, Salzberg SL (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* **1:** RESEARCH0011

Elischewski F, Puhler A, Kalinowski J (1999) Pantothenate production in Escherichia coli K12 by enhanced expression of the panE gene encoding ketopantoate reductase. *J Biotechnol* **75:** 135-146

Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98:** 6742-6746

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403:** 335-338

Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* **9:** 685

Enyeart PJ, Ellington AD (2011) Synthetic biology: a yeast for all reasons. *Nature* **477:** 413-414

Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM (2014) Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mob DNA* **5:** 2

Eskes R, Liu L, Ma H, Chao MY, Dickson L, Lambowitz AM, Perlman PS (2000) Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol* **20:** 8432-8446

Eskes R, Yang J, Lambowitz AM, Perlman PS (1997) Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88:** 865-874

Esnault E, Valens M, Espeli O, Boccard F (2007) Chromosome structuring limits genome plasticity in Escherichia coli. *PLoS Genet* **3:** e226

Fabrizio P, Battistella L, Vardavas R, Gattazzo C, Liou LL, Diaspro A, Dossen JW, Gralla EB, Longo VD (2004) Superoxide is a mediator of an altruistic aging program in Saccharomyces cerevisiae. *J Cell Biol* **166:** 1055-1067

Fajardo A, Martinez JL (2008) Antibiotics as signals that trigger specific bacterial responses. *Curr Opin Microbiol* **11:** 161-167

Falkow S, Citarella RV, Wohlhieter JA (1966) The molecular nature of R-factors. *J Mol Biol* **17:** 102-116

Fangman WL, Novick A (1968) Characterization of 2 Bacterial Mutants with Temperature-Sensitive Synthesis of DNA. *Genetics* **60:** 1-&

Fiegna F, Yu YT, Kadam SV, Velicer GJ (2006) Evolution of an obligate social cheater to a superior cooperator. *Nature* **441:** 310-314

Finkel SE, Johnson RC (1992) The Fis protein: it's not just for DNA inversion anymore. *Mol Microbiol* **6:** 3257-3265

Firer-Sherwood MA, Bewley KD, Mock JY, Elliott SJ (2011) Tools for resolving complexity in the electron transfer networks of multiheme cytochromes c. *Metallomics* **3:** 344-348

Fodor K, Alfoldi L (1976) Fusion of protoplasts of Bacillus megaterium. *Proc Natl Acad Sci U S A* **73:** 2147-2150

Foster KR, Bell T (2012) Competition, not cooperation, dominates interactions among culturable microbial species. *Curr Biol* **22:** 1845-1850

Francis MB, Allen CA, Shrestha R, Sorg JA (2013) Bile acid recognition by the Clostridium difficile germinant receptor, CspC, is important for establishing infection. *PLoS Pathog* **9:** e1003356

Frank SA (1998) *Foundations of social evolution*, Princeton, NJ: Princeton University Press.

Frazier CL, San Filippo J, Lambowitz AM, Mills DA (2003) Genetic manipulation of Lactococcus lactis by using targeted group II introns: generation of stable insertions without selection. *Appl Environ Microbiol* **69:** 1121-1128

Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS, Nealson KH, Osterman AL, Pinchuk G, Reed JL, Rodionov DA, Rodrigues JL, Saffarini DA, Serres MH, Spormann AM, Zhulin IB, Tiedje JM (2008) Towards environmental systems biology of Shewanella. *Nat Rev Microbiol* **6:** 592-603

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*

Fukiya S, Mizoguchi H, Mori H (2004) An improved method for deleting large regions of Escherichia coli K-12 chromosome using a combination of Cre/loxP and lambda Red. *FEMS Microbiol Lett* **234:** 325-331

Gaj T, Gersbach CA, Barbas CF, 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31:** 397-405

Gallagher LA, Shendure J, Manoil C (2011) Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq. *MBio* **2:** e00315-00310

Garcia-Rodriguez FM, Barrientos-Duran A, Diaz-Prado V, Fernandez-Lopez M, Toro N (2011) Use of RmInt1, a group IIB intron lacking the intron-encoded protein endonuclease domain, in gene targeting. *Appl Environ Microbiol* **77:** 854-861

Garcia-Russell N, Harmon TG, Le TQ, Amaladas NH, Mathewson RD, Segall AM (2004) Unequal access of chromosomal regions to each other in Salmonella: probing chromosome structure with phage lambda integrase-mediated long-range rearrangements. *Mol Microbiol* **52:** 329-344

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* **403:** 339-342

Gauger EJ, Leatham MP, Mercado-Lubo R, Laux DC, Conway T, Cohen PS (2007) Role of motility and the flhDC Operon in Escherichia coli MG1655 colonization of the mouse intestine. *Infect Immun* **75:** 3315-3324

Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ (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

Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN et al (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **185:** 5673-5684

Gescher JS, Cordova CD, Spormann AM (2008) Dissimilatory iron reduction in Escherichia coli: identification of CymA of Shewanella oneidensis and NapC of E. coli as ferric reductases. *Mol Microbiol* **68:** 706-719

Gibson DG (2011) Enzymatic Assembly of Overlapping DNA Fragments. *Methods Enzymol* **498:** 349-361

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, 3rd, Smith HO (2008) Complete

chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319:** 1215-1220

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329:** 52-56

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6:** 343-U341

Gilbertson L (2003) Cre-lox recombination: Cre-ative tools for plant biotechnology. *Trends Biotechnol* **21:** 550-555

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, 3rd, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* **103:** 425-430

Glazer AN, Nikaido H (2007) *Microbial biotechnology : fundamentals of applied microbiology*, 2nd edn. Cambridge ; New York: Cambridge University Press.

Goldbeck CP, Jensen HM, TerAvest MA, Beedle N, Appling Y, Hepler M, Cambray G, Mutalik V, Angenent LT, Ajo-Franklin CM (2013) Tuning promoter strengths for improved synthesis and function of electron conduits in Escherichia coli. *ACS Synth Biol* **2:** 150-159

Gong J, Zheng H, Wu Z, Chen T, Zhao X (2009) Genome shuffling: Progress and applications for phenotype improvement. *Biotechnol Adv* **27:** 996-1005

Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6:** 279-289

Gore J, Youk H, van Oudenaarden A (2009) Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459:** 253-256

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat Biotechnol* **18:** 97-100

Goryshin IY, Naumann TA, Apodaca J, Reznikoff WS (2003) Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis. *Genome Res* **13:** 644-653

Gray CH, Tatum EL (1944) X-Ray Induced Growth Factor Requirements in Bacteria. *Proc Natl Acad Sci U S A* **30:** 404-410

Green MR, Sambrook J (2012) *Molecular cloning : a laboratory manual*, 4th edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Greig D, Travisano M (2004) The Prisoner's Dilemma and polymorphism in yeast SUC genes. *Proc Biol Sci* **271 Suppl 3:** S25-26

Guarneros G, Echols H (1970) New mutants of bacteriophage lambda with a specific defect in excision from the host chromosome. *J Mol Biol* **47:** 565-574

Guijo MI, Patte J, del Mar Campos M, Louarn JM, Rebollo JE (2001) Localized remodeling of the Escherichia coli chromosome: the patchwork of segments refractory and tolerant to inversion near the replication terminus. *Genetics* **157:** 1413-1423

Guldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24:** 2519-2524

Gullo M, Caggia C, De Vero L, Giudici P (2006) Characterization of acetic acid bacteria in "traditional balsamic vinegar". *Int J Food Microbiol* **106:** 209-212

Guo F, Gopaul DN, van Duyne GD (1997a) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389:** 40-46

Guo H, Karberg M, Long M, Jones JP, 3rd, Sullenger B, Lambowitz AM (2000) Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289:** 452-457

Guo H, Zimmerly S, Perlman PS, Lambowitz AM (1997b) Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J* **16:** 6835-6848

Haberman S, Ellsworth LD (1940) Lethal and Dissociative Effects of X-Rays on Bacteria. *J Bacteriol* **40:** 483-503

Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR (1989) New method for generating deletions and gene replacements in Escherichia coli. *J Bacteriol* **171:** 4617-4622

Hamilton WD (1963) Evolution of Altruistic Behavior. *Am Nat* **97:** 354-&

Hare RS, Walker SS, Dorman TE, Greene JR, Guzman LM, Kenney TJ, Sulavik MC, Baradaran K, Houseweart C, Yu H, Foldes Z, Motzer A, Walbridge M, Shimer GH, Jr., Shaw KJ (2001) Genetic footprinting in bacteria. *J Bacteriol* **183:** 1694-1706

Harris DR, Pollock SV, Wood EA, Goiffon RJ, Klingele AJ, Cabot EL, Schackwitz W, Martin J, Eggington J, Durfee TJ, Middle CM, Norton JE, Popelars MC, Li H, Klugman SA, Hamilton LL, Bane LB, Pennacchio LA, Albert TJ, Perna NT et al

(2009) Directed evolution of ionizing radiation resistance in Escherichia coli. *J Bacteriol* **191:** 5240-5252

Hartshorne RS, Reardon CL, Ross D, Nuester J, Clarke TA, Gates AJ, Mills PC, Fredrickson JK, Zachara JM, Shi L, Beliaev AS, Marshall MJ, Tien M, Brantley S, Butt JN, Richardson DJ (2009) Characterization of an electron conduit between bacteria and the extracellular environment. *Proc Natl Acad Sci U S A* **106:** 22169-22174

Hayes W (1952) Recombination in Bact. coli K 12; unidirectional transfer of genetic material. *Nature* **169:** 118-119

Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP (2007) The ClosTron: a universal gene knock-out system for the genus Clostridium. *J Microbiol Methods* **70:** 452-464

Hefferin ML, Tomkinson AE (2005) Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair (Amst)* **4:** 639-648

Hemm MR, Paul BJ, Schneider TD, Storz G, Rudd KE (2008) Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol Microbiol* **70:** 1487-1501

Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269:** 400-403

Hershey AD (1971) *The Bacteriophage lambda*, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36:** 39-56

Hida H, Yamada T, Yamada Y (2007) Genome shuffling of Streptomyces sp. U121 for improved production of hydroxycitric acid. *Appl Microbiol Biotechnol* **73:** 1387-1393

Hidaka M, Akiyama M, Horiuchi T (1988) A consensus sequence of three DNA replication terminus sites on the E. coli chromosome is highly homologous to the terR sites of the R6K plasmid. *Cell* **55:** 467-475

Higgins D, Dworkin J (2012) Recent progress in Bacillus subtilis sporulation. *FEMS Microbiol Rev* **36:** 131-148

Hill CW, Gray JA (1988) Effects of chromosomal inversion on cell fitness in Escherichia coli K-12. *Genetics* **119:** 771-778

Hill TM, Pelletier AJ, Tecklenburg ML, Kuempel PL (1988) Identification of the DNA sequence from the E. coli terminus region that halts replication forks. *Cell* **55:** 459-466

Hirota Y, Mordoh J, Jacob F (1970) Process of Cellular Division in Escherichia-Coli .3. Thermosensitive Mutants of Escherichia-Coli Altered in Process of DNA Initiation. *J Mol Biol* **53:** 369-&

Hobbs EC, Astarita JL, Storz G (2010) Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in Escherichia coli: analysis of a bar-coded mutant collection. *J Bacteriol* **192:** 59-67

Hoess RH, Abremski K (1985) Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J Mol Biol* **181:** 351-362

Hoess RH, Wierzbicki A, Abremski K (1986) The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res* **14:** 2287-2300

Hsia CH, Shen MC, Lin JS, Wen YK, Hwang KL, Cham TM, Yang NC (2009) Nattokinase decreases plasma levels of fibrinogen, factor VII, and factor VIII in human subjects. *Nutr Res* **29:** 190-196

Hsu JC (1994) *Multiple Comparisons: Theory and Methods*: Chapman & Hall.

Hu B, Du J, Zou RY, Yuan YJ (2010) An environment-sensitive synthetic microbial ecosystem. *PLoS One* **5:** e10619

Huang LC, Wood EA, Cox MM (1991) A Bacterial Model System for Chromosomal Targeting. *Nucleic Acids Research* **19:** 443-448

Hughes KT, Maloy SR (2007) *Advanced bacterial genetics : use of transposons and phage for genomic engineering*, San Diego, Calif.: Academic Press.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31:** 227-229

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333:** 348-353

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22:** 841-847

Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc Natl Acad Sci U S A* **102:** 15971-15976

Jaacks KJ, Healy J, Losick R, Grossman AD (1989) Identification and characterization of genes controlled by the sporulation-regulatory gene spo0H in Bacillus subtilis. *J Bacteriol* **171:** 4121-4129

Jackson DA, Symons RH, Berg P (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc Natl Acad Sci U S A* **69:** 2904-2909

Jacob F, Monod J (1961) Genetic Regulatory Mechanisms in Synthesis of Proteins. *J Mol Biol* **3:** 318-&

Jacobson JW, Medhora MM, Hartl DL (1986) Molecular structure of a somatically unstable transposable element in Drosophila. *Proc Natl Acad Sci U S A* **83:** 8684-8688

Jayaram M (1985) Two-micrometer circle site-specific recombination: the minimal substrate and the possible role of flanking sequences. *Proc Natl Acad Sci U S A* **82:** 5875-5879

Jensen HM, Albers AE, Malley KR, Londer YY, Cohen BE, Helms BA, Weigele P, Groves JT, Ajo-Franklin CM (2010) Engineering of a synthetic electron conduit in living cells. *Proc Natl Acad Sci U S A* **107:** 19213-19218

Jensen KF (1993) The Escherichia coli K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *J Bacteriol* **175:** 3401-3407

Jia B, Yang JK, Liu WS, Li X, Yan YJ (2010) Homologous overexpression of a lipase from Burkholderia cepacia using the lambda Red recombinase system. *Biotechnol Lett* **32:** 521-526

Jia K, Zhu Y, Zhang Y, Li Y (2011) Group II intron-anchored gene deletion in Clostridium. *PLoS One* **6:** e16693

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31:** 233-239

Jimenez-Zurdo JI, Garcia-Rodriguez FM, Barrientos-Duran A, Toro N (2003) DNA target site requirements for homing in vivo of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. *J Mol Biol* **326:** 413-423

John RP, Gangadharan D, Madhavan Nampoothiri K (2008) Genome shuffling of Lactobacillus delbrueckii mutant and Bacillus amyloliquefaciens through protoplasmic fusion for L-lactic acid production from starchy wastes. *Bioresour Technol* **99:** 8008-8015

Johnson CM, Fisher DJ (2013) Site-specific, insertional inactivation of incA in Chlamydia trachomatis using a group II intron. *PLoS One* **8:** e83989

Johnson ET, Baron DB, Naranjo B, Bond DR, Schmidt-Dannert C, Gralnick JA (2010) Enhancement of survival and electricity production in an engineered bacterium by light-driven proton pumping. *Appl Environ Microbiol* **76:** 4123-4129

Jordan E, Saedler H, Starlinger P (1967) Strong-polar mutations in the transferase gene of the galactose operon in E.coli. *Mol Gen Genet* **100:** 296-306

Jordan E, Saedler H, Starlinger P (1968) O0 and strong-polar mutations in the gal operon are insertions. *Mol Gen Genet* **102:** 353-363

Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S (2006) Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* **188:** 8259-8271

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

Kaern M, Blake WJ, Collins JJ (2003) The engineering of gene regulatory networks. *Annu Rev Biomed Eng* **5:** 179-206

Karas BJ, Jablanovic J, Sun L, Ma L, Goldgof GM, Stam J, Ramon A, Manary MJ, Winzeler EA, Venter JC, Weyman PD, Gibson DG, Glass JI, Hutchison CA, 3rd, Smith HO, Suzuki Y (2013) Direct transfer of whole genomes from bacteria to yeast. *Nat Methods* **10:** 410-412

Karatza P, Frillingos S (2005) Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in Escherichia coli. *Mol Membr Biol* **22:** 251-261

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol* **19:** 1162-1167

Karimova M, Abi-Ghanem J, Berger N, Surendranath V, Pisabarro MT, Buchholz F (2013) Vika/vox, a novel efficient and specific Cre/loxP-like site-specific recombination system. *Nucleic Acids Res* **41:** e37

Karu AE, Sakaki Y, Echols H, Linn S (1975) The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of Escherichia coli. *J Biol Chem* **250:** 7377-7387

Katashkina JI, Hara Y, Golubeva LI, Andreeva IG, Kuvaeva TM, Mashko SV (2009) Use of the lambda Red-recombineering method for genetic engineering of Pantoea ananatis. *BMC Mol Biol* **10:** 34

Keasling JD (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol* **3:** 64-76

Kelly TJ, Jr., Smith HO (1970) A restriction enzyme from Hemophilus influenzae. II. Base sequence of the recognition site. *J Mol Biol* **51:** 393-409

Kelly WJ, Ward LJ, Leahy SC (2010) Chromosomal diversity in Lactococcus lactis and the origin of dairy starter cultures. *Genome Biol Evol* **2:** 729-744

Keravala A, Lee S, Thyagarajan B, Olivares EC, Gabrovsky VE, Woodard LE, Calos MP (2009) Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. *Mol Ther* **17:** 112-120

Kerner A, Park J, Williams A, Lin XN (2012) A programmable Escherichia coli consortium via tunable symbiosis. *PLoS One* **7:** e34032

Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. *Nat Rev Genet* **11:** 367-379

Kim JY, Gum SN, Paik JK, Lim HH, Kim KC, Ogasawara K, Inoue K, Park S, Jang Y, Lee JH (2008) Effects of nattokinase on blood pressure: a randomized, controlled trial. *Hypertens Res* **31:** 1583-1588

Kittelmann M, Stamm WW, Follmann H, Truper HG (1989) Isolation and Classification of Acetic-Acid Bacteria from High Percentage Vinegar Fermentations. *Appl Microbiol Biot* **30:** 47-52

Kleckner N (1981) Transposable elements in prokaryotes. *Annu Rev Genet* **15:** 341-404

Kleckner N, Roth J, Botstein D (1977) Genetic engineering in vivo using translocatable drug-resistance elements. New methods in bacterial genetics. *J Mol Biol* **116:** 125-159

Kmiec E, Holloman WK (1981) Beta protein of bacteriophage lambda promotes renaturation of DNA. *J Biol Chem* **256:** 12636-12639

Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci U S A* **101:** 15573-15578

Kok J, van der Vossen JM, Venema G (1984) Construction of plasmid cloning vectors for lactic streptococci which also replicate in Bacillus subtilis and Escherichia coli. *Appl Environ Microbiol* **48:** 726-731

Kolisnychenko V, Plunkett G, 3rd, Herring CD, Feher T, Posfai J, Blattner FR, Posfai G (2002) Engineering a reduced Escherichia coli genome. *Genome Res* **12:** 640-647

Kolter R, Inuzuka M, Helinski DR (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* **15:** 1199-1208

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

Kubo I, Hosoda K, Suzuki S, Yamamoto K, Kihara K, Mori K, Yomo T (2013) Construction of bacteria-eukaryote synthetic mutualism. *Biosystems* **113:** 66-71

Kumar S, Smith KP, Floyd JL, Varela MF (2011) Cloning and molecular analysis of a mannitol operon of phosphoenolpyruvate-dependent phosphotransferase (PTS) type from Vibrio cholerae O395. *Arch Microbiol* **193:** 201-208

Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. *Science* **342:** 357-360

Lambert JM, Bongers RS, Kleerebezem M (2007) Cre-lox-based system for multiple gene deletions and selectable-marker removal in Lactobacillus plantarum. *Appl Environ Microbiol* **73:** 1126-1135

Lambertsen L, Sternberg C, Molin S (2004) Mini-Tn7 transposons for site-specific tagging of bacteria with fluorescent proteins. *Environ Microbiol* **6:** 726-732

Lambowitz AM, Zimmerly S (2004) Mobile group II introns. *Annu Rev Genet* **38:** 1-35

Lambowitz AM, Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* **3:** a003616

Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci U S A* **96:** 11428-11433

Lampe DJ, Churchill ME, Robertson HM (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J* **15:** 5470-5479

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K et al (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921

Langer SJ, Ghafoori AP, Byrd M, Leinwand L (2002) A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res* **30:** 3067-3077

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* **19:** 2308-2316

Larson JD, Jenkins JL, Schuermann JP, Zhou Y, Becker DF, Tanner JJ (2006) Crystal structures of the DNA-binding domain of Escherichia coli proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. *Protein Sci* **15:** 2630-2641

Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. *Science* **317:** 632-638

Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, Ma L, Noskov VN, Denisova EA, Gibson DG, Assad-Garcia N, Alperovich N, Thomas DW, Merryman C, Hutchison CA, 3rd, Smith HO, Venter JC, Glass JI (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325:** 1693-1696

Lathe WC, 3rd, Snel B, Bork P (2000) Gene context conservation of a higher order than operons. *Trends Biochem Sci* **25:** 474-479

Lederberg EM (1950) Lysogenicity in Escherichia coli strain K-12. *Micriobial Genetics Bulletin* **1:** 5-8

Lederberg EM (1951) Lysogenicity in E. coli K-12. *Genetics* **36:** 560-560

Lederberg J, Lederberg EM, Zinder ND, Lively ER (1951) Recombination analysis of bacterial heredity. *Cold Spring Harb Symp Quant Biol* **16:** 413-443

Lederberg J, Tatum EL (1946a) Gene recombination in Escherichia coli. *Nature* **158:** 558

Lederberg J, Tatum EL (1946b) Novel Genotypes in Mixed Cultures of Biochemical Mutants of Bacteria. *Cold Spring Harb Sym* **11:** 113-114

Lee YK (2006) *Microbial biotechnology : principles and applications*, 2nd edn. Toh Tuck Link, Singapore ; Hackensak, N.J.: World Scientific.

Leenhouts K, Buist G, Bolhuis A, ten Berge A, Kiel J, Mierau I, Dabrowska M, Venema G, Kok J (1996a) A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol Gen Genet* **253:** 217-224

Leenhouts K, Buist G, Bolhuis A, ten Berge A, Kiel J, Mierau I, Dabrowska M, Venema G, Kok J (1996b) A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol Gen Genet* **253:** 217-224

Lehninger AL, Nelson DL, Cox MM (2013) *Lehninger principles of biochemistry*, 6th edn. New York: W.H. Freeman.

Leibig M, Krismer B, Kolb M, Friede A, Gotz F, Bertram R (2008) Marker removal in staphylococci via Cre recombinase and different lox sites. *Appl Environ Microbiol* **74:** 1316-1323

Lenski RE (2011) Evolution in Action: a 50,000-Generation Salute to Charles Darwin. *Microbe* **6:** 30-33

Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-Term Experimental Evolution in Escherichia-Coli .1. Adaptation and Divergence during 2,000 Generations. *Am Nat* **138:** 1315-1341

Lenski RE, Travisano M (1994) Dynamics of Adaptation and Diversification - a 10,000-Generation Experiment with Bacterial-Populations. *P Natl Acad Sci USA* **91:** 6808-6814

Lenz AP, Williamson KS, Pitts B, Stewart PS, Franklin MJ (2008) Localized gene expression in Pseudomonas aeruginosa biofilms. *Appl Environ Microbiol* **74:** 4463-4471

Lesic B, Rahme LG (2008) Use of the lambda Red recombinase system to rapidly generate mutants in Pseudomonas aeruginosa. *BMC Mol Biol* **9:** 20

Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271:** 1247-1254

Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J* **40:** 500-511

Li YH, Tian X (2012) Quorum sensing and bacterial social interactions in biofilms. *Sensors (Basel)* **12:** 2519-2538

Linares JF, Gustafsson I, Baquero F, Martinez JL (2006) Antibiotics as intermicrobial signaling agents instead of weapons. *Proc Natl Acad Sci U S A* **103:** 19484-19489

Lincoln RE, Gowen JW (1942) Mutation of Phytomonas Stewartii by X-Ray Irradiation. *Genetics* **27:** 441-462

Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. *J Bacteriol* **179:** 6228-6237

Liu GR, Liu WQ, Johnston RN, Sanderson KE, Li SX, Liu SL (2006) Genome plasticity and ori-ter rebalancing in Salmonella typhi. *Mol Biol Evol* **23:** 365-371

Liu Q, Li MZ, Leibham D, Cortez D, Elledge SJ (1998) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr Biol* **8:** 1300-1309

Liu SL, Sanderson KE (1996) Highly plastic chromosomal organization in Salmonella typhi. *Proc Natl Acad Sci U S A* **93:** 10303-10308

Liu X, Matsumura P (1994) The FlhD/FlhC complex, a transcriptional activator of the Escherichia coli flagellar class II operons. *J Bacteriol* **176:** 7345-7351

Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25:** 674-681

Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450:** 56-62

Logan BE (2010) Scaling up microbial fuel cells and other bioelectrochemical systems. *Appl Microbiol Biotechnol* **85:** 1665-1671

Lomovskaya O, Lewis K, Matin A (1995) EmrR is a negative regulator of the Escherichia coli multidrug resistance pump EmrAB. *J Bacteriol* **177:** 2328-2334

Louarn JM, Bouche JP, Legendre F, Louarn J, Patte J (1985) Characterization and properties of very large inversions of the E. coli chromosome along the origin-to-terminus axis. *Mol Gen Genet* **201:** 467-476

Lovley DR, Holmes DE, Nevin KP (2004) Dissimilatory Fe(III) and Mn(IV) reduction. *Adv Microb Physiol* **49:** 219-286

Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. *Nat Biotechnol* **27:** 1139-1150

Luchansky JB, Benson AK, Atherly AG (1989) Construction, transfer and properties of a novel temperature-sensitive integrable plasmid for genomic analysis of Staphylococcus aureus. *Mol Microbiol* **3:** 65-78

Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc Natl Acad Sci U S A* **108:** 8617-8622

Lynch MD, Gill RT (2006) Broad host range vectors for stable genomic library construction. *Biotechnol Bioeng* **94:** 151-158

Mah TF (2012) Biofilm-specific antibiotic resistance. *Future Microbiol* **7:** 1061-1072

Malamy MH (1966) Frameshift mutations in the lactose operon of E. coli. *Cold Spring Harb Symp Quant Biol* **31:** 189-201

Malamy MH (1970) Some properties of insertion mutations in the lac operon. In *The Lactose operon*, Beckwith JR, Zisper D (eds), pp 359-373. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory

Malhotra M, Srivastava S (2008) An ipdC gene knock-out of Azospirillum brasilense strain SM and its implications on indole-3-acetic acid biosynthesis and plant growth promotion. *Antonie Van Leeuwenhoek* **93:** 425-433

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339:** 823-826

Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning : a laboratory manual*, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Marchand N, Collins CH (2013) Peptide-based communication system enables Escherichia coli to Bacillus megaterium interspecies signaling. *Biotechnol Bioeng*

Marcia M, Somarowthu S, Pyle AM (2013) Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA* **4:** 14

McClintock B (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36:** 344-355

McMurry JL, Murphy JW, Gonzalez-Pedrajo B (2006) The FliN-FliH interaction mediates localization of flagellar export ATPase FliI to the C ring complex. *Biochemistry* **45:** 11790-11798

Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM (1982) Physical and genetic characterization of symbiotic and auxotrophic mutants of Rhizobium meliloti induced by transposon Tn5 mutagenesis. *J Bacteriol* **149:** 114-122

Mehdi Q, Hockaday TD, Newlands E, el-Kabir DJ (1971) LATS actions of bacterial permeability and protein synthesis. *Proc R Soc Med* **64:** 1268-1269

Mehdiabadi NJ, Jack CN, Farnham TT, Platt TG, Kalla SE, Shaulsky G, Queller DC, Strassmann JE (2006) Social evolution: kin preference in a social microbe. *Nature* **442:** 881-882

Michel F, Ferat JL (1995) Structure and activities of group II introns. *Annu Rev Biochem* **64:** 435-461

Miller JH (1991) *Bacterial genetic systems*, San Diego: Academic Press.

Miller JH (1992) *A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria*, Plainview, N.Y.: Cold Spring Harbor Laboratory Press.

Miller VL, Mekalanos JJ (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

Minorikawa S, Nakayama M (2011) Recombinase-mediated cassette exchange (RMCE) and BAC engineering via VCre/VloxP and SCre/SloxP systems. *Biotechniques* **50:** 235-246

Mohr G, Hong W, Zhang J, Cui GZ, Yang Y, Cui Q, Liu YJ, Lambowitz AM (2013) A targetron system for gene targeting in thermophiles and its application in Clostridium thermocellum. *PLoS One* **8:** e69032

Mohr G, Smith D, Belfort M, Lambowitz AM (2000) Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev* **14:** 559-573

Mondragon-Palomino O, Danino T, Selimkhanov J, Tsimring L, Hasty J (2011) Entrainment of a population of synthetic genetic oscillators. *Science* **333:** 1315-1319

Monod J (1942) *Recherches sur la croissance des cultures bactériennes*, Paris,: Hermann & cie.

Monod J (1949) The Growth of Bacterial Cultures. *Annual Review of Microbiology* **3:** 371-394

Moreau MJ (2013) DNA replication in Escherichia coli: A comprehensive study of the Tus-Ter complex. PhD Thesis, School of Pharmacy and Molecular Sciences, James Cook University,

Moreau MJ, Schaeffer PM (2012a) Differential Tus-Ter binding and lock formation: implications for DNA replication termination in Escherichia coli. *Mol Biosyst* **8:** 2783-2791

Moreau MJ, Schaeffer PM (2012b) A polyplex qPCR-based binding assay for protein-DNA interactions. *Analyst* **137:** 4111-4113

Morse ML, Lederberg EM, Lederberg J (1956) Transduction in Escherichia Coli K-12. *Genetics* **41:** 142-156

Mortelmans K, Zeiger E (2000) The Ames Salmonella/microsome mutagenicity assay. *Mutat Res* **455:** 29-60

Muller HJ (1927) Artificial Transmutation of the Gene. *Science* **66:** 84-87

Muller HJ (1928) The Production of Mutations by X-Rays. *Proc Natl Acad Sci U S A* **14:** 714-726

Muniyappa K, Radding CM (1986) The homologous recombination system of phage lambda. Pairing activities of beta protein. *J Biol Chem* **261:** 7472-7478

Murli S, Kennedy J, Dayem LC, Carney JR, Kealey JT (2003) Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production. *J Ind Microbiol Biotechnol* **30:** 500-509

Muro-Pastor AM, Hess WR (2012) Heterocyst differentiation: from single mutants to global approaches. *Trends Microbiol* **20:** 548-557

Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol* **180:** 2063-2071

Murphy KC (2012) Phage recombinases and their applications. *Adv Virus Res* **83:** 367-414

Myers CR, Myers JM (2002) MtrB is required for proper incorporation of the cytochromes OmcA and OmcB into the outer membrane of Shewanella putrefaciens MR-1. *Appl Environ Microbiol* **68:** 5585-5594

Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat Biotechnol* **31:** 170-174

Nagami Y, Tanaka T (1986) Molecular cloning and nucleotide sequence of a DNA fragment from Bacillus natto that enhances production of extracellular proteases and levansucrase in Bacillus subtilis. *J Bacteriol* **166:** 20-28

Nagy A (2000) Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26:** 99-109

Nahum JR, Harding BN, Kerr B (2011) Evolution of restraint in a structured rock-paper-scissors community. *Proc Natl Acad Sci U S A* **108 Suppl 2:** 10831-10838

Nakashima N, Miyazaki K (2014) Bacterial cellular engineering by genome editing and gene silencing. *Int J Mol Sci* **15:** 2773-2793

Nakayama T (1959) Studies on Acetic Acid-Bacteria .1. Biochemical Studies on Ethanol Oxidation. *J Biochem-Tokyo* **46:** 1217-1225

Nandy S, Srivastava V (2011) Site-specific gene integration in rice genome mediated by the FLP-FRT recombination system. *Plant Biotechnology Journal* **9:** 713-721

Nealson KH, Belz A, McKee B (2002) Breathing metals as a way of life: geobiology in action. *Antonie Van Leeuwenhoek* **81:** 215-222

Neidhardt FC, Curtiss R (1996) *Escherichia coli and Salmonella : cellular and molecular biology*, 2nd edn. Washington, D.C.: ASM Press.

Noah JW, Park S, Whitt JT, Perutka J, Frey W, Lambowitz AM (2006) Atomic force microscopy reveals DNA bending during group II intron ribonucleoprotein particle integration into double-stranded DNA. *Biochemistry* **45:** 12424-12435

Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120:** 621-623

Oehler S, Eismann ER, Kramer H, Muller-Hill B (1990) The three operators of the lac operon cooperate in repression. *EMBO J* **9:** 973-979

Otto M (2012) MRSA virulence and spread. *Cell Microbiol* **14:** 1513-1521

Page RE, Jr., Erber J (2002) Levels of behavioral organization and the evolution of division of labor. *Naturwissenschaften* **89:** 91-106

Pais E, Alexy T, Holsworth RE, Jr., Meiselman HJ (2006) Effects of nattokinase, a pro-fibrinolytic enzyme, on red blood cell aggregation and whole blood viscosity. *Clin Hemorheol Microcirc* **35:** 139-142

Palonen E, Lindstrom M, Karttunen R, Somervuo P, Korkeala H (2011) Expression of signal transduction system encoding genes of Yersinia pseudotuberculosis IP32953 at 28 degrees C and 3 degrees C. *PLoS One* **6:** e25063

Park JM, Jang YS, Kim TY, Lee SY (2010) Development of a gene knockout system for Ralstonia eutropha H16 based on the broad-host-range vector expressing a mobile group II intron. *FEMS Microbiol Lett* **309:** 193-200

Pasteur L (1866) *Études sur le vin, ses maladies, causes qui les provoquent, procédés nouveaux pour le conserver et pour le vieillir*, Paris,: Impr. impériale.

Pasteur L (1876) *Études sur la bière, ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable, avec une théorie nouvelle de la fermentation*, Paris,: Gauthier-Villars.

Patnaik R, Louie S, Gavrilovic V, Perry K, Stemmer WP, Ryan CM, del Cardayre S (2002) Genome shuffling of Lactobacillus for improved acid tolerance. *Nat Biotechnol* **20:** 707-712

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

Pearson MM, Mobley HL (2007) The type III secretion system of Proteus mirabilis HI4320 does not contribute to virulence in the mouse model of ascending urinary tract infection. *J Med Microbiol* **56:** 1277-1283

Peberdy JF (1980) Protoplast fusion—a tool for genetic manipulation and breeding in industrial microorganisms. *Enzyme Microb Tech* **2:** 23-29

Pelosi L, Kuhn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D (2006) Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. *Genetics* **173:** 1851-1869

Perutka J, Wang W, Goerlitz D, Lambowitz AM (2004) Use of computer-designed group II introns to disrupt Escherichia coli DExH/D-box protein and DNA helicase genes. *J Mol Biol* **336:** 421-439

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7:** e37135

Pitts KE, Dobbin PS, Reyes-Ramirez F, Thomson AJ, Richardson DJ, Seward HE (2003) Characterization of the Shewanella oneidensis MR-1 decaheme cytochrome MtrA: expression in Escherichia coli confers the ability to reduce soluble Fe(III) chelates. *J Biol Chem* **278:** 27758-27765

Plante I, Cousineau B (2006) Restriction for gene insertion within the Lactococcus lactis Ll.LtrB group II intron. *RNA* **12:** 1980-1992

Platt T (1981) Termination of transcription and its regulation in the tryptophan operon of E. coli. *Cell* **24:** 10-23

Pomerantsev AP, Sitaraman R, Galloway CR, Kivovich V, Leppla SH (2006) Genome engineering in Bacillus anthracis using Cre recombinase. *Infect Immun* **74:** 682-693

Posfai G, Plunkett G, 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome Escherichia coli. *Science* **312:** 1044-1046

Postma PW, Lengeler JW, Jacobson GR (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57:** 543-594

Preston A (2003) Choosing a cloning vector. *Methods Mol Biol* **235:** 19-26

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Pyle AM, Lambowitz AM (2006) Group II introns: ribozymes that splice RNA and invade DNA. In *The RNA World, 3rd edition*, Gesteland RF, Cech T, Atkins JF (eds), pp 469-505. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press

Qin PZ, Pyle AM (1998) The architectural organization and mechanistic function of group II intron structural elements. *Curr Opin Struct Biol* **8:** 301-308

Quandt EM, Deatherage DE, Ellington AD, Georgiou G, Barrick JE (2013) Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in Escherichia coli. *Proc Natl Acad Sci U S A*

Quandt EM, Deatherage DE, Ellington AD, Georgiou G, Barrick JE (2014) Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in Escherichia coli. *Proc Natl Acad Sci U S A* **111:** 2217-2222

Radding CM (1970) The role of exonuclease and beta protein of bacteriophage lambda in genetic recombination. I. Effects of red mutants on protein structure. *J Mol Biol* **52:** 491-499

Rainey PB, Rainey K (2003) Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425:** 72-74

Rawsthorne H, Turner KN, Mills DA (2006) Multicopy integration of heterologous genes, using the lactococcal group II intron targeted to bacterial insertion sequences. *Appl Environ Microbiol* **72:** 6088-6093

Raymond CS, Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One* **2:** e162

Rebollo JE, Francois V, Louarn JM (1988) Detection and possible role of two large nondivisible zones on the Escherichia coli chromosome. *Proc Natl Acad Sci U S A* **85:** 9391-9395

Rest JS, Mindell DP (2003) Retroids in archaea: phylogeny and lateral origins. *Mol Biol Evol* **20:** 1134-1142

Ricardo D (1817) *On the principles of political economy and taxation*, London,: J. Murray.

Ried JL, Collmer A (1987) An nptI-sacB-sacR cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene* **57:** 239-246

Rocha EP (2008) The organization of the bacterial genome. *Annu Rev Genet* **42:** 211-233

Rodriguez E, Hu Z, Ou S, Volchegursky Y, Hutchinson CR, McDaniel R (2003) Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. *J Ind Microbiol Biotechnol* **30:** 480-488

Rodriguez SA, Yu JJ, Davis G, Arulanandam BP, Klose KE (2008) Targeted inactivation of francisella tularensis genes by group II introns. *Appl Environ Microbiol* **74:** 2619-2626

Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI, Tatusov RL, Szekely LA, Koonin EV (2002) Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res* **30:** 2212-2223

Rosenberg AH, Lade BN, Chui DS, Lin SW, Dunn JJ, Studier FW (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56:** 125-135

Rosner B (2011) *Fundamentals of biostatistics*, 7th edn. Boston: Brooks/Cole, Cengage Learning.

Ross DE, Flynn JM, Baron DB, Gralnick JA, Bond DR (2011) Towards electrosynthesis in shewanella: energetics of reversing the mtr pathway for reductive metabolism. *PLoS One* **6:** e16649

Ross DE, Ruebush SS, Brantley SL, Hartshorne RS, Clarke TA, Richardson DJ, Tien M (2007) Characterization of protein-protein interactions involved in iron reduction by Shewanella oneidensis MR-1. *Appl Environ Microbiol* **73:** 5797-5808

Rossi MS, Paquelin A, Ghigo JM, Wandersman C (2003) Haemophore-mediated signal transduction across the bacterial cell envelope in Serratia marcescens: the inducer and the transported substrate are different molecules. *Mol Microbiol* **48:** 1467-1480

Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ (1999) In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc Natl Acad Sci U S A* **96:** 1645-1650

Rufer AW, Sauer B (2002) Non-contact positions impose site selectivity on Cre recombinase. *Nucleic Acids Res* **30:** 2764-2771

Ruvkun GB, Ausubel FM (1981) A general method for site-directed mutagenesis in prokaryotes. *Nature* **289:** 85-88

Saito H, Inoue T (2009) Synthetic biology with RNA motifs. *Int J Biochem Cell Biol* **41:** 398-404

Salipante SJ, Barlow M, Hall BG (2003) GeneHunter, a transposon tool for identification and isolation of cryptic antibiotic resistance genes. *Antimicrob Agents Chemother* **47:** 3840-3845

Sanchez A, Gore J (2013) Feedback between population and evolutionary dynamics determines the fate of social microbial populations. *PLoS Biol* **11:** e1001547

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274:** 765-768

Santoro SW, Schultz PG (2002) Directed evolution of the site specificity of Cre recombinase. *Proc Natl Acad Sci U S A* **99:** 4185-4190

Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science* **316:** 1912-1915

Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A* **98:** 12712-12717

Sauer B (1996) Multiplex Cre/lox recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. *Nucleic Acids Res* **24:** 4608-4613

Sauer B, McDermott J (2004) DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. *Nucleic Acids Res* **32:** 6086-6095

Savage DF, Way J, Silver PA (2008) Defossiling fuel: how synthetic biology can transform biofuel production. *ACS Chem Biol* **3:** 13-16

Sayeed S, Uzal FA, Fisher DJ, Saputo J, Vidal JE, Chen Y, Gupta P, Rood JI, McClane BA (2008) Beta toxin is essential for the intestinal virulence of Clostridium perfringens type C disease isolate CN3685 in a rabbit ileal loop model. *Mol Microbiol* **67:** 15-30

Scapin G, Grubmeyer C, Sacchettini JC (1994) Crystal structure of orotate phosphoribosyltransferase. *Biochemistry* **33:** 1287-1294

Scapin G, Ozturk DH, Grubmeyer C, Sacchettini JC (1995) The crystal structure of the orotate phosphoribosyltransferase complexed with orotate and alpha-D-5-phosphoribosyl-1-pyrophosphate. *Biochemistry* **34:** 10744-10754

Schaeffer P, Cami B, Hotchkiss RD (1976) Fusion of bacterial protoplasts. *Proc Natl Acad Sci U S A* **73:** 2151-2155

Schellenberg GD, Furlong CE (1977) Resolution of the multiplicity of the glutamate and aspartate transport systems of Escherichia coli. *J Biol Chem* **252:** 9055-9064

Schuetz B, Schicklberger M, Kuermann J, Spormann AM, Gescher J (2009) Periplasmic electron transfer via the c-type cytochromes MtrA and FccA of Shewanella oneidensis MR-1. *Appl Environ Microbiol* **75:** 7789-7796

Scolari VF, Bassetti B, Sclavi B, Lagomarsino MC (2011) Gene clusters reflecting macrodomain structure respond to nucleoid perturbations. *Mol Biosyst* **7:** 878-888

Segall A, Mahan MJ, Roth JR (1988) Rearrangement of the bacterial chromosome: forbidden inversions. *Science* **241:** 1314-1318

Selby CP, Sancar A (1993) Molecular mechanism of transcription-repair coupling. *Science* **260:** 53-58

Senecoff JF, Bruckner RC, Cox MM (1985) The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proc Natl Acad Sci U S A* **82:** 7270-7274

Senecoff JF, Rossmeissl PJ, Cox MM (1988) DNA recognition by the FLP recombinase of the yeast 2 mu plasmid. A mutational analysis of the FLP binding site. *J Mol Biol* **201:** 405-421

Shao L, Hu S, Yang Y, Gu Y, Chen J, Jiang W, Yang S (2007) Targeted gene disruption by use of a group II intron (targetron) vector in Clostridium acetobutylicum. *Cell Res* **17:** 963-965

Shapiro JA (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci U S A* **76:** 1933-1937

Shapiro JA (1998) Thinking about bacterial populations as multicellular organisms. *Annu Rev Microbiol* **52:** 81-104

Shapiro SS, Wilk MB (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52:** 591-&

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309:** 1728-1732

Shi L, Chen B, Wang Z, Elias DA, Mayer MU, Gorby YA, Ni S, Lower BH, Kennedy DW, Wunschel DS, Mottaz HM, Marshall MJ, Hill EA, Beliaev AS, Zachara JM, Fredrickson JK, Squier TC (2006) Isolation of a high-affinity functional protein complex between OmcA and MtrC: Two outer membrane decaheme c-type cytochromes of Shewanella oneidensis MR-1. *J Bacteriol* **188:** 4705-4714

Shi L, Rosso KM, Clarke TA, Richardson DJ, Zachara JM, Fredrickson JK (2012) Molecular Underpinnings of Fe(III) Oxide Reduction by Shewanella Oneidensis MR-1. *Front Microbiol* **3:** 50

Shimada T, Fujita N, Yamamoto K, Ishihama A (2011) Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS One* **6:** e20081

Shong J, Jimenez Diaz MR, Collins CH (2012) Towards synthetic microbial consortia for bioprocessing. *Curr Opin Biotechnol* **23:** 798-802

Short JM, Fernandez JM, Sorge JA, Huse WD (1988) Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res* **16:** 7583-7600

Shou W, Ram S, Vilar JM (2007) Synthetic cooperation in engineered yeast populations. *Proc Natl Acad Sci U S A* **104:** 1877-1882

Shulman MJ, Hallick LM, Echols H, Signer ER (1970) Properties of recombination-deficient mutants of bacteriophage lambda. *J Mol Biol* **52:** 501-520

Siegal ML, Hartl DL (2000) Application of Cre/loxP in Drosophila. Site-specific recombination and transgene coplacement. *Methods Mol Biol* **136:** 487-495

Siegel RW, Jain R, Bradbury A (2001) Using an in vivo phagemid system to identify non-compatible loxP sequences. *FEBS Lett* **505:** 467-473

Sievers M, Sellmer S, Teuber M (1992) Acetobacter-Europaeus Sp-Nov a Main Component of Industrial Vinegar Fermenters in Central-Europe. *Syst Appl Microbiol* **15:** 386-392

Signer ER, Weil J (1968) Recombination in bacteriophage lambda. I. Mutants deficient in general recombination. *J Mol Biol* **34:** 261-271

Singh NN, Lambowitz AM (2001) Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J Mol Biol* **309:** 361-386

Smith C, Song H, You L (2008) Signal discrimination by differential regulation of protein stability in quorum sensing. *J Mol Biol* **382:** 1290-1297

Smith CL, Weiss BL, Aksoy S, Runyen-Janecky LJ (2013) Characterization of the achromobactin iron acquisition operon in Sodalis glossinidius. *Appl Environ Microbiol* **79:** 2872-2881

Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M (2005) Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev* **19:** 2477-2487

Smith HO, Wilcox KW (1970) A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *J Mol Biol* **51:** 379-391

Smith JM (1964) Group selection and kin Selection. *Nature* **201:** 1145-1147

Smith KM, Liao JC (2011) An evolutionary strategy for isobutanol production strain development in Escherichia coli. *Metab Eng* **13:** 674-681

Sokollek SJ, Hertel C, Hammes WP (1998) Description of Acetobacter oboediens sp. nov. and Acetobacter pomorum sp. nov., two new species isolated from industrial vinegar fermentations. *Int J Syst Bacteriol* **48:** 935-940

Srivastava V, Gidoni D (2010) Site-specific gene integration technologies for crop improvement. *In Vitro Cell Dev-Pl* **46:** 219-232

Stahl ML, Pattee PA (1983a) Computer-assisted chromosome mapping by protoplast fusion in Staphylococcus aureus. *J Bacteriol* **154:** 395-405

Stahl ML, Pattee PA (1983b) Confirmation of protoplast fusion-derived linkages in Staphylococcus aureus by transformation with protoplast DNA. *J Bacteriol* **154:** 406-412

Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD (2010) Fis is essential for capsule production in Pasteurella multocida and regulates expression of other important virulence factors. *PLoS Pathog* **6:** e1000750

Stella S, Cascio D, Johnson RC (2010) The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev* **24:** 814-826

Stephanopoulos G (2008) Metabolic engineering: enabling technology for biofuels production. *Metab Eng* **10:** 293-294

Sternberg N, Hamilton D (1981) Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J Mol Biol* **150:** 467-486

Sternberg N, Hamilton D, Austin S, Yarmolinsky M, Hoess R (1981) Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harb Symp Quant Biol* **45 Pt 1:** 297-309

Stockinger P, Kvitsiani D, Rotkopf S, Tirian L, Dickson BJ (2005) Neural circuitry that governs Drosophila male courtship behavior. *Cell* **121:** 795-807

Sumi H, Hamada H, Tsushima H, Mihara H, Muraki H (1987) A novel fibrinolytic enzyme (nattokinase) in the vegetable cheese Natto; a typical and popular soybean food in the Japanese diet. *Experientia* **43:** 1110-1111

Sun N, Zhao H (2013) Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol Bioeng* **110:** 1811-1821

Suzuki E, Nakayama M (2011) VCre/VloxP and SCre/SloxP: new site-specific recombination systems for genome engineering. *Nucleic Acids Res* **39:** e49

Suzuki N, Nonaka H, Tsuge Y, Inui M, Yukawa H (2005a) New multiple-deletion method for the Corynebacterium glutamicum genome, using a mutant lox sequence. *Appl Environ Microbiol* **71:** 8472-8480

Suzuki N, Tsuge Y, Inui M, Yukawa H (2005b) Cre/loxP-mediated deletion system for large genome rearrangements in Corynebacterium glutamicum. *Appl Microbiol Biotechnol* **67:** 225-233

Swingle B, Bao Z, Markel E, Chambers A, Cartinhour S (2010) Recombineering using RecTE from Pseudomonas syringae. *Appl Environ Microbiol* **76:** 4960-4968

Takano T (1966) Behavior of Some Episomal Elements in a Recombination-Deficient Mutant of Escherichia coli. *Japanese Journal of Microbiology* **10:** 201-210

Tanouchi Y, Pai A, Buchler NE, You L (2012) Programming stress-induced altruistic death in engineered bacteria. *Mol Syst Biol* **8:** 626

Tatum EL (1945) X-Ray Induced Mutant Strains of Escherichia Coli. *Proc Natl Acad Sci U S A* **31:** 215-219

Taylor AL, Thoman MS (1964) The Genetic Map of Escherichia Coli K-12. *Genetics* **50:** 659-677

Thomas CM, Smith CA (1987) Incompatibility group P plasmids: genetics, evolution, and use in genetic manipulation. *Annu Rev Microbiol* **41:** 77-101

Thorpe HM, Smith MC (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A* **95:** 5505-5510

Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM (2010) Tertiary architecture of the Oceanobacillus iheyensis group II intron. *RNA* **16:** 57-69

Toor N, Keating KS, Taylor SD, Pyle AM (2008) Crystal structure of a self-spliced group II intron. *Science* **320:** 77-82

Torrens R (1815) *An essay on the external corn trade*, London,: Printed for J. Hatchard.

Travers A, Schneider R, Muskhelishvili G (2001) DNA supercoiling and transcription in Escherichia coli: The FIS connection. *Biochimie* **83:** 213-217

Turan S, Zehe C, Kuehle J, Qiao J, Bode J (2013) Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene* **515:** 1-27

Twort FW (1915) An investigation on the nature of ultra-microscopic viruses. *The Lancet* **186:** 1241-1243

Typas A, Nichols RJ, Siegele DA, Shales M, Collins SR, Lim B, Braberg H, Yamamoto N, Takeuchi R, Wanner BL, Mori H, Weissman JS, Krogan NJ, Gross CA (2008) High-throughput, quantitative analyses of genetic interactions in E. coli. *Nat Methods* **5:** 781-787

Ubeda C, Olivarez NP, Barry P, Wang H, Kong X, Matthews A, Tallent SM, Christie GE, Novick RP (2009) Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Mol Microbiol* **72:** 98-108

Uehara H, Yoneda Y, Yamane K, Maruo B (1974) Regulation of neutral protease productivity in Bacillus subtilis: transformation of high protease productivity. *J Bacteriol* **119:** 82-91

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11:** 636-646

Val ME, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D (2012) Genome engineering in Vibrio cholerae: a feasible approach to address biological issues. *PLoS Genet* **8:** e1002472

Valens M, Penaud S, Rossignol M, Cornet F, Boccard F (2004) Macrodomain organization of the Escherichia coli chromosome. *EMBO J* **23:** 4330-4341

van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A, Loux V, Dervyn R, Bossy R, Bolotin A, Batto JM, Walunas T, Gibrat JF, Bessieres P, Weissenbach J, Ehrlich SD et al (2006) The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution. *Proc Natl Acad Sci U S A* **103:** 9274-9279

van de Putte P, Zwenk H, Rorsch A (1966) Properties of four mutants of Escherichia coli defective in genetic recombination. *Mutat Res* **3:** 381-392

van Kessel JC, Hatfull GF (2007) Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4:** 147-152

van Kessel JC, Hatfull GF (2008) Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol* **67:** 1094-1107

van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6:** 767-772

van Pijkeren JP, Britton RA (2012) High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Res* **40:** e76

van Vliet F, Silva B, van Montagu M, Schell J (1978) Transfer of RP4::Mu plasmids to Agrobacterium tumefaciens. *Plasmid* **1:** 446-455

Volkert FC, Broach JR (1986) Site-specific recombination promotes plasmid amplification in yeast. *Cell* **46:** 541-550

Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M (2003) Stepwise manipulation of DNA specificity in Flp recombinase: Progressively adapting Flp to individual and combinatorial mutations in its target site. *J Mol Biol* **326:** 65-76

Waite AJ, Shou W (2012) Adaptation to a new environment allows cooperators to purge cheaters stochastically. *Proc Natl Acad Sci U S A* **109:** 19079-19086

Wall JD, Krumholz LR (2006) Uranium reduction. *Annu Rev Microbiol* **60:** 149-166

Walter J, Ley R (2011) The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol* **65:** 411-429

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460:** 894-898

Wang Y, Li Y, Pei X, Yu L, Feng Y (2007) Genome-shuffling improved acid tolerance and L-lactic acid volumetric productivity in Lactobacillus rhamnosus. *J Biotechnol* **129:** 510-515

Wang Y, Weng J, Waseem R, Yin X, Zhang R, Shen Q (2012) Bacillus subtilis genome editing using ssDNA with short homology regions. *Nucleic Acids Res* **40:** e91

Watanabe T, Fukasawa T (1961) Episome-mediated transfer of drug resistance in Enterobacteriaceae. I. Transfer of resistance factors by conjugation. *J Bacteriol* **81:** 669-678

Wechsler JA, Gross JD (1971) Escherichia-Coli Mutants Temperature-Sensitive for DNA Synthesis. *Mol Gen Genet* **113:** 273-&

Weiss B, Richardson CC (1967) Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. *Proc Natl Acad Sci U S A* **57:** 1021-1028

White TB, Lambowitz AM (2012) The retrohoming of linear group II intron RNAs in Drosophila melanogaster occurs by both DNA ligase 4-dependent and -independent mechanisms. *PLoS Genet* **8:** e1002534

Whitt JT (2011) DNA Target Site Recognition by the Ll.LtrB Group II Intron RNP. PhD Thesis, University of Texas at Austin,

Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE, Schneider D (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *P Natl Acad Sci USA* **110:** 222-227

Wiesmann KE, Cortes J, Brown MJ, Cutter AL, Staunton J, Leadlay PF (1995) Polyketide synthesis in vitro on a modular polyketide synthase. *Chem Biol* **2:** 583-589

Wilson JW, Figurski DH, Nickerson CA (2004) VEX-capture: a new technique that allows in vivo excision, cloning, and broad-host-range transfer of large bacterial genomic DNA segments. *J Microbiol Methods* **57:** 297-308

Winkler J, Kao KC (2012) Harnessing recombination to speed adaptive evolution in Escherichia coli. *Metab Eng* **14:** 487-495

Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. *Mol Syst Biol* **6:** 407

Wirth D, Gama-Norton L, Riemer P, Sandhu U, Schucht R, Hauser H (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr Opin Biotechnol* **18:** 411-419

Wollman EL, Jacob F, Hayes W (1956) Conjugation and genetic recombination in Escherichia coli K-12. *Cold Spring Harb Symp Quant Biol* **21:** 141-162

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. *P Natl Acad Sci USA* **103:** 9107-9112

Wyckoff RW (1930) The Killing of Certain Bacteria by X-Rays. *J Exp Med* **52:** 435-446

Xu Z, Lee NC, Dafhnis-Calas F, Malla S, Smith MC, Brown WR (2008) Site-specific recombination in Schizosaccharomyces pombe and systematic assembly of a 400kb transgene array in mammalian cells using the integrase of Streptomyces phage phiBT1. *Nucleic Acids Res* **36:** e9

Yang J, Zimmerly S, Perlman PS, Lambowitz AM (1996) Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381:** 332-335

Yao J (2008) Bacterial Gene Targeting Using Group II Intron Ll.LtrB and the Role of Bacterial Host Factors in Ll.LtrB Splicing and Retrohoming. PhD Thesis, University of Texas at Austin,

Yao J, Lambowitz AM (2007) Gene targeting in gram-negative bacteria by use of a mobile group II intron ("Targetron") expressed from a broad-host-range vector. *Appl Environ Microbiol* **73:** 2735-2743

Yao J, Truong DM, Lambowitz AM (2013) Genetic and biochemical assays reveal a key role for replication restart proteins in group II intron retrohoming. *PLoS Genet* **9:** e1003469

Yao J, Zhong J, Fang Y, Geisinger E, Novick RP, Lambowitz AM (2006) Use of targetrons to disrupt essential and nonessential genes in Staphylococcus aureus reveals temperature sensitivity of Ll.LtrB group II intron splicing. *RNA* **12:** 1271-1281

Yoon JH, Kang SS, Mheen TI, Ahn JS, Lee HJ, Kim TK, Park CS, Kho YH, Kang KH, Park YH (2000) Lactobacillus kimchii sp. nov., a new species from kimchi. *Int J Syst Evol Microbiol* **50 Pt 5:** 1789-1795

Youngman P, Perkins JB, Losick R (1984) Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene. *Plasmid* **12:** 1-9

Youngman PJ, Perkins JB, Losick R (1983) Genetic transposition and insertional mutagenesis in Bacillus subtilis with Streptococcus faecalis transposon Tn917. *Proc Natl Acad Sci U S A* **80:** 2305-2309

Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, Lee WS, Kim MS, Kim SC (2002) Minimization of the Escherichia coli genome using a Tn5-targeted Cre/loxP excision system. *Nat Biotechnol* **20:** 1018-1023

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97:** 5978-5983

Zarschler K, Janesch B, Zayni S, Schaffer C, Messner P (2009) Construction of a gene knockout system for application in Paenibacillus alvei CCM 2051T, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. *Appl Environ Microbiol* **75:** 3077-3085

Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WP, del Cardayre SB (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415:** 644-646

Zhong J, Karberg M, Lambowitz AM (2003) Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res* **31:** 1656-1664

Zhou Y, Larson JD, Bottoms CA, Arturo EC, Henzl MT, Jenkins JL, Nix JC, Becker DF, Tanner JJ (2008) Structural basis of the transcriptional regulation of the proline utilization regulon by multifunctional PutA. *J Mol Biol* **381:** 174-188

Zhuang F, Karberg M, Perutka J, Lambowitz AM (2009) EcI5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* **15:** 432-449

Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM (1995a) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83:** 529-538

Zimmerly S, Guo H, Perlman PS, Lambowitz AM (1995b) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82:** 545-554

Zinder ND, Lederberg J (1952) Genetic exchange in Salmonella. *J Bacteriol* **64:** 679-699

Zoraghi R, See RH, Gong H, Lian T, Swayze R, Finlay BB, Brunham RC, McMaster WR, Reiner NE (2010) Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from methicillin-resistant Staphylococcus aureus. *Biochemistry* **49:** 7733-7747

Zoraghi R, Worrall L, See RH, Strangman W, Popplewell WL, Gong H, Samaai T, Swayze RD, Kaur S, Vuckovic M, Finlay BB, Brunham RC, McMaster WR, Davies-Coleman MT, Strynadka NC, Andersen RJ, Reiner NE (2011) Methicillin-resistant Staphylococcus aureus (MRSA) pyruvate kinase as a target for bis-indole alkaloids with antibacterial activities. *J Biol Chem* **286:** 44716-44725

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406-3415

Vita

Peter James Enyeart graduated from Skyline High School in Idaho Falls, Idaho in 1997, and in 2002 graduated *cum laude* from Brigham Young University with a Bachelor's of Science in biochemistry, with minors in music and Japanese. He then worked as a substitute teacher in Las Vegas, Nevada, as an English teacher on the island of Takashima in Nagasaki Prefecture, Japan, and as a patent translator in Osaka, Japan, before entering the Graduate School at the University of Texas at Austin in 2007.

Email: peter.enyeart@utexas.edu

This dissertation was typed by Peter James Enyeart.