Compensatory evolution within proteins:
An experimental assessment

Dissertation

In fulfilment of the requirements for the degree
Doctor rerum naturalium (Dr. rer. nat.)
of the Faculty of Mathematics and Natural Sciences
at Christian-Albrechts University of Kiel

submitted by:

Muhammad Bilal Haider

Department of Evolutionary Genetics
Max Planck Institute for Evolutionary Biology

Plön, Germany, March 2021

Thesis Defense Committee

First Examiner: Dr. Julien Y. Dutheil

Second Examiner: Prof. Dr. Tal Dagan

Thrid Examiner: Prof. Dr. Dietrich Ober

Chairperson: Prof. Dr. Matthias Leippe

Date of the Oral Examination: 4th May, 2021

Zusammenfassung

Interaktionen zwischen den einzelnen Aminosäuren einer Proteinsequenz sind wichtig, um deren Struktur und Funktion zu erhalten. Folglich können Mutationen an einer funktionell bedeutsamen Position durch eine Mutation an einer interagierenden Position kompensiert werden. So kann beispielsweise die Substitution einer Aminosäure von einer kurzen zu einer langen Seitenkette durch eine reziproke Substitution an einer interagierenden Position kompensiert werden. Ein solches Koevolutionsszenario impliziert, dass die erste Mutation eine Fitnessreduktion verursacht, während die kompensierende Mutation diese wiederherstellt. Es wurden mehrere Methoden entwickelt, um koevolvierende Positionen basierend auf Sequenzalignments zu identifizieren. Die Validierung der resultierenden Prognosen stützt sich bisher jedoch lediglich auf indirekte Beweise, wie beispielsweise Kontaktkarten von Proteinuntereinheiten, für welche eine experimentelle Struktur verfügbar ist. In der hier vorgestellten Studie wird ein Datensatz verwendet, der Substitutionen für ein Protein-Sequenz-Alignment auf jedem phylogenetischen Zweig von E.Colie kartiert, um koevolvierende Aminosäuren in homologen bakteriellen Proteinfamilien zu erkennen (CoMap). Unter Berücksichtigung der biochemischen Eigenschaften der Aminosäuren haben wir aus einem Datensatz mit Tausenden von koevolvierenden Gruppen eine Vorauswahl von Kandidaten getroffen. Anschließend haben wir jeweils drei Kandidatengruppen aus drei Proteinen (Elongationsfaktor 4, IspH und YebC-Proteine) ausgewählt, bei denen im phylogenetischen Zweig von Escherichia coli ein Muster von Co-Substitutionen für Ladungs- und Beta-Neigungen zur Kompensation erkennbar ist. Wir haben dann die lokale Fitnesslandschaft experimentell konstruiert, indem wir den ursprünglichen Genotyp von E. coli rekonstruiert und schließlich in Konkurrenz zu den jeweiligen Einzelmutanten, sowie dem rekonstruierten ursprünglichen Genotyp (Doppelmutante) gestellt haben. Wir haben sowohl Konkurrenzexperimente in nährstoffangereichertem LB Medium als auch in M9 minimal Nährmedium (mit Glukose als einzige Kohlenstoffquelle) durchgeführt. Bei den EF4- und YebC-Proteinen konnten wir eine verminderte Fitness bei den Einzelmutanten sowie einen Wiederherstellungspeak im ursprünglichen Zustand innerhalb der lokalen Fitnesslandschaft beobachten. Wir stellen hier die erste experimentelle Einschätzung zur Vorhersagbarkeit von Koevolution bei Proteinen vor. Zusätzlich haben wir einen Peak in einer der Einzelmutanten von IspH-Kandidaten beobachtet, der die Ladung nicht kompensiert. Statt einen Proxy für die, in den meisten Studien erwähnte, Proteinfitness zu verwenden, haben wir hier eine direkte Messung der Fitness in E. coli durchgeführt. Unsere Ergebnisse liefern

experimentelle Beweise für die kompensatorische Fähigkeit von zwei der drei getesteten Kandidaten und unterstreichen das Potenzial von Methoden zur Erkennung von Koevolution für das Verständnis molekularer Evolution. Diese Studie hilft, die Interaktion der Aminosäuren im dreidimensionalen Raum zu verstehen und Mechanismen der Evolution vorherzusagen.
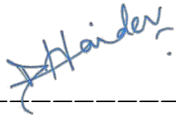
Summary

Amino acids within a protein sequence interact to maintain their structure and function. Consequently, mutations at a given functionally significant position can be potentially compensated by a mutation at an interacting position. For instance, a substitution from an amino acid with a small side chain to an amino-acid with a big side chain can be compensated by a reciprocal substitution at an interacting position. Such a coevolution scenario implies that the first mutation leads to a fitness reduction, while the compensating mutation restores it. Several methods have been developed to detect coevolving positions from sequence alignments. However, the validation of the resulting predictions relies so far only on indirect evidence such as residue contact maps in proteins for which an experimental structure is available. This study used the dataset that mapped substitutions for a protein sequence alignment on each phylogeny branch to detect coevolving amino-acids in bacterial homologous protein families (CoMap). Accounting for the biochemical properties of amino acids, we short-listed potential candidates from a dataset of thousands of coevolving groups for the experimental assessment. We then selected three candidate groups from three proteins (Elongation factor 4, IspH, and YebC proteins) displaying a pattern of co-substitutions in the *Escherichia coli* branch of the phylogeny for charge and beta propensities for compensation. I experimentally reconstructed the local fitness landscape, resurrecting the ancestral genotype in *E. coli* and putting it in competition with single mutants and reconstructed ancestral state (double mutant). I have performed competition experiments in LB Broth nutrient-enriched medium and M9 minimal (glucose as a single carbon source). In EF4 and YebC proteins, I observed a valley of lower fitness in single mutants in the local fitness landscape and restoration peak in the ancestral state. I report the first experimental assessment of the prediction of coevolution within a protein. I have also observed a peak in one of the single mutants of IspH candidates, which compensates for the charge only in one direction of the mutation. We used direct measurement of fitness estimation in *E. coli* rather than using a proxy for fitness, *i.e.*, protein fitness mentioned in most studies. The results of this study provide experimental evidence of the compensating nature in two out of three tested candidates, highlighting the potential of coevolution detection methods as tools to understand molecular evolution. This study helps to understand the interaction of the amino acids in the 3D space and can be used to predict the mechanism of evolution.

Declaration

I hereby declare that:

- Apart from my supervisor's guidance the content and design of the paper is all my work;

- This thesis has not been submitted either partially or wholly as part of a doctoral degree to another examining body and no material has been published or submitted for publication;

- The preparation of this thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation;

- No academic degree has even been withdrawn prior to this thesis.

Plön, 9th March, 2021

—————————

Muhammad Bilal Haider

To all the old folks who have suffered and beaten the COVID19

Acknowledgments

It is not possible to do things all alone. Many cooperative people are around you who help you with the completion of your work. I feel the honor to mention the names of those who deserved to be acknowledged.

First of all, I would like to pay my credit to Dr. Julien Dutheil for his benignant supervision, precious pieces of advice, and guidance in every step of my research. He has been a great support and help. He provided me support and courage to complete difficult tasks during my research experimentations. His special attention towards my task is always remembered, and his friendly behavior made me feel comfortable to ask anything. I also acknowledge Dr. Jenna Gallie for all the help and troubleshooting in the experiments. Discussions with her are always motivating and encouraging. I always felt very appreciated by my thesis advisory committee (TAC), Prof. Dr. Diethard Tautz, and Prof. Dr. Tal Dagan. Scientific discussions with them make me able to work hard and keep myself motivated. I also feel lucky to get a valuable scientific discussion from Diethard and Tal. I also pay my gratitude to the Molecular system's evolution (MSE) group, especially Gustavo, Filipa, and Natasha. Friday evenings with Gustavo will never be forgotten. Their support and help really made me able to deal with the problems during my Ph.D. I acknowledge the laboratory technicians at the Max Planck Institute for Evolutionary Biology, Michaela, Cornelia, and Sven, for their help in the laboratory.

I am also very thankful to Eric Hugoson and my beer partner to make me able to handle the computational part of the project. I learn a lot of coding with his help. Discussions at beer hours were always joyful.

I am also very grateful for the valuable discussions with David Rogers. I am delighted to have friends like Joanna Summers, Michael Barnett, and Norma Rivera during my Ph.D. Loukas Theodosius, Our late-night walks on Klanderstrasse. Their company kept me motivated and late-night discussions over any topic were great times. My great friend Nico Furhmann. All the Friday nights and then lab work on the weekends and finding him as well in the lab was always encouraging. My great friend Thomas Braun remembered me even after leaving Plön.

I would like to extend my thanks to my batchmates Roman, Wiola, Elena Damm (Thanks a lot Elena for the German translation of the summary.), Aditi, Onur, Lizel, Khawla, and Elena Horas for happy times during my stay. Devika and Anuradha for helpful discussion on the experimental part of the project. I am also very grateful to Neel Prabh, Nidhi Singh, and Iqra Kasu; their support and company were very important during my thesis write-up phase. I would like to mention a long list of people who were here and made my stay the best time of my life so

far. I especially pay my thankfulness to everyone Jatin, Maria, Juan, Ezgi, Zawya, Niklas, Runa, Carolina, Artemis, Andrea, Pauline, Jordan, Bram, Andy, Carsten, Johanna, Cecilia, Maryam, Zahra, Mr. Moyni, Lara, Damagoj, Ana, M. Sieber, Ellen, and Mayuresh. I am also thankful to every person who helped me during this piece of important research. I feel lucky to have support of my family. My mom, dad, brothers and sisters. Thanks for being here for me.

Contents

List of Figures

List of Tables

List of Abbreviations

AAindex	Amino acid Index
DCA	Direct coupling analysis
DHFR	Dihydrofolate reductase
DMAPP	Dimethylallyl Pyrophosphate
EF	Elongation factor
GFP	Green Fluorescence Protein
HMBPP	4-Hydroxy-3-methyl-but-2-enyl pyrophosphate
IPP	Isopentenyl Pyrophosphate
IR	Ionization Radiation
IS150	Insertion sequence 150
LB	Luria Bertani
MEP	Non-melvalonate Pathway
MP	Malthusian Parameter
MSA	Multiple Sequence Alignment
MVA	Mevalonate Pathway
PCR	Polymerase chain reaction
PDB	Protein Database
SOE	Splicing Overlap Extension PCR
TA	Tetrazolium Arabinose

Introduction

Chapter 1

Introduction

The theory of systems der Formbildung [1933] applied to biology recognizes several levels of the organization, from macromolecules to ecosystems [Odum and Barrett, 1971]. The principle of synergy states that the properties of a given level are not simply the sum of the properties of the nested levels but include some emerging properties resulting from their interactions. This fundamental principle has important implications for the evolution of the systems: because they interact and share constraints at a higher level, they do not evolve independently. This non-independent evolution is termed *coevolution* and was first described at the species level. John N. Thompson Thompson [1994] defines coevolution as "a reciprocal evolutionary change in interacting species", implying that a change in one species impacts the selection pressure on another species. This change, in turn, can alter the selection pressure on the first one. Classic examples of inter-species coevolution include the evolution of predators and preys, hosts and parasites, competitors, and mutualists. Interacting species introduce selection on each other and continue to reshape their life histories and each other's phenotypic traits. Natural selection drives the process of reciprocal evolutionary change. The coevolutionary arms race between two species is one factor that generates biodiversity in nature and makes a continuous process of evolution [Thompson, 2010].

Coevolution was initially theorized at the species level, and coevolutionary dynamics have also been recognized at the molecular level within genomes. This chapter first describes the mechanisms of coevolution at the molecular level and introduces prediction methods from sequence data. Further, this chapter discusses specifically one mechanism of molecular coevolution, compensatory mutations, and highlights the importance of molecular coevolution in the context of fitness landscapes. Finally, I will introduce how fitness can be assessed experimentally and show how such experiments can be applied to the experimental study of molecular coevolution.

1.1   Molecular coevolution

At the molecular level, coevolution between two loci occurs when a change at the first locus affects the selective pressure at the second one [Atchley et al., 2000]. Lovell and Robertson [2010] therefore, defines coevolution at the molecular level as the "reciprocal evolutionary change in interacting loci". Molecular coevolution has two forms: between molecules and within a molecule, and the focus here is on coevolution within a protein. Several approaches are present that predict the molecular coevolution between proteins and within protein [Tuffery et al., 1999; Weigt et al., 2009; Morgan et al., 2006].

1.1.1   Inter and intra molecular coevolution

Coevolution between two proteins is termed as *Inter-molecular coevolution*. Protein-protein interactions (PPIs) within a cell are important in network biology and have been studied widely to understand the interactions of proteins involved in a particular biology network or pathway [Makino and Gojobori, 2007]. Interaction of ligand and receptor molecules is one of the causes of coevolution in the cell. The other form of molecular coevolution is *intra-molecular coevolution*. The principles mentioned above for interactions between proteins also apply to the interaction of amino acids within an individual protein. The complex interaction of amino acids within a proteins' 3D structure, functional constraints, and 3D contact maps derive coevolution. Several methods have been proposed to predict coevolution within a protein by exploring the correlated patterns and correlated processes in a set of sites [Galtier and Dutheil, 2007]. Methods that detect correlated processes take into account the underlying phylogeny (Figure 1.1). Several papers predict the physical contact of the coevolving residues [Morcos et al., 2011]. The method mentioned above claims the accurate prediction of residue pairs in their dataset using multiple alignment sequences. To understand non-independent evolution in terms of the 3D structure of proteins, one of the mechanisms of predicting evolution is compensatory evolution.

1.1.2   Compensatory evolution

For instance, consider two sites of a protein sequence. A substitution from a positively charged amino acid to a negatively charged amino acid depends on the other amino acid's charge state in the protein sequence. In the said example, the interaction involves the charge of two residues, one is positively charged, and the other is negatively charged. The mutation on the two interacting sites (in amino acids) with the same charge leads to an incompatible interaction, either +/+ or -/-. The interaction can be restored either by a reverse mutation or by a compensatory

mutation at the other site, transforming a -/+ interaction into a +/- interaction. Protein sequence data tells us about the biochemical properties (charge in this example) of the amino acids. So, this non-independence of the amino acids is predictable in protein sequences of all organisms. All amino acids share different physiochemical properties like charge, volume, polarity, and hydrophobicity. In the case of proteins, amino acids' biochemical properties weaken the signal of coevolving sites because of the amino acids' redundant properties. Some substitutions are neutral and do not have any fitness cost. Multiple sequence alignments (MSA) in comparative methods help capture sites' association (correlated patterns). Here, I have only discussed methods that use MSA as input. One class of method uses 3D contact maps of amino acids in the proteins combined with correlated mutations [Fariselli et al., 2001; Bohr et al., 1993; Olmea and Valencia, 1997; Vendruscolo et al., 1997]. They use frequent co-occurrence of states of amino acids at a pair of sites to predict the site-specific coevolution [Gutell et al., 1992; Neher, 1994]. A high rate of false-positive is a problem in these methods, and the reason was the assumption of independence. I argue here that while inferring coevolution in biological sequence data, the sites' shared history influences the coevolution detection signal. Mutations on the interacting sites that have higher selection pressure tend to evolve slowly. Therefore, incorporating evolutionary history into the estimation of coevolution improves the interpretation of the coevolutionary signal [Dutheil et al., 2005].

I have discussed that for accurate prediction, the shared history of the sites and biochemical properties of the amino acids are essential to be considered in an accurate prediction method of coevolving sites within a protein or between proteins. The empirical assessment of the prediction of coevolving residues is missing, and there is no experimental evaluation of these predictions on the fitness of the target organism has been reported. To fill that gap, one needs to know how one can assess coevolving amino acids' impact on the organism's fitness. I have aimed the experimental assessment of predicted coevolving amino acids in this study, and I used the mechanism of compensatory mutations to try to answer this question. The relationship among all genotypes of a gene and their fitness can be visualized in the fitness landscape. A rugged fitness landscape shows local peaks for each genotype of the target gene, and these peaks have a fixed effect on fitness [Van Cleve and Weissman, 2015]. The rugged fitness landscape helps to understand each genotype's fitness of the gene under study and provides the knowledge of the fitness effects of mutations (increase or decrease in the fitness). Much experimental work has been done to map the fitness landscape and analyze its genotypes' fitness peaks. Reconstruction of the genotypes experimentally is also relevant under the hypothesis of

compensatory evolution.

1.1.3 Fitness and Environment

The ability of an individual to survive and reproduce in a given environment is called fitness, and every individual in the population responds differently towards the change in the environment. For example, the individual's growth would not be the same in the environment, and they grow differently if a change happens in that given environment. Hence, the fitness of individual changes with change in the environment leads to the fact that both fitness and environment play essential roles. Fitness can be assessed at any biological level; for instance, successful reproduction of individuals is an example of an individual organism's fitness on a species level. The same principle applies to a protein's fitness: the translation of enough protein molecules after the change happened on its locus. The environment is one of the contributing factors in the dynamics of coevolution, specifically in the species' interactions. Change in the environment might affect the balance in the densities of the interactions among sites. Hence, it is essential to keep the environment in mind while conducting experiments in the context of coevolution. The effects of the interactions between coevolving species may not be the same in two different environments. Due to the change in the environmental conditions, individuals in a population behave differently for survival and reproduction.

1.2 Detecting coevolving sites from sequence alignments

Acquiring the signal of coevolution in a protein is a challenge. For example, there is enormous diversity in the proteins as they are composed of 20 amino acids. These amino acids have side chains involved in biochemical reactions and create diversity in the function and structure of the proteins. In the three-dimensional space of proteins, the state of one amino acid depends on the state of other amino acid/s, and they interact with each other and have structural constraints [Pollock et al., 1999].

1.2.1 Accounting for the history of sequences

The simplest way to measure non-independence is to consider the frequencies of state pairs at two sites and compare them to each state's marginal frequencies. Let us note $f_i(X)$ the frequency of state X at the site i, and $f_j(Y)$ the frequency of state Y at site j. We further note $f_{i,j}(XY)$ the frequency of the state pair XY at the two sites. Under the null hypothesis of independence, the expected frequency of $\hat{f}_{i,j}(XY) = f_i(X) * f_j(Y)$. We can compute all $(f_{i,j})$

Figure 1.1 Correlated patterns on phylogeny (independent and non-independent evolution). The tree on the top right panel with fewer substitutions shows random mutations (red on one site and blue on the other site) present on the tree branches. The left top tree shows a correlated pattern on fewer branches giving information about slow evolution. The left bottom shows fast evolution without any pattern to follow of cosubstitution on the branches of the tree. The right bottom shows the cosubstitutions detected on the multiple branches giving the strong signal of coevolution.
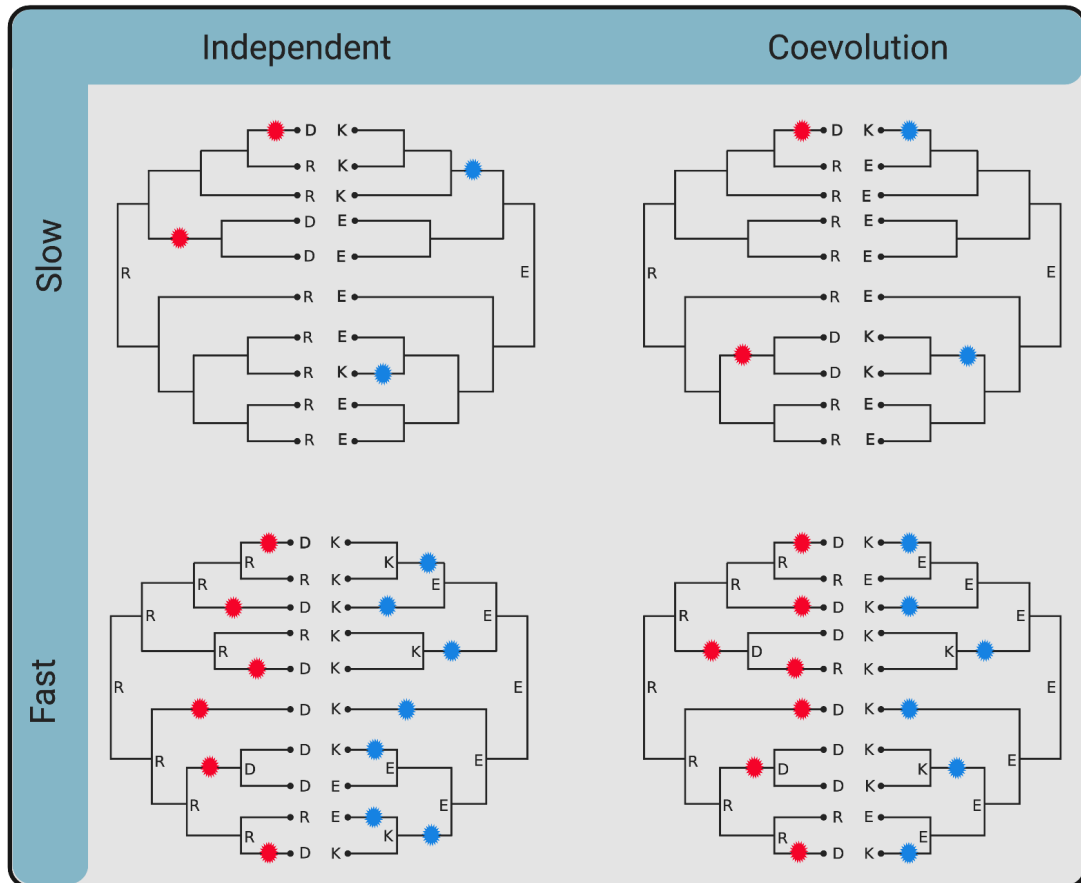
for all pairs of states and compare them to their expected distribution ($f_{i,j}$) using a chi-square test [Larson et al., 2000]. This chi-square method detects covariation based on the observed pair frequencies from a sequence alignment compared to the expectation under the null hypothesis of independence. Another statistic, mutual information (MI), is based on information theory and assesses one random variable's information by observing another random variable. Applied to sequences, MI can effectively tell how much we can predict one amino acid's presence at one site from the amino acid at another site of the sequence [Korber et al., 1993].

However, simple methods like chi-square and MI assume that the data points are independent, which is not the case of biological sequences because of the underlying phylogenetic relationships. Ignoring the shared history of the sequences in the statistics can overestimate the correlation and its significance [Felsenstein, 1985]. Therefore, methods like MI or chi-square

are prone to high false-positive rates of predicted coevolution, as they do not account for the history of the sequences. Accounting for the shared history of the sequences was shown to improve the accuracy of the predictions and reduce the false-positive rate [Dutheil, 2011]. One way to account for phylogeny is to use substitution mapping and look for the cosubstitutions present on the same branch of the phylogeny. Shindyalov et al. [1994] and colleagues pioneered in developing a method to detect pairs of positions with correlated mutations in a protein MSA. The method is based on reconstructing the phylogenetic tree for sequences and statistical analysis of the distribution of mutations in the tree branches. Their goal was to assess the degree of relationship between coevolution of protein residues and spatial proximity. The drawback in this method was that the statistics used in correlated changes have led to limited success in detecting pairs of residues adjacent to the three-dimensional structures. This statistical flaw was then resolved by Tuffery et al. [1999] where they improved their method and showed that the correct detection is possible with the phylogenetic information. Tuffery et al. [1999] method still has a limitation of uncertainty in the reconstruction of the ancestral state, which CoMap method resolved later [Dutheil et al., 2005]. The CoMap method was first benchmarked on RNA datasets and was later extended to protein sequences [Dutheil and Galtier, 2007].

1.2.2   Predicting compensatory mutations in proteins

CoMap is a phylogenetic-based method that accounts for the uncertainty in the ancestral state, includes biochemical properties of amino acids, and a clustering approach that gives robustness in predicting coevolving sites. CoMap was also introduced to measure coevolution to look at compensatory coevolution considering the biochemical properties of amino acids. It is studied that two or more sites in a protein coevolve to maintain structural integrity or optimal function. This co-dependence contributes to the compensatory behavior of amino acids within proteins. The compensatory evolution mentioned earlier, a substitution from an amino acid with a small side chain to a big side chain, can be compensated by a reciprocal big to small substitution at an interacting position [Neher, 1994]. The weighted substitution vector accounting for the physicochemical distance between amino acids is calculated. The weighted substitution vector in this study accounts for the charge, volume, and polarity. Correlations are estimated using the compensation index instead of the Pearson correlation coefficient in the previous studies [Dutheil and Galtier, 2007].

In the study mentioned above, the extended version of CoMap explains the "cosubstitu-tions"; that is, substitution events happen on the same branch of the phylogeny in the context

of amino acids and their physio-chemical properties. In a recent study, this cosubstitution mapping method expanded its limits to a large dataset of bacterial homologous protein families and detected coevolving sites in different bacterial species [Chaurasia and Dutheil in prep]. They included canonical biochemical properties, charge, volume, polarity, and eight indices in their analysis and predicted coevolving sites in the proteins. They have identified a large number of coevolving groups and predicted compensatory evolution in their dataset. Using the fuzzy clustering approach, one can cluster amino acids based on their properties other than the canonical physio-chemical properties (a charge, volume, and polarity) [Saha et al., 2012]. They found eight clusters in the AAindex dataset of amino acids.

1.3 Experimental assessment of fitness landscape

Fitness landscapes capture the relationship of genotype and evolutionary fitness and help to visualize the evolutionary trajectories. Experimental work has contributed to understanding natural fitness landscapes by constructing the mutants of all possible combinations of small sets of mutations. These small sets of mutations can be from a single lineage of an organism or multiple lineages of the same organism, but each lineage is used as a separate treatment to estimate the fitness. As a result, the empirical work has captured the attention of theoretical analyses of the predictability of evolution. Some experimental approaches analyze the fitness landscape to elucidate the relationship of genotypes and phenotypes of a particular gene. High-throughput sequencing techniques have enabled us to produce extensive sequencing data (whole genomes), and it is possible to generate all mutants' combinations under various treatments or conditions that help to analyze the fitness effects of mutations in all mutants. Several studies investigated the mutational effects of multiple mutations in their target gene (intragenic) or set of genes (intergenic). These studies have used different ways of estimating fitness values in several model organisms, such as the catalytic activity of an enzyme, resistance to an antibiotic, growth rate and cell survival, and competition experiments of different genotypes. Some of these studies are detailed below.

1.3.1 Antibiotic resistance as a proxy for fitness

Weinreich et al. [2006] and colleagues developed a quintuple mutant with five point mutations in the beta-lactamase gene and found enhanced bacterial resistance to the ampicillin antibiotic as compared to the wild type. Interestingly, only a single point mutation was responsible for the enhanced antibiotic resistance, and the other four had negligible effects, which means that sign

epistasis mediated pleiotropic effects of mutations on the molecular level. They also estimated the relative probabilities of the possible 120 trajectories followed by natural selection. 102 out of 120 possible mutational trajectories were inaccessible, which explained that in the presence of selection, only a few mutational paths followed to achieve the fitness peak (in terms of resistance against antibiotic) [Weinreich et al., 2006]. Compensatory mutations can restore the deleterious effects of mutations. In another study, for the mobile colistin resistance gene (mcr), one variant, mcr-3.5 of the gene showed neutral and compensatory mutations in the resistance gene and revealed the complexity of the local fitness landscape [Yang et al., 2020].

1.3.2 Fitness using growth rate

In one more study in a transgenic *Saccharomyces falciparum*, Brown et al. [2010] and colleagues explored 48 combinations of six mutations at five amino acids positions of an enzyme Dihydrofolate reductase (DHFR) from *Plasmodium falciparum*. They analyzed the difference in the growth rate of *P. falciparum* in the presence of pyrimethamine drug. Their study found out that those mutations that occurred later in the evolutionary trajectory can compensate for earlier mutations' fitness consequences. Compensatory mutations between later and earlier mutations over time explained the mutational trajectories of the mutations in a particular evolutionary scenario of parasite's resistance and growth. Most of these studies focused on an intragenic landscape (all sites within a gene). These studies gave the qualitative clue of the interactions between amino acids in proteins and included a large genotypic space for the analysis. There is still a challenge in the laboratory to create all the sites' possible mutants in the understudy gene/s. Nevertheless, some studies succeeded in generating data for a large number of mutations.

1.3.3 Fitness of the green fluorescence protein

Another study for the global fitness landscape (including all positions in a gene) of the green fluorescence protein from *Aequorea victoria* explores the effects of mutations either deleterious or beneficial. A large number of derivative genotypes were created by random mutagenesis. As a function for fitness, they measured the fluorescence intensity of green fluorescence proteins (GFP) proteins expressed by all variants. Seventy-five percent of the mutations have deleterious effects, and single mutations cause a low level of fluorescence [Sarkisyan et al., 2016].

1.3.4 Discovering large fitness landscapes

In addition to these studies of fitness landscape where only a few mutational sites in a gene were analyzed, Bank et al. [2016] presented a large intragenic multiallelic fitness landscape of 640 mutants of the Hsp90 protein in *Saccharomyces cerevisiae* under a stress condition (high salinity environment). They observed that the gene's global fitness peak was achieved via four positively epistatic mutations, and mostly negative epistatic mutations prevail in the landscape. Recently, another study from the same group Bank et al. [2016] explored the largest fitness map with 14,160 amino acid variants (44,604 single codon changes) of heat shock protein (Hsp90). The Hsp90 protein aids in the folding and stability of other proteins in the cell. It is an ideal candidate to study fitness under stress conditions. Five different environmental conditions, Temperature, Diamide, Ethanol, Salt, and Nitrogen depletion, were applied to all variants in this study. The growth rate was measured in different stress environments, and growth patterns vary between these conditions suggested an environmental impact on the evolution of Hsp90. Moreover, some variants performed better in one environment and have poor growth in another. The difference in the environments' performance suggested that all mutational trajectories do not follow the same path in every environment [Flynn et al., 2020]. This study was an exploratory study that provided many data regarding the organism's fitness and environmental impact. Although, large fitness maps lose information about the individual effects of mutations at the sites of proteins. It might not elucidate the protein's function or the vital amino acids in the protein that can impact the protein structure. Hence, with such large landscapes, the global fitness landscape could be explained but does not help understand the local fitness landscape where only a few positions are involved within a gene. Beneficial mutations lead to the fixation and have a trajectory to follow that can be explored. Therefore, most studies of the fitness effects of mutations explained beneficial mutations under a particular stress environment like antibiotic resistance or high salinity.

1.3.5 Unveiling deleterious mutations from the fitness landscape

The studies mentioned above provide a glimpse of the beneficial mutations that increase the target organism's fitness or growth. However, these studies do not address deleterious mutations' effects and their contribution to the fitness landscape. Using antibiotic resistance, growth rate, and protein expression as a proxy for fitness can help understand the molecules' fitness (proteins) but does not reveal the organism's fitness because the mutants would not be viable in case of deleterious mutations. Experiments are biased towards studying the beneficial or neu-

tral mutations for obvious experimental reasons. Capturing the complete picture of all types of mutations with their evolutionary history is a challenge and leaves unexplored areas. Information about the deleterious mutational effects on an organism's fitness is missing in the studies mentioned above.

1.4 Objectives of the study

Exploring the fitness effects of mutations in the fitness landscape leads to an open question that how we can detect the evolutionary relationship of interacting positions in a protein and estimate the impact on the organism's fitness in a specific environment. One way is to use coevolving positions in the proteins, mutate these positions experimentally, and measure the organisms' fitness under different environments. I used the method of cosubstitution mapping described by Dutheil and Galtier [2007] to predicted coevolving sites in bacterial homologous protein families inferred [Chaurasia and Dutheil *in prep.*]. I experimentally reconstructed short-listed candidates' local fitness landscape in *Escherichia coli* our model organism in this study. Fitness values of all genotypes under the local fitness landscape of 2x2 (only two positions in a target protein) allow a more profound resolution of landscape and a direct method to measure the fitness effects on an organism instead of using a proxy of fitness. A subset of fitness landscape like 2x2 helps to understand the mutational effects on a protein function affecting an organism's fitness under a particular environment, unlike using indirect ways to assess the protein's fitness.

In competition experiments, it is essential to differentiate between genotypes. Therefore, I used *E. coli* Strain B REL606 and REL607 to differentiate genotypes (single and double mutants or *vice versa*) in competition experiments. REL607 is prototrophic of REL606 Ara- and consumes L (+) arabinose with a mutation at arabinose marker gene Ara+ [Lenski, 1988]. When these strains grow on tetrazolium arabinose (TA) plates, Ara- and Ara+ give red and white colonies, respectively. I used relative fitness measurement of mutant strains and the wild type in our competition experiments. Relative fitness using the Malthusian growth parameter is a well-established method to evaluate competing organisms' fitness levels in the experimental setup (Box 1.4.1). The study aims to explore compensatory mutations' fitness effects in an environment. Our study reports the first direct method to analyze the mutational effects on an organism's fitness under the light of compensatory evolution.

Box 1.4.1 Short-coming in the calculation of relative fitness

Malthusian growth parameter (MP) is the estimation of the exponential growth of an organism. For instance, two strains of a microorganism are competing for resources in a specific environment. Both strains would follow the standard growth curve starting from the exponential phase until the saturation phase over time. MP is calculated when both strains are at their exponential phase. Being at the exponential phase means that they have equal opportunity for the resources in the environment. Calculating MP at exponential growth would assess both strains' fitnesses and help estimate the selection coefficient. In a pairwise competition experiment between two strains of the bacteria, MP of both strains' ratios determines which strain is relatively better in growth under an environment [Lenski et al., 1991]. Primarily competition experiments assess the distribution of mutation fitness effects and use the ratios of the Malthusian parameter. There is an argument that the use of MP (relative fitness) ratios leads to overestimating the selection's strength per-generation time. Overestimation without generation time as a parameter may be corrected by scaling the organism's generation time into the calculation of relative fitness [Chevin, 2011]. This argument is essential because the study group is between different species rather than within the same species. In case of competition within the same species, both genotypes would have more or less the same generation time. According to Lenski [1988], the doubling time (generation time) is a dimensionless quantity and is identical to the Malthusian parameter. In conclusion, the short competition experiments in the laboratory would not affect by overestimating MP because of the short generation time (a few generations) and do not affect the estimation of the relative fitness of two genotypes from the same species.

Materials and Methods

Chapter 2

Materials and Methods

We aim for the experimental investigation of the predicted coevolving groups. To evaluate these predictions, we short-listed potential candidates in *E. coli*. It is a widely used microorganism for genomic manipulation and also has easy handling in the laboratory. After short-listing the candidates, we reconstructed all three candidates' mutants' genotypes and used those confirmed mutants for fitness assay.

2.1   Selection of the candidates

A data set of 1,630 bacterial protein families with at least one three-dimensional structure available was used (Chaurasia and Dutheil in prep). Sequences were aligned and a phylogeny inferred from each family, allowed to infer coevolving positions using the CoMap method [Dutheil and Galtier, 2007]. A relational database (SQL) was developed from the 51,661 coevolving positions found by Chaurasia and Dutheil, including site-specific information. This site-specific information contains the family name, coevolving groups in *E. coli*, quantity change (for example, change of the charge from positive to negative on each site) on *E. coli* branch of the tree, compensation for the biochemical property, PDB ID for the protein family, and branch distance.

We combined the data of predicted coevolving groups and site-specific quantity change for each method (biochemical property). Therefore, two tables were built in the database; one table had the predicted coevolving groups at the *E. coli* branch called the "Groupwise table" for each method, and the other table on the database called the "Sitewise table" had all the sites and their quantity change for each method from the *E. coli* branch (mentioned SQL query schema in 2.1). Python scripts were developed to parse the CoMap output files and insert data into the sqlite3 tables, using the python modules sqlite3.

In order to identify suitable candidate groups, we queried the database with the following

Figure 2.1 The pipeline and table schema used in the sqlite3 database
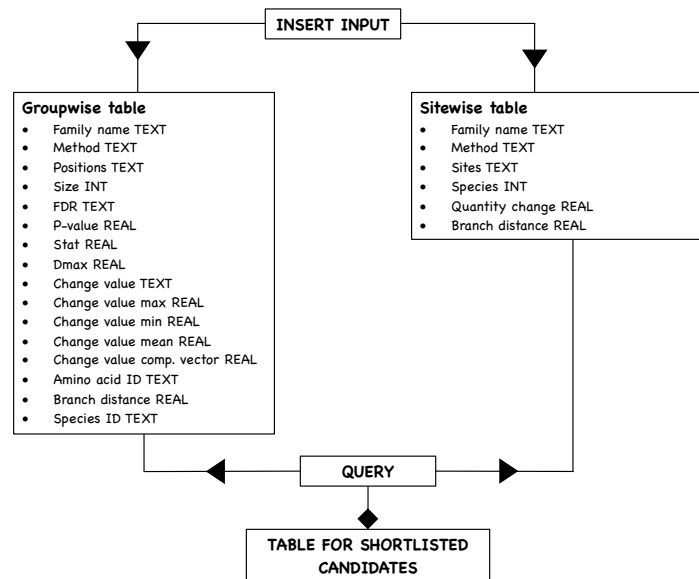
criteria:

- the three-dimensional structure of the candidate protein must be available for *E. coli*

- the candidate coevolving group must involve only two sites

- the statistical significance for coevolution should be high (p-value)

- the two coevolving residues should be in contact in the 3D structure

- close positions to be used to construct mutant DNA fragments

- the branch length is leading to *E. coli* in the corresponding phylogeny must be lower than 1

2.2   Ancestral state reconstruction

The predicted coevolving groups were identified using protein sequences. To create mutant strains, we need a DNA sequence to infer the ancestral state. Because of the redundancy of the genetic code, there are several codons for amino acids. All codons for one amino acid may not be equivalent, as distinct synonymous codons may have distinct fitness [Frumkin et al., 2018]. So we decided to reconstruct the ancestral codon using DNA sequences from the same database (HOGENOM) We used the `query_win` retrieval program to access the HOGENOM database and retrieve the DNA sequences for the selected protein families [Gouy and Delmotte, 2008].

The sequences were filtered to retain only the bacterial sequences used in the previous CoMap analysis. The codon sequence alignment was deduced from the protein sequence alignment using pal2nal [Suyama et al., 2006]. The marginal ancestral sequences were reconstructed using the bppAncestor program from the Bio++ program suite [Dutheil and Boussau, 2008]. The final table after adding codon reconstruction was then used for the short-listing of potential candidates. All scripts used for analyses were written in Python using Biopython. The final list of candidates for the experiments in this study is shown in table 2.1.

2.3 Construction of mutant plasmids

Mutagenic primers were designed for the candidate gene regions from the *E. coli* strain B REL606 genome. One set of the primers (F2 and R2) having predicted substitution was designed for each mutant (both singles and the double mutant) from all the candidates. Another set of primers was designed within the 500 bp upstream and downstream sequences of the candidate genes. Finally, a set of primers was designed to confirm the substitutions into the genome after the allelic replacement experiment. Table 2.4 has information for the sequences of all the primers used in this study. The cloning step of the modified genes was carried out in *E. coli* strain TOP-10 or DH5α-λpir (see table 2.2 for the strains used in this study). The amplified DNA fragments with the substitutions were first cloned in a high copy number entry vector (pCR8/GW/TOPO/TA or pUC57) and then cloned into pKOV-unstuff for the allelic replacement. The plasmids details are given in the table 2.4. All plasmid maps are shown in Supplementary material 8.6.

2.3.1 Genomic DNA extraction

We needed genomic DNA of *E. coli* for the cloning purpose (amplification and introduction of mutations) of candidate genes. Overnight cultures of *E. coli* strain B REL606 were grown in LB liquid with streptomycin 100 μg/ml. The following day, total genomic DNA was extracted from these overnight liquid cultures with the help of the GenElute Bacterial Genomic DNA kit (catalog no. NA2110-1KT). The DNA concentration and integrity were determined on Nanodrop (using 1OD260 = 50 μg ds DNA) and by visualization on 1 $\%$ agarose gel with SYBR green staining.

2.3.2 SOE-PCR amplification

Overlap extension polymerase chain reaction by splicing (SOE-PCR) is one of the PCR-based approaches to introduce mutations in the target DNA sequence. SOE-PCR was performed us-

ing genomic DNA with the Phusion polymerase (Phusion polymerase *pfu* has 3' exonuclease activity and is high fidelity polymerase). The PCR was divided into several steps, described below.

Round 1 PCR

In the first round of the PCR, it was further subdivided into a and b. In 1a PCR, a pair of forward primer F1 with the reverse primer of each mutant-R2 was used. In 1b PCR, forward primer mutant-F2 with reverse primer R1 was used for each mutant. This PCR step gave two amplified fragments for each mutant named m1-a, m1-b, m2-a, m2-b, anc-a, and anc-b. These PCR products were purified using the Qiagen PCR purification kit (catalog no. 28104).

Round 2 PCR

In the second round of PCR, a and b samples for each mutant with the same ratio were used as a DNA template for the PCR amplification. In this step, F1 and R1 primer sets were used to amplify the entire region used for the cloning. PCR products were run on 1 % agarose gel at 80V for 45 minutes and proceeded for the gel elution step to get one single amplified band required for cloning.

2.3.3 Cloning in pCR8/GW/TOPO/TA

Precisely ~1 kb target band was excised from the gel with a sterilized blade and weighed. A standard Qiagen agarose gel elution kit was used to purify the DNA from agarose gel (catalog no. 28704). For the cloning into pCR8/GW/TOPO, it needs A at the 3' end of the PCR product to bind with the T overhangs in the presence of Topoisomerase I covalently bound to the vector. We used *pfu* because of its high fidelity and 3' exonuclease activity. Therefore, we had to add A at the 3' end in another step. The gel-purified samples were incubated with the *Taq* polymerase for 20 minutes and were directly used for the cloning. After adding the polyA tail, ligation into pCR8 was done and used for the transformation into chemically competent cells of *E. coli* strain DH5α. A vial of competent cells was thawed on the ice for 10 minutes. 4 μl of ligation mixture was used to incubate with the cells for 10 minutes in ice. Cells were then heat shocked at 42° for 30 seconds and immediately transferred to the ice for 2 minutes. Autoclaved LB volume of 250 μl was added and incubated at 37°C for 1 hour and plated on the spectinomycin plates (100 μg/ml). 25 μl, 50 μl, 100 μl, and the rest of the volume were used to spread on the plates using autoclaved glass beads. Plates were incubated at 37°C overnight. The following day, few single

colonies were used from 100 μl plates for each mutant and used for the overnight cultures for the plasmid isolation using 100 μg/ml working concentration of spectinomycin in the liquid LB.

2.3.4   Confirmation of clones

Plasmids were isolated from overnight grown cultures. 5 ml of culture was used for the miniprep using the Qiagen miniprep kit (catalog no. 27104). The confirmation of the successful DNA insert was carried out by digestion of the plasmids with the restriction enzyme *BglII* which has the site in the primers. The digestion reaction mixture was prepared, and samples were incubated at 37°C for at least 2 hours. Samples were run on 1 % agarose gel for 30 minutes. Positive plasmids from restriction analysis were used for the Sanger sequencing to see the substitutions in the sequences before cloning into the pKOV-unstuff plasmid.

2.3.5   Cloning into pKOV-unstuff

pKOV-unstuff was used in this study for the allelic replacement. pKOV-unstuff is the modified plasmid of pKOV with the removal of 3 kb fragment "stuffer sequence". It has a size of 5.7 kb with chloramphenicol antibiotic selection marker and pSC101-ts temperature-sensitive origin of replication. With the temperature-sensitive origin of replication, cells can be screened for only chromosomal integration of the vector. BglII restriction site in the MCS was used for the cloning. Overnight cultures of pKOV-unstuff in *E. coli* strain TOP-10 were grown at 30°C and used to isolate pKOV-unstuff. The plasmid was isolated using the Qiagen miniprep kit and restricted with *BglII* to get it linearized. Then it was treated with CIAP (calf intestine alkaline phosphatase) to avoid self-ligation. The linear plasmid was incubated for 1.5 hours more at 37°C and inactivated the enzyme at 85°C for 5 minutes. The fragments were also digested with BglII from respective pCR8 vectors and ligated into pKOV-unstuff with the insert:vector volume ratio of 1:1. The ligation mixture was incubated at room temperature for 15 minutes and then at 4°C overnight. The ligation mixture was directly used for transformation in strains B REL606 and REL607.

2.3.6   Constructs from GenScript

In the case of the *lepA* gene candidate, the distance between coevolving positions was 100 amino acids, and constructs were *de novo* synthesized (GenScript). Plasmids with substitutions (both singles and double) were used to transform the B strains directly. Results were confirmed using Sanger sequencing.

2.3.7 Chemically competent cells

The fresh plate of *E. coli* strains B REL606, and REL607 were streaked from the glycerol stock. A single pure colony was picked and incubated in liquid LB medium in a 37°C incubator with shaking. The next day 3 mL of the overnight grown culture was added to 300 mL LB medium in a flask and placed in the shaking incubator at 37°C until an OD600 of 0.8 was reached. The culture flask was kept on ice for 30 minutes. After this, the culture was transferred to sterile disposable 50 mL falcon tubes in sterile conditions and spun at 4000 rpm at 4°C for 10 minutes to pellet down the cells. The pellet was dissolved in 20 mL of 0.1 M $MgCl_2$ and was centrifuged at 4000 rpm at 4°C. The pellet was again dissolved gently in 20 mL of 0.1 M $CaCl_2$, kept on ice for 30 minutes, and centrifuged under the same conditions. After discarding the supernatant, the pellet was dissolved in 10 mL of $CaCl_2$, centrifuged with the same conditions. Lastly, the pellet was re-suspended in 3-5 mL of 0.1 M $CaCl_2$ and filtered glycerol with a ratio of 3:1, respectively. The competent cells were frozen in aliquots of 100 μL at -80°C freezer.

2.4 Reconstruction of ancestral genotypes

The confirmed plasmids for all the mutants of all three of the genes under study were used to transform chemically competent cells of *E. coli* B REL606 and REL607 strains. Plates were incubated at 30°C for 36 to 48 hours. The colonies were picked to perform crossing over steps in the allelic replacement phase.

1x Crossover for the allelic replacement

A single colony was picked from the LB-agar plate for all the mutants and re-suspended in the liquid LB (pre-heated at 43°C). All dilutions down to 10^{-5} were plated on pre-heated LB-agar plates and placed in 43°C incubator for 36-48 hours. Checked all the plates for the colonies and at least three distant colonies from dilution 10^{-3} or 10-4 were carefully selected and used for the second Crossover. The schema for generating the mutant strains is shown in the figure 2.2 (A).

2x Crossover for the allelic replacement

Colonies from 1x Crossover were re-suspended in freshly prepared -NaCl LB medium (without NaCl) for all the mutants. All the dilutions (10^{-3} - 10^{-5}) were plated and incubated at 30°C. These plates presumably have a ratio of 50:50 for the wild type and mutant. Eight colonies from the most diluted plate were used to purify on LB with streptomycin agar plates. Colony PCR was

done using check primers and proceeded for the Sanger sequencing.

Verification of mutant strains

Substitutions for each mutant in both genetic backgrounds (*E. coli* Strain B REL606 and REL607) were confirmed with the help of Sanger sequencing. All mutant strains were stored as glycerol stocks in -80°C freezer for long-term storage. All the strains with their genotypes are shown in the table 2.2.

2.5 Fitness assays

2.5.1 Environments for the competition experiments

Two environments were used in this study to estimate the fitness of the mutants. We included the nutrient-enriched medium and glucose limiting medium. The composition of both media is given below (Also table 2.5 for the composition of media and buffers used in this study).

LB (Luria-Bertani) medium

Nutrient enriched medium LB was used for the competition experiments. The composition of LB liquid medium used in this study for 1 liter was 10 g of Tryptone, 1 g of Yeast extract, and 5 g of Sodium Chloride (NaCl). LB medium was one competing environment that was used in this study. LB agar plates were also made with the same recipe with the addition of 16 g agar.

M9 minimal medium

Another competing environment was the M9 minimal medium with the concentration of 10 % glucose and 2.5X M9 salts (Sigma catalog no. M6030) including ($MgSO_4$, Na_2HPO_4, KH_2PO_4, NH_4Cl, and NaCl). For the TA plates 5 %, TTC (Triphenyltetrazolium chloride) was used in the recipe of agar plates for 1 liter (Tryptone 10 g, Yeast extract 1 g, NaCl 5 g, Agar 16 g, and Arabinose 10 g per liter of medium).

2.5.2 Competition Experiments

The strains were streaked from the glycerol stocks on LB-agar plates and incubated at 37°C overnight in a shaking incubator. The following day, a single colony was inoculated and grown in a 5 ml standard culture tube overnight using the LB liquid medium or reduced M9 medium. After 16-24 hours, the volume of 100 μl from each strain was mixed with the 100 μl of competing

strain (for example, 100 μl of wildtype strain was mixed with 100 μl of double mutant). 4 ml of LB medium or M9 minimal medium was inoculated with 4 μl of inoculum from well-mixed (vortex both cultures properly) overnight culture tube. This time point was set as T0 (start of the competition experiment). Also, 20 μl of newly inoculated culture was used to make dilutions down to 10^{-3}. This dilution was plated on the TA plate and incubated at 37°C overnight. All the dilutions were made in 96-well plates with the final volume of 200 μl in the minor dilution. Colonies for T0 were counted on the next day, and 20 μl was used to make dilution (10^{-5} for M9 minimal medium and 10^{-6} for LB medium) and plated on TA agar plates for 24hour time interval. Another fresh 4 ml of LB medium or M9 minimal medium was prepared and inoculated with 80 μl from 24 hours old competition for the transfer of a 48 hours competition. The next day, 20 μl was used to make dilution (10^{-5} for M9 minimal medium and 10^{-6} for LB medium) and plated on TA agar plates for the 48 hours time interval colony count. Colonies from the 24 hours time interval were counted for all the combinations of competition experiments. Another fresh 4 ml of LB medium or M9 minimal medium was prepared and inoculated with 80 μl from 48 hours old competition for the transfer of 72 hours competition. Colonies from the 48 hours time interval were counted for all the combinations of competition experiments. At the end of the 72 hours time interval, the competition experiment was finished, and 20 μl was used to make dilution and plated on TA agar plates for a 72 hours time interval, and next-day colonies were counted for all the competition experiments. All the dilutions were made in the sterilized 96-well plates, and 100 μl volume was used to plate for all the plating on TA plates. The experimental setup is shown in the illustration 2.2 (B). We used relative fitness to estimate the fitness of mutants compared with wild type and double mutant. The relative fitness $r_{wt,m}$ of the wild type wt to the mutant m is given by the formula:

$$r_{wt,m} = \frac{M_{wt}}{M_m} = \frac{\frac{F_{wt}(t0)}{F_{wt}(t24)}}{\frac{F_m(t0)}{F_m(t24)}}, \tag{2.1}$$

where M_{wt} and M_m are the Malthusian parameters of the wild type and mutant, and $F_{wt}(t)$ and $F_m(t)$ are the colony counts at time t of the wild type and mutant, respectively.

2.5.3 Growth and storage conditions of strains

E. coli strains for the cloning purposes were grown in Luria-Bertani (LB) broth or with agar for the agar plates (BERTANI, 1951) for 16-18 hours at 37°C in the shaking incubator. These growth conditions were for the standard growth of the strains except specific conditions that were applied for the later experiments (for example, 2x Crossover in allelic replacement needed 43°C

Figure 2.2 A. The illustration for the construction of mutant genotypes. B. Competition experiment layout

A

Overlap extension PCR

Cloning into plasmid

Allelic replacement
(homologous recombination)

Confirmation of mutant strain
with Sanger sequencing

B

Streaking of all strains from stocks

overnight cultures in medium

equal ratio of both strains to start
competition

competition till 72 hours with two
transfers

colony count at each time interval
(0-72 hours)
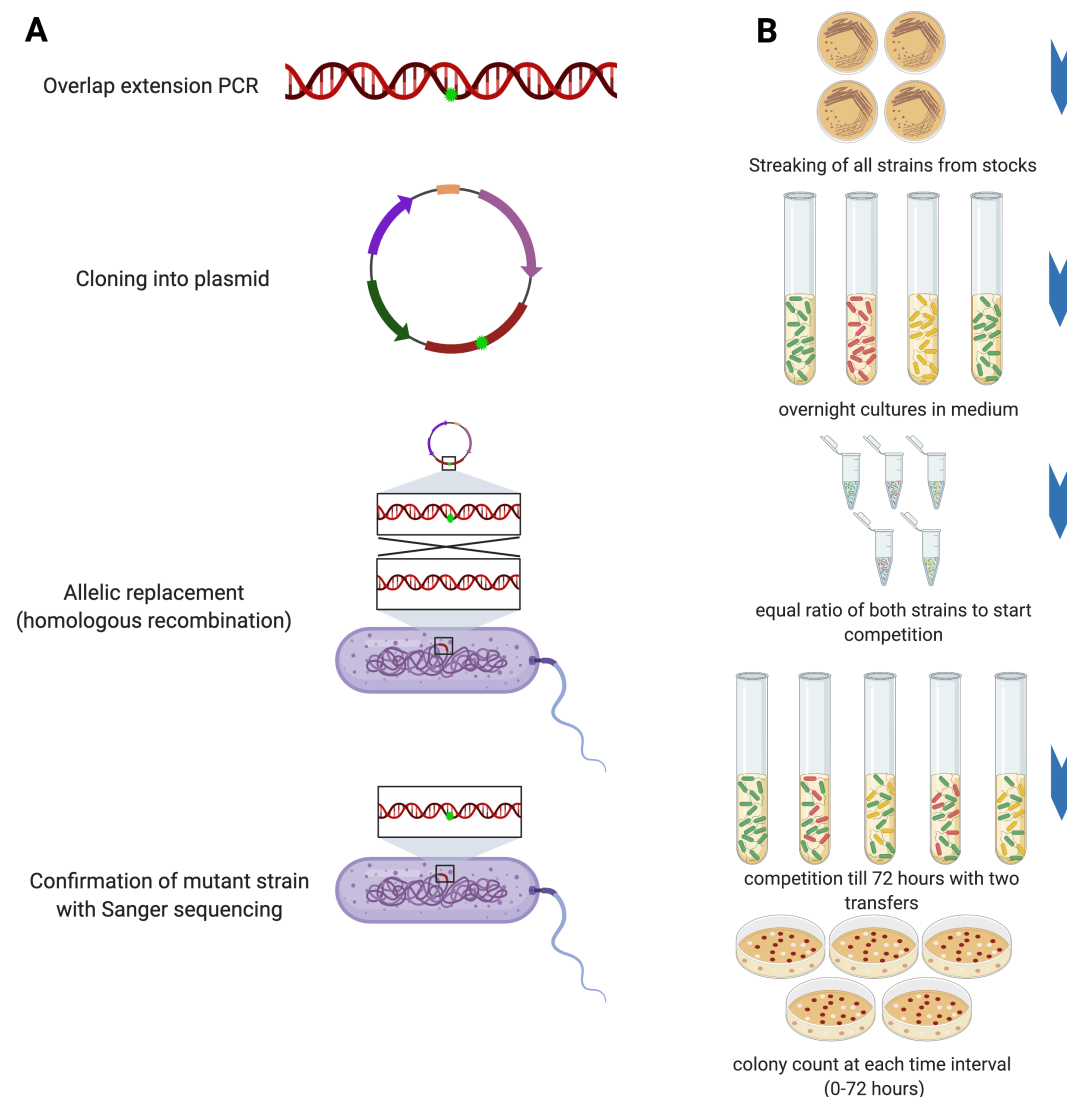
). Glycerol stocks for all the mutants constructed in this study were stored in -80°C freezer for long-term storage. Overnight liquid cultures of the mutant strains were mixed with autoclaved 50 % glycerol in the same ratio (1:1) and immediately stored in -80°C freezer.

2.6 Expression of the candidate genes

We assessed the expression of candidate genes with PCR amplification from total extracted RNA. Total RNA was used to synthesize cDNA, and this cDNA was used as a template to amplify candidate genes (isolated from all mutant strains for all the candidate genes).

2.6.1 RNA isolation

Overnight liquid cultures of all the mutants (3 mutants for each candidate and wildtype) were grown at 37°C. 0.5 ml of the liquid culture was used to mix it with 1 ml of RNAprotect Bacterial Reagent (Qiagen, catalog no. 76506). The tubes were vortexed and centrifuged at 5000 x g for 10 minutes. The pellet was used to mix it with 100 μl of TE buffer containing lysozyme in it. 10 μl of proteinase K was also added to the sample and homogenized with the help of a pipette tip. Samples were incubated at room temperature for 10 minutes (vortex it in intervals). Samples were used for the RNA isolation with the described protocol in the RNeasy Plus kit (Qiagen, catalog no. 17134). At the end of the protocol, the eluted volume was used to mixed it with 8 M Lithium Chloride (LiCl) and spun the microtubes at 10000 rpm for 5 minutes at 4°C. The pellet was washed with 70 % ethanol and dried at room temperature. The dried pellet was dissolved in RNA-free water. This additional step of precipitation with LiCl made sure that gDNA was not present in the sample. The concentration was determined on the Nanodrop and proceeded for cDNA synthesis. The remaining RNA samples were stored at -80°C freezer for long-term storage.

2.6.2 cDNA synthesis

RNA samples were used to make cDNA synthesis using the QuantiTect Reverse Transcription Kit with the described protocol (catalog no. 205311). The cDNA was used as a template to amplify the candidate genes in PCR. The PCR amplification was done using the gene-specific primers and Phusion polymerase. 1 % agarose gel was prepared to assess the amplified gene band size with the DNA marker.

2.7 Whole-genome Sequencing

Overnight cultures of all mutant strains with the wildtype and ancestral strains (the strain used to make the mutant strains) were used to grow in LB liquid with streptomycin 100μg/ml. The following day, these liquid cultures were used for the extraction of genomic DNA using the GenElute Bacterial Genomic DNA kit (catalog no. NA2110-1KT). The DNA concentration was determined on Nanodrop. The whole-genomes of all the mutants were sequenced using Next Generation Illumina sequencing. The read length was 150 x 2 with a coverage of 100x. We mapped the sequencing reads of all mutant strains against REL606 (the reference genome for *E. coli strain B*) using breseq [Jeong et al., 2009]. Mapped sequences of all the mutants were

analyzed to identify candidate point mutations using the `breseq` pipeline [Deatherage and Barrick, 2014].

Table 2.1:

Short-listed candidate groups in the genes for the experimental evaluation. The highlighted candidates are analyzed in this study.

Serial no	Gene	Biochemical index	Ancestral states	*E. coli* positions
1	trans editing enzyme ProX	Ind1	TYR84;GLN151	LEU84;GLU151
2	ribosomal back translocase	Ind1	THR66;ASN426	CYS66;SER426
3	ribosomal back translocase	Ind2	THR66;MET333	CYS66;GLN333
4	phosphofructokinase	Ind4	CYS123;TYR217	PHE123;SER217
5	Serine Hydroxymethyl transferase	Ind4	ALA97;ARG249	ARG97;GLY249
6	Lysyl-tRNA synthetase	Ind5	VAL141;GLU175	ALA141;THR175
7	Protein yebC	Ind7	ASN111;ILE112	THR111;LEU112
8	trans editing enzyme ProX	23;112	LYS23;ILE112	MET23;ALA112
9	trans editing enzyme ProX	117;125	LEU117;ARG125	ALA117;VAL125
10	ribosomal back translocase lepA	Charge	LYS104;GLU208	SER104;GLN208
11	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	Charge	ARG69;GLU72	ASN69;ALA72
12	Serine Hydroxymethyl transferase	Volume	ALA97;ARG249	ARG97;GLY249
13	RNA polymerase associated protein	Volume	VAL797;GLU835	ILE797;ASP835

Table 2.2: Bacterial strains used in this study

Name	Genotype	Source
E. coli DH5α-λpir	supE44, δlacU169, hsdR17, recA1, endA1, gyrA96, thi- 1, relA1, λpir	Life Technologies, [Simon et al., 1983]
E. coli TOP10	F- mcrA δ(mrr-hsdRMS-mcrBC) ϕ80lacZ δM15 δlacX74 recA1 araD139 δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG	Life Technologies, [Simon et al., 1983]
E. coli B REL606	The ancestral strain used in Richard Lenski long term evolution experiment (Ara$^+$ Red colonies on the TA indicator plates).	[Daegelen et al., 2009]
E. coli B REL607	Derived strain of B REL606 having two point mutations at arabinose marker and recA (Ara$^-$ White colonies on TA indicator plates).	[Daegelen et al., 2009]
Wildtype-606 Ara$^+$	The wildtype strain of present sequence of *E. coli* Strain B REL606 used in the allelic replacement experiment as a method control. Red colonies on the TA indicator plates.	In this study
Wildtype-607 Ara$^-$	The wildtype strain of present sequence of *E. coli* Strain B REL607 used in the allelic replacement experiment as a method control. White colonies in the TA indicator plates.	In this study

Name	Genotype	Source
N24R-IspH Ara$^+$	Asparagine was replaced with Arginine (predicted coevolving amino acids) in *ispH* gene called as single mutant in the background of *E. coli* strain B REL606. Red colonies on the TA indicator plates.	In this study
N24R-IspH Ara$^-$	Asparagine was replaced with Arginine (predicted coevolving amino acids) in *ispH* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study
A27E-IspH Ara$^+$	Alanine was replaced with Glutamic acid (predicted coevolving amino acids) in *ispH* gene called as a single mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study
N24R-A27E-IspH Ara$^+$	The inferred ancestral state having both substitutions in *ispH* gene also called as double mutant in both the background of *E.coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study
N24R-A27E-IspH Ara$^-$	The inferred ancestral state having both substitutions in *ispH* gene called as double mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study

Name	Genotype	Source
S66K-LepA Ara$^+$	Serine was replaced with Lysine (predicted coevolving amino acids) in *lepA* gene called as a single mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study
S66K-LepA Ara$^-$	Serine was replaced with Lysine (predicted coevolving amino acids) in *lepA* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study
Q170E-LepA Ara$^+$	Glutamine was replaced with Glutamic acid (predicted coevolving amino acids) in *lepA* gene called as a single mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates	In this study
Q170E-LepA Ara$^-$	Glutamine was replaced with Glutamic acid (predicted coevolving amino acids) in *lepA* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study
S66K-Q170E-LepA Ara$^+$	The inferred ancestral state having both substitutions in *lepA* gene called as a double mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study

Name	Genotype	Source
S66K-Q170E-LepA Ara⁻	The inferred ancestral state having both substitutions in *lepA* gene called as a double mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study
T65N-YebC Ara⁺	Threonine was replaced with Asparagine (predicted coevolving amino acids) in *yebC* gene called as a single mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study
T65N-YebC Ara⁻	Threonine was replaced with Asparagine (predicted coevolving amino acids) in *yebC* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study
L66I-YebC Ara⁺	Leucine was replaced with Isoleucine (predicted coevolving amino acids) in *yebC* gene called as a single mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study
L66I-YebC Ara⁻	Leucine was replaced with Isoleucine (predicted coevolving amino acids) in *yebC* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study

Name	Genotype	Source
T65N-L66I-YebC Ara$^+$	The inferred ancestral state having both substitutions in *yebC* gene called as a double mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.	In this study
T65N-L66I-YebC Ara$^-$	The inferred ancestral state having both substitutions in *yebC* gene called as double mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.	In this study.

Table 2.3: DNA Primers used in this study

Name	Description	Reference
IspH-F1	GAagatctCGTCCTGGTGCACTTCACG	for the step 1 of SOE-PCR
IspH-R1	GAagatctCGTCGATCACATCAGACGTGTCATC	for the step 1 of SOE-PCR
IspH-mut-m1-F2	CATTGTTGAACGCGCGCTGGCCATTTAC	for the step 2 PCR with substitution
IspH-mut-m1-R2	GTAAAATGGCCAGCGCGCGTTCAACAATG	for the step 2 PCR with substitution
IspH-mut-m2-F2	CATTGTTGAAAACGCGCTGGAGATTTAC	for the step 2 PCR with substitution
IspH-mut-m2-R2	GTAAAATCTCCAGCGCGTTTCAACAATG	for the step 2 PCR with substitution
IspH-mut-anc-F2	CATTGTTGAACGCGCGCTGGAGATTTAC	for the step 2 PCR with substitution
IspH-mut-anc-R2	GTAAAATCTCCAGCGCGCGTTCAACAATG	for the step 2 PCR with substitution
YebC-F1	GaagatctACAGGCAGGGCTATCTGGAG	for the step 1 PCR
YebC-R1	GaagatctCACGCGTCTCCGTTTGCTGTTTAG	for the step 1 PCR
YebC-mut-m1-F2	CATGACCCGCGACAACCTGAAC	for the step 2 PCR with substitution
YebC-mut-m1-R2	GTTCAGGTTGTCGCGGGTCATG	for the step 2 PCR with substitution
YebC-mut-m2-F2	GACCCGCGACACAATTAACCGC	for the step 2 PCR with substitution
YebC-mut-m2-R2	GCGGTTAATTGTGTCGCGGGTC	for the step 2 PCR with substitution
YebC-mut-anc-F2	GCGACAACATTAACCGCGCAATTG	for the step 2 PCR with substitution
YebC-mut-anc-R2	CAATTGCGCGGTTAATGTTGTCGC	for the step 2 PCR with substitution
LepA-F	GATCATAGCTCACATTGAC	Check primers for sanger sequencing
LepA-R	CGTATACCTGCGGGTTTGAC	Check primers for sanger sequencing

Table 2.4: Plasmids used in this study

Name	Description	Source
pCR8/ GW/ TOPO	Entry vector for the cloning of substituted fragment of DNA to be used for the confirmation of substitutions.	Life Technologies, [Simon et al., 1983]
pUC57	Entry vector for the cloning of substituted fragment of DNA to be used for the confirmation of substitutions.	GenScript Biotech
pKOV-unstuff	Cloning vector for the homologous recombination using antibiotic selection and SacB counter selection.	Jenna Gallie's group, [Link et al., 1997]
pCR8-IspH_N24R_A27E	Entry vector for the N24R_A27E substituted gene fragment of *ispH* gene to be used for the confirmation of substitutions.	In this study
pCR8-IspH_N24R	Entry vector for the N24R substituted gene fragment of *ispH* gene to be used for the confirmation of substitutions.	In this study
pCR8-IspH_A27E	Entry vector for the A27E substituted gene fragment of *ispH* gene to be used for the confirmation of substitutions.	In this study
pCR8-YebC_T65N_L66I	Entry vector for the T65N_L66I substituted gene fragment of *yebC* gene to be used for the confirmation of substitutions.	In this study
pCR8-YebC_T65N	Entry vector for the T65N substituted gene fragment of *yebC* gene to be used for the confirmation of substitutions.	In this study

Name	Description	Source
pCR8-YebC_L66I	Entry vector for the L66I substituted gene fragment of *yebC* gene to be used for the confirmation of substitutions.	In this study
pUC57-S66K_Q170E	Entry vector for the S66K_Q170E substituted gene fragment of *lepA* gene to be used for the confirmation of substitutions.	In this study
pUC57-S66K	Entry vector for the S66K substituted gene fragment of *lepA* gene to be used for the confirmation of substitutions.	In this study
pUC57-Q170E	Entry vector for the Q170E substituted gene fragment of *lepA* gene to be used for the confirmation of substitutions.	In this study
pKOV-IspH_N24R_A27E	pKOV_unstuff vector for the N24R_A27E substituted gene fragment of *ispH* gene to be used for the allelic replacement experiment.	In this study
pKOV-IspH_N24R	pKOV_unstuff vector for the N24R substituted gene fragment of *ispH* gene to be used for the allelic replacement experiment.	In this study
pKOV-IspH_A27E	pKOV_unstuff vector for the A27E substituted gene fragment of *ispH* gene to be used for the allelic replacement experiment	In this study
pKOV-YebC_T65N_L66I	pKOV_unstuff vector for the T65N_L66I substituted gene fragment of *yebC* gene to be used for the allelic replacement experiment.	In this study

Name	Description	Source
pKOV-YebC_T65N	pKOV_unstuff vector for the T65N substituted gene fragment of *yebC* gene to be used for the allelic replacement experiment.	In this study
pKOV-YebC_L66I	pKOV_unstuff vector for the L66I substituted gene fragment of *yebC* gene to be used for the allelic replacement experiment.	In this study
pKOV-S66K_Q170E	pKOV_unstuff vector for the S66K_Q170E substituted gene fragment of *lepA* gene to be used for the allelic replacement experiment.	In this study
pKOV-S66K	pKOV_unstuff vector for the S66K substituted gene fragment of *lepA* gene to be used for the allelic replacement experiment.	In this study
pKOV-Q170E	pKOV_unstuff vector for the Q170E substituted gene fragment of *lepA* gene to be used for the allelic replacement experiment.	In this study

Table 2.5 Growth media and buffers used in this study

Name	Ingredients	Application
TAE buffer (1X)	40 mM Tris acetate, 2 mM EDTA, pH 8	Life Technologies, Agarose gel running buffer
LB media for 1 L	20 g LB-dry powder (Invitrogen, catalog no. 12780052), ddH20 to 1 L, pH 7.0	LB broth liquid
LB agar plates	32 g LB-dry powder(Invitrogen, catalog no. 22700025), ddH20 to 1 L	LB agar plates
TA indicator agar plates for 1 L	Tryptone 10 g, Yeast extract 1 g, Sodium chloride 5 g, Agar 16 g, Arabinose 10 g, and TTC (5%) 1 mL	TA indicator agar plates
M9 minimal media for 1 L	5x Salts 200ml (Merck, catalog. no M6030), 20% glucose 20ml, 1 M $MgSO_4$ 2 ml, 1 M $CaCl_2$ 0.1 ml	Defined growth medium
Washing buffers	0.1 M $CaCl_2$ and 0.1 M $MgCl_2$	Chemically Competent cells
DNA loading buffer	30% (v/v) glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF	DNA loading buffer for gel electrophoresis

Results

Chapter 3

Charge-compensating mutation in the EF4 protein

The first candidate that we tested experimentally is detected in the *lepA* genes, which encodes the elongation factor-F protein. The predicted coevolving sites in this gene are Ser (S) at position 66 and Gln (Q) at position 170 from *E. coli*. These positions were predicted as coevolving because of charge compensation. After reviewing the known structural and functional properties of the *lepA* gene, I report the bioinformatic evidence for coevolution at these positions. I then report the competition experiment results between the reconstructed possible ancestors and the wild type strain and the resulting inferred local fitness landscape.

3.1 Role of Elongation factor F in translation

The *lepA* gene encodes the elongation factor 4 protein. It is involved in the back translocation of tRNAs on erroneous translocated ribosomes. The process of back translocation occurs in the elongation cycle of the protein translation. The function of ribosomes is categorized into four phases: initiation, elongation, termination of the proteins, and the recycling phase. In the recycling phase, ribosomes are separated into their subunits so that the small subunit initiates the re-entry into the subsequent initiation phase [Michel and Baranov, 2013]. Each phase is regulated by specific proteins termed as factors. While the translation phases significantly differ in the three domains of life, Bacteria, Archaea, and Eukarya, the elongation phase is similar in all three domains and is a crucial phase of the translation process. The elongation phase consists of a cycle of biochemical reactions called the "elongation cycle" (Figure 3.1). The foremost step is to prolong the nascent polypeptide by one amino acid. The elongation cycle is regulated by two factors: EF-Tu and EF-G in bacteria and EF1 and EF2 in Archaea and Eukarya. EF-Tu transports an aminoacyl-tRNA (aa-tRNA) in the ternary complex aa-tRNA—EF-Tu—GTP to the decoding center of the ribosomal A site (A for aminoacyl-tRNA) on the small ribosomal

Figure 3.1 Elongation cycle. The illustration is created by `https://biorender.com/`
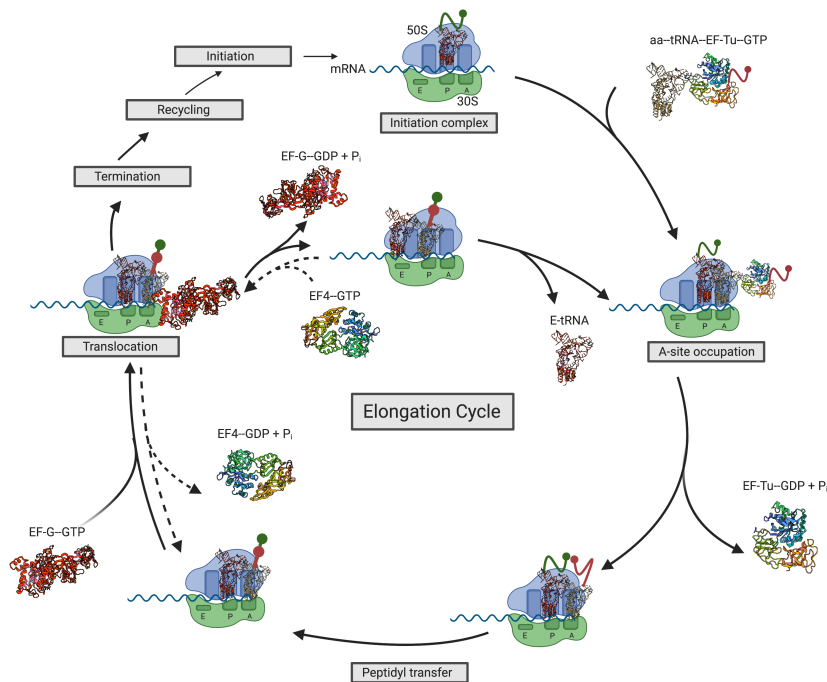
subunit. After this decoding, EF-Tu hydrolyzes GTP and leaves the ribosome like EF-Tu–GDP. The attached aa-tRNA integrates fully into the A site of the ribosome. The peptide bond forms in the next step and does not require a translation factor. During this next step, the peptidyl-tRNA residue at the ribosomal P site splits from the peptidyl-tRNA and transfers to the aa-tRNA. The splitting gives the peptidyl-tRNA residing at the A site and being prolonged by one amino acid. The third step in the elongation cycle is the translocation reaction and operates by EF-G—GTP. The tRNA–mRNA complex moves along on a codon length on the ribosome (moving the peptidyl-tRNA from A site to P site and the deacylated tRNA from the P site to E site, the Exit site).

In bacteria, the third unique elongation factor termed as elongation factor 4 "EF-4" has been reported by March and Inouye [1985]. First, it was called LepA in *E. coli* based on its function. The *lepA* gene is present upstream of the Lep protein on the Lep operon. The Lep protein is a peptidase that cleaves the signal peptide on the N terminus of proteins after translocation through the membrane. Whereas knockout of *lepA* had not shown any significant reduction in protein transportation [March and Inouye, 1985]. Dibb and Wolfe [1986] and colleagues have reported no phenotype of *lepA* in *E. coli* under different growth environments. They also showed that no growth effects were found in the knockout strain of LepA. However, in a medically

important pathogen *Helicobacter pylori* 10 genes were analyzed for the disease spread (stomach *ulcera*), including the *lepA gene*. These ten genes are essential for survival in the low-pH of the stomach [Bijlsma et al., 2000]. The work mentioned above attracted the attention to study the function and structure of the *lepA* gene.

The elongation factor EF4 encoded by the *lepA* gene was found to be a highly conserved protein in all bacteria [Qin et al., 2006; Liu et al., 2010]. Functional studies of EF4 revealed the following points:

1. the ribosome-bounded EF4 has a GTPase activity [Liu et al., 2010],

2. EF4 is responsible for the back-translocation of post-translational complexes (tRNA2–mRNA) from E and P sites to P and A sites, respectively,

3. EF4 increases the active fraction of newly translated proteins [Qin et al., 2006].

Earlier studies characterized the function of EF4 as preventing the misincorporation of the amino acid during translocation [Qin et al., 2006]. The authors in the study, as mentioned above, suggested developing a homologous model of EF4 in other species based on the domain similarity of the known structure of EF G. Pech et al. [2011] later suggested that EF4 played a role in back-translocation and recognition of stalled ribosomes, remobilization, and reactivation of the protein synthesis. This study was done using a stress condition of a high concentration of Mg^{+2}, which showed that EF4 could not perform its proper function.

Crucial genes such as *lepA* are good candidates because of their vital role in the cellular processes. *lepA* plays an essential role in the cellular translationary machinery and an adequate candidate for this coevolutionary predictions' study. The predicted coevolving residues in EF4 are at amino acids positions 66 and 170 in the *E. coli* structure (PDB ID:3CB4), with amino acids Ser and Gln, respectively. The pair was predicted to coevolve because of charge compensation. We reconstructed the ancestral states for both positions using the DNA sequence and accounted for the codon usage bias (See Chapter 2, section 2.2). The reconstructed amino acids' ancestral states are LYS at position 66 and Glu acid at position 170, leading to the possible evolutionary scenarios displayed in Figure 3.2. Strains containing the putative ancestral genotypes were reconstructed in *E. coli* Strain B:

- the strain with the ancestral states at the two positions was reconstructed by introducing the double mutation S66K-Q170E,

- the two possible intermediate genotypes were reconstructed by introducing each of the two single mutations S66K and Q170E.

Figure 3.2 Reconstructed ancestral states and putative compensatory mutations on the *E. coli* lineage for the *lepA* gene. Inferred ancestral states are displayed on the left, and the *E. coli* (wild type) sequence on the right. The arrow represents the putative mutations and their order at both sites. The three reconstructed strains used in this study are indicated as a double mutant, single mutant 1, and single mutant 2.
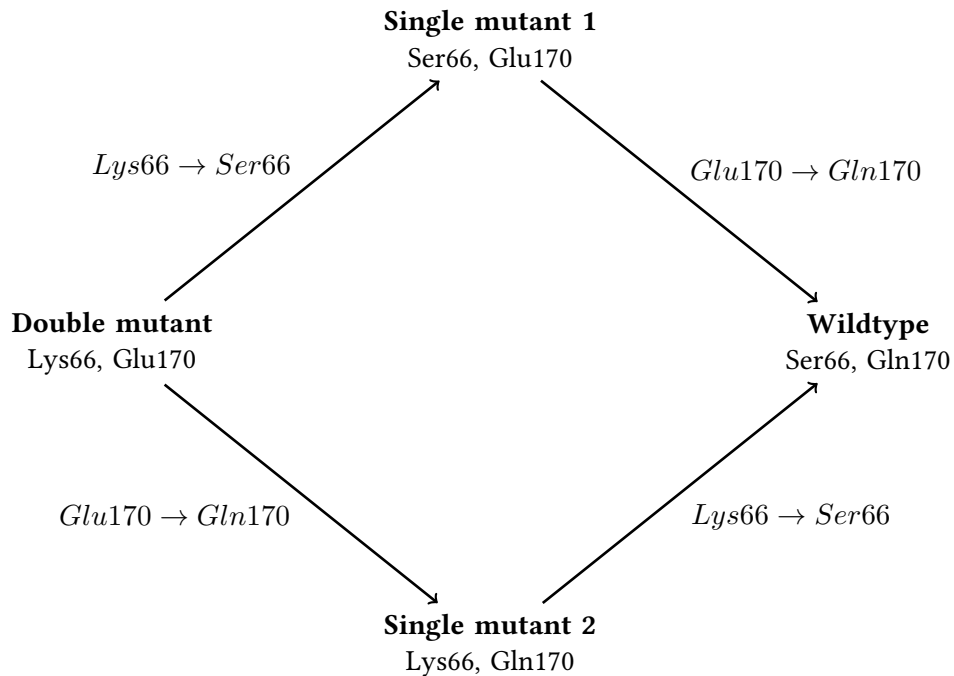
We then analyzed the fitness of both single mutants in two competition environments, one in the M9 minimal medium and the other in the LB nutrient enriched medium. We estimated the relative fitness of single mutants by comparing them with the double mutant and wild type for 72 hours. We then inferred the local fitness landscape for the two positions.

3.2 3D structure, Phylogenetic analysis, and evidence for coevolution by compensation

The 3D structure of the EF4 protein is shown in Figure 3.3. EF4 is a 599 amino acid protein with a molecular weight of 67 kDa. The 2.8Å resolution crystal structure of EF4 from *E. coli* [Evans et al., 2008] revealed the structural homology of individual domains of LepA with EF G, except for the 130 amino acid long domain IV. The primary distance between these two coevolving positions was 104 amino acids. However, these two positions were in close proximity in the 3D structure with a distance of 8.91Å between alpha carbons. Two hundred seventy protein sequences were included in the analysis of coevolution for this bacterial homologous protein family. We found eighteen compensatory mutations for charge biochemical property on the

Figure 3.3 3D structure of EF 4 chain A (PDB ID: 3CB4). Orange highlighted residues (SER66 and GLN170) are the predicted coevolving amino acids.
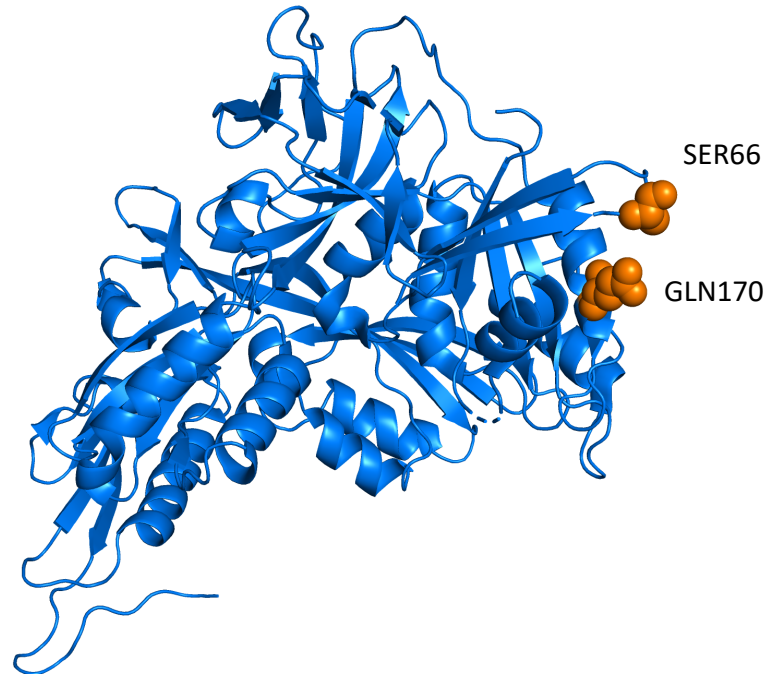
branches of the phylogenetic tree of *lepA* gene, *i.e.*, positively charged amino acid compensated by negatively charged amino acid and *vice versa* (Table 3.1). Sequences of all the *E. coli* strains were included and analyzed the frequencies of amino acids present on these positions. Ser and Gln at positions 66 and 170 (coordinates in PDB ID: 3CB4) are highly conserved in all 260 of the

Table 3.1 Frequency of observed amino-acid state pairs among bacterial species. The highlighted one is the pair observe in *E. coli* strains

Pairs of Amino acids	Frequency in all Species	Frequency in *E. coli* Strains
S;Q	**1**	**260**
K;E	167	0
K;D	14	0
D;K	13	0
R;E	10	0
Other Pairs, each in low frequency	65	0

E. coli strains.

We built the compensogram and phylogenetic tree for EF4 at sites 66 and 170 (Figure 3.4). Blue highlighted branches indicate negative change, i.e., the charge is reduced ($+ \rightarrow -$), and red highlighted branches indicate more positive change, i.e., the charge increases ($- \rightarrow +$). The figure shows evidence for coevolution by charge compensation for positions 66 and 170. In this figure 3.4 **A** shows the changes of charge, as inferred by weighted substitution mapping, on the phylogenetic tree of the protein family (bacterial species names have been omitted for clarity. The coevolution signal is widespread along the phylogeny, suggesting that the constraint acting on these sites is conserved throughout the bacteria. The majority of observed mutations involve a change from positive to negative at site 66, compensated by a mutation from negative to positive at site 170. The inferred compensating mutations on the *E. coli* branch are unique and not observed on any other branches of the tree, the Ser and Gln state pair being only present in *E. coli*. Interestingly, no polymorphism was observed in the 261 strains of *E. coli* for which a sequence was available for the *ispH* gene. The conserved sequences in all strains of *E. coli* suggests that the mutation are fixed in the population.

Table 3.2 Mutant strains genotypes for the EF4 candidate.

Strain	Genotype
Wildtype-606 Ara⁺	The wildtype strain of present sequence of *E. coli* Strain B REL606 used in the allelic replacement experiment as wild type. Red colonies on the TA indicator plates.
Wildtype-607 Ara⁻	The wildtype strain of present sequence of *E. coli* Strain B REL607 used in the allelic replacement experiment as wild type. White colonies in the TA indicator plates.
S66K-LepA Ara⁻	Serine was replaced with Lysine (predicted coevolving amino acids) in *lepA* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.
Q170E-LepA Ara⁻	Glutamine was replaced with Glutamic acid (predicted coevolving amino acids) in *lepA* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.
S66K-Q170E-LepA Ara⁺	The inferred ancestral state having both substitutions in *lepA* gene called as a double mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.

Figure 3.4 A). The phylogenetic tree of the EF4 protein family used in the CoMap analysis. Blue highlighted branches show negative change (more negative charge) at position 66. The red highlighted branches show the more positive charge at position 170. **B).** Compensogram: Branches are sorted according to the strength of the compensation signal, and only the top branches are represented. The red point indicates the compensation index for the branch, and the blue bars show the null distribution obtained for each branch by resampling all sites in the alignment. Point: mean, error bar: 95% interval of the distribution. The green line shows the expected average compensation after randomizing the branches independently for the two sites, and the green dashed line if the corresponding 95% quantile of this distribution
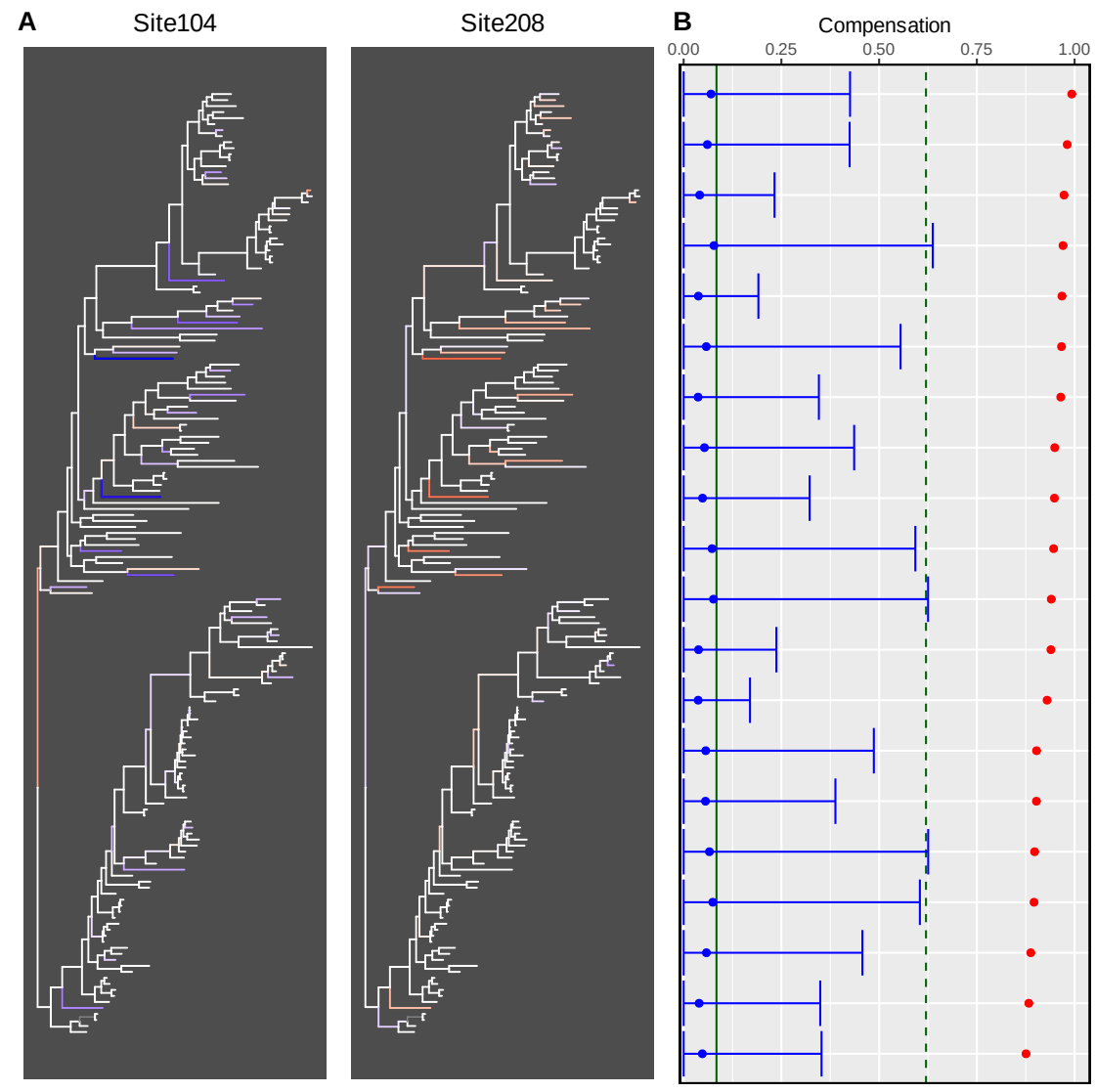
3.3 Construction of mutant strains

As mutations are far in the primary sequences, we designed the mutant DNA fragments with the desired mutations: a single mutant with a mutation from Ser to Lys at position 66, another single mutant mutation from Gln to Glu acid at position 170, and both mutations in the double mutant (see Figure 3.2). The GenScript Biotech then synthesized these fragments in high copy

number plasmids (pUK57-kan). Sanger sequencing results for the positive clones confirmed the presence of the desired substitution of the EF4 mutants. The scarless allelic replacement method successfully gave us all three mutant strains (Table 3.2 shows the strain name and their genotype). Whole-genome sequencing also confirmed no off-target substitutions after the homologous recombination step of generating mutants (Table 3.3). Breseq is a computational pipeline that allows detecting mutations in a given sample using a reference genome [Deatherage and Barrick, 2014]. It identified desired mutations in all three mutants, and an additional signal of the IS150 non-coding region was also detected in wildtype and S66K-LepA samples (Detailed table is shown in 8.2). The IS150 insertion is a non-coding insertion sequence. Breseq did not report any mutation in coding regions of all generated mutants, suggesting that any difference in phenotype between the tested strains may be unambiguously attributed to the introduced mutations. After 2x crossing over in allelic replacement, it generates two strains with mutations and without mutations. The strain without mutations is used as a wild type in the competition assays Chapter 2, section 2.4).

Table 3.3 Columns from the breseq output is shown here to show the desired mutations in the mutant strains.

Strain	Mutation	Gene	%[1]	Description
Wildtype-606	N/A	N/A	98.7	Repeat region
Wildtype-607	T $\rightarrow$ C	*araA*	89.4	L-arabinose isomerase
	A $\rightarrow$ G	*recD*	89.4	Exonuclease V (RecBCD complex) alpha chain
S66K-LepA-607	T $\rightarrow$ C	*araA*	98.5	L-arabinose isomerase
	A $\rightarrow$ G	*recD*	98.5	Exonuclease V (RecBCD complex) alpha chain
	3 bp $\rightarrow$ CCT	*lepA*	98.5	GTP-binding protein
Q170E-LepA-607	T $\rightarrow$ C	*araA*	98.6	L-arabinose isomerase
	A $\rightarrow$ G	*recD*	98.6	Exonuclease V (RecBCD complex) alpha chain
	G $\rightarrow$ C	*lepA*	98.6	GTP-binding protein
S66K-Q170E-LepA-606	3 bp $\rightarrow$ CCT	*lepA*	98.4	GTP-binding protein
	2bp $\rightarrow$ AG	*lepA*	98.4	GTP-binding protein

[1] % of the genome where reads could be mapped

3.4 Expression of mutant strains

In order to validate that potential phenotypic differences are not due to a lack of expression of the LepA protein in the reconstructed strains, the expression of *lepA* mutant genes was confirmed in all strains. For this, we extracted the whole RNA from all the mutants and wild type and synthesized cDNA. I checked for the gDNA contamination as PCR-based can give us false-positive results. So, we sensitive kit to eliminate gDNA from the RNA samples (See Chapter 2, section 2.6). PCR Amplification using *lepA* gene-specific primers gave us the band size of 800 bp from the mutant strains S66K, Q170E, S66K-Q170E, and wild type. The positive results of PCR showed that the *lepA* gene is being expressed in all the mutant strains, at least at the transcriptional level (Supplement Figure 8.1).

3.5 Competition experiments

We used two environments for the competition experiments: the cells were grown on the LB broth medium where nutrients were not limiting in the first environment. The second environment was the M9 minimal medium, a defined medium with a single carbon source (glucose). This medium is a stress condition used in competition experiments as it increases competition between the strains because of the limited resources. Smaller fitness effects might therefore be detected in such stressful environment. Since the *lepA* gene is directly involved in protein translation, however, differences in fitness effects might not be detected if the translation is not the limiting factor.

3.5.1 LB Broth medium

Competition experiments were conducted in five combinations of competing strains, and each combination was repeated eight times. These combinations were S66K-LepA vs. S66K-Q170E-LepA (single mutant 2 competing against double mutant), Q170E-LepA vs. S66K-Q170E-LepA (single mutant 1 competing against double mutant), S66K-LepA vs. Wild type (one single mutant competing against wild type), Q170E-LepA vs. Wild type (other single mutant competing against wild type. Finally, we tested S66K-Q170E-LepA vs. wild type. If the two mutations $Glu \rightarrow Gln$ at position 66 and $\rightarrow$ at position 170 are compensating, we expect the fitness of the single mutants to be lower than that of both the wildtype and double mutant. Comparing the fitness of the wildtype with that of the double mutant can further inform us about the magnitude of the compensation: if the wildtype has a similar fitness to that of the double mutant,

Figure 3.5 Competition experiments in LB Broth medium; Relative fitness is shown on the Y-axis, and three time points for the plating are on the x-axis. Lighter color boxplots show the combination of single mutants vs. double mutant. Darker color boxplots show the wild type vs. single mutants and wild type vs. double mutant. The red horizontal line is the reference line for zero fitness difference. One sample t-test is used for the statistics here (The mean of population $\mu = 1$). See Table 3.4 for exact p-values, and supp Table 8.3 and 8.4 for the raw data.
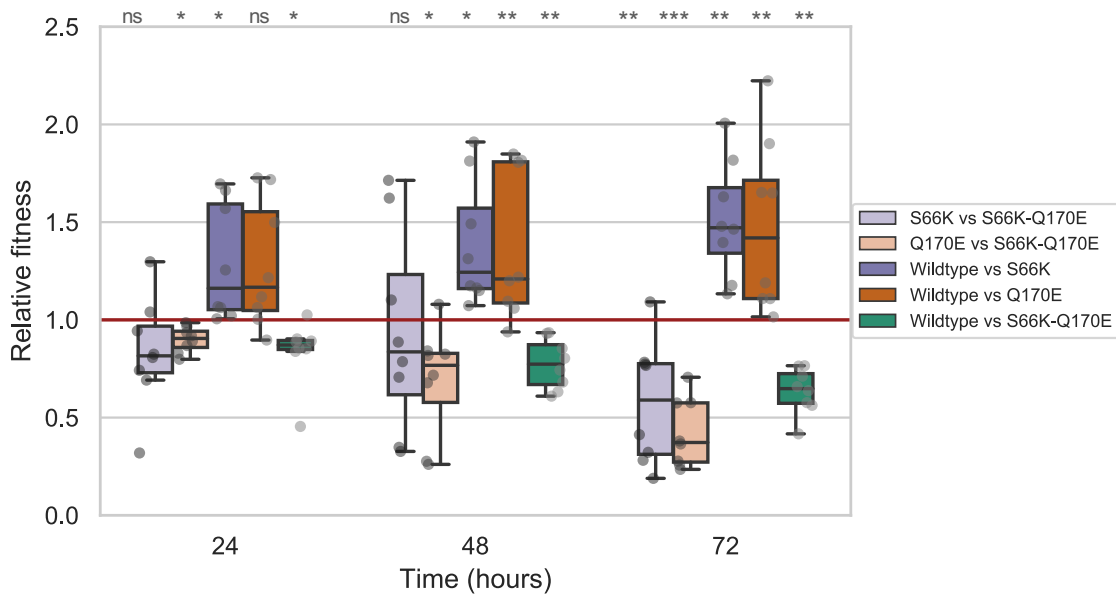
the mutations are perfectly compensating. If it is lower than the wild type, the second mutation was only partially compensating if it is higher than the wild type; the second mutation was overcompensating.

Each combination was plated on TA indicator plates and used for colony counts at 0, 24, 48, and 72 hours. The single mutant S66K-LepA competed against S66K-Q170E-LepA and the wild type for up to 72 hours. S66K-LepA against double mutant S66K-Q170E-LepA had no significant fitness difference at 24 hours and 48 hours (Figure 3.5). Later at 72 hours, single mutant S66K-LepA showed significantly lower fitness (*p value = 0.0072057*). Also, the wild type against S66K-LepA showed significantly higher fitness at all time points. Other single mutant Q170E-LepA showed significantly lower fitness at all time points (Figure 3.5). Both single mutants had a lower relative fitness compared to the double mutant or wild type. The double mutant had higher relative fitness than the wild type, suggesting that the second mutation was only partially compensating.

3.5.2 M9 minimal medium

We used the same experimental setup and a similar number of combinations of strains for competition experiments in M9 minimal as in the LB-Broth medium. M9 minimal media has a

Figure 3.6 Competition experiments in M9 minimal medium; Relative fitness is shown on the Y-axis, and three time points for the plating are on the x-axis. Lighter color boxplots show the combination of single mutants vs. double mutant. Darker color boxplots show the wild type vs. single mutants and wild type vs. double mutant. The red horizontal line is the reference line for zero fitness difference. One sample t-test is used for the statistics here (The mean of population $\mu = 1$). See Table 3.4 for exact p-values, and supp Table 8.3 and 8.4 for the raw data.
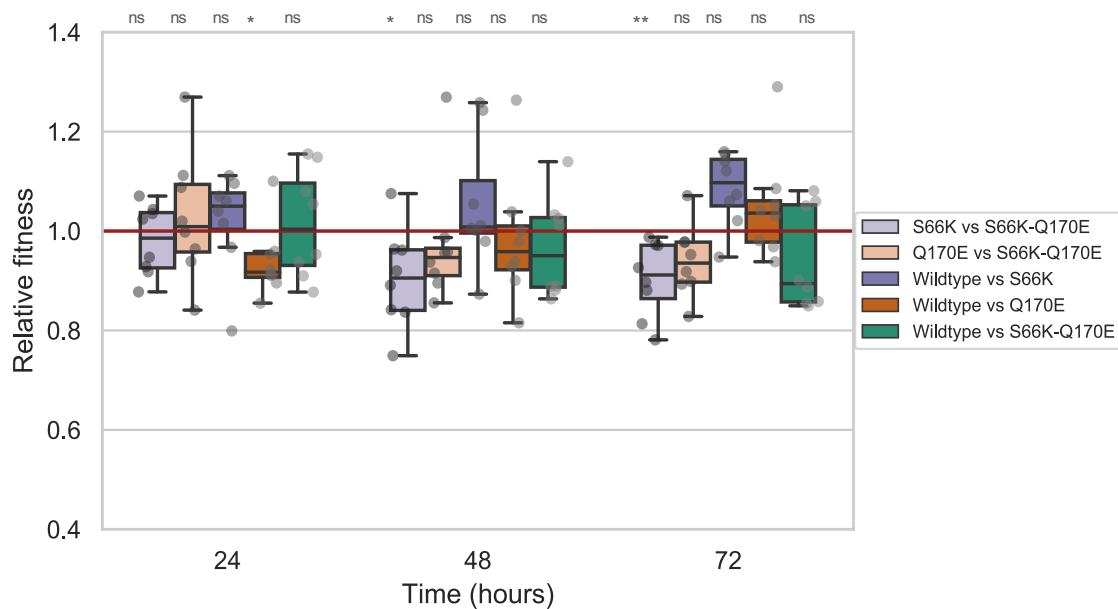
defined concentration of glucose and salts (See Chapter 2). The single mutant S66K-LepA has competed against the double mutant S66K-Q170E-LepA and the wild type for 72 hours. For the competition experiments against double mutant S66K-Q170E-LepA, there was no significant difference in fitness at 24 hours (Figure 3.6). Later at 48 and 72 hours, single mutant S66K showed significantly lower fitness (*p values 0.0303313 and 0.00904312* respectively). Other single mutant Q170E-LepA does not show significantly lower fitness at 72 hours. Both single mutants were lower in fitness as expected. In the competition against wild type, both single mutants showed non-significant fitness differences. However, they followed the same trend as in the LB broth medium; that is, they had lower relative fitness than that of both the double mutant and wild type. The double mutant showed a slightly higher fitness but not significant (*p value = 0.1566335*). The difference in the fitness values was not very large but followed the trends according to the assumption of compensatory mutations, i.e., both single mutants S66K-LepA and Q170E-LepA were lower in fitness when they competed against ancestral state and wild type.

3.6 Analysis of charge compensation

The EF4 protein was shown to play a role in an essential cellular process, the back translocation of the ribosomes in the translational machinery[Qin et al., 2006]. Mutations at functionally constrained sites can affect the stability of protein's 3D structure and adversely impact the function [Worth et al., 2009]. Mutations at other sites of the protein can compensate for this instability of the 3D structure [Moore et al., 2000]. As a result, compensatory mutations can be fixed in the population [Szamecz et al., 2014]. We identified two positions that underwent potentially compensating mutations throughout the bacterial phylogeny, including mutations present on *E. coli* branch. We show that the two single mutations led to a decrease in fitness when tested separately, but this effect was reduced or removed when the two mutations were present, there-

Table 3.4 Mean relative fitness and p values for the competition experiments in LB Broth and M9 minimal media

Competing strains	Time	Medium	Mean Rel. fitness	p value
S66K-LepA vs. S66K-Q170E-LepA	24 hours	LB Broth	0.833261	0.14057
S66K-LepA vs. S66K-Q170E-LepA	48 hours	LB Broth	0.936567	0.74062
S66K-LepA vs. S66K-Q170E-LepA	72 hours	LB Broth	0.577658	0.00720
Q170E-LepA vs. S66K-Q170E-LepA	24 hours	LB Broth	0.898993	0.00314
Q170E-LepA vs. S66K-Q170E-LepA	48 hours	LB Broth	0.686918	0.01693
Q170E-LepA vs. S66K-Q170E-LepA	72 hours	LB Broth	0.421516	3.41972
Wildtype vs. S66K-LepA	24 hours	LB Broth	1.292538	0.02880
Wildtype vs. S66K-LepA	48 hours	LB Broth	1.385992	0.01142
Wildtype vs. S66K-LepA	72 hours	LB Broth	1.512685	0.00184
Wildtype vs. Q170E-LepA	24 hours	LB Broth	1.279975	0.04508
Wildtype vs. Q170E-LepA	48 hours	LB Broth	1.372884	0.02841
Wildtype vs. Q170E-LepA	72 hours	LB Broth	1.481042	0.01776
Wildtype vs. S66K-Q170E-LepA	24 hours	LB Broth	0.838456	0.02814
Wildtype vs. S66K-Q170E-LepA	48 hours	LB Broth	0.773631	0.00154
Wildtype vs. S66K-Q170E-LepA	72 hours	LB Broth	0.636666	5.06e-05
S66K-LepA vs. S66K-Q170E-LepA	24 hours	M9 minimal	0.980562	0.46358
S66K-LepA vs. S66K-Q170E-LepA	48 hours	M9 minimal	0.904946	0.03033
S66K-LepA vs. S66K-Q170E-LepA	72 hours	M9 minimal	0.903771	0.00904
Q170E-LepA vs. S66K-Q170E-LepA	24 hours	M9 minimal	1.028889	0.54727
Q170E-LepA vs. S66K-Q170E-LepA	48 hours	M9 minimal	0.971783	0.54968
Q170E-LepA vs. S66K-Q170E-LepA	72 hours	M9 minimal	0.939766	0.05183
Wildtype vs. S66K-LepA	24 hours	M9 minimal	1.019987	0.59000
Wildtype vs. S66K-LepA	48 hours	M9 minimal	1.053200	0.29319
Wildtype vs. S66K-LepA	72 hours	M9 minimal	1.084751	0.01438
Wildtype vs. Q170E-LepA	24 hours	M9 minimal	0.938541	0.04881
Wildtype vs. Q170E-LepA	48 hours	M9 minimal	0.983182	0.72898
Wildtype vs. Q170E-LepA	72 hours	M9 minimal	1.048511	0.24893
Wildtype vs. S66K-Q170E-LepA	24 hours	M9 minimal	1.014359	0.37327
Wildtype vs. S66K-Q170E-LepA	48 hours	M9 minimal	0.966546	0.94870
Wildtype vs. S66K-Q170E-LepA	72 hours	M9 minimal	0.942817	0.15663

fore demonstrating their compensating nature. The reconstructed fitness landscape for the two positions showed "peaks" for the double mutant and wild type and the single mutants "valleys".

The reduction of fitness of single mutations might be attributed to an instability of the protein structure, preventing it from performing its function correctly. In this case, the function of the EF4 is involvement in the back translocation of the posttranslational complex. Thus, it might be affected by the single mutation, and the process of translation is perturbed overall. Restoration of the double mutant (ancestral state) fitness suggests that EF4 regained the proper function and structure in the double mutant and increased the fitness in the competition environment.

Furthermore, We tested fitness in two environments that gave us different relative fitness. M9 minimal media has a limited concentration of salts and a single carbon source that is glucose. There is not a large difference in the fitness because *E. coli* strain B has been adapted to this medium for all the experimental work of [Lenski, 1988]. Protein translation is also compromised in the M9 minimal stress environment as cells only translate the necessary proteins to perform in the stress conditions. As we see in the LB Broth environment, the fitness values of single mutants drop significantly. Given that LB Broth is enriched in nutrients, we speculate that the strains have enough nutrients to grow on, and both competitors (any combination of competition experiments used in our experimental setup) were performing their cellular processes at an optimal level. In the view of lower fitness of single mutants, we hypothesize that the single mutants behave poorly in growth because the protein translation is not having a proper back translocation (that provides misincorporation of amino acids in the elongation cycle) [Michel and Baranov, 2013].

Charge as a biochemical property contributes substantially to the amino acid coevolution [Chakrabarti and Panchenko, 2010]. In *E. coli*, our competition results suggest that the ancestral state is as stable as the present *E. coli* whereas the present sequence does not have any charge on the tested amino acids (Ser at 66 and Gln at 170). All strains of *E. coli* have conserved amino acids on these positions (Table 3.1). The reconstructed ancestral state has a positive and negative charge on both predicted coevolving sites. The wild type *E. coli* has zero charges on both sites. Both neutral sites maybe the reason that the double mutant has slightly higher fitness than the wild type. *E. coli* might not have a charge on these positions, and both amino acids might be fixed in the *E. coli* populations due to an unknown environmental change and were stable for the EF4 to perform the optimal function. These laboratory conditions do not represent the actual environment and hard to replicate in the natural environment. It is noted

that a suitable stressful environment to assess fitness difference is a challenge in these kinds of experiments. One can try a different range of available laboratory media but still not represent the natural environment. It needs time to find appropriate stress to analyze large differences in fitness. Interestingly the predicted coevolving group in EF4 has perfect compensation for the charge in other species (Table 3.1). The perfect charge compensation in these bacterial species opens the future aspects to study coevolving groups in these bacterial species. There might be more exciting results regarding fitness differences in reconstructed ancestral states in these other species than *E. coli* Strain B. There are experimental limitations; for instance, different colony colors to be used in competition assay are there in the case of other bacterial species.

Chapter 4

Charge-compensating mutation in the IspH protein

The *ispH* gene encodes a 4-hydroxy-3-methylbut-2-enyl diphosphate reductase. The predicted coevolving sites in this gene are Asn (N) at position 24 and Alanine (A) at position 27 in the *E. coli* sequence. These positions were predicted to be coevolving because of charge compensation. After reviewing the functional properties of the *ispH* gene, we report the bioinformatic evidence for coevolution. As for the *lepA* gene (see Chapter 3), we reconstructed the possible ancestral states of the *ispH* at the two coevolving positions and assessed their relative fitness by competing them with the wild type strain in order to infer the local fitness landscape.

4.1 Role of IspH in Isoprenoid Biosynthesis Pathway

IspH is an enzyme involved in the non-mevalonate pathway (MEP). MEP pathway synthesizes Isoprenoids that are one of the biggest class of naturally occurring compounds [Chandran et al., 2011]. The second candidate group of predicted coevolving amino acids is in the *ispH* gene. I first describe the MEP pathway and the role of 4-hydroxy-3-methylbut-2-enyl diphosphate reductase in an organism.

The five-carbon hydrocarbon called isoprene (2-methyl-1,3-butadiene) is the isoprenoid building unit of the diverse organic compounds. These compounds play an essential role in the growth and survival of prokaryotes. Cell wall and membrane synthesis, conversion of light into chemical energy, electron transport, and derivatives of secondary metabolites are a few processes where isoprenoids are present. Metabolic engineering of bacteria helps to understand the function and structure of the desired enzyme. Various studies, for instance, Ajikumar et al. [2008]; Kirby and Keasling [2009]; Misawa [2011]; Ward et al. [2018] reported the derivates of enzymes that are involved in the synthesis of isoprenoids.

Isopentenyl pyrophosphate (IPP) or its isomer, dimethylallyl pyrophosphate (DMAPP), is

Figure 4.1 Illustration shows the steps involved in MEP pathways. Three molecules of acetyle-CoA MEP start with the precursors; Glyceraldehyde 3-phosphate (G3P) pyruvate [Rohmer et al., 1996]. Intermediate DXP (1-deoxy-D-xylulose-5-phosphate) converts into MEP (2-C-methylerythritol-4-phosphate) in the presence of reductoisomerase (IspC, MEP synthase) [Hale et al., 2012]. MEP converts into 4-diphosphocytidyl-2-C-methylerythritol (CDP-Me) intermediate in the presence of IspD. CDP-Me further converts into 2-C-Methylerythritol-2,4-cyclodiphosphate (CDP-MEP) in the regulation of IspE. CDP-MEP converts into hydroxymethylbutenyl 4-diphosphate (HMBPP) in the presence of IspG. HMBPP converts into isopentenyl diphosphate (IPP), and then isoprene is synthesized or be used as a precursor for higher molecular height isoprenoids (for example, terpenoids). The Illustration is adopted from Ward et al. [2018] and created by `https://biorender.com/`.
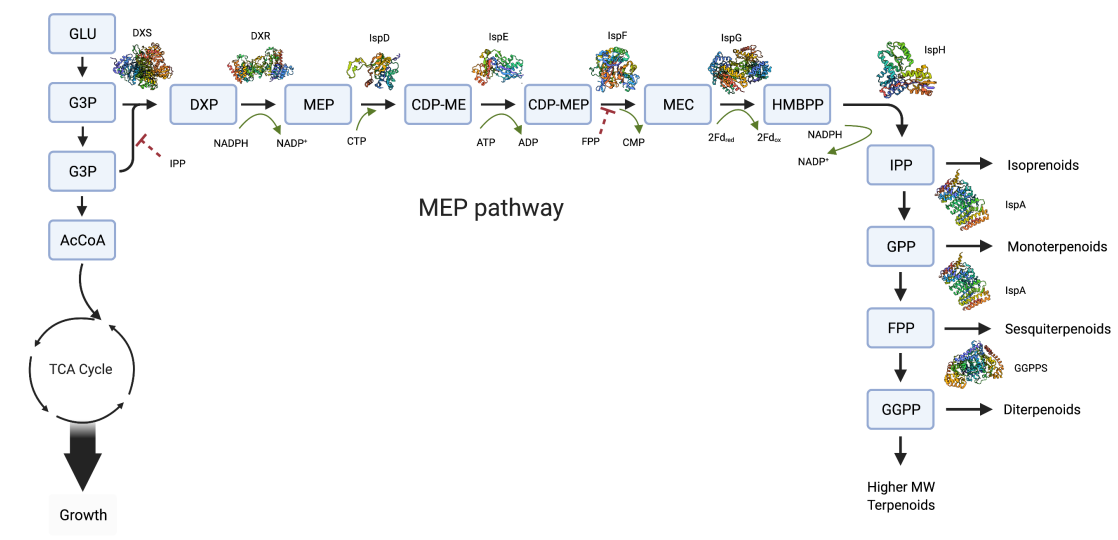
MEP pathway

the isoprenoid precursor and synthesized in the mevalonate pathway (MVA) in all organisms. There is an alternative pathway non-mevalonate pathway (MEP) present in microorganisms [Lichtenthaler et al., 1997; Phillips et al., 2008]. The MEP pathway generates only IPP. Only one pathway is used for the biosynthesis of isoprenoids in one organism. Most of the bacteria use the non-mevalonate pathway (sometimes called mevalonate-independent pathway) [Boronat and Rodríguez-Concepciön, 2014; Frank and Groll, 2017].

The *ispH* genes belong to the MEP pathway, where it encodes the 4-hydroxy-3methyl-2-enyl diphosphate reductase, which is involved in the last step of the biosynthesis pathway, responsible for converting HMBPP into IPP. One of the candidate genes that we analyzed in this study is the *ispH* gene (previously named *lytB*). The most accepted explanation for the function of IspH is that the IspH protein requires a [4Fe-4S] cluster for catalytic activity [Seemann et al., 2009]. The mechanism catalyzed by IspH requires the removal of the hydroxyl group transfer of two electrons from the [4Fe-4S] cluster and protonation of an intermediate allylic anion [Rohdich et al., 2003]. The deletion of this gene leads to cell lysis and cell death, suggesting that it is essential for the growth of bacterial cells [McAteer et al., 2001]. Another study by

Figure 4.2 Reconstructed ancestral states and putative compensatory mutations on the *E. coli* lineage for the *ispH* gene. Inferred ancestral states are displayed on the left, and the *E. coli* (wild type) sequence on the right. The arrow represents the putative mutations and their order at both sites. The three reconstructed strains used in this study are indicated as a double mutant, single mutant 1, and single mutant 2.
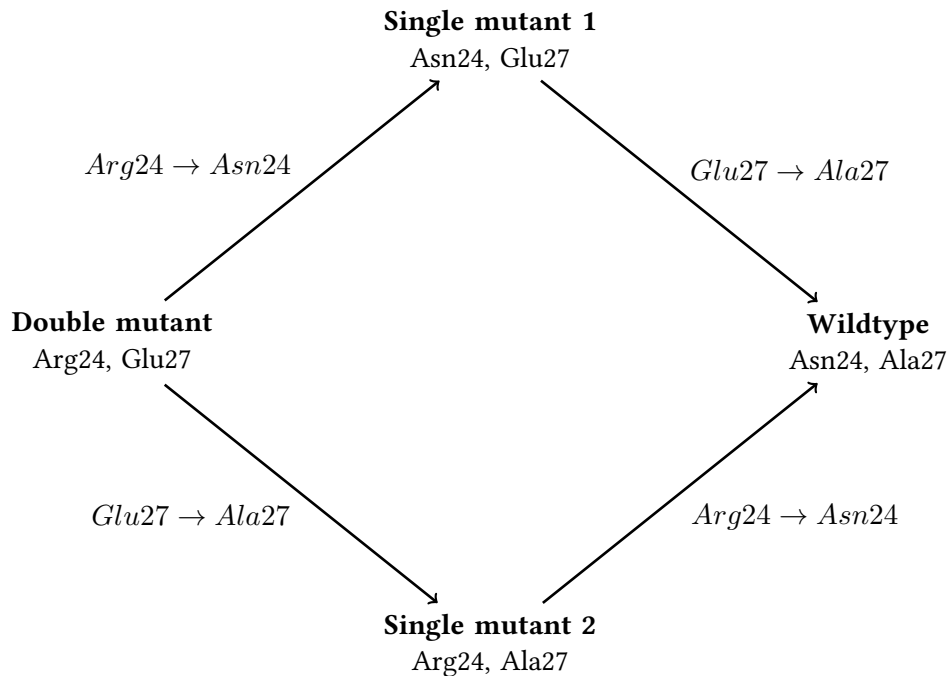
Altincicek et al. [2001] showed that bacterial cells are only viable in the medium supplemented with mevalonate and emphasized the crucial role in the MEP pathway.

Because of its essentiality, *isph* is a good candidate to study prediction such as coevolving amino acids. A small change in one of the isoprenoid pathway genes might perturb the pathway and have fitness effects. The pair was predicted for the charge compensation. The predicted coevolving residues in IspH are at sites 24 and 27 with the amino acids coordinates Asn and Alanine, respectively (PDB ID:3KE8). We reconstructed the ancestral state for both positions using the DNA sequence and accounted for the ancestral reconstruction for codon (See Chapter 2, section 2.2). The reconstructed amino acids' ancestral states are Arg at position 24 and Glutamic acid at position 27, leading to the possible evolutionary scenarios displayed in Figure 4.2. Strains containing the putative ancestral genotypes were reconstructed in *E. coli* Strain B:

- the strain with the ancestral states at the two positions was reconstructed by introducing the double mutation N24R-A27E,

- the two possible intermediate genotypes were reconstructed by introducing each of the two single mutations N24R and A27E.

Figure 4.3 3D structure of IspH 4 chain A (PDB ID: 3KE8). Orange highlighted residues are the predicted coevolving amino acids.
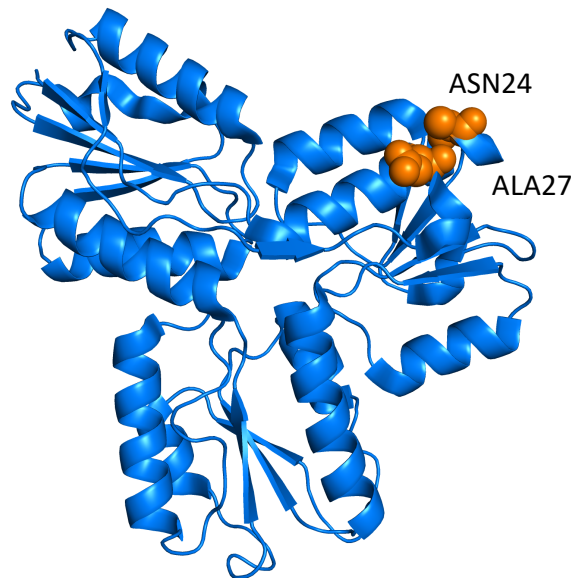
We then analyzed the fitness of both single mutants in two competition environments, one in the M9 minimal medium and the other in the LB nutrient enriched medium. We estimated the relative fitness of single mutants by comparing them with the double mutant and wild type for 72 hours. We then inferred the local fitness landscape for the two positions.

4.2 3D structure, Phylogenetic analysis, and evidence for coevolution by compensation

The 3D structure of the IspH protein is shown in Figure 4.3. IspH is a 316 amino acid protein with a molecular weight of 35 kDa. The IspH protein is characterized by paramagnetic resonance spectroscopy after reconstitution of the purified protein [Wolff et al., 2003]. The primary distance between the two coevolving positions was three amino acids. However, these two positions were close in the 3D structure, and the distance was 5.79Å between two alpha carbons.

We built the corresponding compensogram and phylogenetic tree for IspH at sites 24 and 27

(See section 3.2 Chapter 3). The reconstructed weighted substitution maps at the two detection positions reveal multiple branches where a strong compensation signal can be observed (Figure 4.4). The coevolution signal is widespread along the phylogeny, suggesting that the constraint acting on these sites is conserved throughout the bacteria. The majority of observed mutations involve a change from positive to negative at site 24, compensated by a mutation from negative to positive at site 27. The inferred compensating mutations on the *E. coli* branch are unique and not observed on any other branches of the tree, the Asn and Ala state pair being only present in *E. coli*. Interestingly, no polymorphism was observed in the 261 strains of *E. coli* for which a sequence was available for the *ispH* gene. The conserved sequences in all strains of *E. coli* suggests that the mutation are fixed in the population (Table 4.1).

4.3   Construction of mutant strains

We designed DNA primers with the desired mutations and used Splicing Overlap Extension PCR (SOE-PCR) to amplify DNA fragments. The amplification fragments were sequenced in order to confirm the presence of the desired substitutions (see Chapter2 section 2.3). The successful fragments were cloned in the high copy number plasmid pCR8/TOPO/TA. Sanger sequencing results for the positive clones confirmed the desired mutations of the IspH mutants in pKOV_unstuff (Plasmid used for the homologous recombination). Three mutants were made: a single mutant with a mutation from Asn to Arg at position 24 (N24R-IspH), another single mutation from Alanine to Glutamic acid at position 27 (A27E-IspH), and both mutations in the double mutant (N24R-A27E-IspH). The scarless allelic replacement method successfully gave us all three mutant strains (Table 4.2 shows the strain name and its genotype). Whole-genome se-

Table 4.1 Frequency of observed amino-acid state pairs among bacterial species. The highlighted one is observed in *E. coli* strains

Pairs of Amino acids	Frequency in all Species	Frequency in *E. coli* Strains
N;A	**1**	**261**
R;E	60	0
L;K	33	0
K;E	22	0
R;D	17	0
Other pairs, each in low frequency	88	0

Figure 4.4 A). The phylogenetic tree of the IspH protein family used in the CoMap analysis. Blue highlighted branches show negative change at position 24 on the protein sequence of *E. coli* and the red highlighted branches show the positive change at position 27 at the protein sequence of *E. coli*. **B).** Corresponding Compensogram for IspH.
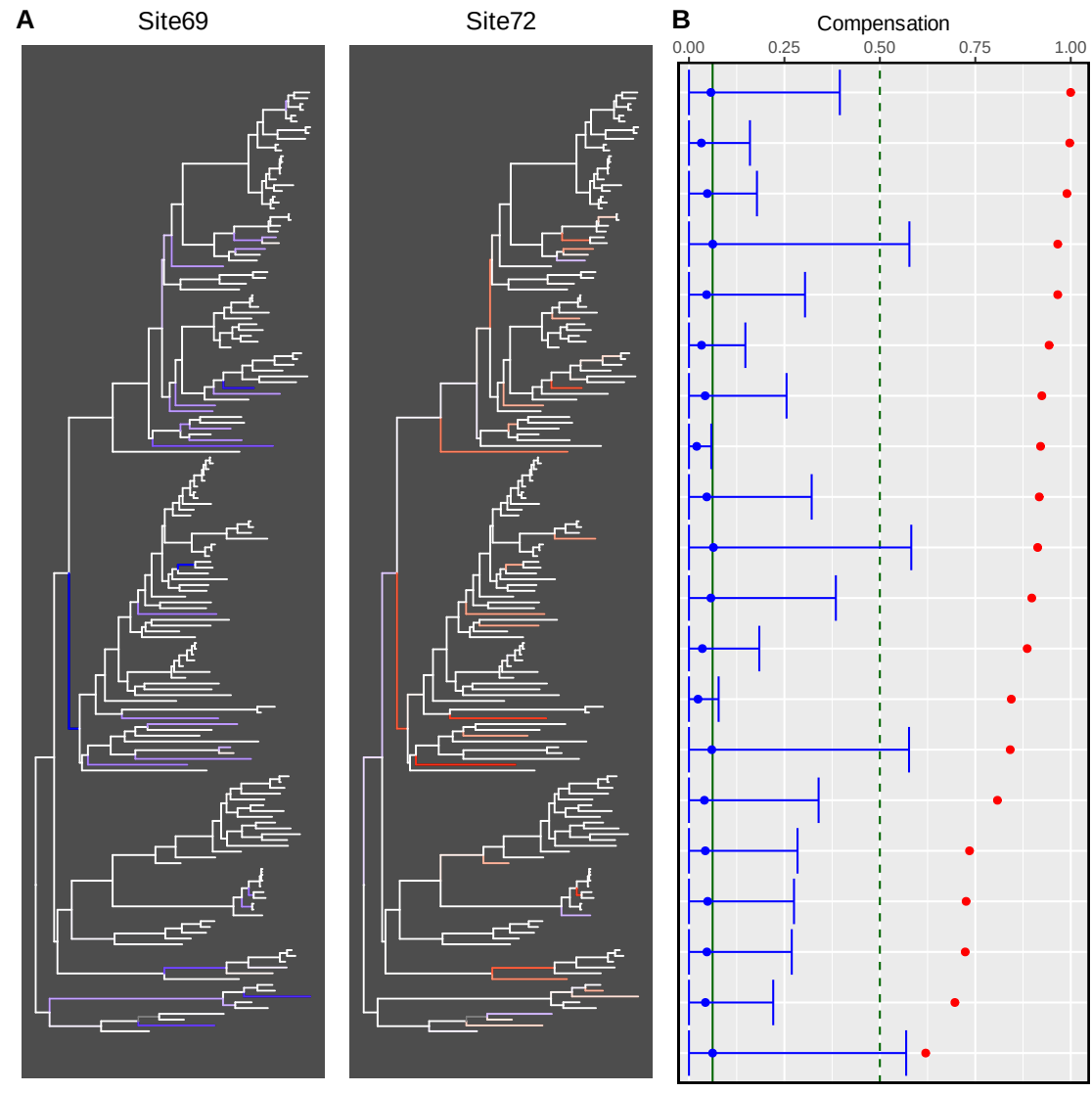
quencing also confirmed no off-target substitutions after the homologous recombination step of generating mutants (Table 4.3). Breseq is a computational pipeline to detect mutations in the given sample using the reference genome [Deatherage and Barrick, 2014]. We mapped the sequencing reads from fastq files of all mutant strains of IspH against REL606 (the reference genome for *E. coli strain B*) using breseq [Jeong et al., 2009]. An additional prediction of IS150 non-coding sequence was detected in wildtype-606, A27E-IspH-606, and N24R-A27E-IspH-606 breseq mapping output (Detailed table is shown in 8.2). The IS150 region is a non-coding insertion sequence. No mutations in all the mutants' coding regions, including wild type, allowed us to proceed with the mutants for the competition experiments. During allelic replacement, two

strains are generated, one with the mutation and one that has no mutation. The strain without mutation was used as a wild type in the competition experiments.

4.4   Expression of mutant strains

To validate that potential phenotypic differences are not due to a lack of expression of the IspH protein in the reconstructed strains, the expression of *ispH* mutant genes were confirmed in all strains. For this, we extracted the whole RNA from all the mutants and wild type and synthesized cDNA from it without any genomic DNA contamination (See Chapter 2, section 2.6). PCR Amplification using *ispH* gene-specific primers gave us the expected band size of 1000 bp from the mutant strains N24R, A27E, N24R-A27E, and wild type. The positive results of PCR showed that the *ispH* gene is being expressed in all the mutant strains, at least at the transcriptional level (Supplement Figure 8.2).

4.5   Competition experiments

We used the Malthusian parameter to calculate the relative fitness compared with the other strain (wild type or ancestral state in this study) described in [Lenski et al., 1991]. Relative fitness

Table 4.2 Mutant strains genotypes for the IspH candidate.

Strain	Genotype
Wildtype-606 Ara$^+$	The wildtype strain of present sequence of *E. coli* Strain B REL606 used in the allelic replacement experiment as wild type. Red colonies on the TA indicator plates.
Wildtype-607 Ara$^-$	The wildtype strain of present sequence of *E. coli* Strain B REL607 used in the allelic replacement experiment as wild type. White colonies in the TA indicator plates.
N24R-IspH Ara$^-$	Asparagine was replaced with Arginine (predicted coevolving amino acids) in *ispH* gene called as single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.
A27E-IspH Ara$^-$	Alanine was replaced with Glutamic acid (predicted coevolving amino acids) in *ispH* gene called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.
N24R-A27E-IspH Ara$^+$	The inferred ancestral state having both substitutions in *ispH* gene called as double mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.

greater than 1 means the one strain is better in fitness and vice versa. We estimated the relative fitness of single mutants compared to the double mutant and wild type (See Box 1.4.1). We used two environments for the competition experiments: one environment or medium was LB broth with enough nutrients for the growth of competing strains (nutrient-enriched medium). The reason for choosing LB Broth is that competing strains should grow without compromising the translation of the proteins. The other medium was the M9 minimal medium, a defined medium with a single carbon source (glucose). This medium is widely used in the laboratory for competition experiments to have limited resources for growth.

4.5.1 LB Broth medium

We analyzed five combinations of competing strains, and each combination was run in eight replicates. These combinations were N24R-IspH vs. N24R-A27E-IspH, A27E-IspH vs. N24R-A27E-IspH, N24R-IspH vs. Wild type, A27E-IspH vs. Wild type, and N24R-A27E-IspH vs. wild type. Each combination was plated on TA indicator plates and used for colony counts at 0 hours, 24 hours, 48 hours, and 72 hours. We used the Malthusian parameter (See Chapter 1 section

Table 4.3 Columns from the breseq output is shown here to show the desired mutations in the mutant strains.

Strain	Mutation	Gene	%[1]	Description
Wildtype-606	N/A	N/A	98.8	Repeat region
Wildtype-607	T → C	*araA*	89.4	L-arabinose isomerase
	A → G	*recD*	89.4	Exonuclease V (RecBCD complex) alpha chain
N24R-IspH-607	T → C	*araA*	98.3	L-arabinose isomerase
	A → G	*recD*	98.3	Exonuclease V (RecBCD complex) alpha chain
	2 bp → CG	*ispH*	98.3	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
A27E-IspH-606	2bp → AG	*ispH*	98.6	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
N24R-A27E-IspH-606	2 bp → CG	*ispH*	98.5	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
	2bp → AG	*ispH*	98.5	4-hydroxy-3-methylbut-2-enyl diphosphate reductase

[1] % of the genome where reads could be mapped

Figure 4.5 Competition experiments in LB Broth medium. Relative fitness is shown on the y-axis, and was estimated at three time points shown on the x-axis. Lighter color boxplots show the combination of single mutants vs. double mutant. Darker color boxplots show the wild type vs. single mutants and wild type vs. double mutant. The red horizontal line is the reference line for zero fitness difference (relative fitness of 1). One sample t-test is used for the statistics here (The mean of population $\mu = 1$). See Table 4.4 for exact p-values, and supp Table 8.5 and 8.6 for the raw data.
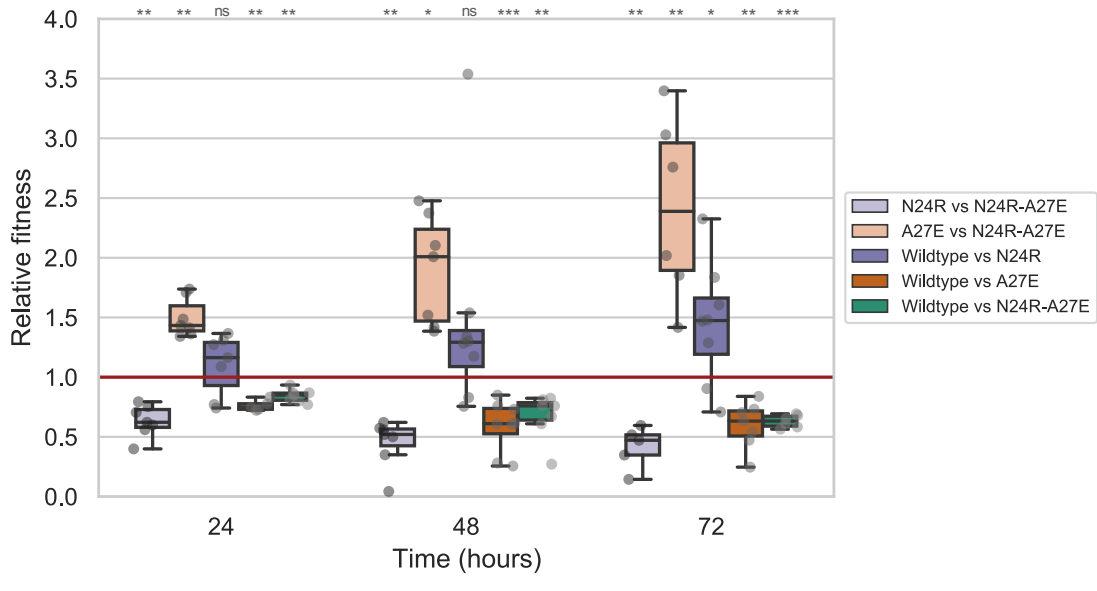
1.4.1) to calculate the relative fitness described in Lenski et al. [1991]. We used the LB Broth medium as an environment with unlimited resources. The single mutant N24R-IspH competed against N24R-A27E-IspH and the wild type for up to 72 hours. N24R had a consistent significantly lower fitness difference when competed against the double mutant at all time intervals. N24R also had a lower fitness at all time points when competed against the wild type (relative fitness greater than one on the wildtype-IspH vs. N24R-IspH comparison). Under coevolution with compensation scenario, we expect to observe a lower fitness in both single mutants when they compete against inferred ancestral state or wild type. However, the single mutant A27E showed significantly higher fitness at all time points when it was competing against the inferred ancestral state (double mutant). It also showed a significantly higher fitness at all time intervals in the A27E-IspH vs. wild type comparison. The higher fitness of the A27E strain is slightly different results under the assumption of compensation for the charge. Furthermore, the double mutant showed significantly higher fitness as compared to the wild type. High variation in the replicates was observed in LB broth medium for both single mutants competing against double mutant or wild type (Figure 4.5).

Figure 4.6 Competition experiments in M9 minimal medium; Relative fitness is shown on the Y-axis, and three time points for the plating is on the x-axis. Legend as in figure 4.6. See Table 4.4 for exact p-values, and supp Table 8.6 for the raw data.
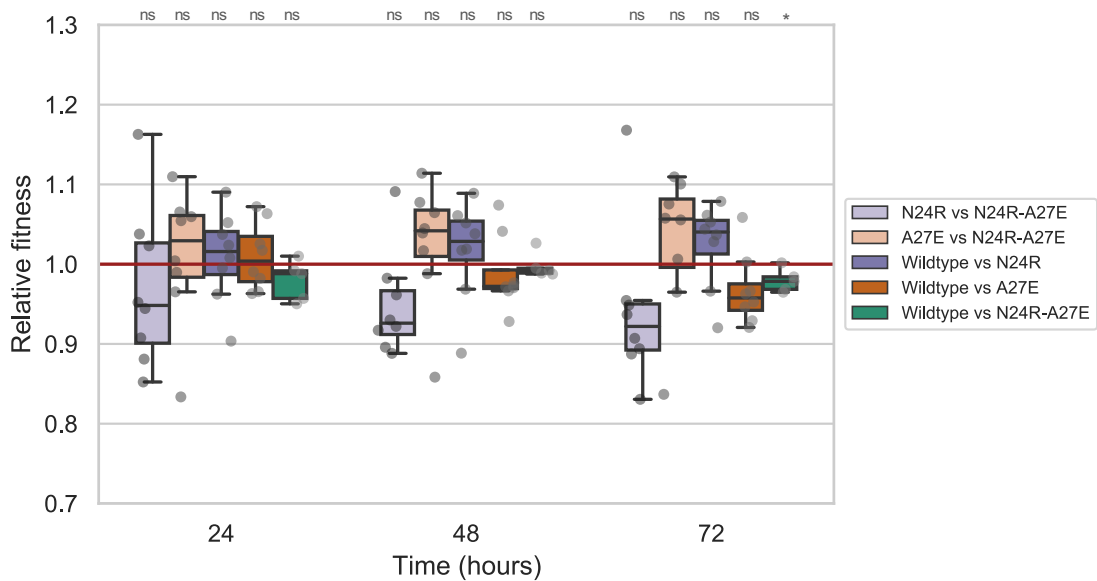
4.5.2 M9 minimal medium

We used the same combinations of competing strains for the M9 minimal medium as in the LB Broth medium. M9 minimal media has a defined concentration of glucose and salts (See chapter 2). The single mutant N24R-IspH had competed against N24R-A27E-IspH and the wild type for 72 hours. For the competition experiments for N24R-IspH against the double mutant N24R-A27E, there were no significant differences in fitness at 24 hours, 48, and 72 hours points. Other single mutant A27E-IspH did not show significantly lower fitness at all time intervals. Although, it was the same trend as followed in LB Broth medium. Single mutant A27E-IspH has higher fitness compared to the ancestral state. In the competition against wild type, both single mutants N24R-IspH and A27E-IspH had shown non-significant fitness differences. The double mutant showed slightly higher fitness and significance at 72 hours (*p value = 0.0348961*). The difference in the fitness values was lower than those observed for the LB Broth medium.

4.6 Analysis of charge compensation

Several studies for the functional understanding of the *ispH* explained its essential role in the MEP pathway and showed that IspH is involved in converting HMBPP to IPP [Seemann et al., 2009]. Mutations at functionally constrained sites can adversely affect the protein's structure and have an impact on its function [Worth et al., 2009]. One or several other mutations (s) can

compensate for this instability of the 3D structure at other sites of the protein [Moore et al., 2000]. As a result, compensatory mutations can be fixed in the population. We used competition assays to observe the fitness effects of mutants strains using compensatory evolution (compensation for the charge in this candidate). In the IspH candidate group of coevolving amino acids, the single mutant N24R showed lower fitness in the presence of wild type or inferred ancestral state, and in the light of these results, the hypothesis is that the single mutation at position 24 is not stable for the IspH protein to perform its function correctly. I have observed lower fitness in the single mutants in Chapter 3. However, the other single mutant A27E-IspH does not follow the expected local fitness landscape under the assumption of compensatory evolution. The higher fitness of the single mutant A27E in the competition experiment against the inferred

Table 4.4 Mean relative fitness and p values for the competition experiments in LB Broth and M9 minimal media

Competing strains	Time	Medium	Mean Rel. fitness	p value
N24R-IspH vs. N24R-A27E-IspH	24 hours	LB Broth	0.633294	0.00034
N24R-IspH vs. N24R-A27E-IspH	48 hours	LB Broth	0.452433	0.00034
N24R-IspH vs. N24R-A27E-IspH	72 hours	LB Broth	0.415184	0.00175
A27E-IspH vs. N24R-A27E-IspH	24 hours	LB Broth	1.497481	0.00018
A27E-IspH vs. N24R-A27E-IspH	48 hours	LB Broth	1.898148	0.00200
A27E-IspH vs. N24R-A27E-IspH	72 hours	LB Broth	2.412369	0.00629
Wildtype vs. N24R-IspH	24 hours	LB Broth	1.101791	0.32868
Wildtype vs. N24R-IspH	48 hours	LB Broth	1.470224	0.17265
Wildtype vs. N24R-IspH	72 hours	LB Broth	1.451822	0.04036
Wildtype vs. A27E-IspH	24 hours	LB Broth	0.762346	0.00028
Wildtype vs. A27E-IspH	48 hours	LB Broth	0.588311	0.00099
Wildtype vs. A27E-IspH	72 hours	LB Broth	0.595301	0.00156
Wildtype vs. N24R-A27E-IspH	24 hours	LB Broth	0.842901	0.00025
Wildtype vs. N24R-A27E-IspH	48 hours	LB Broth	0.672698	0.00412
Wildtype vs. N24R-A27E-IspH	72 hours	LB Broth	0.630746	1.24e-05
N24R-IspH vs. N24R-A27E-IspH	24 hours	M9 minimal	0.970114	0.42960
N24R-IspH vs. N24R-A27E-IspH	48 hours	M9 minimal	0.948525	0.06147
N24R-IspH vs. N24R-A27E-IspH	72 hours	M9 minimal	0.940870	0.13886
A27E-IspH vs. N24R-A27E-IspH	24 hours	M9 minimal	1.010252	0.74403
A27E-IspH vs. N24R-A27E-IspH	48 hours	M9 minimal	1.025389	0.38483
A27E-IspH vs. N24R-A27E-IspH	72 hours	M9 minimal	1.025811	0.44408
Wildtype vs. N24R-IspH	24 hours	M9 minimal	1.009059	0.66815
Wildtype vs. N24R-IspH	48 hours	M9 minimal	1.016630	0.47830
Wildtype vs. N24R-IspH	72 hours	M9 minimal	1.023521	0.25216
Wildtype vs. A27E-IspH	24 hours	M9 minimal	1.009934	0.52447
Wildtype vs. A27E-IspH	48 hours	M9 minimal	0.987666	0.47861
Wildtype vs. A27E-IspH	72 hours	M9 minimal	0.967396	0.07722
Wildtype vs. N24R-A27E-IspH	24 hours	M9 minimal	0.979440	0.14034
Wildtype vs. N24R-A27E-IspH	48 hours	M9 minimal	0.997965	0.79174
Wildtype vs. N24R-A27E-IspH	72 hours	M9 minimal	0.979510	0.03489

ancestral state and wild type proposed that the single mutation at the amino acid position 27 (A → E) leads stable structure of the IspH protein than the wild type inferred ancestral state. The A27E-IspH mutant results maybe because we tested the inferred ancestral state in the genomic background of *E. coli* and it might follow different local fitness landscape in the case of testing these mutations in other bacterial species. In our experimental setup, we only study the predicted coevolving amino acids and their interaction. There might be more beneficial mutations in the case of the A27E mutant that led to the fitness peak in the local fitness landscape. The mutations that might occur on the other sites are not studied in this analysis. The single mutant N24R-IspH has lower fitness against wild type and inferred ancestral state explains partially or one direction mutation trajectory. It also explains the direction of the mutation that led to the wild type of *E. coli* present sequence. In IspH candidate, we can determine the trajectory of the mutations, particularly in the case of two positions (24 and 27 in IspH). Furthermore, the double mutant having the compensatory mutation restores the function and stabilizes the IspH protein structure, leading to the restoration of the competition experiment's fitness. The local fitness landscape that has been reconstructed in our study clearly shows the valley in the single mutation N24R-IspH (lower fitness) and restoration in the fitness in the inferred ancestral state with only one direction.

The double mutant has slightly higher fitness than the wild type of the present *E. coli*. The third peak in the local fitness landscape does not support the assumption of compensatory evolution in the testing pair of the coevolving positions in IspH. The higher fitness of the double mutant is slightly different from the expected results. Charge as a biochemical property substantially contributes to the residue coevolution Chakrabarti and Panchenko [2010]. In *E. coli*, all strains have conserved amino acids on these predicted coevolving positions, indicating that the ancestral state might be more stable than the wild type. If double mutant certainly higher in fitness than the wild type, one may wonder how the latter could have been fixed in the population (see table 4.1). One possible explanation for these results is that there might be a stress condition in *E. coli* that could cause these mutations, reduce the population size and allow the fixation of a less fit state by chance in *E. coli*. Before the possible stress condition, the ancestral state was more stable than the wild type. Another aspect is that these laboratory conditions do not represent the actual environment, and stress conditions happened back in time. Interestingly the difference in fitness is higher in LB broth as compared to M9 minimal. It could be because the stable strain consumes more nutrients and successful in growth in the presence of a nutrient-enriched medium. It could also suggest that fitness differences can be observed not only in the stress conditions.

Chapter 5

Beta propensities-compensating mutations in the YebC protein

The *yebC* gene encodes the YebC protein, with the putative function of transcriptional regulation. The predicted coevolving sites in this gene are Thr (T) at position 65 and Leu (L) at position 66 from *E. coli.* β-strand is a stretch of the polypeptide, and the conformational propensity of each amino acid in the secondary structures of proteins is beta propensities. These positions were identified as coevolving for compensation for the beta propensity. After reviewing the known structural and functional properties of the *yebC* gene, we report the bioinformatic evidence for coevolution. We then report the competition experiment results between the reconstructed possible ancestors and the wild type strain and the resulting inferred local fitness landscape.

5.1 Role of YebC in stress response

yebC is a gene that regulates the transcription in a stress condition, for instance ionization radiation [Byrne et al., 2014a]. Various studies had explored the role of YebC in different bacterial species. Stress conditions regulate the expression of many proteins in the organisms to maintain genome integrity. Variation in the expression of specific proteins is one way to evolve organisms to understand the mechanism of evolution. One of the target mechanisms is DNA repair in extreme conditions. Cells protect the DNA replication process and repair the caused damage because of the stress condition. Exposure to ionization radiation (IR) is a typical laboratory stress condition to study how cellular components respond against the applied stress [Mikkelsen and Wardman, 2003; Collins et al., 2005]. Several studies showed that after exposure to IR, DNA damage is one of the key factors that affect cell survival, specifically double-stranded

breaks [Sargentini and Smith, 1985; Repar et al., 2010; Slade et al., 2009; Krisch et al., 1978]. The adaptive evolution in the DNA repair systems clearly shows that this system is the substantial contributor to an extreme level of IR resistance [Byrne et al., 2014b].

It is hard to determine the single system responsible for the IR resistance in microorganisms due to microorganisms' complex metabolism. Byrne et al. [2014a] and his colleagues studied the *E. coli* K-12 strain MG1655 using directed evolution and acquired phenotype under the exposure of IR. This study's acquired phenotype is explained by mutations in *recA, dnaB* and *yfjK* genes. Also, one of the genes in the mentioned study *yebC* has dropped in the survival of the cells approximately two to three orders of magnitude. This study's conclusion explains the collective effect of a group of genes responsible for the decline in survival. In *Edwardsiella piscicida*, YebC is identified as a virulence regulator and involved in the disease development of the fish [Wei et al., 2018]. The study mentioned above also concluded that YebC regulates the expression of the type III secretion system (T3SS). T3SS is the macromolecular protein structure present in several gram-negative pathogenic bacteria. T3SS is a needle-like structure used to detect eukaryotic cells and help infect them [Gaytán et al., 2016]. Another study by Brown et al. [2017] has shown that YebC acts as a transcriptional repressor of key proteolytic genes in *Lactobacillus delbrueckii* subsp. *lactis*. The proteolytic system is essential in *L. delbrueckii* for the flavor development of fermented products. From the aforementioned studies, YebC plays essentially as a transcriptional regulator in most bacterial species. YebC is a 246 amino acid protein with a molecular weight of 30.47kDa.

The third candidate group that we tested was identified in the *yebC* gene. The predicted co-evolving residues in YebC are at sites 65 and 66 with the amino acids Thr and Leu, respectively (PDB ID:1KON). Other than the canonical properties (charge, volume, and polarity), amino acids display several physicochemical properties that contribute to the 3D structure stability of proteins. Saha et al. [2012] studied these properties using the AAindex database of known 544 amino acid (AA) indices. These properties or features were clustered in 8 indices, *i.e.*, electric charge, hydrophobicity, alpha, and turn properties, physiochemical properties, residue properties composition, beta propensities, and intrinsic properties. The coevolving group in YebC protein was predicted to coevolve for the beta propensity that is β-strand conformational propensity in the secondary structure [Fujiwara et al., 2012]. The pair was predicted for the beta propensities. The PDB coordinates for YebC in *E. coli* are Thr at site 65 and Leu at position 66 (PDB ID:1KON). We reconstructed the ancestral state for both positions using the DNA sequence (See Chapter 2, section 2.2). The reconstructed amino acids' ancestral states are Asn

Figure 5.1 Reconstructed ancestral states and putative compensatory mutations on the *E. coli* lineage for the *yebC* gene. Inferred ancestral states are displayed on the left, and the *E. coli* (wild type) sequence on the right. The arrow represents the putative mutations and their order at both sites. The three reconstructed strains used in this study are indicated as a double mutant, single mutant 1, and single mutant 2.
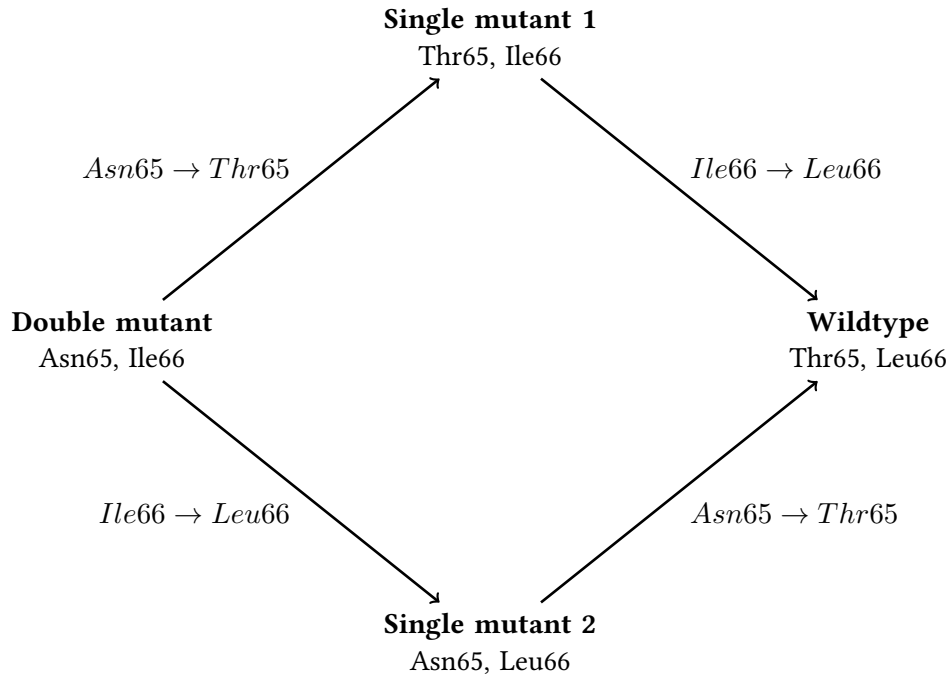
at position 65 and Ile at position 66, leading to the possible evolutionary scenarios displayed in Figure 5.1. Strains containing the putative ancestral genotypes were reconstructed in *E. coli* Strain B:

- the strain with the ancestral states at the two positions was reconstructed by introducing the double mutation T65N-L66I,

- the two possible intermediate genotypes were reconstructed by introducing each of the two single mutations T65N and L66I.

We then analyzed the fitness of both single mutants in two competition environments, one in the M9 minimal medium and the other in the LB nutrient enriched medium. We estimated the relative fitness of single mutants by comparing them with the double mutant and wild type for 72 hours. We then inferred the local fitness landscape for the two positions.

5.2 3D structure, Phylogenetic analysis, and compensation

Figure 5.2 showed the 3D structure of YebC protein. These two predicted coevolving positions are adjacent in the primary distance and were close in the 3D structure. The distance was 4.43

Figure 5.2 3D structure of YebC chain A (PDB ID: 1KON). Orange highlighted residues are the predicted coevolving amino acids.
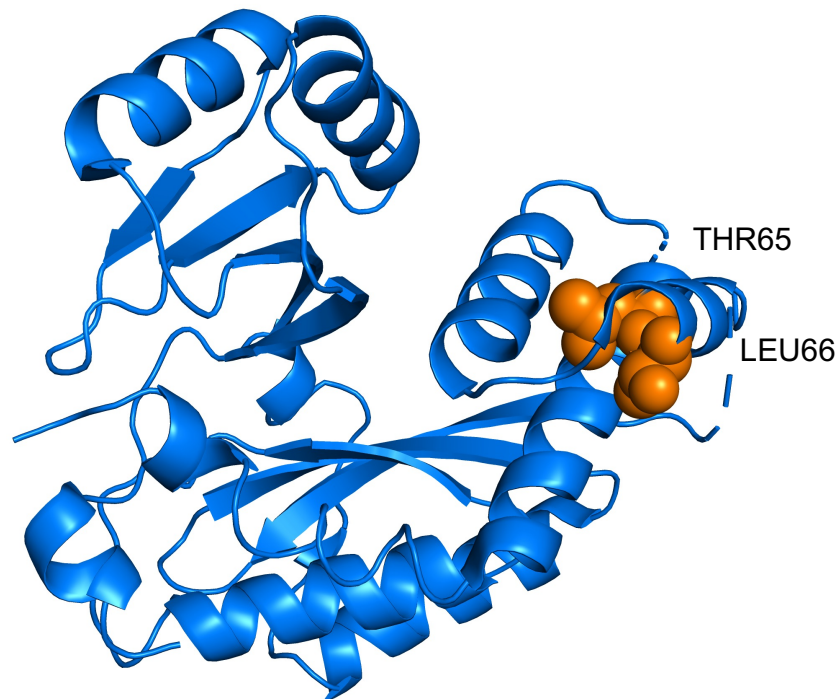
Å between two alpha carbons. The corresponding compensogram and phylogenetic tree were built as mentioned in Chapter 3 section 3.2. Multiple branches can be observed with the strength of the compensation signal. The coevolving positions are inferred by the weighted substitution mapping (Figure 5.3). The signal of coevolution is widespread along the phylogeny suggesting the presence of constraints acting on these two sites. The majority of observed compensating mutations involve beta-propensities biochemical property. The Thr and Leu are observed on *E. coli* branch and one more branch. Interestingly, there were not a lot of sequences present for the *yebC* gene in *E. coli* but still no polymorphism was observed in 27 sequences of *E. coli* strains. This suggested that the mutations are fixed in the population (Table 5.1).

5.3 Construction of mutant strains

We used the same method as in chapter 3 and 4 to generate clones for all the mutant strains (See Chapter 4 section 4.3). The successful fragments were cloned in high copy number plasmid pCR8/TOPO/TA. Sanger sequencing results for the positive clones confirmed the presence of the desired mutations of the YebC mutants in pKOV_unstuff (Plasmid used for the homologous

recombination). Three mutants were made: a single mutant with a mutation from Thre to Asn at position 65 (T65N-YebC), another single mutation from Leu to Ile at position 66 (L66I-YebC), and both mutations in the double mutant (T65N-L66I-YebC). Sanger sequencing results for the positive clones confirmed the desired mutations of the YebC mutants (See Chapter 2 section 2.4). The scarless allelic replacement method successfully gave us all three mutant strains (Table 5.2 shows the strain name and its genotype). Whole-genome sequencing also confirmed no off-target substitutions after the homologous recombination step of generating mutants (Table 5.3). We mapped the sequencing reads of all mutant strains of YebC against REL606 (the reference genome for *E. coli* strain B) using breseq [Jeong et al., 2009]. An additional signal of IS150 non-coding region was detected in wildtype-606 in the breseq mapping output (Detailed table is shown in Supplementary 8.2). The IS150 non-coding insertion sequence was found in this candidate as well (also found in other candidates). No mutations in all the mutants' coding regions, including wild type, allowed us to proceed with the mutants for the competition experiments. The strain that does not have the desired mutation/s after crossing over experiments was used as a wild type in the competition experiments.

5.4 Expression of mutant strains

In order to validate that potential phenotypic differences are not due to a lack of expression of the YebC protein in the reconstructed strains, the expression of *yebC* mutant genes was confirmed in all strains. For this, we extracted the whole RNA from all the mutants and wild type and synthesized cDNA from it without any genomic DNA contamination (See Chapter 2, section 2.6). PCR Amplification using *yebC* gene-specific primers gave us the band size of 800 bp from the mutant strains T65N, L66I, T65N-L66I, and wild type. The positive results of

Table 5.1 Frequency of observed amino-acid state pairs among bacterial species and *E. coli* strains

Pairs of Amino acids	Frequency in all Species	Frequency in *E. coli* Strains
T;L	2	27
N;I	223	0
N;V	14	0
T;I	12	0
Other pairs, each in low frequency	16	0

Figure 5.3 A). The phylogenetic tree of the YebC protein family used in the CoMap analysis. Blue highlighted branches show the positive change at position 65 of *E. coli*). The red highlighted branches show the negative change at position 66 **B).** Corresponding compensogram for YebC, legend as Figure 3.4
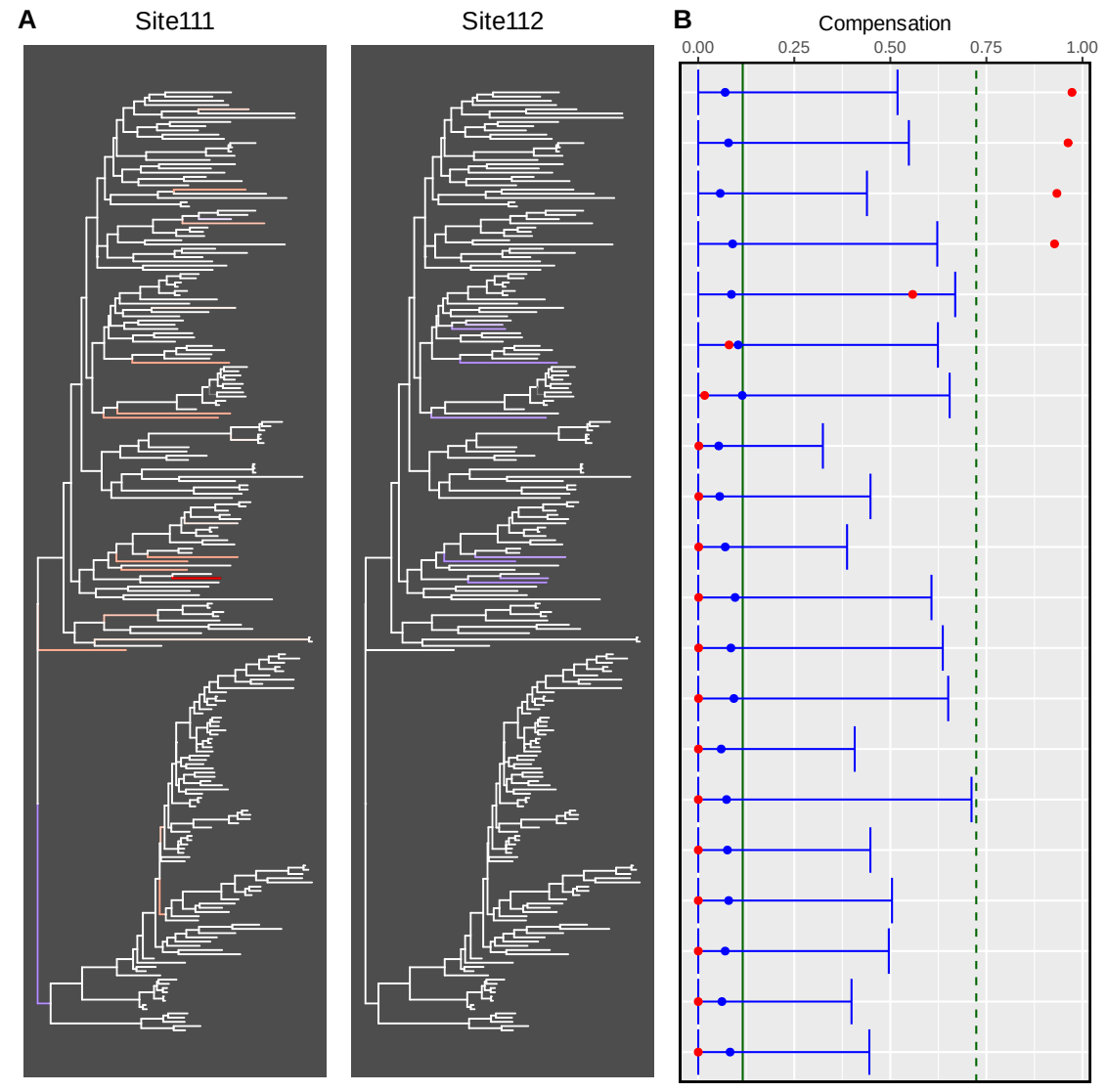
PCR showed that the *yebC* gene is being expressed in all the mutant strains, at least at the transcriptional level (Supplement Figure 8.3).

5.5   Competition experiments

We used the Malthusian parameter to calculate the relative fitness of single mutants compared with the other strain (wild type or ancestral state in this study) described in [Lenski et al., 1991]. Relative fitness greater than 1 means the one strain is better in fitness and *vice versa*. We estimated the relative fitness of single mutants compared to the double mutant and wildtype. We

used two environments for the competition experiments: one environment or medium was LB broth with enough nutrients for the growth of competing strains (nutrient-enriched medium). The rationale of using LB Broth and M9 media is explained in previous chapters (See chapter 3 section 3.5).

5.5.1 LB Broth medium

We analyzed five combinations of competing strains, and each combination had eight replicates. These combinations were T65N-YebC vs. T65N-L66I-YebC, L66I-YebC vs. T65N-L66I-YebC, T65N-YebC vs. Wild type, L66I-YebC vs. Wild type, and T65N-L66I-YebC vs. wild type as a control. Each combination was plated on TA indicator plates and used for colony counts at 0 hours, 24 hours, 48 hours, and 72 hours. We used LB Broth medium as an environment for plenty of resources for growth that was available for both competing strains. The single mutant T65N-YebC competed against T65N-L66I-YebC and the wild type for up to 72 hours separately. T65N-YebC against double mutant T65N-L66I-YebC had a significantly lower fitness difference at 72 hour time intervals (*p value = 0.00289*). Also, in the combination of wild type vs. T65N-YebC, the wild type showed significantly higher fitness at all time points. In the competition experiments between L66I-YebC and double mutant (ancestral state), L66I-YebC had only sig-

Table 5.2 Mutant strains genotypes for the YebC candidate.

Strain	Genotype
Wildtype-606 Ara$^+$	The wildtype strain of present sequence of *E. coli* Strain B REL606 used in the allelic replacement experiment as wild type. Red colonies on the TA indicator plates.
Wildtype-607 Ara$^-$	The wildtype strain of present sequence of *E. coli* Strain B REL607 used in the allelic replacement experiment as wild type. White colonies in the TA indicator plates.
T65N-YebC Ara$^-$	Threonine was replaced with Aspargine (predicted coevolving amino acids) in *yebC* called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.
L66I-YebC Ara$^-$	Leucine was replaced with Isoleucine acid (predicted coevolving amino acids) in *yebC* called as a single mutant in the background of *E. coli* Strain B REL607. White colonies on the TA indicator plates.
T65N-L66I-YebC Ara$^+$	The inferred ancestral state having both substitutions in *yebC* called as a double mutant in the background of *E. coli* Strain B REL606. Red colonies on the TA indicator plates.

nificantly lower fitness at 48 hours time intervals (*p value = 0.0003418*). At the same time, fitness difference is not significant in the competition of wildtype *vs.* L66I-YebC. The assumption of the compensation considering beta propensities was to observe lower fitness in both single mutants when they compete against inferred ancestral state or wild type. However, Both single mutant T65N-YebC and L66I-YebC showed lower fitness than double mutant or wild type but not statistically significant. The double mutant showed significantly higher fitness as compared to the wild type at all time intervals. High variation in the replicates was observed in LB broth medium for both single mutants competing against double mutant or wild type (Figure 5.4).

5.5.2 M9 minimal medium

We used the same number of combinations for the M9 minimal medium as in the LB Broth medium. M9 minimal media has a defined concentration of glucose and salts (See the materials and methods). The single mutant T65N-YebC had competed against T65N-L66I-YebC or the

Table 5.3 Selective columns from the breseq output is shown here to show the acquired mutations in the mutant strains.

Strain	Mutation	Gene	%[1]	Description
Wildtype-606	N/A	N/A	98.8	Repeat region
Wildtype-607	T → C	*araA*	89.4	L-arabinose isomerase
	A → G	recD	98.7	Exonuclease V (RecBCD complex) alpha chain
T65N-YebC-607	T → C	*araA*	98.7	L-arabinose isomerase
	A → G	*recD*	98.5	Exonuclease V (RecBCD complex) alpha chain
	Δ 1 bp	*yebC*	98.7	hypothetical protein
	+ T	*yebC*	98.7	hypothetical protein
L66I-YebC-607	T → C	*araA*	98.7	L-arabinose isomerase
	A → G	*recD*	98.7	Exonuclease V (RecBCD complex) alpha chain
	C → A	*yebC*	98.7	hypothetical proteins
T65N-L66I-606	Δ 1 bp	*yebC*	98.7	hypothetical protein
	+ T	*yebC*	98.7	hypothetical protein
	C → A	*yebC*	98.7	hypothetical proteins

[1] % of the genome where reads could be mapped

Figure 5.4 Competition experiments in LB Broth medium; Relative fitness is shown on the y-axis, and three time points for the ploting are on the x-axis. Legend as Figure 3.5
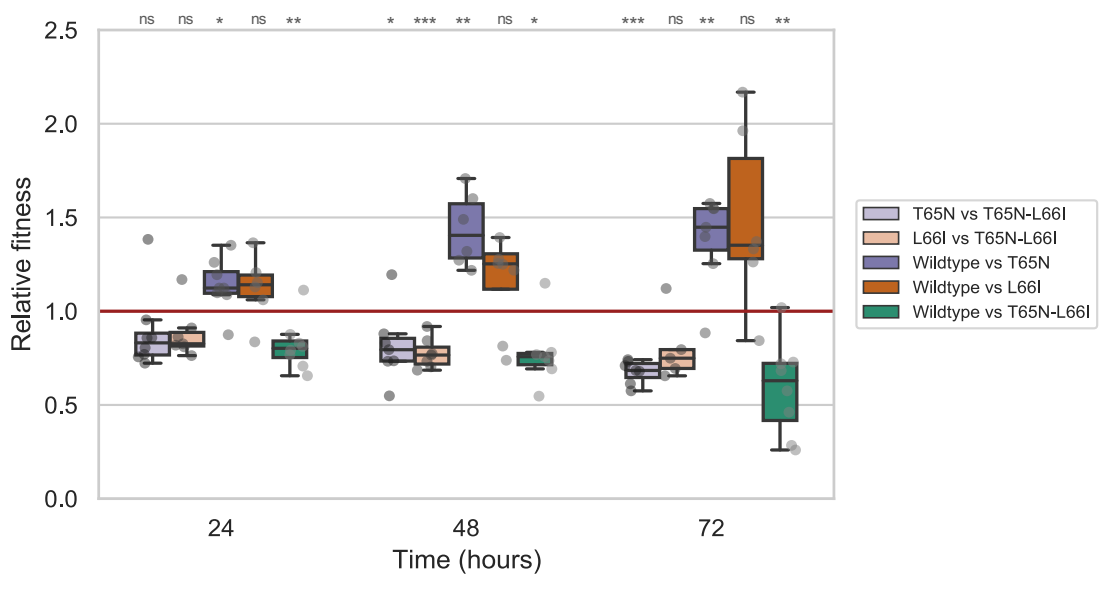
wild type for 72 hours. For the competition experiments for T65N-YebC against double mutant T65N-L66I, there were only significant differences in fitness at 72 hours (*p value = 0.0028936*). The competition of wild type and T65N-YebC also had marginal significance at 72 hours time interval *p value = 0.0439525*. The other single mutant L66I-YebC vs. double mutant T65N-L66I-YebC had a significant fitness difference at all time intervals. The competition of wild type and L66I-YebC had significant fitness differences at 48 and 72 hours. For 72 hours, both single mutants T65N-YebC and L66I-YebC showed significantly lower fitness than wild type or ancestral state, which fulfills the assumption of compensatory mutation. The ancestral state had higher fitness than the wild type at 72 hours.

5.6   Analysis of beta propensities compensation

The *yebC* gene is characterized as a transcriptional regulator in many bacterial species [Brown et al., 2017; Gaytán et al., 2016; Wei et al., 2018]. The versatile regulatory role in expressing essential transcriptional genes makes it a good candidate for genomic manipulation. Our results in this chapter show mutations on functionally important positions can have fitness effects on the organismal fitness of *E. coli*. In the YebC candidate group of coevolving amino acids, the single mutants T65N-YebC and L66I-YebC showed lower fitness in the presence of wild type or inferred ancestral state. From our results of competition experiments of single mutants *vs.* wild type or inferred ancestral state, we hypothesize that the mutations at position 65 and position

Figure 5.5 Competition experiments in M9 minimal medium; Relative fitness is shown on the y-axis, and three time points for the plating are on the x-axis. Legend as Figure 3.6
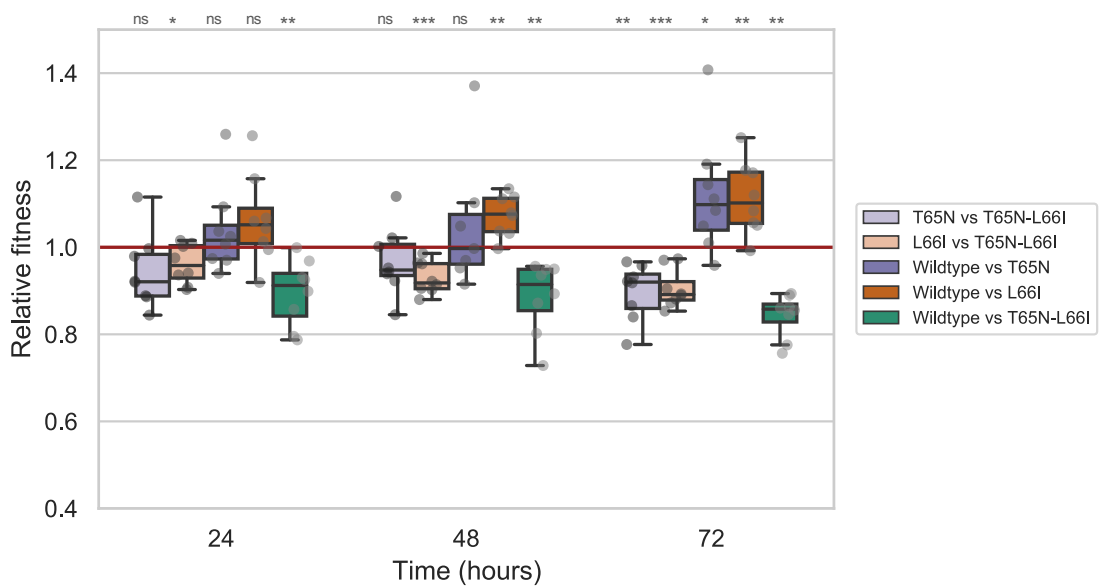
66 are not stable for the YebC protein to regulate the transcription [Brown et al., 2017]. So far, there is no information about these sites for the structural importance. We suggest that more exploration is needed for the structural aspects of these positions. For instance, sites present on active sites can have high fitness differences as they have functional constraints more than other protein sites. The results of this candidate are following the same pattern as EF4 candidate (See chapter 3 section 3.6). Furthermore, the double mutant having the compensatory mutation restores the function and stabilizes the YebC protein structure. It is a challenge to predict which mutation happens first to determine the trajectory of the mutations in the present wild type sequence.

The YebC candidate's reconstructed local fitness landscape shows the valley (lower in fitness) in the case of both single mutants T65N-YebC and L66I-YebC and restoration of fitness in ancestral state T65N-L66I-YebC. The T65N-L66I-YebC has slightly higher fitness than the wild type of the present *E. coli*. The higher fitness of the double mutant is slightly different from the assumption of compensatory evolution. Higher fitness in the ancestral state might be because the ancestral state was stable, and unknown environmental changes introduced these mutations in *E. coli* and the YebC protein is stable to perform their function as in the present sequence. An important point to keep in mind is that we tested this candidate group in the *E. coli* genomic background. Other included species in the prediction can show different results. Using *E. coli* helps to understand the compensation for the biochemical property that our results suggest and experimentally assess the predictions. It might acquire more mutations in

the inferred ancestral that we have not included in our analysis. Fewer sequences are present for the *yebC* in *E. coli*, suggesting that YebC is not well explored and has many gaps to fill in with the experimental data. It is still noted that the difference in fitness is not very high in M9 minimal media as the strain of *E coli* is adapted to the M9 minimal medium for experimental purposes. A suitable stressful environment such as ionization radiation or high temperature can give us more insight into explaining our results.

Along with the results of EF4 (See chapter 3), the results of the YebC candidate provide a second case study for the experimental evidence of the compensatory nature of the tested mutations. In YebC, beta propensities were used as biochemical property for the compensation. The yebC candidate results also highlight the importance of experimental support of the

Table 5.4 Mean relative fitness and p values for the competition experiments in LB Broth and M9 minimal media

Competing strains	Time	Medium	Mean Rel. fitness	p value
T65N-YebC vs. T65N-L66I-YebC	24 hours	LB Broth	0.888445	0.18168
T65N-YebC vs. T65N-L66I-YebC	48 hours	LB Broth	0.816725	0.04918
T65N-YebC vs. T65N-L66I-YebC	72 hours	LB Broth	0.676488	8.79e-06
L66I-YebC vs. T65N-L66I-YebC	24 hours	LB Broth	0.880129	0.05756
L66I-YebC vs. T65N-L66I-YebC	48 hours	LB Broth	0.774723	0.00034
L66I-YebC vs. T65N-L66I-YebC	72 hours	LB Broth	0.803157	0.07663
Wildtype vs. T65N-YebC	24 hours	LB Broth	1.139337	0.02619
Wildtype vs. T65N-YebC	48 hours	LB Broth	1.434702	0.00286
Wildtype vs. T65N-YebC	72 hours	LB Broth	1.379008	0.00643
Wildtype vs. L66I-YebC	24 hours	LB Broth	1.125067	0.13983
Wildtype vs. L66I-YebC	48 hours	LB Broth	1.375336	0.18297
Wildtype vs. L66I-YebC	72 hours	LB Broth	1.490042	0.05740
Wildtype vs. T65N-L66I-YebC	24 hours	LB Broth	0.819238	0.00744
Wildtype vs. T65N-L66I-YebC	48 hours	LB Broth	0.775981	0.01766
Wildtype vs. T65N-L66I-YebC	72 hours	LB Broth	0.591452	0.00258
T65N-YebC vs. T65N-L66I-YebC	24 hours	M9 Minimal	0.943980	0.10542
T65N-YebC vs. T65N-L66I-YebC	48 hours	M9 Minimal	0.967849	0.29405
T65N-YebC vs. T65N-L66I-YebC	72 hours	M9 Minimal	0.897316	0.00289
L66I-YebC vs. T65N-L66I-YebC	24 hours	M9 Minimal	0.961270	0.04648
L66I-YebC vs. T65N-L66I-YebC	48 hours	M9 Minimal	0.929472	0.00099
L66I-YebC vs. T65N-L66I-YebC	72 hours	M9 Minimal	0.904707	0.00049
Wildtype vs. T65N-YebC	24 hours	M9 Minimal	1.038179	0.32089
Wildtype vs. T65N-YebC	48 hours	M9 Minimal	1.050902	0.41583
Wildtype vs. T65N-YebC	72 hours	M9 Minimal	1.119451	0.04395
Wildtype vs. L66I-YebC	24 hours	M9 Minimal	1.038179	0.12355
Wildtype vs. L66I-YebC	48 hours	M9 Minimal	1.063861	0.00356
Wildtype vs. L66I-YebC	72 hours	M9 Minimal	1.112755	0.00673
Wildtype vs. T65N-L66I-YebC	24 hours	M9 Minimal	0.895317	0.00630
Wildtype vs. T65N-L66I-YebC	48 hours	M9 Minimal	0.885866	0.00580
Wildtype vs. T65N-L66I-YebC	72 hours	M9 Minimal	0.842455	4.63e-05

coevolution predictions to understand molecular evolution and interpret the function and 3D structure of the proteins.

Discussion

Chapter 6

Discussion

In this thesis, I reported the experimental assessment of the fitness effect of predicted compensatory mutations. Coevolving amino acids were computationally predicted using previously developed bioinformatic methods that unravel the coevolutionary process from sequence alignments and corresponding phylogenies. I have shown the results of three selected candidate coevolving pairs with a strong coevolution signal and used the model bacterial system *E. coli* to test the coevolution prediction experimentally, assessing fitness effects of mutations using competition assays after resurrecting the ancestral genotypes.

In this chapter, I will further discuss the following points:

- evidence for the compensatory mutations using competition experiments

- fitness difference of mutants analyzed using local fitness landscape

- and the environment as a contributing factor in the experiments.

In particular, we will answer the following questions:

- what did we learn about the compensatory mutations?

- How can local fitness landscapes help to predict the evolutionary history and fitness effects?

- How does this study help to interpret the impact of mutations on the 3D structure?

- How does the environment affect the experimental setup?

6.1   Evidence for compensatory mutations

Compensatory mutations are the results of the interaction among amino acids in the protein sequences. There are studies that Considerable experimental evidence is present to support

compensatory mutations [Burch and Chao, 1999; Davis et al., 2009; Estes and Lynch, 2003]. The studies mentioned above did not include the biochemical property of amino acids in their analyses. To recall the compensation for a biochemical property: a mutation from an amino acid with a small side chain to a big side chain can be compensated by a reciprocal big to small mutation at an interacting position. These interacting positions coevolve with a protein to maintain the structure of the protein [Neher, 1994]. In the competition experiments, two out of three predicted coevolved candidate groups (EF4 and YebC) showed that coevolving amino acids undergo compensatory mutations in *E. coli* genes. The lower fitness values in the single mutants and fitness restoration of double mutant suggest that single mutants were not as fit as wild type or double mutant in a competition environment. The lower fitness values of the single mutants might be because of the unstable structure of the proteins, the Elongation factor, and YebC that impacted the function of the proteins. These proteins have interactions with other proteins in the cell, and because of the protein-protein interactions, the cellular processes were not at the optimal level, resulting in lower fitness in single mutants [Beerenwinkel et al., 2007]. Two or more mutations may interact, and their combined effect on fitness might be less or greater than their individual effects. Deviation from the expected effects of individual mutations is called epistasis. Intra-molecular epistasis may contribute towards the path of the evolution [Bridgham et al., 2009; Salverda et al., 2011]. The direction of the mutations, that is, which mutation happened first, mutation at position 66 in *lepA*, or mutation at position 170 in *lepA* happened first? In this case, it is hard to determine which mutation happened first as both single mutations were lower in fitness (See Chapter 3 3.6). It is the same for the YebC candidate (see Chapter 5). In the case of these two candidates, the fitness landscape demonstrates epistasis between the tested mutations but does not allow the prediction of the exact evolutionary path. Epistasis determines the coevolutionary dynamics, and studying the evolutionary history of the sites can help to interpret epistasis.

The results of the IspH candidate did not show a similar trend of compensatory mutations on the inferred local fitness landscape (See Chapter 4). One single mutant had lower fitness than both the wild type and ancestral state. In this case, there is compensation in one direction, *i.e.*, if site 24 mutated first, the mutation at site 27 might have compensated for this mutation and led to the wild type. Conversely, the single mutant with state A27E-IspH at position 27 had higher fitness than both the ancestral state and wild type in our experiment. In this case, the mutation at position 27 had a positive effect, whatever the state at position 24 (see Figure 6.1, and that the mutations at the two positions are therefore additive, their fitness effect is

Figure 6.1 Expected local fitness landscape vs. observed local fitness landscape for EF4, IspH, and YebC. The green bars show the sequence of wild type and double mutant. The brown bar shows single mutant s1 and the purple bar shows the other single mutant s2.

Observed local fitness landscape in EF4

Expected local fitness landscape

Observed local fitness landscape in IspH

Observed local fitness landscape in YebC
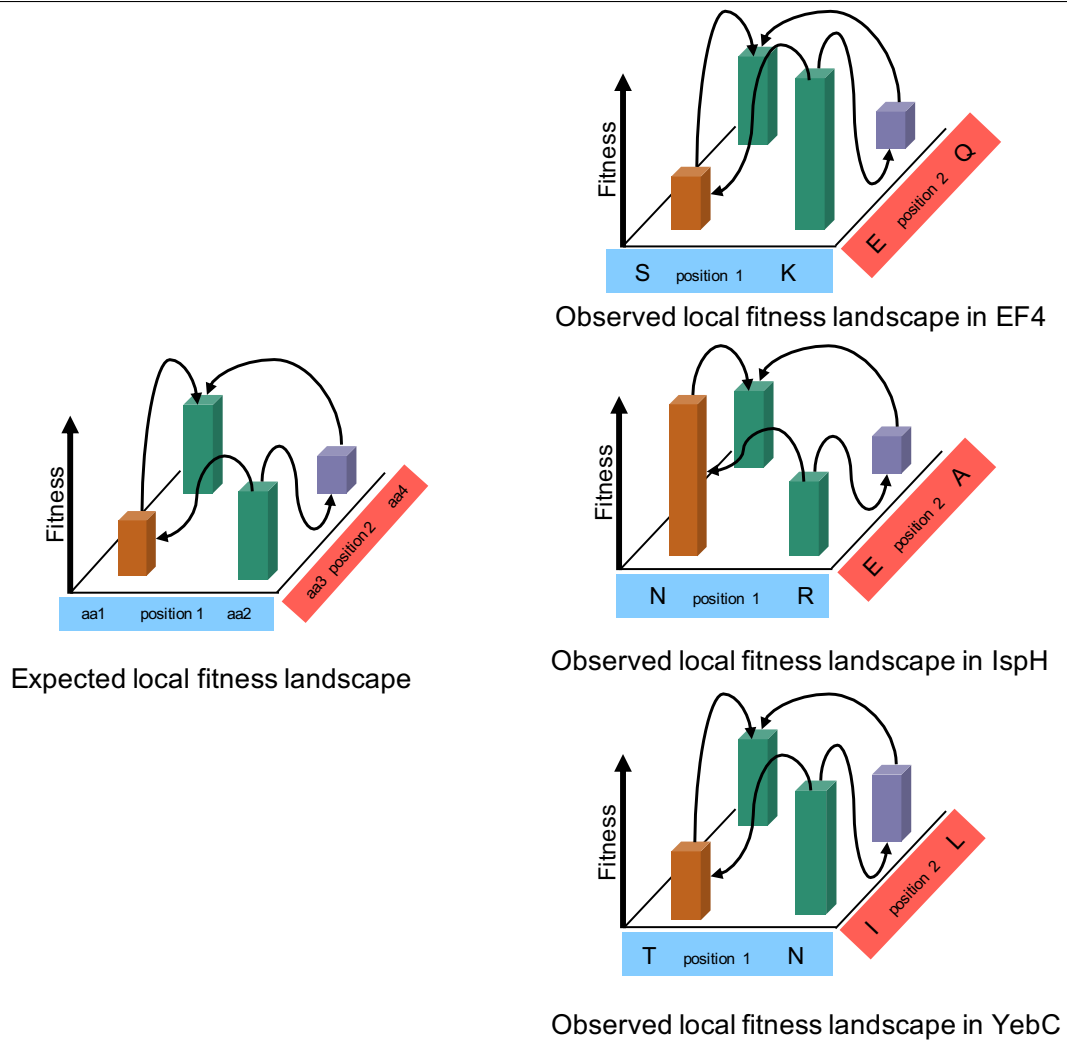
independent of the order in which the mutations occurred.

6.2 Local fitness landscape

The fitness landscape is a visual representation of the genotype-phenotype relationship and the fitness [de Visser et al., 2018]. I applied the term *local fitness landscape* to a subspace of the global fitness landscape, for a group of two coevolving positions, with two alleles each. This study has 2 x 2 local fitness landscapes with two peaks (present sequence and inferred ancestral state) and two valleys (intermediate mutants). It is important to mention here that the peaks for wild type and inferred ancestral state are from the 2 x 2 fitness landscape due to the competition assay's experimental setup. The fitness assay using competition experiments

involves two competing strains. Let us consider a bigger fitness landscape of 4 positions (more than two loci); combinations of competition experiments would be 2 x 2 x 2 x 2 = 32 to carry out in one environment to test the predicted candidate. This number is too large to be evaluated experimentally in the laboratory. To resolve this problem, we would need high-throughput techniques where automatization is involved. Several studies have used large fitness landscapes to study the global landscape [Flynn et al., 2020]. The global fitness landscape can give you a complex network of interactions at different positions. Using the local fitness landscape, we can narrow down the interactions of two positions and see the fitness effects of selected positions. In this study, the coevolving positions were studied in the local fitness landscape, and we observed the genotype-phenotype relationship of predicted coevolving positions (Figure 6.1). I argue here for such questions as predicting coevolving predictions; it helps consider the local fitness landscape compared to studying the global landscape. Also, it depends on the nature of the question as well to choose local or global fitness landscape for the study of fitness effects.

Most of the studies use protein fitness as a proxy to estimate the fitness of the organism [Brown et al., 2010; Sarkisyan et al., 2016; Bank et al., 2016]. The studies mentioned above estimate the fitness of the proteins. It might help explain the protein fitness but would not show the fitness effects due to protein-protein interactions in the organism's cells. We address this problem by estimating the fitness of the organism directly. We generated the mutations into the genome of the *E. coli* rather than expressing the genes in plasmids and estimate the fitness of the protein. The fitness effects of this study's results highlight the complex protein network involved in the fitness of the organisms. When we only mutate a few positions in one of the genes, we observe different behavior of the competing strain in the environment. Thus, for the fitness assays, direct fitness estimation can explain a better picture of fitness effects.

6.3   Interactions in 3D structure

The three-dimensional protein architecture represents the folding of the proteins and attracts structure biologists towards understanding the protein's function. Studying interactions of amino acids in the proteins is an exciting problem whose explanation might help solve the proteins' folding. Some methods try to predict the distance and contact of the amino acids in the 3D structure of the proteins [Fariselli et al., 2001; Bohr et al., 1993; Olmea and Valencia, 1997; Vendruscolo et al., 1997]. Morcos et al. [2011] developed a method to predict the contacts of the amino acids from sequence data (DCA). The method mentioned above suggested the accuracy of the prediction for protein conformation, protein complex formation, and global

protein structure. These methods are based on bioinformatics predictions and lack experimental assessment. Various prediction methods suggested that significant proportions of coevolving positions are in contact in the 3D structure. In this study, the selected coevolving positions were in contact in the 3D structure, and they might be true coevolving residues. Based on our results of fitness assay for all of the three candidates, we can hypothesize that the proteins were not stable and because of the conformational change in the proteins. Conformational change might affect the function of the protein and exerted an effect on the interactions with the other proteins (protein-protein interactions) in the cell. For instance, charge as a biochemical property contributes substantially to the residue coevolution [Chakrabarti and Panchenko, 2010]. Change in charge of the protein might have conformational change and affect the solubility of the protein. The coevolving positions that were studied here are not present in the active site.

Not all coevolving positions are necessarily in contact. Distant sites in distinct protein domains were also suggested to be coevolving [Anishchenko et al., 2017]. However, evolution is stronger on the amino acids that are physically interacting than long distances in the network. Thus, the fitness effects in the mutants in the candidate genes might have stronger evolutionary pressure because of contact in the 3D structure.

6.4 Environment as contributing factor

Individuals do not respond identically towards the same environment [Flynn et al., 2020]. Finding an appropriate environment for assessing an organism's fitness in laboratory conditions is a challenge. Natural environments for testing effects of mutations are not known and hard to replicate in laboratory conditions. I used two environments, a defined M9 minimal media with limited resources and a single carbon source as a stress environment. We have observed less variation in the technical replicates of the competition experiments. The less variation is might be because *E. coli* Strain B is adapted to the M9 minimal media as a stress environment for all the experimental studies. The difference in the fitness is also not large as mutants, and wild type both were growing at the optimal growth level. Whereas LB Broth is enriched with sufficient nutrients for both competing strains in the fitness assay, I used this medium as a non-stress environment. There are multiple options in selecting environments. For instance, knockout of *ispH* genes are only viable in the medium supplemented by mevalonate [Altincicek et al., 2001]. Such a specific medium can be used to assess the fitness effects. One of the causes of mutations is environmental stress. By using the M9 minimal medium, I hypothesize that stress conditions and cells are compromised for growth. Notably, the results in the IspH

candidate can also be interpreted as the mutants may behave differently in a different environ-
ment. Furthermore, we can extend the range of environments to observe fitness differences in
competition experiments. Another critical aspect of the laboratory environments is that they
are not the real conditions in the natural environment, and it is difficult to answer questions
such as compensatory evolution within a protein from an organism accurately. The realistic
way is to have an estimation of the fitness difference in a given laboratory environment.

6.5 Conclusion and Outlook

We have analyzed three candidates from a list of 13 short-listed candidates. The results of
the three candidates showed us the compensation for the biochemical properties in the fitness
assay. Results have shown us the evidence of compensatory mutations and observed the intra-
molecular coevolution experimentally. The present experimental setup has proved to test the
fitness of the organism in the competition experiments. However, there is still a need for a pre-
cise estimation of fitness. Classical methods like colony counting in competition experiments
are time-consuming. One way to improve is to tag the cells with fluorescence proteins (GFP) and
use high-through-put approaches like Flow cytometry (FACS). FACS can help to count cells in
a shorter time with high precision. Further analysis of the thermodynamics of the 3D structure
of the candidate proteins would be an exciting aspect of this study. The 3D structure of the mu-
tant proteins can be analyzed *in-silico* using homology modeling. More analysis of 3D structure
would predict the change in the 3D structure of the candidate proteins because of the predicted
compensatory mutations. The future goal would be to assess the protein fitness of candidates
used in this study. By including protein fitness in the study, would help to understand the fit-
ness effects on the protein level and can also help to identify the protein-protein interaction.
The complex protein-protein interactions lead to the lower fitness of the single mutants. For
instance, using methods, protein purification and X-ray crystallography or nuclear magnetic
resonance spectroscopy could highlight the structural changes in these proteins because of the
compensation of biochemical properties. More laboratory experiments can also be included in
the existing data set of competition experiments as I have shown thirteen potential candidates
suitable for the experimental evaluation. The eukaryotic protein data set is also applicable for
the CoMap to detect coevolution. The increasing amount of sequencing data provides an op-
portunity to enhance the coevolution analysis dataset and increase the prediction's precision.
An improvement in the relative fitness estimation would be an excellent addition to the present
study and increase the accuracy of fitness calculations.

I report the first experimental assessment of the prediction of coevolution within a protein. Our results provide experimental evidence of the compensating nature of the tested mutations, highlighting the potential of coevolution detection methods as tools to understand molecular evolution. In the light of *E. coli* as a model system, other microorganisms can also be interesting to study the coevolution experimentally. This study's findings highlight the importance of experimental assessment of coevolving amino acids within proteins and aid the ideas of testing coevolution (inter-molecular or intra-molecular) experimentally.

Bibliography

Chapter 7

Bibliography

Bibliography

P. K. Ajikumar, K. Tyo, S. Carlsen, O. Mucha, T. H. Phon, and G. Stephanopoulos. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Molecular pharmaceutics*, 5(2):167–190, 2008.

B. Altincicek, A.-K. Kollas, M. Eberl, J. Wiesner, S. Sanderbrand, M. Hintz, E. Beck, and H. Jomaa. Lytb, a novel gene of the 2-c-methyl-d-erythritol 4-phosphate pathway of isoprenoid biosynthesis in escherichia coli. *FEBS letters*, 499(1-2):37–40, 2001.

I. Anishchenko, S. Ovchinnikov, H. Kamisetty, and D. Baker. Origins of coevolution between residues distant in protein 3d structures. *Proceedings of the National Academy of Sciences*, 114 (34):9122–9127, 2017.

W. R. Atchley, K. R. Wollenberg, W. M. Fitch, W. Terhalle, and A. W. Dress. Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Molecular biology and evolution*, 17(1):164–178, 2000. ISSN 1537-1719.

C. Bank, S. Matuszewski, R. T. Hietpas, and J. D. Jensen. On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences*, 113(49):14085 LP – 14090, dec 2016. doi: 10.1073/pnas.1612676113. URL `http://www.pnas.org/content/113/49/14085.abstract`.

N. Beerenwinkel, L. Pachter, B. Sturmfels, S. F. Elena, and R. E. Lenski. Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC evolutionary biology*, 7(1):1–12, 2007.

J. J. Bijlsma, M. Lie-A-Ling, I. C. Nootenboom, C. M. Vandenbroucke-Grauls, and J. G. Kusters. Identification of loci essential for the growth of helicobacter pylori under acidic conditions. *The Journal of infectious diseases*, 182(5):1566–1569, 2000.

J. Bohr, H. Bohr, S. Brunak, R. M. Cotterill, H. Fredholm, B. Lautrup, and S. B. Petersen. Protein structures from distance inequalities. *Journal of molecular biology*, 231(3):861–869, 1993.

A. Boronat and M. Rodríguez-Concepciön. Terpenoid biosynthesis in prokaryotes. *Biotechnology of isoprenoids*, pages 3–18, 2014.

J. T. Bridgham, E. A. Ortlund, and J. W. Thornton. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, 461(7263):515–519, 2009.

K. M. Brown, M. S. Costanzo, W. Xu, S. Roy, E. R. Lozovsky, and D. L. Hartl. Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Molecular biology and evolution*, 27(12):2682–2690, dec 2010. ISSN 1537-1719 (Electronic). doi: 10.1093/molbev/msq160.

L. Brown, J. M. Villegas, M. Elean, S. Fadda, F. Mozzi, L. Saavedra, and E. M. Hebert. Yebc, a putative transcriptional factor involved in the regulation of the proteolytic system of lactobacillus. *Scientific reports*, 7(1):1–11, 2017.

C. L. Burch and L. Chao. Evolution by small steps and rugged landscapes in the rna virus ϕ6. *Genetics*, 151(3):921–927, 1999.

R. T. Byrne, S. H. Chen, E. A. Wood, E. L. Cabot, and M. M. Cox. Escherichia coli genes and pathways involved in surviving extreme exposure to ionizing radiation. *Journal of bacteriology*, 196(20):3534–3545, 2014a.

R. T. Byrne, A. J. Klingele, E. L. Cabot, W. S. Schackwitz, J. A. Martin, J. Martin, Z. Wang, E. A. Wood, C. Pennacchio, L. A. Pennacchio, et al. Evolution of extreme resistance to ionizing radiation via genetic adaptation of dna repair. *Elife*, 3:e01322, 2014b.

S. Chakrabarti and A. R. Panchenko. Structural and functional roles of coevolved sites in proteins. *PloS one*, 5(1):e8591, 2010.

S. S. Chandran, J. T. Kealey, and C. D. Reeves. Microbial production of isoprenoids. *Process biochemistry*, 46(9):1703–1710, 2011.

L.-M. Chevin. On measuring selection in experimental evolution. *Biology letters*, 7(2):210–213, apr 2011. ISSN 1744-957X (Electronic). doi: 10.1098/rsbl.2010.0580.

C. Collins, X. Zhou, R. Wang, M. C. Barth, T. Jiang, J. A. Coderre, and P. C. Dedon. Differential oxidation of deoxyribose in dna by γ and α-particle radiation. *Radiation research*, 163(6):654–662, 2005.

P. Daegelen, F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). *Journal of molecular biology*, 394(4):634–643, dec 2009. ISSN 1089-8638 (Electronic). doi: 10.1016/j.jmb.2009.09.022.

B. H. Davis, A. F. Y. Poon, and M. C. Whitlock. Compensatory mutations are repeatable and clustered within proteins. *Proceedings of the Royal Society B: Biological Sciences*, 276(1663): 1823–1827, 2009. ISSN 0962-8452.

J. A. G. de Visser, S. F. Elena, I. Fragata, and S. Matuszewski. The utility of fitness landscapes and big data for predicting evolution, 2018.

D. E. Deatherage and J. E. Barrick. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in molecular biology (Clifton, N.J.)*, 1151:165–188, 2014. ISSN 1940-6029. doi: 10.1007/978-1-4939-0554-6_ 12. URL `https://pubmed.ncbi.nlm.nih.gov/24838886https://www. ncbi.nlm.nih.gov/pmc/articles/PMC4239701/`.

K. T. der Formbildung. Borntraeger 1928. *English: Modern Theories of Development: An Introduction to Theoretical Biology, Oxford University Press, New York: Harper*, 1933.

N. Dibb and P. Wolfe. lep operon proximal gene is not required for growth or secretion by escherichia coli. *Journal of bacteriology*, 166(1):83–87, 1986.

J. Dutheil and B. Boussau. Non-homogeneous models of sequence evolution in the bio++ suite of libraries and programs. *BMC evolutionary biology*, 8(1):1–12, 2008.

J. Dutheil and N. Galtier. Detecting groups of coevolving positions in a molecule: a clustering approach. *BMC Evolutionary Biology*, 7(1):242, nov 2007. ISSN 1471-2148. doi: 10.1186/ 1471-2148-7-242. URL `https://doi.org/10.1186/1471-2148-7-242`.

J. Dutheil, T. Pupko, A. Jean-Marie, and N. Galtier. A model-based approach for detecting coevolving positions in a molecule. *Molecular biology and evolution*, 22(9):1919–1928, 2005. ISSN 1537-1719.

J. Y. Dutheil. Detecting coevolving positions in a molecule: why and how to account for phylogeny. *Briefings in Bioinformatics*, 13(2):228–243, sep 2011. ISSN 1467-5463. doi: 10.1093/bib/bbr048. URL `https://doi.org/10.1093/bib/bbr048`.

S. Estes and M. Lynch. Rapid fitness recovery in mutationally degraded lines of caenorhabditis elegans. *Evolution*, 57(5):1022–1030, 2003.

R. N. Evans, G. Blaha, S. Bailey, and T. A. Steitz. The structure of lepa, the ribosomal back translocase. *Proceedings of the National Academy of Sciences*, 105(12):4673–4678, 2008.

P. Fariselli, O. Olmea, A. Valencia, and R. Casadio. Prediction of contact maps with neural networks and correlated mutations. *Protein engineering*, 14(11):835–843, 2001.

J. Felsenstein. Phylogenies and the Comparative Method. *The American Naturalist*, 125(1):1–15, 1985. ISSN 00030147, 15375323. URL http://www.jstor.org/stable/2461605.

J. M. Flynn, A. Rossouw, P. Cote-Hammarlof, I. Fragata, D. Mavor, C. r. Hollins, C. Bank, and D. N. Bolon. Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *eLife*, 9, mar 2020. ISSN 2050-084X (Electronic). doi: 10.7554/eLife.53810.

A. Frank and M. Groll. The methylerythritol phosphate pathway to isoprenoids. *Chemical Reviews*, 117(8):5675–5703, 2017.

I. Frumkin, M. J. Lajoie, C. J. Gregg, G. Hornung, G. M. Church, and Y. Pilpel. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proceedings of the National Academy of Sciences*, 115(21):E4940–E4949, 2018.

K. Fujiwara, H. Toda, and M. Ikeguchi. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC structural biology*, 12(1):1–15, 2012.

N. Galtier and J. Dutheil. Coevolution within and between genes. *Gene and Protein Evolution*, 3:1–12, 2007.

M. O. Gaytán, V. I. Martínez-Santos, E. Soto, and B. González-Pedrajo. Type three secretion system in attaching and effacing pathogens. *Frontiers in cellular and infection microbiology*, 6:129, 2016.

M. Gouy and S. Delmotte. Remote access to acnuc nucleotide and protein sequence databases at pbil. *Biochimie*, 90(4):555–562, 2008.

R. R. Gutell, A. Power, G. Z. Hertz, E. J. Putz, and G. D. Stormo. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Research*, 20(21):5785–5795, nov 1992. ISSN 0305-1048. doi: 10.1093/nar/20.21.5785. URL https://doi.org/10.1093/nar/20.21.5785.

I. Hale, P. M. O'Neill, N. G. Berry, A. Odom, and R. Sharma. The mep pathway and the development of inhibitors as potential anti-infective agents. *MedChemComm*, 3(4):418–433, 2012.

H. Jeong, V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S.-H. Choi, A. Couloux, S.-W. Lee, S. H. Yoon, L. Cattolico, et al. Genome sequences of escherichia coli b strains rel606 and bl21 (de3). *Journal of molecular biology*, 394(4):644–652, 2009.

J. Kirby and J. D. Keasling. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annual review of plant biology*, 60:335–355, 2009.

B. T. Korber, R. M. Farber, D. H. Wolpert, and A. S. Lapedes. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proceedings of the National Academy of Sciences*, 90(15):7176–7180, 1993. ISSN 0027-8424.

R. Krisch, F. Krasin, and C. Sauri. Dna breakage, repair, and lethality accompanying 125i decay in microorganisms. *Current topics in radiation research quarterly*, 12(1-4):355–368, 1978.

S. M. Larson, A. A. Di Nardo, and A. R. Davidson. Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. *Journal of molecular biology*, 303(3):433–446, 2000. ISSN 0022-2836.

R. E. Lenski. Experimental studies of pleiotropy and epistasis in Escherichia coli. II. Compensation for maladaptive effects associated with resistance to virus T4. *Evolution*, 42(3):433–440, 1988. ISSN 0014-3820.

R. E. Lenski, M. R. Rose, S. C. Simpson, and S. C. Tadler. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, 138(6):1315–1341, 1991. ISSN 00030147, 15375323. URL `http://www.jstor.org/stable/2462549`.

H. K. Lichtenthaler, M. Rohmer, and J. Schwender. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiologia plantarum*, 101(3):643–652, 1997.

A. J. Link, D. Phillips, and G. M. Church. Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characteri-

zation. *Journal of bacteriology*, 179(20):6228–6237, oct 1997. ISSN 0021-9193 (Print). doi: 10.1128/jb.179.20.6228-6237.1997.

H. Liu, D. Pan, M. Pech, and B. S. Cooperman. Interrupted catalysis: the ef4 (lepa) effect on back-translocation. *Journal of molecular biology*, 396(4):1043–1052, 2010.

S. C. Lovell and D. L. Robertson. An integrated view of molecular coevolution in protein–protein interactions. *Molecular biology and evolution*, 27(11):2567–2575, 2010. ISSN 1537-1719.

T. Makino and T. Gojobori. Evolution of protein-protein interaction network. *Gene and protein evolution*, 3:13–29, 2007.

P. E. March and M. Inouye. Characterization of the lep operon of escherichia coli. identification of the promoter and the gene upstream of the signal peptidase i gene. *Journal of Biological Chemistry*, 260(12):7206–7213, 1985.

S. McAteer, A. Coulson, N. McLennan, and M. Masters. The lytb gene of escherichia coli is essential and specifies a product needed for isoprenoid biosynthesis. *Journal of bacteriology*, 183(24):7403–7407, 2001.

A. M. Michel and P. V. Baranov. Ribosome profiling: a hi-def monitor for protein synthesis at the genome-wide scale. *Wiley Interdisciplinary Reviews: RNA*, 4(5):473–490, 2013.

R. B. Mikkelsen and P. Wardman. Biological chemistry of reactive oxygen and nitrogen and radiation-induced signal transduction mechanisms. *Oncogene*, 22(37):5734–5754, 2003.

N. Misawa. Pathway engineering for functional isoprenoids. *Current opinion in biotechnology*, 22(5):627–633, 2011.

F. B.-G. Moore, D. E. Rozen, and R. E. Lenski. Pervasive compensatory adaptation in escherichia coli. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1442):515–522, 2000.

F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49): E1293–E1301, 2011.

D. H. Morgan, D. M. Kristensen, D. Mittelman, and O. Lichtarge. Et viewer: an application for predicting and visualizing functional sites in protein structures. *Bioinformatics*, 22(16): 2049–2050, 2006.

E. Neher. How frequent are correlated changes in families of protein sequences? *Proceedings of the National Academy of Sciences*, 91(1):98–102, 1994. ISSN 0027-8424.

E. P. Odum and G. W. Barrett. *Fundamentals of ecology*, volume 3. Saunders Philadelphia, 1971.

O. Olmea and A. Valencia. Improving contact predictions by the combination of correlated mutations and other sources of sequence information. *Folding and Design*, 2:S25–S32, 1997.

M. Pech, Z. Karim, H. Yamamoto, M. Kitakawa, Y. Qin, and K. H. Nierhaus. Elongation factor 4 (ef4/lepa) accelerates protein synthesis at increased mg2+ concentrations. *Proceedings of the National Academy of Sciences*, 108(8):3199–3203, 2011.

M. A. Phillips, P. León, A. Boronat, and M. Rodríguez-Concepción. The plastidial mep pathway: unified nomenclature and resources. *Trends in plant science*, 13(12):619–623, 2008.

D. D. Pollock, W. R. Taylor, and N. Goldman. Coevolving protein residues: maximum likelihood identification and relationship to structure. *Journal of molecular biology*, 287(1):187–198, 1999. ISSN 0022-2836.

Y. Qin, N. Polacek, O. Vesper, E. Staub, E. Einfeldt, D. N. Wilson, and K. H. Nierhaus. The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosome. *Cell*, 127(4):721–733, nov 2006. ISSN 0092-8674 (Print). doi: 10.1016/j.cell.2006.09.037.

J. Repar, S. Cvjetan, D. Slade, M. Radman, D. Zahradka, and K. Zahradka. Reca protein assures fidelity of dna repair and genome stability in deinococcus radiodurans. *DNA repair*, 9(11): 1151–1161, 2010.

F. Rohdich, F. Zepeck, P. Adam, S. Hecht, J. Kaiser, R. Laupitz, T. Gräwert, S. Amslinger, W. Eisenreich, and A. Bacher. The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proceedings of the National Academy of Sciences*, 100(4):1586–1591, 2003. ISSN 0027-8424.

M. Rohmer, M. Seemann, S. Horbach, S. Bringer-Meyer, and H. Sahm. Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. *Journal of the American Chemical Society*, 118(11):2564–2566, 1996.

I. Saha, U. Maulik, S. Bandyopadhyay, and D. Plewczynski. Fuzzy clustering of physicochemical and biochemical properties of amino acids. *Amino Acids*, 43(2):583–594, 2012. ISSN 0939-4451.

M. L. Salverda, E. Dellus, F. A. Gorter, A. J. Debets, J. Van Der Oost, R. F. Hoekstra, D. S. Tawfik, and J. A. G. de Visser. Initial mutations direct alternative pathways of protein evolution. *PLoS Genet*, 7(3):e1001321, 2011.

N. J. Sargentini and K. C. Smith. Growth-medium-dependent repair of dna single-strand and double-strand breaks in x-irradiated escherichia coli. *Radiation research*, 104(1):109–115, 1985.

K. S. Sarkisyan, D. A. Bolotin, M. V. Meer, D. R. Usmanova, A. S. Mishin, G. V. Sharonov, D. N. Ivankov, N. G. Bozhanova, M. S. Baranov, O. Soylemez, N. S. Bogatyreva, P. K. Vlasov, E. S. Egorov, M. D. Logacheva, A. S. Kondrashov, D. M. Chudakov, E. V. Putintseva, I. Z. Mamedov, D. S. Tawfik, K. A. Lukyanov, and F. A. Kondrashov. Local fitness landscape of the green fluorescent protein. *Nature*, 533(7603):397–401, may 2016. ISSN 1476-4687 (Electronic). doi: 10.1038/nature17995.

M. Seemann, K. Janthawornpong, J. Schweizer, L. H. Bö ttger, A. Janoschka, A. Ahrens-Botzong, E. N. Tambou, O. Rotthaus, A. X. Trautwein, and M. Rohmer. Isoprenoid biosynthesis via the mep pathway: in vivo mossbauer spectroscopy identifies a [4fe-4s] 2+ center with unusual coordination sphere in the lytb protein. *Journal of the American Chemical Society*, 131(37): 13184–13185, 2009.

I. N. Shindyalov, N. A. Kolchanov, and C. Sander. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Engineering, Design and Selection*, 7(3):349–358, mar 1994. ISSN 1741-0126. doi: 10.1093/protein/7.3.349. URL `https://doi.org/10.1093/protein/7.3.349`.

R. Simon, U. Priefer, and A. Pühler. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology*, 1(9):784–791, 1983. ISSN 1546-1696. doi: 10.1038/nbt1183-784. URL `https://doi.org/10.1038/nbt1183-784`.

D. Slade, A. B. Lindner, G. Paul, and M. Radman. Recombination and replication in dna repair of heavily irradiated deinococcus radiodurans. *Cell*, 136(6):1044–1055, 2009.

M. Suyama, D. Torrents, and P. Bork. Pal2nal: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, 34(suppl_2):W609–W612, 2006.

B. Szamecz, G. Boross, D. Kalapis, K. Kovács, G. Fekete, Z. Farkas, V. Lázár, M. Hrtyan, P. Kemmeren, M. J. G. Koerkamp, et al. The genomic landscape of compensatory evolution. *PLoS Biol*, 12(8):e1001935, 2014.

J. N. Thompson. *The coevolutionary process*. University of Chicago Press, 1994. ISBN 0226797597.

J. N. Thompson. Four Central Points About Coevolution. *Evolution: Education and Outreach*, 3(1):7–13, 2010. ISSN 1936-6434. doi: 10.1007/s12052-009-0200-x. URL https://doi.org/10.1007/s12052-009-0200-x.

P. Tuffery, M. Durand, and P. Darlu. How possible is the detection of correlated mutations? *Theoretical Chemistry Accounts*, 101(1):9–15, 1999.

J. Van Cleve and D. B. Weissman. Measuring ruggedness in fitness landscapes. *Proceedings of the National Academy of Sciences*, 112(24):7345 LP – 7346, jun 2015. doi: 10.1073/pnas.1507916112. URL http://www.pnas.org/content/112/24/7345.abstract.

M. Vendruscolo, E. Kussell, and E. Domany. Recovery of protein structure from contact maps. *Folding and Design*, 2(5):295–306, 1997.

V. C. Ward, A. O. Chatzivasileiou, and G. Stephanopoulos. Metabolic engineering of Escherichia coli for the production of isoprenoids. *FEMS Microbiology Letters*, 365(10), mar 2018. ISSN 15746968. doi: 10.1093/femsle/fny079. URL https://doi.org/10.1093/femsle/fny079.

L. Wei, Y. Wu, H. Qiao, W. Xu, Y. Zhang, X. Liu, and Q. Wang. Yebc controls virulence by activating t3ss gene expression in the pathogen edwardsiella piscicida. *FEMS microbiology letters*, 365(14):fny137, 2018.

M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa. Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1):67–72, 2009.

D. M. Weinreich, N. F. Delaney, M. A. DePristo, and D. L. Hartl. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, 312(5770):111 LP – 114, apr 2006. doi: 10.1126/science.1123539. URL http://science.sciencemag.org/content/312/5770/111.abstract.

M. Wolff, M. Seemann, B. T. S. Bui, Y. Frapart, D. Tritsch, A. G. Estrabot, M. Rodré guez Concepción, A. Boronat, A. Marquet, and M. Rohmer. Isoprenoid biosynthesis via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (LytB/IspH) from Escherichia coli is a [4Fe−4S] protein. *FEBS letters*, 541(1-3):115−120, 2003. ISSN 1873-3468.

C. L. Worth, S. Gong, and T. L. Blundell. Structural and functional constraints in the evolution of protein families. *Nature Reviews Molecular Cell Biology*, 10(10):709−720, 2009.

Q. E. Yang, C. MacLean, A. Papkou, M. Pritchard, L. Powell, D. Thomas, D. O. Andrey, M. Li, B. Spiller, W. Yang, and T. R. Walsh. Compensatory mutations modulate the competitiveness and dynamics of plasmid-mediated colistin resistance in Escherichia coli clones. *The ISME Journal*, 14(3):861−865, 2020. ISSN 1751-7370. doi: 10.1038/s41396-019-0578-6. URL `https://doi.org/10.1038/s41396-019-0578-6`.

Supplementary Materials

Chapter 8

Supplements

Figure 8.1 Expression of *lepA* mutant strains; Lane 1 is the wild type. Lane 2 and 3 are single mutants S66K and Q170E, respectively. Lane 4 is S66K-Q170E, the double mutant. M is a ladder (NEB, N3232L).
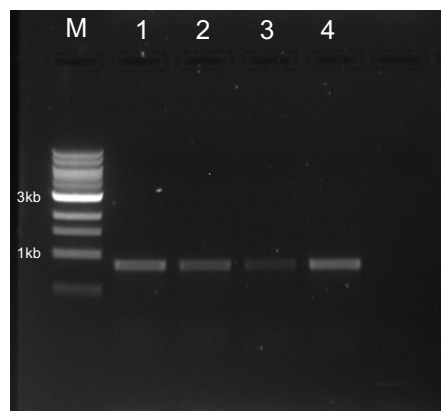
Figure 8.2 Expression of *ispH* mutant strains; Lane 1 is the wild type. Lane 2 and 3 are single mutants N24R and A27E, respectively. Lane 4 is N24R-A27E, the double mutant. M is a ladder (NEB, N3232L).
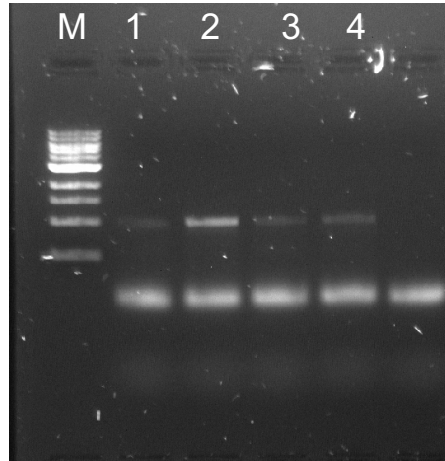
Figure 8.3 Expression of *yebC* mutant strains; Lane 1 is the wild type. Lane 2 and 3 are single mutants T65N and L66I, respectively. Lane 4 is T65N-L66I, the double mutant. M is a ladder (NEB, N3232L).
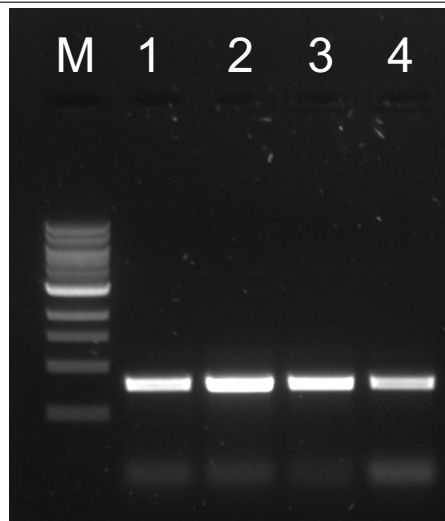
Table 8.1 Antibiotics used in this study

Name	Working concentration
Spectinomycin	100μL
Streptomycin	100μL
Chloramphenicol	20μL
Kanamycin	100μL

Figure 8.4 pKOV_unstuff plasmid maps
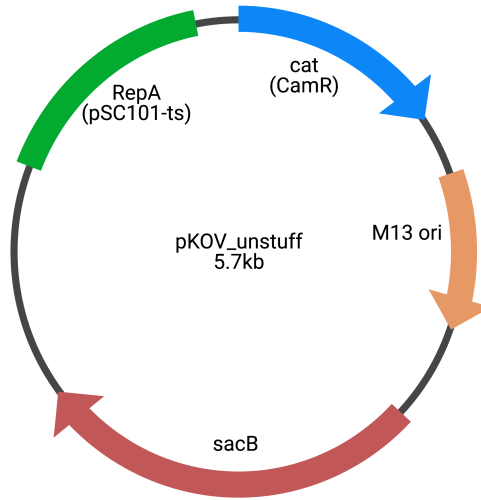
Figure 8.5 pUC57-kan maps
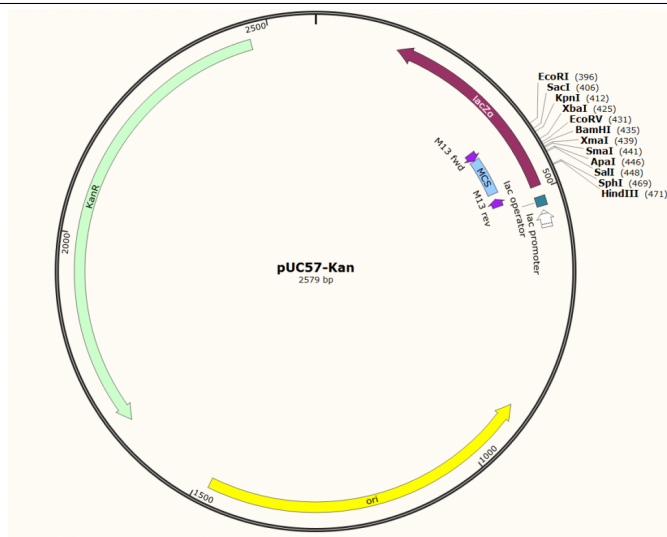
Figure 8.6 All plasmid maps for *lepA* gene used for homologous recombination
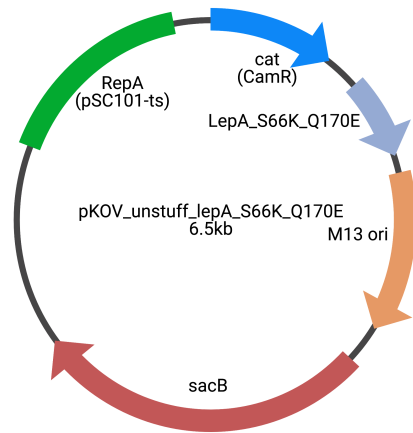
(a) Plasmid map of S66K-Q170E
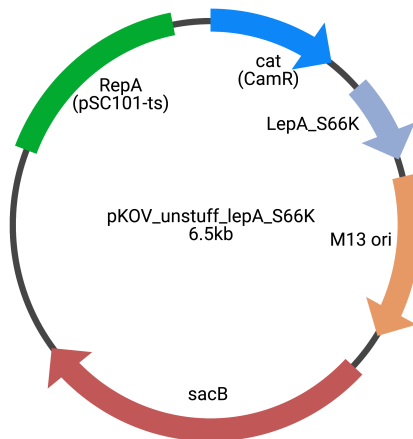
(b) Plasmid map of S66K
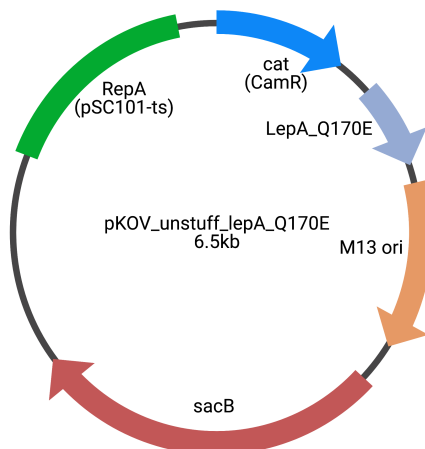
(c) Plasmkd map of Q170E

Figure 8.7 All plasmid maps for *isph* gene used for homologous recombination
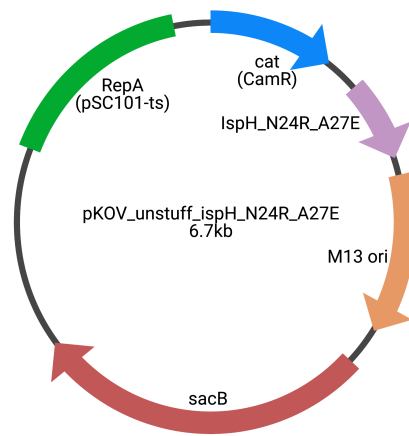
(a) Plasmid map of N24R-A27E
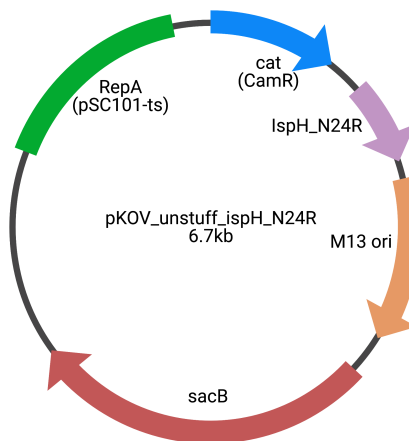
(b) Plasmid map of N24R
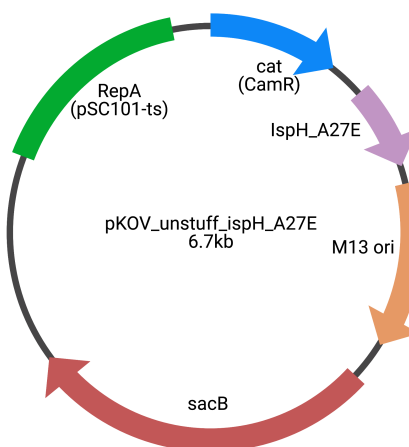
(c) Plasmkd map of A27E

Figure 8.8 All plasmid maps for *isph* gene used for homologous recombination
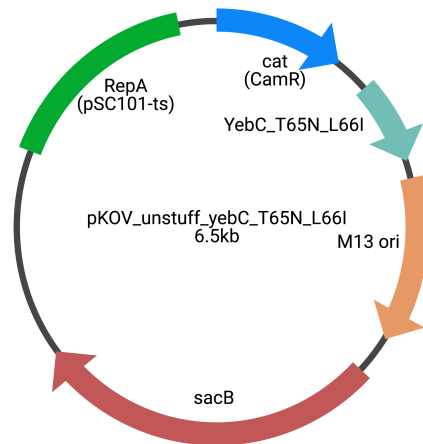
(a) Plasmid map of T65N-L66I
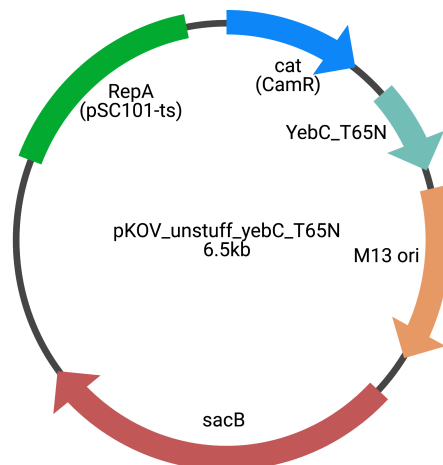
(b) Plasmid map of T65N
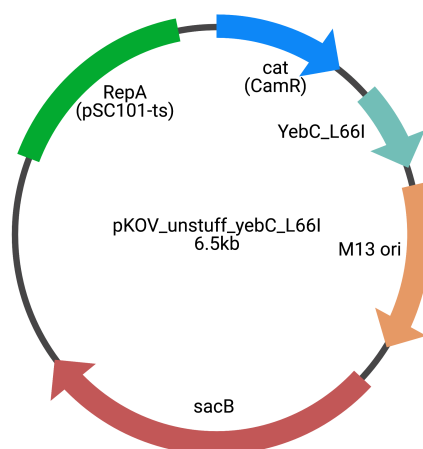
(c) Plasmkd map of L66I

Table 8.2: Overview of the whole-genome sequencing results analyzed by breseq with default settings

Strain	Mutation	Position	Annotation	Gene	Junction[1]	%[2]	Description
Wildtype-606	N/A	= 3894996	Noncoding (1443/1443 nt)	IS150	Yes	98.8	Repeat region
	N/A	3902278 =	Coding 142/1428 nt	*yieO*	Yes	98.8	Predicted multidrug or homocysteine efflux system
Wildtype-607	T − > C	70,867	D92G (GAC > GGC)	*araA*	No	89.4	L-arabinose isomerase
	A − > G	2,847,052	V10A (GTT > GCT)	*recD*	No	89.4	Exonuclease V (RecBCD complex) alpha chain
N24R-IspH-607	T − > C	70,867	D92G (GAC > GGC)	*araA*	No	89.4	L-arabinose isomerase
	A − > G	2,847,052	V10A (GTT > GCT)	*recD*	No	89.4	Exonuclease V (RecBCD complex) alpha chain
	2 bp > CG	30,418	Coding (70-71/951 nt)	*ispH*	Yes	89.4	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
A27E-IspH-606	2 bp > AG	30, 428	Coding (80-81/951 nt)	*ispH*	No	98.6	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
N24R-A27E-IspH-606	2 bp > CG	30,418	Coding (70-71/951 nt)	*ispH*	Yes	92.3	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
	2 bp > AG	30, 428	Coding (80-81/951 nt)	*ispH*	No	92.3	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
S66K-LepA-607	T − > C	70,867	D92G (GAC > GGC)	*araA*	No	98.5	L-arabinose isomerase
	A − > G	2,847,052	V10A (GTT > GCT)	*recD*	No	98.5	Exonuclease V (RecBCD complex) alpha chain

Strain	Mutation	Position	Annotation	Gene	Junction[1]	%[2]	Description
	3 bp > CCT	2,628,420	Coding (196-198/1800 nt)	*lepA*	Yes	98.5	GTP-binding protein LepA
Q170E-LepA-607	T − > C	70,867	D92G (GAC > GGC)	*araA*	No	98.5	L-arabinose isomerase
	A − > G	2,847,052	V10A (GTT > GCT)	*recD*	No	98.5	Exonuclease V (RecBCD complex) alpha chain
	G > C	2,628,110	Q170E (CAG > GAG)	*lepA*	No	98.5	GTP-binding protein LepA
S66K-Q170E-LepA-606	3 bp > CCT	2,628,420	Coding (196-198/1800 nt)	*lepA*	No	98.4	GTP-binding protein LepA
	G > C	2,628,110	Q170E (CAG > GAG)	*lepA*	No	98.4	GTP-binding protein LepA
T65N-YebC-606	Δ 1 bp	1,926,724	Coding (198/741 nt)	*yebC*	No	98.7	hypothetical protein
	+ T	1,926,729	Coding (193/741 nt)	*yebC*	No	98.7	hypothetical protein
L66I-YebC-606	C > A	1,926,726	L66I (CTG > TTTG)	*yebC*	No	98.7	Hypothetical protein
T65N-L66I-YebC-606	Δ 1 bp	1,926,724	Coding (198/741 nt)	*yebC*	No	98.7	hypothetical protein
	+ T	1,926,729	Coding (193/741 nt)	*yebC*	No	98.7	hypothetical protein
	C > A	1,926,726	L66I (CTG > TTTG)	*yebC*	No	98.7	Hypothetical protein

[1] Unassigned new junction prediction

[2] % of the genome where reads could be mapped

Table 8.3: . Colony counts for the *lepA* gene in LB medium. WT: wildtype. M1: S66K-LepA single mutant . M2: Q170E-LepA other single mutant . WM1M2: double mutant S66K-Q170E-LepA.

nameA	nameB	countA	countB	time	replicate
WT	WM1M2	2.90E+05	4.90E+05	0	1
WT	WM1M2	3.30E+05	4.80E+05	0	2
WT	WM1M2	2.90E+05	2.60E+05	0	3
WT	WM1M2	3.40E+05	4.70E+05	0	4
WT	WM1M2	3.90E+05	4.50E+05	0	5
WT	WM1M2	4.40E+05	2.40E+05	0	6
WT	WM1M2	2.20E+05	4.90E+05	0	7
WT	WM1M2	4.00E+05	4.20E+05	0	8
WT	WM1M2	1.10E+08	1.90E+08	24	1
WT	WM1M2	1.20E+08	3.80E+08	24	2
WT	WM1M2	0.00E+00	2.90E+08	24	3
WT	WM1M2	3.00E+07	4.00E+08	24	4
WT	WM1M2	4.00E+07	5.60E+08	24	5
WT	WM1M2	1.00E+07	5.30E+08	24	6
WT	WM1M2	2.10E+08	4.30E+08	24	7
WT	WM1M2	2.00E+07	3.70E+08	24	8
WT	WM1M2	9.00E+07	5.00E+08	48	1
WT	WM1M2	0.00E+00	4.50E+08	48	2
WT	WM1M2	0.00E+00	5.80E+08	48	3
WT	WM1M2	0.00E+00	3.50E+08	48	4
WT	WM1M2	0.00E+00	4.00E+08	48	5
WT	WM1M2	0.00E+00	3.80E+08	48	6
WT	WM1M2	1.00E+08	4.00E+08	48	7
WT	WM1M2	0.00E+00	4.60E+08	48	8
WT	WM1M2	2.00E+07	4.40E+08	72	1
WT	WM1M2	0.00E+00	2.90E+08	72	2
WT	WM1M2	0.00E+00	3.10E+08	72	3
WT	WM1M2	0.00E+00	4.90E+08	72	4
WT	WM1M2	1.00E+07	4.90E+08	72	5
WT	WM1M2	0.00E+00	2.70E+08	72	6
WT	WM1M2	6.00E+07	5.40E+08	72	7
WT	WM1M2	0.00E+00	4.60E+08	72	8
WT	M1	3.60E+05	4.90E+05	0	1
WT	M1	3.80E+05	2.30E+05	0	2
WT	M1	4.50E+05	2.40E+05	0	3
WT	M1	3.20E+05	2.50E+05	0	4
WT	M1	2.40E+05	2.50E+05	0	5
WT	M1	4.40E+05	2.20E+05	0	6
WT	M1	3.50E+05	2.20E+05	0	7
WT	M1	3.90E+05	1.50E+05	0	8
WT	M1	3.20E+08	9.00E+07	24	1
WT	M1	4.10E+08	1.10E+08	24	2
WT	M1	4.50E+08	6.00E+07	24	3
WT	M1	4.40E+08	5.00E+07	24	4
WT	M1	4.10E+08	5.00E+07	24	5
WT	M1	3.40E+08	1.70E+08	24	6
WT	M1	2.30E+08	6.00E+07	24	7
WT	M1	3.70E+08	3.00E+07	24	8
WT	M1	3.90E+08	9.00E+07	48	1
WT	M1	3.50E+08	1.00E+07	48	2

nameA	nameB	countA	countB	time	replicate
WT	M1	5.50E+08	2.00E+07	48	3
WT	M1	2.20E+08	0.00E+00	48	4
WT	M1	3.90E+08	3.00E+07	48	5
WT	M1	3.70E+08	5.00E+07	48	6
WT	M1	4.20E+08	1.30E+08	48	7
WT	M1	4.20E+08	1.00E+07	48	8
WT	M1	0.00E+00	0.00E+00	72	1
WT	M1	2.70E+08	0.00E+00	72	2
WT	M1	3.60E+08	1.00E+07	72	3
WT	M1	4.80E+08	1.00E+07	72	4
WT	M1	3.40E+08	0.00E+00	72	5
WT	M1	4.60E+08	6.00E+07	72	6
WT	M1	6.50E+08	8.00E+07	72	7
WT	M1	3.50E+08	0.00E+00	72	8
WT	M2	3.10E+05	4.90E+05	0	1
WT	M2	3.10E+05	3.00E+05	0	2
WT	M2	3.80E+05	4.80E+05	0	3
WT	M2	3.20E+05	2.10E+05	0	4
WT	M2	3.00E+05	4.20E+05	0	5
WT	M2	3.80E+05	2.70E+05	0	6
WT	M2	2.80E+05	2.80E+05	0	7
WT	M2	5.50E+05	4.30E+05	0	8
WT	M2	3.60E+08	3.00E+07	24	1
WT	M2	4.00E+08	1.00E+08	24	2
WT	M2	3.30E+08	3.00E+07	24	3
WT	M2	3.00E+08	1.00E+07	24	4
WT	M2	3.00E+08	1.00E+07	24	5
WT	M2	2.00E+08	1.20E+08	24	6
WT	M2	2.60E+08	2.00E+07	24	7
WT	M2	7.00E+07	5.00E+07	24	8
WT	M2	2.70E+08	1.00E+07	48	1
WT	M2	2.60E+08	2.00E+07	48	2
WT	M2	4.30E+08	1.00E+07	48	3
WT	M2	2.20E+08	0.00E+00	48	4
WT	M2	3.50E+08	2.00E+07	48	5
WT	M2	3.00E+08	1.00E+08	48	6
WT	M2	4.40E+08	0.00E+00	48	7
WT	M2	4.40E+08	2.00E+07	48	8
WT	M2	3.10E+08	2.00E+07	72	1
WT	M2	3.20E+08	1.00E+07	72	2
WT	M2	3.20E+08	0.00E+00	72	3
WT	M2	1.30E+08	0.00E+00	72	4
WT	M2	5.10E+08	0.00E+00	72	5
WT	M2	3.70E+08	1.00E+08	72	6
WT	M2	5.00E+08	1.00E+07	72	7
WT	M2	4.00E+07	0.00E+00	72	8

Table 8.4: . Colony counts for the *lepA* gene in m9 medium. WT: wildtype . M1: S66K-LepA single mutant . M2: Q170E-LepA other single mutant. WM1M2: double mutant S66K-Q170E-LepA.

nameA	nameB	countA	countB	time	replicate
WT	WM1M2	4.10E+05	8.90E+05	0	1
WT	WM1M2	8.20E+05	8.10E+05	0	2
WT	WM1M2	1.05E+06	1.07E+06	0	3
WT	WM1M2	6.80E+05	1.05E+06	0	4
WT	WM1M2	1.02E+06	8.60E+05	0	5
WT	WM1M2	7.50E+05	1.04E+06	0	6
WT	WM1M2	4.30E+05	8.20E+05	0	7
WT	WM1M2	7.30E+05	6.60E+05	0	8
WT	WM1M2	4.40E+07	5.10E+07	24	1
WT	WM1M2	3.30E+07	4.70E+07	24	2
WT	WM1M2	3.50E+07	4.50E+07	24	3
WT	WM1M2	2.90E+07	3.40E+07	24	4
WT	WM1M2	2.70E+07	3.60E+07	24	5
WT	WM1M2	3.70E+07	4.20E+07	24	6
WT	WM1M2	3.10E+07	3.40E+07	24	7
WT	WM1M2	5.20E+07	5.80E+07	24	8
WT	WM1M2	3.10E+07	6.40E+07	48	1
WT	WM1M2	3.60E+07	5.70E+07	48	2
WT	WM1M2	2.80E+07	4.30E+07	48	3
WT	WM1M2	4.60E+07	6.40E+07	48	4
WT	WM1M2	3.70E+07	5.50E+07	48	5
WT	WM1M2	3.50E+07	4.30E+07	48	6
WT	WM1M2	2.60E+07	3.00E+07	48	7
WT	WM1M2	2.60E+07	3.80E+07	48	8
WT	WM1M2	2.50E+07	1.09E+08	72	1
WT	WM1M2	6.00E+07	9.50E+07	72	2
WT	WM1M2	4.80E+07	9.20E+07	72	3
WT	WM1M2	6.00E+07	7.20E+07	72	4
WT	WM1M2	4.90E+07	8.20E+07	72	5
WT	WM1M2	7.10E+07	7.00E+07	72	6
WT	WM1M2	3.50E+07	5.40E+07	72	7
WT	WM1M2	3.40E+07	5.00E+07	72	8
WT	M1	9.20E+05	6.80E+05	0	1
WT	M1	1.02E+06	9.30E+05	0	2
WT	M1	1.26E+06	1.08E+06	0	3
WT	M1	1.01E+06	9.30E+05	0	4
WT	M1	8.20E+05	7.70E+05	0	5
WT	M1	1.02E+06	1.11E+06	0	6
WT	M1	1.12E+06	7.50E+05	0	7
WT	M1	9.30E+05	3.90E+05	0	8
WT	M1	8.20E+07	4.70E+07	24	1
WT	M1	7.10E+07	5.50E+07	24	2
WT	M1	6.60E+07	4.00E+07	24	3
WT	M1	8.60E+07	9.20E+07	24	4
WT	M1	6.80E+07	4.10E+07	24	5
WT	M1	4.10E+07	3.50E+07	24	6
WT	M1	7.00E+07	4.40E+07	24	7
WT	M1	4.50E+07	5.00E+07	24	8
WT	M1	2.40E+07	1.90E+07	48	1
WT	M1	1.00E+08	7.20E+07	48	2

nameA	nameB	countA	countB	time	replicate
WT	M1	9.90E+07	8.10E+07	48	3
WT	M1	4.10E+07	3.70E+07	48	4
WT	M1	4.70E+07	2.00E+07	48	5
WT	M1	2.70E+07	1.50E+07	48	6
WT	M1	2.70E+07	1.80E+07	48	7
WT	M1	1.30E+07	8.00E+06	48	8
WT	M1	7.40E+07	5.00E+07	72	1
WT	M1	1.05E+08	5.40E+07	72	2
WT	M1	9.90E+07	6.30E+07	72	3
WT	M1	1.04E+08	5.80E+07	72	4
WT	M1	8.00E+07	4.00E+07	72	5
WT	M1	7.50E+07	4.60E+07	72	6
WT	M1	5.40E+07	2.90E+07	72	7
WT	M1	5.70E+07	3.00E+07	72	8
WT	M2	1.18E+06	7.20E+05	0	1
WT	M2	8.10E+05	5.30E+05	0	2
WT	M2	1.04E+06	6.40E+05	0	3
WT	M2	1.10E+06	7.40E+05	0	4
WT	M2	7.10E+05	1.04E+06	0	5
WT	M2	8.80E+05	6.20E+05	0	6
WT	M2	9.70E+05	4.40E+05	0	7
WT	M2	8.40E+05	4.80E+05	0	8
WT	M2	1.45E+08	1.40E+08	24	1
WT	M2	6.10E+07	4.80E+07	24	2
WT	M2	6.20E+07	5.40E+07	24	3
WT	M2	5.10E+07	5.00E+07	24	4
WT	M2	5.70E+07	5.60E+07	24	5
WT	M2	6.00E+07	5.20E+07	24	6
WT	M2	6.30E+07	5.80E+07	24	7
WT	M2	5.00E+07	4.60E+07	24	8
WT	M2	4.70E+07	2.50E+07	48	1
WT	M2	3.80E+07	3.20E+07	48	2
WT	M2	2.10E+07	1.80E+07	48	3
WT	M2	2.50E+07	1.80E+07	48	4
WT	M2	5.60E+07	3.30E+07	48	5
WT	M2	4.70E+07	3.30E+07	48	6
WT	M2	4.30E+07	4.60E+07	48	7
WT	M2	4.40E+07	3.40E+07	48	8
WT	M2	1.09E+08	5.30E+07	72	1
WT	M2	8.60E+07	4.90E+07	72	2
WT	M2	6.80E+07	4.80E+07	72	3
WT	M2	8.60E+07	6.30E+07	72	4
WT	M2	1.16E+08	5.40E+07	72	5
WT	M2	8.70E+07	5.10E+07	72	6
WT	M2	7.30E+07	4.40E+07	72	7
WT	M2	6.40E+07	2.60E+07	72	8
M1M2	M1	1.06E+06	1.10E+06	0	1
M1M2	M1	1.02E+06	8.80E+05	0	2
M1M2	M1	8.80E+05	5.40E+05	0	3
M1M2	M1	8.30E+05	9.20E+05	0	4
M1M2	M1	1.12E+06	1.28E+06	0	5
M1M2	M1	9.60E+05	9.60E+05	0	6
M1M2	M1	5.60E+05	6.30E+05	0	7
M1M2	M1	6.10E+05	4.50E+05	0	8

nameA	nameB	countA	countB	time	replicate
M1M2	M1	6.00E+07	3.80E+07	24	1
M1M2	M1	6.50E+07	6.20E+07	24	2
M1M2	M1	6.50E+07	5.40E+07	24	3
M1M2	M1	7.10E+07	9.20E+07	24	4
M1M2	M1	4.80E+07	4.50E+07	24	5
M1M2	M1	6.20E+07	4.60E+07	24	6
M1M2	M1	3.90E+07	3.10E+07	24	7
M1M2	M1	4.50E+07	4.00E+07	24	8
M1M2	M1	5.70E+07	4.30E+07	48	1
M1M2	M1	3.30E+07	3.70E+07	48	2
M1M2	M1	4.90E+07	2.60E+07	48	3
M1M2	M1	4.60E+07	2.70E+07	48	4
M1M2	M1	4.10E+07	1.90E+07	48	5
M1M2	M1	2.90E+07	2.00E+07	48	6
M1M2	M1	3.70E+07	2.10E+07	48	7
M1M2	M1	2.50E+07	1.60E+07	48	8
M1M2	M1	6.80E+07	4.30E+07	72	1
M1M2	M1	7.70E+07	5.90E+07	72	2
M1M2	M1	7.40E+07	4.30E+07	72	3
M1M2	M1	5.80E+07	4.70E+07	72	4
M1M2	M1	8.60E+07	3.80E+07	72	5
M1M2	M1	7.00E+07	4.50E+07	72	6
M1M2	M1	7.00E+07	3.20E+07	72	7
M1M2	M1	5.40E+07	3.50E+07	72	8
M1M2	M2	9.90E+05	7.60E+05	0	1
M1M2	M2	1.12E+06	7.00E+05	0	2
M1M2	M2	1.14E+06	8.30E+05	0	3
M1M2	M2	8.60E+05	6.60E+05	0	4
M1M2	M2	6.60E+05	8.70E+05	0	5
M1M2	M2	7.50E+05	6.40E+05	0	6
M1M2	M2	6.00E+05	2.10E+05	0	7
M1M2	M2	3.70E+05	3.90E+05	0	8
M1M2	M2	4.90E+07	5.30E+07	24	1
M1M2	M2	6.50E+07	6.40E+07	24	2
M1M2	M2	8.00E+07	5.00E+07	24	3
M1M2	M2	7.10E+07	5.40E+07	24	4
M1M2	M2	7.20E+07	4.50E+07	24	5
M1M2	M2	5.70E+07	5.30E+07	24	6
M1M2	M2	4.40E+07	4.90E+07	24	7
M1M2	M2	5.40E+07	4.20E+07	24	8
M1M2	M2	6.60E+07	3.90E+07	48	1
M1M2	M2	5.90E+07	3.50E+07	48	2
M1M2	M2	6.50E+07	3.10E+07	48	3
M1M2	M2	4.50E+07	2.90E+07	48	4
M1M2	M2	4.30E+07	3.10E+07	48	5
M1M2	M2	3.40E+07	2.10E+07	48	6
M1M2	M2	2.00E+07	1.80E+07	48	7
M1M2	M2	1.90E+07	1.70E+07	48	8
M1M2	M2	9.20E+07	5.70E+07	72	1
M1M2	M2	7.90E+07	4.50E+07	72	2
M1M2	M2	8.60E+07	4.40E+07	72	3
M1M2	M2	6.90E+07	4.80E+07	72	4
M1M2	M2	9.00E+07	5.10E+07	72	5
M1M2	M2	7.90E+07	4.10E+07	72	6

nameA	nameB	countA	countB	time	replicate
M1M2	M2	9.00E+07	4.50E+07	72	7
M1M2	M2	6.40E+07	4.00E+07	72	8

Table 8.5: . Colony counts for the *ispH* gene in LB medium. WT: wildtype. M1: N24R-IspH single mutant . M2: A27E-IspH other single mutant . WM1M2: double mutant N24R-A27E-IspH.

nameA	nameB	countA	countB	time	replicate
WT	WM1M2	1.80E+06	1.50E+06	0	1
WT	WM1M2	2.60E+06	2.30E+06	0	2
WT	WM1M2	2.00E+06	2.20E+06	0	3
WT	WM1M2	2.20E+06	1.80E+06	0	4
WT	WM1M2	3.20E+06	1.90E+06	0	5
WT	WM1M2	2.20E+06	1.30E+06	0	6
WT	WM1M2	2.30E+06	1.60E+06	0	7
WT	WM1M2	2.20E+06	1.70E+06	0	8
WT	WM1M2	1.90E+08	3.60E+08	24	1
WT	WM1M2	0.00E+00	1.10E+08	24	2
WT	WM1M2	2.70E+08	4.20E+08	24	3
WT	WM1M2	1.60E+08	2.50E+08	24	4
WT	WM1M2	1.70E+08	3.30E+08	24	5
WT	WM1M2	2.00E+08	3.30E+08	24	6
WT	WM1M2	2.00E+08	4.30E+08	24	7
WT	WM1M2	3.50E+08	6.00E+08	24	8
WT	WM1M2	0.00E+00	1.70E+08	48	1
WT	WM1M2	1.00E+07	3.30E+08	48	2
WT	WM1M2	1.60E+08	4.50E+08	48	3
WT	WM1M2	1.20E+08	3.40E+08	48	4
WT	WM1M2	7.00E+07	3.00E+08	48	5
WT	WM1M2	1.80E+08	4.30E+08	48	6
WT	WM1M2	1.20E+08	2.10E+08	48	7
WT	WM1M2	1.00E+08	5.00E+08	48	8
WT	WM1M2	0.00E+00	4.10E+08	72	1
WT	WM1M2	0.00E+00	8.10E+08	72	2
WT	WM1M2	5.00E+07	5.50E+08	72	3
WT	WM1M2	1.10E+08	5.70E+08	72	4
WT	WM1M2	9.00E+07	4.50E+08	72	5
WT	WM1M2	5.00E+07	3.30E+08	72	6
WT	WM1M2	1.00E+08	5.10E+08	72	7
WT	WM1M2	1.10E+08	4.80E+08	72	8
WT	M1	2.80E+06	3.70E+06	0	1
WT	M1	2.00E+06	3.50E+06	0	2
WT	M1	4.50E+06	2.70E+06	0	3
WT	M1	1.80E+06	2.10E+06	0	4
WT	M1	1.10E+06	3.20E+06	0	5
WT	M1	2.60E+06	2.60E+06	0	6
WT	M1	3.00E+06	2.40E+06	0	7
WT	M1	2.90E+06	2.80E+06	0	8
WT	M1	4.30E+08	2.80E+08	24	1
WT	M1	2.50E+08	1.20E+08	24	2
WT	M1	1.70E+08	3.00E+08	24	3
WT	M1	6.00E+07	0.00E+00	24	4
WT	M1	5.00E+07	5.50E+08	24	5
WT	M1	3.50E+08	1.10E+08	24	6
WT	M1	3.30E+08	1.80E+08	24	7
WT	M1	4.60E+08	1.50E+08	24	8
WT	M1	3.80E+08	9.00E+07	48	1
WT	M1	5.80E+08	2.40E+08	48	2

nameA	nameB	countA	countB	time	replicate
WT	M1	2.30E+08	3.10E+08	48	3
WT	M1	4.50E+08	1.00E+07	48	4
WT	M1	5.00E+07	5.00E+08	48	5
WT	M1	3.00E+08	1.00E+08	48	6
WT	M1	5.10E+08	1.90E+08	48	7
WT	M1	4.80E+08	1.50E+08	48	8
WT	M1	4.70E+08	9.00E+07	72	1
WT	M1	5.00E+08	1.50E+08	72	2
WT	M1	2.90E+08	2.70E+08	72	3
WT	M1	3.40E+08	2.00E+07	72	4
WT	M1	4.00E+07	5.10E+08	72	5
WT	M1	4.00E+08	1.30E+08	72	6
WT	M1	3.50E+08	6.00E+07	72	7
WT	M1	3.00E+07	1.00E+07	72	8
WT	M2	2.90E+06	2.70E+06	0	1
WT	M2	4.00E+06	2.20E+06	0	2
WT	M2	1.10E+06	1.20E+06	0	3
WT	M2	3.50E+06	1.50E+06	0	4
WT	M2	2.30E+06	1.60E+06	0	5
WT	M2	2.60E+06	2.70E+06	0	6
WT	M2	3.50E+06	1.40E+06	0	7
WT	M2	2.60E+06	2.30E+06	0	8
WT	M2	0.00E+00	1.40E+08	24	1
WT	M2	1.70E+08	3.30E+08	24	2
WT	M2	1.30E+08	3.70E+08	24	3
WT	M2	2.00E+08	3.80E+08	24	4
WT	M2	0.00E+00	5.00E+07	24	5
WT	M2	0.00E+00	4.00E+07	24	6
WT	M2	2.30E+08	4.60E+08	24	7
WT	M2	1.60E+08	4.60E+08	24	8
WT	M2	0.00E+00	2.30E+08	48	1
WT	M2	9.00E+07	3.70E+08	48	2
WT	M2	1.30E+08	3.30E+08	48	3
WT	M2	1.10E+08	4.10E+08	48	4
WT	M2	1.00E+07	3.00E+08	48	5
WT	M2	1.00E+07	5.20E+08	48	6
WT	M2	1.90E+08	3.30E+08	48	7
WT	M2	1.50E+08	4.80E+08	48	8
WT	M2	1.00E+07	4.10E+08	72	1
WT	M2	1.70E+08	3.70E+08	72	2
WT	M2	1.50E+08	4.20E+08	72	3
WT	M2	8.00E+07	4.80E+08	72	4
WT	M2	0.00E+00	5.60E+08	72	5
WT	M2	4.00E+07	8.80E+08	72	6
WT	M2	8.50E+08	3.46E+09	72	7
WT	M2	8.00E+07	5.20E+08	72	8
M1M2	M1	1.80E+06	3.10E+06	0	1
M1M2	M1	1.90E+06	2.70E+06	0	2
M1M2	M1	1.60E+06	3.30E+06	0	3
M1M2	M1	1.30E+06	1.70E+06	0	4
M1M2	M1	1.40E+06	2.80E+06	0	5
M1M2	M1	1.30E+06	3.10E+06	0	6
M1M2	M1	1.60E+06	2.50E+06	0	7
M1M2	M1	1.40E+06	2.00E+06	0	8

nameA	nameB	countA	countB	time	replicate
M1M2	M1	4.20E+07	2.20E+07	24	1
M1M2	M1	2.80E+07	1.80E+07	24	2
M1M2	M1	3.80E+07	2.20E+07	24	3
M1M2	M1	4.50E+07	7.00E+06	24	4
M1M2	M1	3.50E+07	1.70E+07	24	5
M1M2	M1	8.00E+06	0.00E+00	24	6
M1M2	M1	2.70E+07	2.10E+07	24	7
M1M2	M1	3.20E+07	2.40E+07	24	8
M1M2	M1	4.60E+07	1.90E+07	48	1
M1M2	M1	3.90E+07	1.30E+07	48	2
M1M2	M1	3.80E+07	1.00E+07	48	3
M1M2	M1	5.70E+07	2.00E+06	48	4
M1M2	M1	4.00E+07	1.50E+07	48	5
M1M2	M1	3.20E+07	0.00E+00	48	6
M1M2	M1	5.30E+07	2.20E+07	48	7
M1M2	M1	6.50E+07	1.80E+07	48	8
M1M2	M1	2.40E+07	0.00E+00	72	1
M1M2	M1	1.00E+06	0.00E+00	72	2
M1M2	M1	5.10E+07	1.10E+07	72	3
M1M2	M1	6.70E+07	3.00E+06	72	4
M1M2	M1	3.60E+07	1.50E+07	72	5
M1M2	M1	5.00E+07	0.00E+00	72	6
M1M2	M1	5.70E+07	2.10E+07	72	7
M1M2	M1	3.40E+07	9.00E+06	72	8
M1M2	M2	3.40E+06	2.20E+06	0	1
M1M2	M2	3.40E+06	2.70E+06	0	2
M1M2	M2	2.90E+06	2.30E+06	0	3
M1M2	M2	2.40E+06	2.40E+06	0	4
M1M2	M2	2.80E+06	1.10E+06	0	5
M1M2	M2	1.90E+06	1.30E+06	0	6
M1M2	M2	1.20E+06	1.30E+06	0	7
M1M2	M2	2.30E+06	2.00E+06	0	8
M1M2	M2	2.70E+07	3.70E+07	24	1
M1M2	M2	2.00E+07	3.30E+07	24	2
M1M2	M2	1.90E+07	3.40E+07	24	3
M1M2	M2	1.40E+07	3.30E+07	24	4
M1M2	M2	1.70E+07	2.40E+07	24	5
M1M2	M2	0.00E+00	9.00E+06	24	6
M1M2	M2	1.60E+07	4.20E+07	24	7
M1M2	M2	1.50E+07	5.20E+07	24	8
M1M2	M2	1.20E+07	5.00E+07	48	1
M1M2	M2	1.20E+07	3.40E+07	48	2
M1M2	M2	1.60E+07	2.60E+07	48	3
M1M2	M2	1.70E+07	4.70E+07	48	4
M1M2	M2	1.60E+07	4.30E+07	48	5
M1M2	M2	0.00E+00	1.50E+07	48	6
M1M2	M2	1.60E+07	4.70E+07	48	7
M1M2	M2	9.00E+06	5.10E+07	48	8
M1M2	M2	9.00E+06	4.20E+07	72	1
M1M2	M2	1.50E+07	5.40E+07	72	2
M1M2	M2	0.00E+00	8.00E+06	72	3
M1M2	M2	7.00E+06	4.60E+07	72	4
M1M2	M2	2.10E+07	4.60E+07	72	5
M1M2	M2	3.00E+06	6.20E+07	72	6

nameA	nameB	countA	countB	time	replicate
M1M2	M2	1.60E+07	5.10E+07	72	7
M1M2	M2	6.00E+06	5.20E+07	72	8

Table 8.6: . Colony counts for the *ispH* gene in m9 medium. WT: wildtype. M1: N24R-IspH single mutant . M2: A27E-IspH other single mutant . WM1M2: double mutant N24R-A27E-IspH.

nameA	nameB	countA	countB	time	replicate
WT	M1M2	6.20E+05	7.40E+05	0	1
WT	WM1M2	7.20E+05	8.10E+05	0	2
WT	WM1M2	6.70E+05	5.40E+05	0	3
WT	WM1M2	5.00E+05	5.20E+05	0	4
WT	WM1M2	1.90E+05	2.50E+05	0	5
WT	WM1M2	NA	NA	0	6
WT	WM1M2	NA	NA	0	7
WT	WM1M2	NA	NA	0	8
WT	WM1M2	9.10E+08	1.01E+09	24	1
WT	WM1M2	7.30E+08	1.12E+09	24	2
WT	WM1M2	6.00E+08	6.90E+08	24	3
WT	WM1M2	1.01E+09	1.12E+09	24	4
WT	WM1M2	6.00E+08	8.70E+08	24	5
WT	WM1M2	NA	NA	24	6
WT	WM1M2	NA	NA	24	7
WT	WM1M2	NA	NA	24	8
WT	WM1M2	1.99E+09	2.53E+09	48	1
WT	WM1M2	1.92E+09	2.25E+09	48	2
WT	WM1M2	2.42E+09	2.14E+09	48	3
WT	WM1M2	1.83E+09	2.11E+09	48	4
WT	WM1M2	1.64E+09	1.71E+09	48	5
WT	WM1M2	NA	NA	48	6
WT	WM1M2	NA	NA	48	7
WT	WM1M2	NA	NA	48	8
WT	WM1M2	1.81E+09	2.89E+09	72	1
WT	WM1M2	1.64E+09	2.09E+09	72	2
WT	WM1M2	1.58E+09	1.64E+09	72	3
WT	WM1M2	1.61E+09	2.00E+09	72	4
WT	WM1M2	1.52E+09	1.97E+09	72	5
WT	WM1M2	NA	NA	72	6
WT	WM1M2	NA	NA	72	7
WT	WM1M2	NA	NA	72	8
WT	M1	9.40E+05	9.30E+05	0	1
WT	M1	7.00E+05	8.50E+05	0	2
WT	M1	7.50E+05	1.30E+06	0	3
WT	M1	8.60E+05	5.80E+05	0	4
WT	M1	4.70E+05	5.20E+05	0	5
WT	M1	5.60E+05	6.30E+05	0	6
WT	M1	4.30E+05	5.30E+05	0	7
WT	M1	2.90E+05	8.00E+04	0	8
WT	M1	1.00E+09	7.70E+08	24	1
WT	M1	9.50E+08	1.09E+09	24	2
WT	M1	1.12E+09	1.06E+09	24	3
WT	M1	9.80E+08	8.70E+08	24	4
WT	M1	8.80E+08	1.01E+09	24	5
WT	M1	1.25E+09	9.60E+08	24	6
WT	M1	1.13E+09	1.16E+09	24	7
WT	M1	1.30E+09	8.80E+08	24	8
WT	M1	1.33E+09	1.15E+09	48	1
WT	M1	9.60E+08	1.03E+09	48	2

nameA	nameB	countA	countB	time	replicate
WT	M1	1.03E+09	9.90E+08	48	3
WT	M1	1.31E+09	1.12E+09	48	4
WT	M1	1.21E+09	9.10E+08	48	5
WT	M1	1.08E+09	9.20E+08	48	6
WT	M1	1.00E+09	7.90E+08	48	7
WT	M1	9.80E+08	7.50E+08	48	8
WT	M1	1.32E+09	1.07E+09	72	1
WT	M1	1.29E+09	1.20E+09	72	2
WT	M1	1.23E+09	1.39E+09	72	3
WT	M1	1.23E+09	1.07E+09	72	4
WT	M1	1.63E+09	1.28E+09	72	5
WT	M1	1.45E+09	1.10E+09	72	6
WT	M1	1.66E+09	1.12E+09	72	7
WT	M1	2.95E+09	1.81E+09	72	8
WT	M2	4.10E+05	6.50E+05	0	1
WT	M2	1.02E+06	6.60E+05	0	2
WT	M2	1.14E+06	7.70E+05	0	3
WT	M2	9.60E+05	7.40E+05	0	4
WT	M2	7.60E+05	7.30E+05	0	5
WT	M2	7.60E+05	5.00E+05	0	6
WT	M2	6.40E+05	5.80E+05	0	7
WT	M2	1.70E+05	4.20E+05	0	8
WT	M2	9.40E+08	9.40E+08	24	1
WT	M2	1.14E+09	6.20E+08	24	2
WT	M2	9.50E+08	8.30E+08	24	3
WT	M2	1.15E+09	9.50E+08	24	4
WT	M2	8.90E+08	1.10E+09	24	5
WT	M2	1.08E+09	8.10E+08	24	6
WT	M2	9.50E+08	7.60E+08	24	7
WT	M2	7.70E+08	1.08E+09	24	8
WT	M2	1.51E+09	1.73E+09	48	1
WT	M2	1.45E+09	1.17E+09	48	2
WT	M2	2.26E+09	1.83E+09	48	3
WT	M2	2.25E+09	2.27E+09	48	4
WT	M2	1.68E+09	2.02E+09	48	5
WT	M2	1.62E+09	1.93E+09	48	6
WT	M2	1.69E+09	1.91E+09	48	7
WT	M2	1.45E+09	1.92E+09	48	8
WT	M2	1.07E+09	1.66E+09	72	1
WT	M2	1.57E+09	1.35E+09	72	2
WT	M2	1.63E+09	1.58E+09	72	3
WT	M2	1.38E+09	1.99E+09	72	4
WT	M2	1.57E+09	1.97E+09	72	5
WT	M2	1.72E+09	2.04E+09	72	6
WT	M2	1.38E+09	1.93E+09	72	7
WT	M2	2.31E+09	3.37E+09	72	8
M1M2	M1	7.60E+05	1.06E+06	0	1
M1M2	M1	8.20E+05	1.18E+06	0	2
M1M2	M1	1.11E+06	2.02E+06	0	3
M1M2	M1	1.76E+06	1.66E+06	0	4
M1M2	M1	1.38E+06	1.35E+06	0	5
M1M2	M1	4.80E+05	6.70E+05	0	6
M1M2	M1	4.50E+05	6.40E+05	0	7
M1M2	M1	4.20E+05	1.90E+05	0	8

nameA	nameB	countA	countB	time	replicate
M1M2	M1	9.40E+07	8.40E+07	24	1
M1M2	M1	9.60E+07	1.06E+08	24	2
M1M2	M1	9.30E+07	8.80E+07	24	3
M1M2	M1	7.10E+07	7.70E+07	24	4
M1M2	M1	1.10E+08	1.19E+08	24	5
M1M2	M1	9.60E+07	1.04E+08	24	6
M1M2	M1	1.08E+08	8.00E+07	24	7
M1M2	M1	8.40E+07	9.00E+07	24	8
M1M2	M1	1.67E+08	1.49E+08	48	1
M1M2	M1	1.88E+08	1.85E+08	48	2
M1M2	M1	1.75E+08	1.88E+08	48	3
M1M2	M1	2.05E+08	1.61E+08	48	4
M1M2	M1	2.02E+08	1.81E+08	48	5
M1M2	M1	1.97E+08	1.72E+08	48	6
M1M2	M1	2.35E+08	1.66E+08	48	7
M1M2	M1	1.97E+08	1.56E+08	48	8
M1M2	M1	2.19E+08	1.17E+08	72	1
M1M2	M1	1.70E+08	1.49E+08	72	2
M1M2	M1	1.64E+08	1.70E+08	72	3
M1M2	M1	1.93E+08	1.47E+08	72	4
M1M2	M1	1.66E+08	1.27E+08	72	5
M1M2	M1	1.68E+08	1.62E+08	72	6
M1M2	M1	1.82E+08	1.37E+08	72	7
M1M2	M1	2.79E+08	3.76E+08	72	8
M1M2	M2	7.40E+05	1.26E+06	0	1
M1M2	M2	4.60E+05	8.50E+05	0	2
M1M2	M2	1.00E+06	9.40E+05	0	3
M1M2	M2	1.28E+06	8.90E+05	0	4
M1M2	M2	8.60E+05	7.10E+05	0	5
M1M2	M2	6.60E+05	7.10E+05	0	6
M1M2	M2	9.30E+05	5.90E+05	0	7
M1M2	M2	5.20E+05	4.90E+05	0	8
M1M2	M2	7.30E+07	1.06E+08	24	1
M1M2	M2	7.90E+07	6.20E+07	24	2
M1M2	M2	5.10E+07	6.20E+07	24	3
M1M2	M2	9.10E+07	1.01E+08	24	4
M1M2	M2	9.50E+07	8.00E+07	24	5
M1M2	M2	8.60E+07	8.80E+07	24	6
M1M2	M2	1.14E+08	9.40E+07	24	7
M1M2	M2	6.30E+07	7.90E+07	24	8
M1M2	M2	1.66E+08	2.65E+08	48	1
M1M2	M2	2.07E+08	1.61E+08	48	2
M1M2	M2	1.53E+08	1.80E+08	48	3
M1M2	M2	1.74E+08	1.77E+08	48	4
M1M2	M2	2.11E+08	2.16E+08	48	5
M1M2	M2	2.00E+08	2.37E+08	48	6
M1M2	M2	1.83E+08	2.12E+08	48	7
M1M2	M2	1.46E+08	1.98E+08	48	8
M1M2	M2	1.07E+08	1.53E+08	72	1
M1M2	M2	1.52E+08	1.09E+08	72	2
M1M2	M2	1.37E+08	1.71E+08	72	3
M1M2	M2	1.34E+08	1.55E+08	72	4
M1M2	M2	1.54E+08	1.88E+08	72	5
M1M2	M2	1.50E+08	1.67E+08	72	6

nameA	nameB	countA	countB	time	replicate
M1M2	M2	1.70E+08	1.82E+08	72	7
M1M2	M2	2.70E+08	3.60E+08	72	8

Table 8.7: . Colony counts for the *yebC* gene in LB medium. WT: wildtype. M1: T56N-YebC single mutant. M2: L66I-YebC other single mutant . WM1M2: double mutant T65N-L66I-YebC.

nameA	nameB	countA	countB	time	replicate
WT	WM1M2	2.40E+06	2.60E+06	0	1
WT	WM1M2	3.00E+06	1.00E+06	0	2
WT	WM1M2	2.70E+06	1.70E+06	0	3
WT	WM1M2	3.20E+06	1.10E+06	0	4
WT	WM1M2	2.80E+06	1.90E+06	0	5
WT	WM1M2	2.10E+06	1.60E+06	0	6
WT	WM1M2	3.20E+06	2.00E+06	0	7
WT	WM1M2	2.90E+06	1.90E+06	0	8
WT	WM1M2	8.00E+07	3.70E+08	24	1
WT	WM1M2	1.50E+08	3.90E+08	24	2
WT	WM1M2	2.40E+08	3.80E+08	24	3
WT	WM1M2	4.30E+08	9.00E+07	24	4
WT	WM1M2	2.20E+08	4.10E+08	24	5
WT	WM1M2	3.00E+08	4.60E+08	24	6
WT	WM1M2	1.80E+08	3.80E+08	24	7
WT	WM1M2	1.60E+08	3.00E+08	24	8
WT	WM1M2	5.00E+07	6.70E+08	48	1
WT	WM1M2	1.90E+08	4.00E+08	48	2
WT	WM1M2	2.00E+08	4.60E+08	48	3
WT	WM1M2	5.70E+08	1.00E+08	48	4
WT	WM1M2	1.70E+08	4.30E+08	48	5
WT	WM1M2	0.00E+00	1.20E+08	48	6
WT	WM1M2	1.70E+08	4.40E+08	48	7
WT	WM1M2	2.00E+08	4.30E+08	48	8
WT	WM1M2	3.00E+07	6.20E+08	72	1
WT	WM1M2	1.20E+08	6.10E+08	72	2
WT	WM1M2	1.40E+08	5.50E+08	72	3
WT	WM1M2	4.80E+08	1.50E+08	72	4
WT	WM1M2	1.90E+08	6.20E+08	72	5
WT	WM1M2	1.00E+07	6.50E+08	72	6
WT	WM1M2	1.00E+07	1.10E+08	72	7
WT	WM1M2	1.60E+08	5.00E+08	72	8
WT	M1	2.30E+06	3.90E+06	0	1
WT	M1	1.60E+06	2.30E+06	0	2
WT	M1	1.90E+06	2.50E+06	0	3
WT	M1	1.80E+06	2.40E+06	0	4
WT	M1	4.00E+05	1.30E+06	0	5
WT	M1	2.50E+06	3.10E+06	0	6
WT	M1	1.50E+06	3.00E+06	0	7
WT	M1	1.80E+06	2.00E+05	0	8
WT	M1	2.80E+08	2.80E+08	24	1
WT	M1	3.90E+08	1.80E+08	24	2
WT	M1	2.60E+08	2.30E+08	24	3
WT	M1	2.70E+08	2.30E+08	24	4
WT	M1	6.00E+07	4.00E+08	24	5
WT	M1	3.20E+08	1.80E+08	24	6
WT	M1	3.80E+08	1.80E+08	24	7
WT	M1	1.50E+08	0.00E+00	24	8
WT	M1	3.50E+08	9.00E+07	48	1
WT	M1	4.10E+08	1.80E+08	48	2
WT	M1	2.80E+08	1.10E+08	48	3
WT	M1	1.90E+08	1.10E+08	48	4

nameA	nameB	countA	countB	time	replicate
WT	M1	0.00E+00	2.40E+08	48	5
WT	M1	2.60E+08	7.00E+07	48	6
WT	M1	2.50E+08	6.00E+07	48	7
WT	M1	3.90E+08	0.00E+00	48	8
WT	M1	6.50E+08	1.50E+08	72	1
WT	M1	5.50E+08	1.50E+08	72	2
WT	M1	5.80E+08	1.30E+08	72	3
WT	M1	6.40E+08	2.60E+08	72	4
WT	M1	1.00E+08	6.70E+08	72	5
WT	M1	4.60E+08	9.00E+07	72	6
WT	M1	5.00E+08	1.20E+08	72	7
WT	M1	7.10E+08	0.00E+00	72	8
WT	M2	2.10E+06	2.50E+06	0	1
WT	M2	8.00E+05	1.50E+06	0	2
WT	M2	2.00E+06	1.90E+06	0	3
WT	M2	1.40E+06	1.90E+06	0	4
WT	M2	2.00E+05	1.20E+06	0	5
WT	M2	2.20E+06	2.10E+06	0	6
WT	M2	2.00E+06	1.70E+06	0	7
WT	M2	2.70E+06	1.90E+06	0	8
WT	M2	3.90E+08	1.90E+08	24	1
WT	M2	5.10E+08	1.70E+08	24	2
WT	M2	3.20E+08	1.70E+08	24	3
WT	M2	3.00E+08	2.00E+08	24	4
WT	M2	3.00E+07	4.80E+08	24	5
WT	M2	3.50E+08	2.50E+08	24	6
WT	M2	1.10E+08	0.00E+00	24	7
WT	M2	3.70E+08	2.00E+08	24	8
WT	M2	1.60E+08	8.00E+07	48	1
WT	M2	2.40E+08	9.00E+07	48	2
WT	M2	2.20E+08	9.00E+07	48	3
WT	M2	5.20E+08	2.73E+09	48	4
WT	M2	1.00E+07	2.40E+08	48	5
WT	M2	2.30E+08	8.00E+07	48	6
WT	M2	4.50E+08	1.00E+07	48	7
WT	M2	3.90E+08	1.00E+08	48	8
WT	M2	5.80E+08	1.70E+08	72	1
WT	M2	5.30E+08	3.00E+07	72	2
WT	M2	5.10E+08	1.36E+09	72	3
WT	M2	5.10E+08	1.40E+08	72	4
WT	M2	5.00E+07	2.00E+07	72	5
WT	M2	9.00E+07	0.00E+00	72	6
WT	M2	5.80E+08	0.00E+00	72	7
WT	M2	5.60E+08	1.30E+08	72	8
M1M2	M1	2.50E+06	2.10E+06	0	1
M1M2	M1	1.50E+06	2.20E+06	0	2
M1M2	M1	1.80E+06	3.10E+06	0	3
M1M2	M1	9.00E+05	2.20E+06	0	4
M1M2	M1	1.40E+06	2.50E+06	0	5
M1M2	M1	1.90E+06	3.20E+06	0	6
M1M2	M1	2.20E+06	4.70E+06	0	7
M1M2	M1	1.70E+06	4.00E+05	0	8
M1M2	M1	4.70E+08	3.10E+08	24	1
M1M2	M1	3.70E+08	2.50E+08	24	2

nameA	nameB	countA	countB	time	replicate
M1M2	M1	4.30E+08	2.10E+08	24	3
M1M2	M1	5.00E+07	5.70E+08	24	4
M1M2	M1	5.20E+08	1.80E+08	24	5
M1M2	M1	4.50E+08	2.00E+08	24	6
M1M2	M1	2.80E+08	3.00E+08	24	7
M1M2	M1	5.20E+08	4.00E+07	24	8
M1M2	M1	4.70E+08	2.10E+08	48	1
M1M2	M1	5.50E+08	2.40E+08	48	2
M1M2	M1	5.30E+08	2.00E+08	48	3
M1M2	M1	9.00E+07	5.40E+08	48	4
M1M2	M1	4.70E+08	1.80E+08	48	5
M1M2	M1	4.30E+08	2.90E+08	48	6
M1M2	M1	6.90E+08	1.10E+08	48	7
M1M2	M1	8.00E+08	0.00E+00	48	8
M1M2	M1	4.20E+08	4.00E+07	72	1
M1M2	M1	6.30E+08	1.60E+08	72	2
M1M2	M1	5.30E+08	2.10E+08	72	3
M1M2	M1	6.10E+08	1.90E+08	72	4
M1M2	M1	4.80E+08	1.80E+08	72	5
M1M2	M1	6.00E+08	1.60E+08	72	6
M1M2	M1	5.60E+08	1.40E+08	72	7
M1M2	M1	7.00E+08	0.00E+00	72	8
M1M2	M2	8.00E+05	2.30E+06	0	1
M1M2	M2	2.10E+06	2.30E+06	0	2
M1M2	M2	1.80E+06	3.20E+06	0	3
M1M2	M2	1.10E+06	2.80E+06	0	4
M1M2	M2	1.70E+06	2.10E+06	0	5
M1M2	M2	2.60E+06	3.20E+06	0	6
M1M2	M2	1.20E+06	2.20E+06	0	7
M1M2	M2	1.70E+06	2.40E+06	0	8
M1M2	M2	2.60E+08	1.90E+08	24	1
M1M2	M2	4.90E+08	3.30E+08	24	2
M1M2	M2	3.60E+08	3.10E+08	24	3
M1M2	M2	8.00E+07	4.20E+08	24	4
M1M2	M2	5.60E+08	2.30E+08	24	5
M1M2	M2	5.60E+08	2.60E+08	24	6
M1M2	M2	4.60E+08	3.00E+08	24	7
M1M2	M2	4.00E+07	0.00E+00	24	8
M1M2	M2	5.40E+08	2.00E+08	48	1
M1M2	M2	4.70E+08	2.20E+08	48	2
M1M2	M2	6.50E+08	2.40E+08	48	3
M1M2	M2	9.00E+07	1.60E+08	48	4
M1M2	M2	4.60E+08	1.60E+08	48	5
M1M2	M2	6.50E+08	2.20E+08	48	6
M1M2	M2	6.30E+08	1.80E+08	48	7
M1M2	M2	9.80E+08	0.00E+00	48	8
M1M2	M2	6.20E+08	1.80E+08	72	1
M1M2	M2	6.50E+08	2.20E+08	72	2
M1M2	M2	1.40E+08	0.00E+00	72	3
M1M2	M2	1.30E+08	5.90E+08	72	4
M1M2	M2	8.70E+08	1.60E+08	72	5
M1M2	M2	6.00E+07	0.00E+00	72	6
M1M2	M2	5.60E+08	2.20E+08	72	7
M1M2	M2	8.00E+08	0.00E+00	72	8

Table 8.8: . Colony counts for the *yebC* gene in m9 medium. WT: wildtype. M1: T65N-YebC single mutant . M2: L66I-YebC other single mutant . WM1M2: double mutant T65N-L66I-YebC.

nameA	nameB	countA	countB	time	replicate
WT	WM1M2	1.22E+06	1.16E+06	0	1
WT	WM1M2	1.31E+06	1.17E+06	0	2
WT	WM1M2	1.56E+06	1.43E+06	0	3
WT	WM1M2	1.42E+06	1.23E+06	0	4
WT	WM1M2	1.43E+06	1.22E+06	0	5
WT	WM1M2	1.43E+06	1.34E+06	0	6
WT	WM1M2	8.90E+05	9.10E+05	0	7
WT	WM1M2	1.10E+06	9.60E+05	0	8
WT	WM1M2	1.02E+08	1.12E+08	24	1
WT	WM1M2	8.30E+07	1.01E+08	24	2
WT	WM1M2	5.40E+07	1.29E+08	24	3
WT	WM1M2	7.20E+07	9.70E+07	24	4
WT	WM1M2	7.10E+07	1.16E+08	24	5
WT	WM1M2	8.40E+07	7.90E+07	24	6
WT	WM1M2	4.80E+07	1.37E+08	24	7
WT	WM1M2	6.90E+07	8.40E+07	24	8
WT	WM1M2	8.20E+07	1.29E+08	48	1
WT	WM1M2	9.10E+07	1.02E+08	48	2
WT	WM1M2	5.30E+07	1.81E+08	48	3
WT	WM1M2	8.20E+07	1.29E+08	48	4
WT	WM1M2	1.40E+08	1.52E+08	48	5
WT	WM1M2	1.01E+08	1.15E+08	48	6
WT	WM1M2	6.30E+07	1.84E+08	48	7
WT	WM1M2	1.36E+08	1.65E+08	48	8
WT	WM1M2	1.33E+08	2.21E+08	72	1
WT	WM1M2	6.30E+07	1.14E+08	72	2
WT	WM1M2	6.20E+07	1.86E+08	72	3
WT	WM1M2	8.90E+07	1.51E+08	72	4
WT	WM1M2	1.02E+08	1.79E+08	72	5
WT	WM1M2	1.05E+08	1.95E+08	72	6
WT	WM1M2	5.70E+07	1.94E+08	72	7
WT	WM1M2	1.44E+08	2.30E+08	72	8
WT	M1	1.14E+06	1.02E+06	0	1
WT	M1	1.03E+06	1.02E+06	0	2
WT	M1	8.40E+05	7.10E+05	0	3
WT	M1	1.00E+06	7.90E+05	0	4
WT	M1	7.80E+05	5.80E+05	0	5
WT	M1	1.01E+06	7.30E+05	0	6
WT	M1	1.02E+06	5.90E+05	0	7
WT	M1	8.30E+05	9.90E+05	0	8
WT	M1	9.40E+07	7.20E+07	24	1
WT	M1	1.08E+08	4.10E+07	24	2
WT	M1	9.90E+07	9.50E+07	24	3
WT	M1	1.10E+08	8.40E+07	24	4
WT	M1	1.12E+08	9.70E+07	24	5
WT	M1	1.07E+08	6.90E+07	24	6
WT	M1	9.20E+07	7.10E+07	24	7
WT	M1	1.05E+08	8.30E+07	24	8
WT	M1	1.28E+08	9.20E+07	48	1
WT	M1	1.26E+08	3.40E+07	48	2
WT	M1	1.05E+08	9.00E+07	48	3
WT	M1	0.00E+00	0.00E+00	48	4

nameA	nameB	countA	countB	time	replicate
WT	M1	1.14E+08	9.90E+07	48	5
WT	M1	9.80E+07	8.90E+07	48	6
WT	M1	1.17E+08	1.05E+08	48	7
WT	M1	1.08E+08	8.20E+07	48	8
WT	M1	2.26E+08	1.04E+08	72	1
WT	M1	1.93E+08	4.20E+07	72	2
WT	M1	1.38E+08	7.00E+07	72	3
WT	M1	1.64E+08	8.70E+07	72	4
WT	M1	1.33E+08	9.40E+07	72	5
WT	M1	1.54E+08	8.80E+07	72	6
WT	M1	1.61E+08	1.16E+08	72	7
WT	M1	2.39E+08	1.15E+08	72	8
WT	M2	1.06E+06	1.24E+06	0	1
WT	M2	8.70E+05	1.12E+06	0	2
WT	M2	1.02E+06	8.80E+05	0	3
WT	M2	1.43E+06	1.26E+06	0	4
WT	M2	1.08E+06	1.02E+06	0	5
WT	M2	1.04E+06	1.23E+06	0	6
WT	M2	1.13E+06	8.30E+05	0	7
WT	M2	5.90E+05	5.00E+05	0	8
WT	M2	1.16E+08	1.01E+08	24	1
WT	M2	1.08E+08	5.20E+07	24	2
WT	M2	8.60E+07	7.60E+07	24	3
WT	M2	1.08E+08	9.00E+07	24	4
WT	M2	8.40E+07	6.20E+07	24	5
WT	M2	1.10E+08	6.90E+07	24	6
WT	M2	7.70E+07	8.20E+07	24	7
WT	M2	7.50E+07	5.20E+07	24	8
WT	M2	1.21E+08	8.80E+07	48	1
WT	M2	1.06E+08	8.30E+07	48	2
WT	M2	1.20E+08	7.30E+07	48	3
WT	M2	1.17E+08	9.00E+07	48	4
WT	M2	1.10E+08	8.80E+07	48	5
WT	M2	1.27E+08	8.50E+07	48	6
WT	M2	1.10E+08	8.20E+07	48	7
WT	M2	1.24E+08	7.30E+07	48	8
WT	M2	1.56E+08	8.80E+07	72	1
WT	M2	1.54E+08	7.00E+07	72	2
WT	M2	1.37E+08	8.10E+07	72	3
WT	M2	1.46E+08	1.03E+08	72	4
WT	M2	1.29E+08	7.30E+07	72	5
WT	M2	1.84E+08	1.00E+08	72	6
WT	M2	1.48E+08	1.13E+08	72	7
WT	M2	2.20E+08	1.36E+08	72	8
M1M2	M1	1.02E+06	9.70E+05	0	1
M1M2	M1	1.23E+06	1.50E+06	0	2
M1M2	M1	1.41E+06	8.80E+05	0	3
M1M2	M1	9.80E+05	1.02E+06	0	4
M1M2	M1	1.13E+06	1.12E+06	0	5
M1M2	M1	9.80E+05	8.40E+05	0	6
M1M2	M1	8.20E+05	9.70E+05	0	7
M1M2	M1	8.70E+05	1.02E+06	0	8
M1M2	M1	7.50E+07	5.10E+07	24	1
M1M2	M1	8.50E+07	6.40E+07	24	2

nameA	nameB	countA	countB	time	replicate
M1M2	M1	8.40E+07	8.40E+07	24	3
M1M2	M1	7.50E+07	7.70E+07	24	4
M1M2	M1	1.03E+08	7.10E+07	24	5
M1M2	M1	6.10E+07	4.80E+07	24	6
M1M2	M1	9.80E+07	5.50E+07	24	7
M1M2	M1	4.80E+07	3.60E+07	24	8
M1M2	M1	1.27E+08	9.60E+07	48	1
M1M2	M1	1.38E+08	8.10E+07	48	2
M1M2	M1	1.15E+08	1.20E+08	48	3
M1M2	M1	1.47E+08	1.15E+08	48	4
M1M2	M1	1.27E+08	1.27E+08	48	5
M1M2	M1	1.24E+08	1.18E+08	48	6
M1M2	M1	1.35E+08	1.17E+08	48	7
M1M2	M1	1.21E+08	9.70E+07	48	8
M1M2	M1	1.94E+08	1.29E+08	72	1
M1M2	M1	1.67E+08	6.80E+07	72	2
M1M2	M1	1.81E+08	9.60E+07	72	3
M1M2	M1	1.48E+08	1.04E+08	72	4
M1M2	M1	1.62E+08	1.07E+08	72	5
M1M2	M1	1.50E+08	1.04E+08	72	6
M1M2	M1	1.71E+08	8.60E+07	72	7
M1M2	M1	2.10E+08	1.18E+08	72	8
M1M2	M2	1.22E+06	1.34E+06	0	1
M1M2	M2	9.00E+05	1.15E+06	0	2
M1M2	M2	1.10E+06	9.30E+05	0	3
M1M2	M2	8.30E+05	6.60E+05	0	4
M1M2	M2	1.03E+06	1.06E+06	0	5
M1M2	M2	1.03E+06	1.27E+06	0	6
M1M2	M2	7.00E+05	9.60E+05	0	7
M1M2	M2	6.60E+05	5.70E+05	0	8
M1M2	M2	8.60E+07	1.01E+08	24	1
M1M2	M2	1.15E+08	9.40E+07	24	2
M1M2	M2	1.14E+08	8.60E+07	24	3
M1M2	M2	1.32E+08	1.06E+08	24	4
M1M2	M2	9.80E+07	7.70E+07	24	5
M1M2	M2	8.60E+07	6.90E+07	24	6
M1M2	M2	8.30E+07	8.40E+07	24	7
M1M2	M2	6.00E+07	5.40E+07	24	8
M1M2	M2	1.50E+08	1.09E+08	48	1
M1M2	M2	1.31E+08	9.20E+07	48	2
M1M2	M2	1.56E+08	1.23E+08	48	3
M1M2	M2	1.47E+08	9.60E+07	48	4
M1M2	M2	1.49E+08	1.04E+08	48	5
M1M2	M2	1.55E+08	1.19E+08	48	6
M1M2	M2	1.57E+08	1.27E+08	48	7
M1M2	M2	1.23E+08	8.80E+07	48	8
M1M2	M2	1.87E+08	1.18E+08	72	1
M1M2	M2	1.68E+08	1.10E+08	72	2
M1M2	M2	1.59E+08	1.16E+08	72	3
M1M2	M2	1.83E+08	8.70E+07	72	4
M1M2	M2	1.98E+08	1.16E+08	72	5
M1M2	M2	1.58E+08	1.07E+08	72	6
M1M2	M2	1.87E+08	1.13E+08	72	7
M1M2	M2	2.16E+08	1.60E+08	72	8

Table 8.9: Relative fitness of competition experiments for the *lepA* gene candidate in LB medium.

Strain	RF	Time
S66K vs S66K-Q170E	0.31900	24
S66K vs S66K-Q170E	1.04030	24
S66K vs S66K-Q170E	1.29734	24
S66K vs S66K-Q170E	0.80736	24
S66K vs S66K-Q170E	0.74203	24
S66K vs S66K-Q170E	0.82456	24
S66K vs S66K-Q170E	0.69174	24
S66K vs S66K-Q170E	0.94376	24
Q170E vs S66K-Q170E	0.82982	24
Q170E vs S66K-Q170E	0.79832	24
Q170E vs S66K-Q170E	0.96542	24
Q170E vs S66K-Q170E	0.89104	24
Q170E vs S66K-Q170E	0.91889	24
Q170E vs S66K-Q170E	0.86860	24
Q170E vs S66K-Q170E	0.93410	24
Q170E vs S66K-Q170E	0.98576	24
S66K vs S66K-Q170E	1.62299	48
S66K vs S66K-Q170E	1.71392	48
S66K vs S66K-Q170E	1.10190	48
S66K vs S66K-Q170E	0.78613	48
S66K vs S66K-Q170E	0.34780	48
S66K vs S66K-Q170E	0.70648	48
S66K vs S66K-Q170E	0.32693	48
S66K vs S66K-Q170E	0.88639	48
Q170E vs S66K-Q170E	0.67807	48
Q170E vs S66K-Q170E	0.27642	48
Q170E vs S66K-Q170E	1.07947	48
Q170E vs S66K-Q170E	0.82452	48
Q170E vs S66K-Q170E	0.84131	48
Q170E vs S66K-Q170E	0.26063	48
Q170E vs S66K-Q170E	0.81801	48
Q170E vs S66K-Q170E	0.71691	48
S66K vs S66K-Q170E	0.28167	72
S66K vs S66K-Q170E	0.76681	72
S66K vs S66K-Q170E	1.09201	72
S66K vs S66K-Q170E	0.77356	72
S66K vs S66K-Q170E	0.18914	72
S66K vs S66K-Q170E	0.41269	72
S66K vs S66K-Q170E	0.32236	72
S66K vs S66K-Q170E	0.78303	72
Q170E vs S66K-Q170E	0.70637	72
Q170E vs S66K-Q170E	0.23503	72
Q170E vs S66K-Q170E	0.38024	72
Q170E vs S66K-Q170E	0.57620	72
Q170E vs S66K-Q170E	0.27739	72
Q170E vs S66K-Q170E	0.25684	72
Q170E vs S66K-Q170E	0.57501	72
Q170E vs S66K-Q170E	0.36504	72
Wildtype vs S66K	1.06282	24
Wildtype vs S66K	1.56998	24
Wildtype vs S66K	1.69567	24
Wildtype vs S66K	1.00549	24

Strain	RF	Time
Wildtype vs S66K	1.25535	24
Wildtype vs S66K	1.06795	24
Wildtype vs S66K	1.66229	24
Wildtype vs S66K	1.02074	24
Wildtype vs Q170E	0.89671	24
Wildtype vs Q170E	1.49864	24
Wildtype vs Q170E	1.06295	24
Wildtype vs Q170E	1.00293	24
Wildtype vs Q170E	1.21595	24
Wildtype vs Q170E	1.72672	24
Wildtype vs Q170E	1.71792	24
Wildtype vs Q170E	1.11798	24
Wildtype vs S66K	1.15009	48
Wildtype vs S66K	1.91021	48
Wildtype vs S66K	1.49140	48
Wildtype vs S66K	1.17425	48
Wildtype vs S66K	1.31304	48
Wildtype vs S66K	1.07285	48
Wildtype vs S66K	1.81263	48
Wildtype vs S66K	1.16345	48
Wildtype vs Q170E	0.93849	48
Wildtype vs Q170E	1.84856	48
Wildtype vs Q170E	1.21870	48
Wildtype vs Q170E	1.05974	48
Wildtype vs Q170E	1.20004	48
Wildtype vs Q170E	1.81621	48
Wildtype vs Q170E	1.80607	48
Wildtype vs Q170E	1.09525	48
Wildtype vs S66K	1.39596	72
Wildtype vs S66K	2.00629	72
Wildtype vs S66K	1.62975	72
Wildtype vs S66K	1.46399	72
Wildtype vs S66K	1.47868	72
Wildtype vs S66K	1.13331	72
Wildtype vs S66K	1.81707	72
Wildtype vs S66K	1.17644	72
Wildtype vs Q170E	1.01521	72
Wildtype vs Q170E	1.65177	72
Wildtype vs Q170E	1.10912	72
Wildtype vs Q170E	1.10866	72
Wildtype vs Q170E	2.22317	72
Wildtype vs Q170E	1.90170	72
Wildtype vs Q170E	1.64935	72
Wildtype vs Q170E	1.18936	72
Wildtype vs S66K-Q170E	0.88704	24
Wildtype vs S66K-Q170E	0.85565	24
Wildtype vs S66K-Q170E	0.89179	24
Wildtype vs S66K-Q170E	0.85181	24
Wildtype vs S66K-Q170E	0.45468	24
Wildtype vs S66K-Q170E	0.83863	24
Wildtype vs S66K-Q170E	0.90266	24
Wildtype vs S66K-Q170E	1.02539	24
Wildtype vs S66K-Q170E	0.85383	48
Wildtype vs S66K-Q170E	0.80355	48

Strain	RF	Time
Wildtype vs S66K-Q170E	0.74326	48
Wildtype vs S66K-Q170E	0.93442	48
Wildtype vs S66K-Q170E	0.60966	48
Wildtype vs S66K-Q170E	0.63182	48
Wildtype vs S66K-Q170E	0.93087	48
Wildtype vs S66K-Q170E	0.68162	48
Wildtype vs S66K-Q170E	0.71195	72
Wildtype vs S66K-Q170E	0.76537	72
Wildtype vs S66K-Q170E	0.57657	72
Wildtype vs S66K-Q170E	0.76358	72
Wildtype vs S66K-Q170E	0.41682	72
Wildtype vs S66K-Q170E	0.65843	72
Wildtype vs S66K-Q170E	0.63789	72
Wildtype vs S66K-Q170E	0.56272	72

Table 8.10: Relative fitness of competition experiments for the *lepA* gene candidate in M9 minimal medium.

Strain	RF	Time
S66K vs S66K-Q170E	0.87765	24
S66K vs S66K-Q170E	1.02416	24
S66K vs S66K-Q170E	1.07042	24
S66K vs S66K-Q170E	1.03510	24
S66K vs S66K-Q170E	0.94729	24
S66K vs S66K-Q170E	0.92838	24
S66K vs S66K-Q170E	0.91814	24
S66K vs S66K-Q170E	1.04335	24
Q170E vs S66K-Q170E	1.08787	24
Q170E vs S66K-Q170E	1.11192	24
Q170E vs S66K-Q170E	0.96409	24
Q170E vs S66K-Q170E	0.99796	24
Q170E vs S66K-Q170E	0.84096	24
Q170E vs S66K-Q170E	1.01982	24
Q170E vs S66K-Q170E	1.26949	24
Q170E vs S66K-Q170E	0.93900	24
S66K vs S66K-Q170E	0.91997	48
S66K vs S66K-Q170E	1.07537	48
S66K vs S66K-Q170E	0.96383	48
S66K vs S66K-Q170E	0.84165	48
S66K vs S66K-Q170E	0.74928	48
S66K vs S66K-Q170E	0.89098	48
S66K vs S66K-Q170E	0.83674	48
S66K vs S66K-Q170E	0.96174	48
Q170E vs S66K-Q170E	0.93768	48
Q170E vs S66K-Q170E	0.98684	48
Q170E vs S66K-Q170E	0.89537	48
Q170E vs S66K-Q170E	0.95586	48
Q170E vs S66K-Q170E	0.85552	48
Q170E vs S66K-Q170E	0.91525	48
Q170E vs S66K-Q170E	1.26934	48
Q170E vs S66K-Q170E	0.95839	48
S66K vs S66K-Q170E	0.88096	72
S66K vs S66K-Q170E	0.97256	72
S66K vs S66K-Q170E	0.98770	72
S66K vs S66K-Q170E	0.92624	72
S66K vs S66K-Q170E	0.78109	72
S66K vs S66K-Q170E	0.89699	72
S66K vs S66K-Q170E	0.81349	72
S66K vs S66K-Q170E	0.97113	72
Q170E vs S66K-Q170E	0.95270	72
Q170E vs S66K-Q170E	0.97820	72
Q170E vs S66K-Q170E	0.91840	72
Q170E vs S66K-Q170E	0.97760	72
Q170E vs S66K-Q170E	0.82824	72
Q170E vs S66K-Q170E	0.89322	72
Q170E vs S66K-Q170E	1.07118	72
Q170E vs S66K-Q170E	0.89858	72
Wildtype vs S66K	1.06003	24
Wildtype vs S66K	1.03995	24
Wildtype vs S66K	1.09597	24
Wildtype vs S66K	0.96736	24

Strain	RF	Time
Wildtype vs S66K	1.11145	24
Wildtype vs S66K	1.07035	24
Wildtype vs S66K	1.01554	24
Wildtype vs S66K	0.79924	24
Wildtype vs Q170E	0.91292	24
Wildtype vs Q170E	0.95906	24
Wildtype vs Q170E	0.92168	24
Wildtype vs Q170E	0.91061	24
Wildtype vs Q170E	1.10020	24
Wildtype vs Q170E	0.95324	24
Wildtype vs Q170E	0.85500	24
Wildtype vs Q170E	0.89562	24
Wildtype vs S66K	0.97938	48
Wildtype vs S66K	1.05429	48
Wildtype vs S66K	1.01077	48
Wildtype vs S66K	1.00547	48
Wildtype vs S66K	1.24301	48
Wildtype vs S66K	1.25823	48
Wildtype vs S66K	1.00140	48
Wildtype vs S66K	0.87305	48
Wildtype vs Q170E	1.03869	48
Wildtype vs Q170E	0.93847	48
Wildtype vs Q170E	0.90069	48
Wildtype vs Q170E	0.97872	48
Wildtype vs Q170E	1.26337	48
Wildtype vs Q170E	1.00086	48
Wildtype vs Q170E	0.81548	48
Wildtype vs Q170E	0.92916	48
Wildtype vs S66K	1.02089	72
Wildtype vs S66K	1.14098	72
Wildtype vs S66K	1.07325	72
Wildtype vs S66K	1.12132	72
Wildtype vs S66K	1.15954	72
Wildtype vs S66K	1.15396	72
Wildtype vs S66K	1.06038	72
Wildtype vs S66K	0.94769	72
Wildtype vs Q170E	1.05281	72
Wildtype vs Q170E	1.03057	72
Wildtype vs Q170E	0.96822	72
Wildtype vs Q170E	0.98083	72
Wildtype vs Q170E	1.29022	72
Wildtype vs Q170E	1.04170	72
Wildtype vs Q170E	0.93828	72
Wildtype vs Q170E	1.08546	72
Wildtype vs S66K-Q170E	1.15498	24
Wildtype vs S66K-Q170E	0.90989	24
Wildtype vs S66K-Q170E	0.93783	24
Wildtype vs S66K-Q170E	1.07919	24
Wildtype vs S66K-Q170E	0.87727	24
Wildtype vs S66K-Q170E	1.05412	24
Wildtype vs S66K-Q170E	1.14850	24
Wildtype vs S66K-Q170E	0.95308	24
Wildtype vs S66K-Q170E	1.01173	48
Wildtype vs S66K-Q170E	0.88909	48

Strain	RF	Time
Wildtype vs S66K-Q170E	0.88896	48
Wildtype vs S66K-Q170E	1.02535	48
Wildtype vs S66K-Q170E	0.86363	48
Wildtype vs S66K-Q170E	1.03252	48
Wildtype vs S66K-Q170E	1.13957	48
Wildtype vs S66K-Q170E	0.88150	48
Wildtype vs S66K-Q170E	0.85494	72
Wildtype vs S66K-Q170E	0.90098	72
Wildtype vs S66K-Q170E	0.85817	72
Wildtype vs S66K-Q170E	1.05964	72
Wildtype vs S66K-Q170E	0.84958	72
Wildtype vs S66K-Q170E	1.08103	72
Wildtype vs S66K-Q170E	1.05060	72
Wildtype vs S66K-Q170E	0.88759	72

Table 8.11: Relative fitness of competition experiments for the *ispH* gene candidate in LB medium.

Strain	RF	Time
N24R vs N24R-A27E	0.77689	24
N24R vs N24R-A27E	0.70492	24
N24R vs N24R-A27E	0.85669	24
N24R vs N24R-A27E	1.03510	24
N24R vs N24R-A27E	1.08221	24
N24R vs N24R-A27E	0.58602	24
N24R vs N24R-A27E	0.21466	24
N24R vs N24R-A27E	0.66364	24
A27E vs N24R-A27E	1.16573	24
A27E vs N24R-A27E	1.10252	24
A27E vs N24R-A27E	1.25520	24
A27E vs N24R-A27E	2.30872	24
A27E vs N24R-A27E	1.10030	24
A27E vs N24R-A27E	4.46950	24
A27E vs N24R-A27E	3.08983	24
A27E vs N24R-A27E	1.07776	24
N24R vs N24R-A27E	0.72152	48
N24R vs N24R-A27E	0.69924	48
N24R vs N24R-A27E	0.77419	48
N24R vs N24R-A27E	0.84022	48
N24R vs N24R-A27E	0.75667	48
N24R vs N24R-A27E	0.37902	48
N24R vs N24R-A27E	0.37852	48
N24R vs N24R-A27E	0.60901	48
A27E vs N24R-A27E	1.27908	48
A27E vs N24R-A27E	1.32561	48
A27E vs N24R-A27E	1.58567	48
A27E vs N24R-A27E	4.27266	48
A27E vs N24R-A27E	1.19345	48
A27E vs N24R-A27E	5.20600	48
A27E vs N24R-A27E	3.52051	48
A27E vs N24R-A27E	1.29113	48
N24R vs N24R-A27E	0.64807	72
N24R vs N24R-A27E	0.74858	72
N24R vs N24R-A27E	0.57923	72
N24R vs N24R-A27E	0.73131	72
N24R vs N24R-A27E	0.70213	72
N24R vs N24R-A27E	0.28579	72
N24R vs N24R-A27E	0.18046	72
N24R vs N24R-A27E	0.45006	72
A27E vs N24R-A27E	1.81564	72
A27E vs N24R-A27E	1.43428	72
A27E vs N24R-A27E	1.52454	72
A27E vs N24R-A27E	4.52448	72
A27E vs N24R-A27E	1.62356	72
A27E vs N24R-A27E	4.46950	72
A27E vs N24R-A27E	3.19912	72
A27E vs N24R-A27E	1.26097	72
Wildtype vs N24R	0.90086	24
Wildtype vs N24R	1.78114	24
Wildtype vs N24R	1.81510	24
Wildtype vs N24R	1.54552	24

Strain	RF	Time
Wildtype vs N24R	2.34911	24
Wildtype vs N24R	1.52350	24
Wildtype vs N24R	1.70279	24
Wildtype vs N24R	1.49301	24
Wildtype vs A27E	0.54004	24
Wildtype vs A27E	0.62811	24
Wildtype vs A27E	0.90845	24
Wildtype vs A27E	0.48921	24
Wildtype vs A27E	0.84783	24
Wildtype vs A27E	0.48957	24
Wildtype vs A27E	0.95815	24
Wildtype vs A27E	0.88665	24
Wildtype vs N24R-A27E	0.64518	24
Wildtype vs N24R-A27E	0.48025	24
Wildtype vs N24R-A27E	0.89399	24
Wildtype vs N24R-A27E	0.80359	24
Wildtype vs N24R-A27E	0.91254	24
Wildtype vs N24R-A27E	0.90796	24
Wildtype vs N24R-A27E	1.17055	24
Wildtype vs N24R-A27E	0.64292	24
Wildtype vs N24R	0.85235	48
Wildtype vs N24R	1.84030	48
Wildtype vs N24R	2.37907	48
Wildtype vs N24R	2.20407	48
Wildtype vs N24R	2.55085	48
Wildtype vs N24R	2.43798	48
Wildtype vs N24R	2.17343	48
Wildtype vs N24R	1.29308	48
Wildtype vs A27E	0.46800	48
Wildtype vs A27E	0.44373	48
Wildtype vs A27E	0.78790	48
Wildtype vs A27E	0.54499	48
Wildtype vs A27E	0.85802	48
Wildtype vs A27E	0.73733	48
Wildtype vs A27E	0.87953	48
Wildtype vs A27E	0.67004	48
Wildtype vs N24R-A27E	0.51622	48
Wildtype vs N24R-A27E	0.43826	48
Wildtype vs N24R-A27E	0.82883	48
Wildtype vs N24R-A27E	0.73381	48
Wildtype vs N24R-A27E	0.83589	48
Wildtype vs N24R-A27E	0.87620	48
Wildtype vs N24R-A27E	1.08695	48
Wildtype vs N24R-A27E	0.47348	48
Wildtype vs N24R	0.88324	72
Wildtype vs N24R	1.93041	72
Wildtype vs N24R	2.22357	72
Wildtype vs N24R	2.00381	72
Wildtype vs N24R	2.50955	72
Wildtype vs N24R	2.37607	72
Wildtype vs N24R	1.83522	72
Wildtype vs N24R	1.68235	72
Wildtype vs A27E	0.47671	72
Wildtype vs A27E	0.43456	72

Strain	RF	Time
Wildtype vs A27E	0.71152	72
Wildtype vs A27E	0.47128	72
Wildtype vs A27E	0.79769	72
Wildtype vs A27E	0.45934	72
Wildtype vs A27E	0.76098	72
Wildtype vs A27E	0.71743	72
Wildtype vs N24R-A27E	0.56914	72
Wildtype vs N24R-A27E	0.47177	72
Wildtype vs N24R-A27E	0.81019	72
Wildtype vs N24R-A27E	0.64993	72
Wildtype vs N24R-A27E	0.77091	72
Wildtype vs N24R-A27E	1.01811	72
Wildtype vs N24R-A27E	0.51231	72

Table 8.12: Relative fitness of competition experiments for the *ispH* gene candidate in M9 minimal medium.

Strain	RF	Time
N24R vs N24R-A27E	0.90759	24
N24R vs N24R-A27E	0.94439	24
N24R vs N24R-A27E	0.85231	24
N24R vs N24R-A27E	1.03776	24
N24R vs N24R-A27E	1.02298	24
N24R vs N24R-A27E	0.95216	24
N24R vs N24R-A27E	0.88098	24
N24R vs N24R-A27E	1.16274	24
A27E vs N24R-A27E	0.96532	24
A27E vs N24R-A27E	0.83359	24
A27E vs N24R-A27E	1.06541	24
A27E vs N24R-A27E	1.10968	24
A27E vs N24R-A27E	1.00421	24
A27E vs N24R-A27E	0.98973	24
A27E vs N24R-A27E	1.05452	24
A27E vs N24R-A27E	1.05956	24
N24R vs N24R-A27E	0.91715	48
N24R vs N24R-A27E	0.93007	48
N24R vs N24R-A27E	0.89584	48
N24R vs N24R-A27E	0.96151	48
N24R vs N24R-A27E	0.98239	48
N24R vs N24R-A27E	0.92202	48
N24R vs N24R-A27E	0.88817	48
N24R vs N24R-A27E	1.09103	48
A27E vs N24R-A27E	0.98809	48
A27E vs N24R-A27E	0.85836	48
A27E vs N24R-A27E	1.04461	48
A27E vs N24R-A27E	1.07746	48
A27E vs N24R-A27E	1.03909	48
A27E vs N24R-A27E	1.01693	48
A27E vs N24R-A27E	1.11400	48
A27E vs N24R-A27E	1.06458	48
N24R vs N24R-A27E	0.83056	72
N24R vs N24R-A27E	0.90705	72
N24R vs N24R-A27E	0.88734	72
N24R vs N24R-A27E	0.95449	72
N24R vs N24R-A27E	0.94868	72
N24R vs N24R-A27E	0.93686	72
N24R vs N24R-A27E	0.89400	72
N24R vs N24R-A27E	1.16797	72
A27E vs N24R-A27E	0.96490	72
A27E vs N24R-A27E	0.83681	72
A27E vs N24R-A27E	1.05763	72
A27E vs N24R-A27E	1.10943	72
A27E vs N24R-A27E	1.07540	72
A27E vs N24R-A27E	1.00633	72
A27E vs N24R-A27E	1.10047	72
A27E vs N24R-A27E	1.05552	72
Wildtype vs N24R	1.03731	24
Wildtype vs N24R	1.00792	24
Wildtype vs N24R	1.09027	24
Wildtype vs N24R	0.96242	24

Strain	RF	Time
Wildtype vs N24R	0.99515	24
Wildtype vs N24R	1.05209	24
Wildtype vs N24R	1.02378	24
Wildtype vs N24R	0.90354	24
Wildtype vs A27E	1.06333	24
Wildtype vs A27E	1.02538	24
Wildtype vs A27E	0.96314	24
Wildtype vs A27E	0.99033	24
Wildtype vs A27E	0.96555	24
Wildtype vs A27E	0.98227	24
Wildtype vs A27E	1.01737	24
Wildtype vs A27E	1.07210	24
Wildtype vs N24R-A27E	1.01007	24
Wildtype vs N24R-A27E	0.95710	24
Wildtype vs N24R-A27E	0.95030	24
Wildtype vs N24R-A27E	0.99164	24
Wildtype vs N24R-A27E	0.98809	24
Wildtype vs N24R	1.01892	48
Wildtype vs N24R	1.01743	48
Wildtype vs N24R	1.08887	48
Wildtype vs N24R	0.96865	48
Wildtype vs N24R	1.05170	48
Wildtype vs N24R	1.03817	48
Wildtype vs N24R	1.06088	48
Wildtype vs N24R	0.88843	48
Wildtype vs A27E	1.04118	48
Wildtype vs A27E	0.97049	48
Wildtype vs A27E	0.97667	48
Wildtype vs A27E	0.96648	48
Wildtype vs A27E	0.97166	48
Wildtype vs A27E	0.92810	48
Wildtype vs A27E	0.97274	48
Wildtype vs A27E	1.07401	48
Wildtype vs N24R-A27E	0.99224	48
Wildtype vs N24R-A27E	0.99485	48
Wildtype vs N24R-A27E	0.98881	48
Wildtype vs N24R-A27E	0.98758	48
Wildtype vs N24R-A27E	1.02634	48
Wildtype vs N24R	1.02827	72
Wildtype vs N24R	1.03674	72
Wildtype vs N24R	1.06133	72
Wildtype vs N24R	0.96615	72
Wildtype vs N24R	1.04390	72
Wildtype vs N24R	1.05278	72
Wildtype vs N24R	1.07871	72
Wildtype vs N24R	0.92028	72
Wildtype vs A27E	1.00276	72
Wildtype vs A27E	0.96270	72
Wildtype vs A27E	0.95263	72
Wildtype vs A27E	0.92069	72
Wildtype vs A27E	0.96618	72
Wildtype vs A27E	0.92911	72
Wildtype vs A27E	0.94650	72
Wildtype vs A27E	1.05860	72

Strain	RF	Time
Wildtype vs N24R-A27E	0.96481	72
Wildtype vs N24R-A27E	0.98413	72
Wildtype vs N24R-A27E	0.96845	72
Wildtype vs N24R-A27E	0.97847	72
Wildtype vs N24R-A27E	1.00168	72

Table 8.13: Relative fitnesses of competition experiments for the *yebC* gene candidate in LB medium.

Strain	RF	Time
T65N vs T65N-L66I	1.44458	24
T65N vs T65N-L66I	1.36962	24
T65N vs T65N-L66I	1.36445	24
T65N vs T65N-L66I	1.08728	24
T65N vs T65N-L66I	1.06682	24
L66I vs T65N-L66I	1.10584	24
L66I vs T65N-L66I	1.17521	24
L66I vs T65N-L66I	1.12138	24
L66I vs T65N-L66I	1.15332	24
L66I vs T65N-L66I	1.80631	24
L66I vs T65N-L66I	2.98780	24
L66I vs T65N-L66I	1.04726	24
T65N vs T65N-L66I	1.45664	48
T65N vs T65N-L66I	1.61898	48
T65N vs T65N-L66I	1.42460	48
T65N vs T65N-L66I	2.87771	48
T65N vs T65N-L66I	1.32736	48
T65N vs T65N-L66I	1.41412	48
L66I vs T65N-L66I	1.34006	48
L66I vs T65N-L66I	1.15629	48
L66I vs T65N-L66I	1.15752	48
L66I vs T65N-L66I	1.23767	48
L66I vs T65N-L66I	1.85878	48
L66I vs T65N-L66I	2.20035	48
L66I vs T65N-L66I	1.19576	48
T65N vs T65N-L66I	1.62183	72
T65N vs T65N-L66I	1.33778	72
T65N vs T65N-L66I	1.69045	72
T65N vs T65N-L66I	4.02292	72
T65N vs T65N-L66I	2.89474	72
T65N vs T65N-L66I	1.26128	72
T65N vs T65N-L66I	2.61249	72
T65N vs T65N-L66I	1.40510	72
L66I vs T65N-L66I	1.44038	72
L66I vs T65N-L66I	1.33845	72
L66I vs T65N-L66I	1.34068	72
L66I vs T65N-L66I	1.42888	72
L66I vs T65N-L66I	2.44053	72
L66I vs T65N-L66I	3.67736	72
L66I vs T65N-L66I	5.35378	72
L66I vs T65N-L66I	1.15794	72
Wildtype vs T65N	0.74110	24
Wildtype vs T65N	0.77276	24
Wildtype vs T65N	0.59646	24
Wildtype vs T65N	0.95154	24
Wildtype vs T65N	0.94620	24
Wildtype vs T65N	0.97868	24
Wildtype vs T65N	0.86164	24
Wildtype vs L66I	0.84375	24
Wildtype vs L66I	0.85396	24
Wildtype vs L66I	0.94665	24
Wildtype vs L66I	0.82660	24

Strain	RF	Time
Wildtype vs L66I	0.90969	24
Wildtype vs L66I	0.97275	24
Wildtype vs T65N-L66I	1.23562	24
Wildtype vs T65N-L66I	0.87374	24
Wildtype vs T65N-L66I	1.60658	24
Wildtype vs T65N-L66I	1.02362	24
Wildtype vs T65N-L66I	1.14065	24
Wildtype vs T65N	0.79129	48
Wildtype vs T65N	0.78548	48
Wildtype vs T65N	0.47960	48
Wildtype vs T65N	0.73481	48
Wildtype vs T65N	0.87612	48
Wildtype vs T65N	0.94332	48
Wildtype vs T65N	0.71407	48
Wildtype vs L66I	0.83586	48
Wildtype vs L66I	0.67166	48
Wildtype vs L66I	0.78061	48
Wildtype vs L66I	0.90708	48
Wildtype vs L66I	0.82482	48
Wildtype vs L66I	0.89787	48
Wildtype vs L66I	0.99844	48
Wildtype vs T65N-L66I	1.24293	48
Wildtype vs T65N-L66I	2.08835	48
Wildtype vs T65N-L66I	0.98059	48
Wildtype vs T65N-L66I	1.08094	48
Wildtype vs T65N-L66I	1.20148	48
Wildtype vs T65N	0.66871	72
Wildtype vs T65N	0.65883	72
Wildtype vs T65N	0.78013	72
Wildtype vs T65N	0.78387	72
Wildtype vs T65N	0.82428	72
Wildtype vs L66I	0.85755	72
Wildtype vs L66I	0.72212	72
Wildtype vs L66I	0.76857	72
Wildtype vs L66I	0.79795	72
Wildtype vs L66I	0.54658	72
Wildtype vs L66I	0.75160	72
Wildtype vs L66I	0.77587	72
Wildtype vs L66I	0.96014	72
Wildtype vs T65N-L66I	1.36890	72
Wildtype vs T65N-L66I	2.07865	72
Wildtype vs T65N-L66I	1.10274	72
Wildtype vs T65N-L66I	2.06117	72
Wildtype vs T65N-L66I	1.14894	72
Wildtype vs T65N-L66I	1.19899	72

Table 8.14: Relative fitness of competition experiments for the *yebC* gene candidate in M9 minimal medium.

Strain	RF	Time
T65N vs T65N-L66I	0.92196	24
T65N vs T65N-L66I	0.88615	24
T65N vs T65N-L66I	1.11534	24
T65N vs T65N-L66I	0.99684	24
T65N vs T65N-L66I	0.91952	24
T65N vs T65N-L66I	0.97930	24
T65N vs T65N-L66I	0.84412	24
T65N vs T65N-L66I	0.88860	24
L66I vs T65N-L66I	1.01573	24
L66I vs T65N-L66I	0.90789	24
L66I vs T65N-L66I	0.97544	24
L66I vs T65N-L66I	1.00194	24
L66I vs T65N-L66I	0.94076	24
L66I vs T65N-L66I	0.90289	24
L66I vs T65N-L66I	0.93637	24
L66I vs T65N-L66I	1.00915	24
T65N vs T65N-L66I	0.95241	48
T65N vs T65N-L66I	0.84508	48
T65N vs T65N-L66I	1.11678	48
T65N vs T65N-L66I	0.94302	48
T65N vs T65N-L66I	1.00188	48
T65N vs T65N-L66I	1.02160	48
T65N vs T65N-L66I	0.93905	48
T65N vs T65N-L66I	0.92297	48
L66I vs T65N-L66I	0.91415	48
L66I vs T65N-L66I	0.87983	48
L66I vs T65N-L66I	0.98591	48
L66I vs T65N-L66I	0.96196	48
L66I vs T65N-L66I	0.92195	48
L66I vs T65N-L66I	0.90551	48
L66I vs T65N-L66I	0.90247	48
L66I vs T65N-L66I	0.96399	48
T65N vs T65N-L66I	0.93183	72
T65N vs T65N-L66I	0.77664	72
T65N vs T65N-L66I	0.96648	72
T65N vs T65N-L66I	0.92171	72
T65N vs T65N-L66I	0.91826	72
T65N vs T65N-L66I	0.95784	72
T65N vs T65N-L66I	0.83983	72
T65N vs T65N-L66I	0.86594	72
L66I vs T65N-L66I	0.88986	72
L66I vs T65N-L66I	0.87214	72
L66I vs T65N-L66I	0.97036	72
L66I vs T65N-L66I	0.90467	72
L66I vs T65N-L66I	0.89287	72
L66I vs T65N-L66I	0.88094	72
L66I vs T65N-L66I	0.85333	72
L66I vs T65N-L66I	0.97349	72
Wildtype vs T65N	1.03651	24
Wildtype vs T65N	1.25957	24
Wildtype vs T65N	0.97408	24
Wildtype vs T65N	1.00727	24

Strain	RF	Time
Wildtype vs T65N	0.97022	24
Wildtype vs T65N	1.02507	24
Wildtype vs T65N	0.93981	24
Wildtype vs T65N	1.09289	24
Wildtype vs L66I	1.06712	24
Wildtype vs L66I	1.25625	24
Wildtype vs L66I	0.99461	24
Wildtype vs L66I	1.01306	24
Wildtype vs L66I	1.06002	24
Wildtype vs L66I	1.15748	24
Wildtype vs L66I	0.91913	24
Wildtype vs L66I	1.04322	24
Wildtype vs T65N-L66I	0.96850	24
Wildtype vs T65N-L66I	0.93062	24
Wildtype vs T65N-L66I	0.78725	24
Wildtype vs T65N-L66I	0.89887	24
Wildtype vs T65N-L66I	0.85735	24
Wildtype vs T65N-L66I	0.99911	24
Wildtype vs T65N-L66I	0.79527	24
Wildtype vs T65N-L66I	0.92557	24
Wildtype vs T65N	1.04865	48
Wildtype vs T65N	1.37078	48
Wildtype vs T65N	0.99711	48
Wildtype vs T65N	0.96981	48
Wildtype vs T65N	0.95246	48
Wildtype vs T65N	0.91523	48
Wildtype vs T65N	1.10227	48
Wildtype vs L66I	1.11151	48
Wildtype vs L66I	1.11548	48
Wildtype vs L66I	1.07908	48
Wildtype vs L66I	1.03181	48
Wildtype vs L66I	1.03724	48
Wildtype vs L66I	1.13441	48
Wildtype vs L66I	0.99678	48
Wildtype vs L66I	1.07310	48
Wildtype vs T65N-L66I	0.89313	48
Wildtype vs T65N-L66I	0.94916	48
Wildtype vs T65N-L66I	0.72831	48
Wildtype vs T65N-L66I	0.87175	48
Wildtype vs T65N-L66I	0.95004	48
Wildtype vs T65N-L66I	0.95624	48
Wildtype vs T65N-L66I	0.80231	48
Wildtype vs T65N-L66I	0.93599	48
Wildtype vs T65N	1.14378	72
Wildtype vs T65N	1.40756	72
Wildtype vs T65N	1.11122	72
Wildtype vs T65N	1.08470	72
Wildtype vs T65N	1.00998	72
Wildtype vs T65N	1.04903	72
Wildtype vs T65N	0.95841	72
Wildtype vs T65N	1.19092	72
Wildtype vs L66I	1.17112	72
Wildtype vs L66I	1.25175	72
Wildtype vs L66I	1.08356	72

Strain	RF	Time
Wildtype vs L66I	1.05048	72
Wildtype vs L66I	1.11993	72
Wildtype vs L66I	1.17679	72
Wildtype vs L66I	0.99212	72
Wildtype vs L66I	1.05627	72
Wildtype vs T65N-L66I	0.89366	72
Wildtype vs T65N-L66I	0.84581	72
Wildtype vs T65N-L66I	0.75645	72
Wildtype vs T65N-L66I	0.86024	72
Wildtype vs T65N-L66I	0.85542	72
Wildtype vs T65N-L66I	0.86265	72
Wildtype vs T65N-L66I	0.77573	72
Wildtype vs T65N-L66I	0.88969	72