Characterization of psychrophilic alleles of essential genes as means of generating temperature-sensitive strains of mesophilic organisms

by

Jaroslaw Pankowski
BSc, University of Warsaw, 2006
MSc, University of Warsaw, 2009

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

© Jaroslaw Pankowski, 2016
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

Characterization of psychrophilic alleles of essential genes as means of generating temperature-sensitive strains of mesophilic organisms

by

Jaroslaw Pankowski
BSc, University of Warsaw, 2006
MSc, University of Warsaw, 2011

Supervisory Committee

Dr. Francis Nano, Department of Biochemistry and Microbiology
Supervisor

Dr. Christopher Upton, Department of Biochemistry and Microbiology
Departmental Member

Dr. Martin Boulanger, Department of Biochemistry and Microbiology
Departmental Member

Dr. Kim Juniper, School of Earth and Ocean Sciences
Outside Member

Abstract

Supervisory Committee
Francis Nano, Department of Biochemistry and Microbiology
Supervisor
Dr. Christopher Upton, Department of Biochemistry and Microbiology
Departmental Member
Dr. Martin Boulanger, Department of Biochemistry and Microbiology
Departmental Member
Dr. Kim Juniper, School of Earth and Ocean Sciences
Outside Member

Essential genes are involved in control of the basic metabolism of their host. These genes encode elements involved is such crucial processes as DNA replication, transcription, translation or biosynthesis of important molecules. What makes essential genes unique is the fact that they cannot be lost from the genome. If any of them becomes inactivated it would result in inevitable death of an organism. Because of their role they can be efficiently used to control the survival of genetically modified organisms. Specific regulatory mechanisms can be applied to modulate the activity of essential genes, which prevents an organism from growing at determined conditions. Such mechanisms are called “kill switches” and have been developed in recent years as a response to significant development in the field of molecular biology.

Proteins encoded by psychrophilic organisms are characterized by decreased resistance to thermal denaturation. This is believed to be a result of adaptation to low-temperature environment, where mutations that destabilize the protein structure are not selected against. For these reasons they often cannot perform their functions at moderate temperatures, which are typical for mesophilic organisms. At the same time psychrophilic proteins do not display any inhibition at permissive conditions.

Use of psychrophilic alleles of essential genes has been proposed as a method of rendering modified organisms incapable of surviving at elevated temperatures. This allows generation of attenuated strains of pathogenic bacteria or generally safe versions of laboratory organisms. A temperature-sensitive organism can be created by substituting a single essential gene in mesophilic organism with its psychrophilic homologue. This can be facilitated by using the host's native recombination system or through the use of plasmid based allele shuffling mechanisms.

The objective of this work was to analyze a number of psychrophilic alleles of various essential genes for their ability to cause temperature-sensitive phenotype in mesophilic bacterium *Francisella novicida*. The special attention has been placed on investigating psychrophilic alleles of bacterial DNA ligase. Furthermore a selected psychrophilic strain has been characterized as a potential source of multiple temperature-sensitive alleles of essential genes. Finally the secondary focus was to develop a simple and robust mechanism allowing efficient exchange of alleles of essential genes in the mesophilic host.

Table of Contents

Supervisory Committee ii

Abstract iii

Table of Contents v

List of Tables x

List of Figures xi

Acknowledgments xiii

Dedication xiv

Chapter 1 Introduction 1

Synthetic biology 1

Definition 1

Modularity 2

Biological devices 3

Biosafety devices 6

Beginnings 6

Current situation 7

Genetic biosafety mechanisms 8

Essential genes 10

Definition 10

Minimal genome of prokaryotes 11

Minimal genome of eukaryotes 14

Identification 15

Utilizing essential genes in biosafety control mechanisms 16

Examples of essential genes 19

DNA ligase 20

Role 20

Organisation 21

Catalysis 25

Methionyl-tRNA formyltransferase 26

Translation initiation 26

Role of formyltransferase 28

Tyrosyl-tRNA synthetase 29

Role 29

Structure 31

Porphobilinogen deaminase 31

Role 32

Structure 34

Peptidylt tRNA-hydrolase 34

Role 34

Organisation and catalysis 36

Temperature-sensitive enzymes and psychrophiles. 37

Psychrophilic organisms 38

Psychrophilic adaptation 39

Psychrophilic enzymes 40

Psychrophilic genomes 43

Allele exchange ... 44

Overview ... 44

BAC vectors and plasmid shuffling ... 44

Chromosomal recombination ... 46

Chapter 2 Materials and methods ... 49

Strains and growth conditions ... 49

Cloning and insertion of psychrophilic *ligA* alleles into the *F. novicida* chromosome. 52

Cloning and insertion of other psychrophilic alleles into the *F. novicida* chromosome. ... 53

Determination of restrictive temperatures of *F. novicida*-TS strains. ... 56

Reversion frequency determination. ... 56

Temperature shift experiments. ... 56

Yeast growth experiments ... 57

Colwellia sp. C1 DNA sequencing. ... 57

Colwellia sp. C1 sequence annotation and submission. ... 57

GenBank Accession numbers. ... 57

Generation of *E. coli* Δpth strain ... 58

Assembly of vector shuffling system. ... 58

Cloning the pth sequence into pBeloBAC11 derived vectors. ... 59

Primers. ... 59

Chapter 3 Collection of temperature-sensitive alleles of DNA ligase ... 65

Introduction ... 65

Results and discussion ... 68

Isolation of *ligA* alleles from psychrophilic bacteria. 68

Comparative analysis of the *ligA* deduced amino acid sequences 81

Growth properties of *F. novicida* harboring psychrophilic *ligA* alleles.................... 86

Mutation frequencies to temperature resistance of *F. novicida* strains carrying *ligA* alleles. 88

Psychrophilic *ligA* alleles that support *S. cerevisiae* viability. 88

Chapter 4 Analysis of temperature resistant mutants 93

Introduction.. 93

Results and discussion 97

Mutations that convert TS *ligA* gene products into temperature-resistant forms. 97

Growth properties of *F. novicida* harboring psychrophilic *ligA* alleles................. 107

Temperature resistant mutations in the context of *ligA* sequence similarity. 110

Discussion 110

Chapter 5 Analysis of genome from extreme psychrophilic *Colwellia* sp. C1.............. 113

Introduction.. 113

Results and discussion 117

Characterization of partial genomic sequence of *Colwellia* sp. C1. 117

Generation of *Francisella novicida hemC* TS strain. 122

Frequency of mutation to temperature resistant phenotype in *F. novicida hemC* TS strain.. 123

Conclusions.. 125

Chapter 6 Screening of psychrophilic alleles of multiple essential genes 127

Introduction.. 127

Results and discussion 128
Generation of TS strains. 128
Frequency of mutation to temperature resistance of *F. novicida* strains carrying temperature sensitive alleles. 133
Conclusions 133
Chapter 7 Plasmid based system for efficient allele exchange 135
Introduction 135
Results and discussion 137
Chapter 8 Summary, conclusions and future directions 141
Summary 141
Conclusions 144
Future directions 146
Bibliography 148

List of Tables

Table 1 Characteristics of selected bacterial genomes. 13

Table 2 Primers used in this study. 60

Table 3 Bacterial origin of psychrophilic *ligA* alleles, restrictive temperature and reversion frequencies in F. novicida strains harboring these alleles. 71

Table 4 Examples of mutations found in temperature-resistant mutants.......................... 98

Table 5 Effect of mutations on predicted RBS strength. 104

Table 6 Stability of *ligA* alleles mRNA 106

Table 7 Analysis of bacterial growth and RBS strength. 108

Table 8 Characteristic of analyzed fragments of genomic DNA from *Colwellia* sp. C1 119

Table 9 Possible essential genes identified in the genome of *Colwellia* sp. C1. 121

Table 10 Essential genes investigated in search for temperature sensitive alleles 130

List of Figures

Figure 1 Example of a complex genetic circuit. 5

Figure 2 Organisation of DNA ligases in representatives of different domains of life. ... 24

Figure 3 Properties of enzymes from different tropic groups of microorganisms........... 42

Figure 4 Locations where psychrophilic strains were isolated. 51

Figure 5 Map of vector pRS426KmSacB 54

Figure 6 Creation of constructs with psychrophilic alleles... 55

Figure 7 Structure of typical psychrophilic gene substitution into the *F. novicida* chromosome. 69

Figure 8 Temperature sensitive phenotype of *Francisella novicida* 72

Figure 9 Phenotype of merodiploid strains 73

Figure 10 Complementation of temperature sensitive phenotype 76

Figure 11 Growth pattern of mesophilic *F. novicida* strains. 77

Figure 12 Hybrid form of C1 allele of DNA ligase. 79

Figure 13 Temperature sensitive phenotype of *Salmonella enterica*............................... 80

Figure 14 Sequence alignment of LigA protein sequences 84

Figure 15 Evolutionary relation of DNA ligase alleles. 85

Figure 16 Growth characteristics of *F. novicida* strains harboring psychrophilic *ligA* alleles 87

Figure 17 Temperature sensitive phenotype of *Saccharomyces cerevisiae*...................... 90

Figure 18 Growth pattern of yeast strains... 91

Figure 19 Common ancestry of enzymes adapted to extreme conditions 94

Figure 20 Relation between temperature resistant phenotype and observed mutations ... 99

Figure 21 Location of mutations that converted TS psychrophilic alleles into ones that produced temperature-resistant products 101

Figure 22 Location of mutations that converted TS psychrophilic alleles into temperature-resistant products, within the local structures of protein 102

Figure 23 Secondary structure of mRNA from *ligA* alleles... 105

Figure 24 RBS strength and generation time.. 109

Figure 25 Predicted ORFs found in the sequenced fragments of *Colwellia* sp. C1 genome 118

Figure 26 Hybrid form of aminoacyl-tRNA synthetase. 132

Acknowledgments

In the first place I would like to thank my supervisor – prof. Nano, who guided me during my work as a grad student.

I would also like to mention my labmates - Sheila Potter, Stephanie Puckett, Stephanie Pearce, Daniel Kemp, Crystal Pinto, Alicia Parfett, Kurtis Vallee, Ryan Woloschuk and Kathleen McLeod – they have managed to withstand working in the same room as I for up to 5 years.

Dedication

To my parents, for repeatedly forcing me to actually do something with my life.

Chapter 1

Introduction

Synthetic biology

Definition

Synthetic biology is a term for an interdisciplinary branch of science focused on analysing and creating biological devices. This involves researching the basic elements (modules) of natural systems and adapting them to synthetic ones. It could be said that synthetic biology is an engineering approach to biological systems. This involves initially reverse engineering such systems to isolate individual elements that create them and to gain insight into relations between these elements, followed by using this knowledge to control their processes and adapt them to specific functions. This would involve creating organisms that could facilitate more efficient bioremediation, synthesis of organic compounds or prevention of diseases.

The term "synthetic biology" has been remarked a few times in publications from the Twentieth century. One of the first uses of this term is attributed to the Polish geneticist Waclaw Szybalski who, in 1974, mentioned synthetic biology as a stage of development in molecular biology that comes after the "descriptive phase" and acts as a mean of creating "new better control circuits" for organisms (1). However the actual development of synthetic biology only began in the Twenty-first century, after the progress was made by traditional molecular biology. In addition the early development of synthetic biology was enhanced by the "omics" sciences (genomics, transcriptomics and

proteomics) that allowed large scale analysis of the basic components of living organisms. It was also dependant on the establishment of technologies allowing cheaper sequencing and synthesis of DNA fragments (2–4). Therefore while the concept of synthetic biology did exist for at least several decades, only within the last 15 years the progress of technology allowed actual implementation of these ideas.

Modularity

The synthetic biology's approach to living systems is characterized by the interest in defining individual parts – modules, which create more complex elements (5). In terms of synthetic biology a part would be any DNA sequence fulfilling a certain role that can be preserved after such element is isolated from its background. This means that modules can be transferred to a new system and perform the same function in a predictable manner. Some examples of basic biological parts would be sequences that directly control gene expression like promoters (6, 7), terminators (8) and ribosome binding sites (RBSs) (9). Other basic elements can be involved in encoding protein or protein domains that have been characterized (10). The important feature of modular elements is that their activity is predictable when moved between different genetic backgrounds (11).

There has been a considerable effort put into characterizing as many biological parts as possible. Currently the biggest collection of elements characterized for use in synthetic biology is the Registry of Standard Biological Parts, which was founded in 2003 at the Massachusetts Institute of Technology (12). At this moment the registry contains approximately 9000 unique elements. This includes basic components (promoters, coding sequences) as well as complete devices consisting of multiple

individual elements. These complex elements were created to perform a specific function in the cells of their host e.g. biosynthesis of selected compounds or response to specific environmental stimuli. Information about individual collections of biological parts have been also published in the scientific literature (8). There are also standardized methods that facilitate assembly of basic components into more complex devices (13). It is likely that in time more collections of biological devices will be created in response to developments in the field of synthetic biology.

Biological devices

Assembled biological devices consist of multiple regulatory and effector parts (11). The fundamental criterion for a device is its ability to regulate a specific process within the cell. The simplest device would consist of a single operon controlling expression of a protein or a regulatory RNA. Such simple devices are often used as reporters that produce a visible output e.g. fluorescent protein. More complex devices might provide control over elaborate mechanism. They can also integrate input from several stimuli at the same time (Fig. 1). With the need to control different types of complex processes, most synthetic biology devices must include multiple individual elements.

Complex synthetic biology devices can give robust control over multiple genes that, together, encode a biochemical pathway (14). Commonly studied control devices include biological toggle switches and oscillators (15, 16). The interest in these types of devices stems from the requirement to control cellular activities that generate outcomes dictated by the investigator. Some examples of such outcome are biological systems that

act as biosensors, are used to deliver therapeutics or synthesize biofuels (17, 18). Some of the created synthetic biology devices are linked the cell's own metabolic network and allow control over its basic processes. The progress in development of genetic devices allows their incorporation into the genetic background of the host organism.

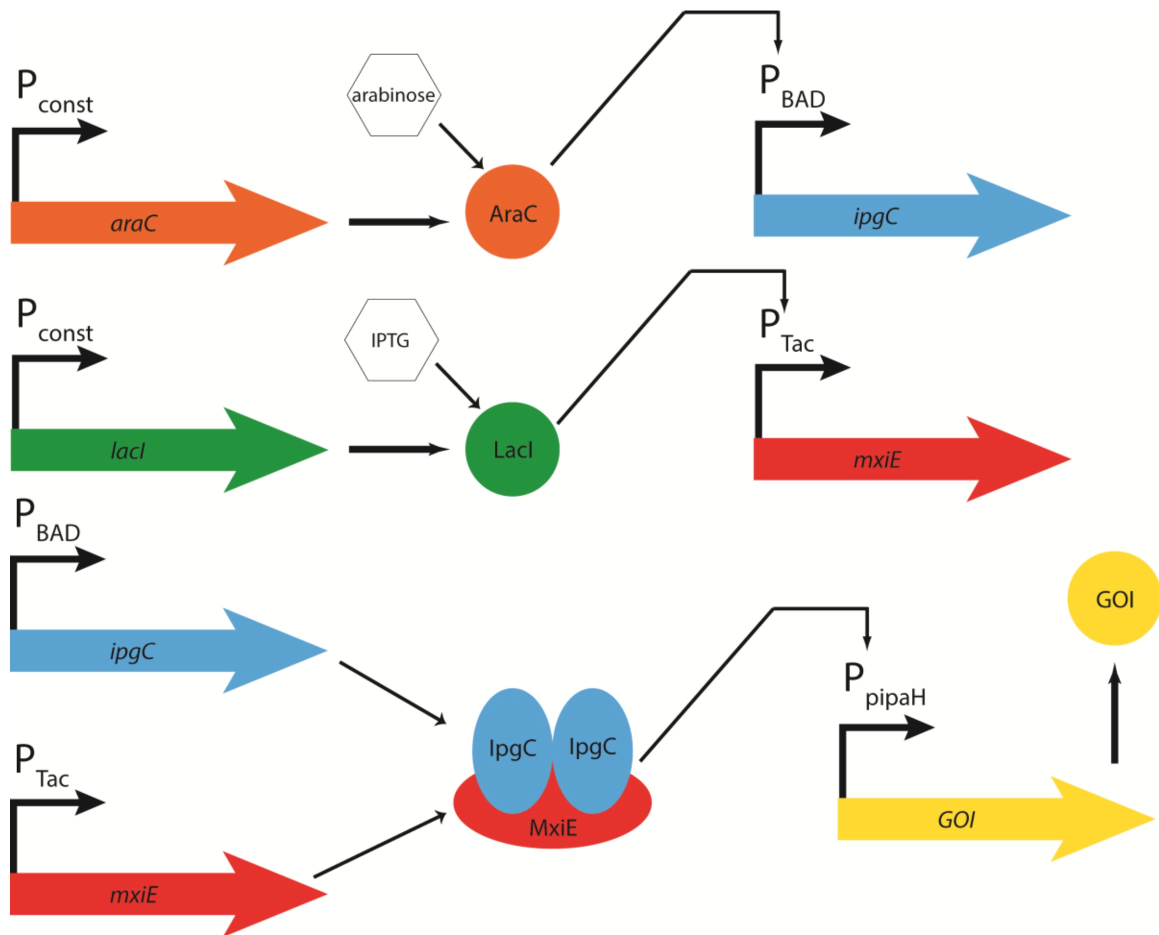

Figure 1 Example of a complex genetic circuit.

This schematic shows an example of genetic circuit. The presented device can produce an output, resulting in expression of a gene of interest (GOI), in response to the presence of two signals (arabinose and IPTG induction) recognized by the appropriate regulatory factors (AraC and LacI regulators). The regulators are constantly being expressed in order to detect the presence of their respective signals. Only when both signals are present at the same time the input is processed. This allows the expression of IpgC chaperone responsible for proper folding of MxiE transcription factor. That element controls the expression of a final product. This kind of genetic circuit has been presented to generate the predicted output (19).

Biosafety devices

Beginnings

In February 1975 a conference held at Asilomar State Beach in California discussed the possible hazards that might result from the recent development in biotechnology and regulations that have to be implemented for public safety. One of the discoveries that prompted this event was an experiment conducted by Paul Berg at the Stanford University. In this experiment, DNA of the monkey/human virus SV40 was recombined with the DNA of bacteriophage lambda. The obtained construct would be capable of replicating in the cells of *Escherichia coli* and easily generating many copies of the viral genome (20). This was recognized as a potential threat for human health. In response to this, the Asilomar conference suggested a number of means that could limit the possible danger caused by similar experiments (21). This includes necessity for containment facilities adequate to the risk caused by the subject of research. In addition to using appropriate physical barriers preventing the unwanted spread of modified organisms it was also recommended to use specific biological barriers. This includes use of laboratory organisms, such as for cloning purposes, which would be unlikely to survive in the natural environment. Another important element would be preparation of the cloning vectors capable of replicating only in a narrow range of host species. These measures were later implemented to minimalize the potential risks for humans and environment.

Current situation

The progress of biological sciences has led to development of more precise and high throughput methods of modifying the living organisms. This includes the possibility of synthesizing large fragments of DNA, even entire genomes (22). It also allows creating functional sequences that would not be found among the living organisms outside of the laboratory. These new methods give possibility of generating robust genetically modified organisms (GMOs) that could establish their presence in the natural environment and introduce the possessed sequences to the wild species. This gave a rise to a necessity of developing better methods of limiting the chances of survival of GMOs outside of the regulated conditions and to prevent the chances of gene transfer between them and the environmental species.

Limitation of both survival of laboratory strains and the horizontal gene transfer are the focus of newly developed means of regulating the growth of GMOs. The risk of horizontal gene transfer (HGT) comes from vectors that can move across species via natural transformation, transduction or conjugation. To address this issue, several mechanisms have been proposed. One approach is to create vectors that lack elements essential for their proper replication (23). These elements are transferred into the chromosome of the host strain, effectively limiting the replication of such vectors to the modified strain. Another possibility is to use conditionally lethal cassettes that are not dangerous to the host strain, but upon transfer to a new organism would kill it (24, 25). These means can be efficiently utilized to limit the chance of unwanted gene transfer from the GMOs.

Genetic biosafety mechanisms

Preventing GMOs from surviving in the natural environment requires modifications that allow the strain to propagate at the permissive conditions, such as in the laboratory culture, but prevents them from living at the restrictive ones. One way to achieve this, in similar fashion to the HGT prevention, is to use a controlled expression of a lethal gene. Under controlled growth conditions the lethal gene's expression is constantly being repressed by an external regulatory factor. Upon release to the environment this element is no longer present, which allows expression of the lethal gene (26). The other commonly used method to achieve this is creation of an auxotrophic mutant. Such a strain can be created by simply deleting a gene responsible for biosynthesis of specific compound that is necessary for survival. The modified strain can grow in the laboratory conditions when its growth medium is supplemented with the appropriate growth factors. However in the natural environment, it is highly unlikely that concentration of the growth factor would be sufficient to allow the mutant's propagation. One example of such a strategy is a *Salmonella typhimurium* strain lacking active aspartate β-semialdehyde dehydrogenase, which makes it incapable of surviving without the supplementation with diaminopimelic acid (27). Finally, instead of deleting the gene that is crucial for the organism's survival, it is possible to simply regulate its expression (or regulate the activity of the gene's product). For example, one could block the formation of an essential metabolite at certain conditions, thus controlling the strain's survival.

Recently an elaborate system for control of *E. coli* growth was described. Its aim was to generate a strain that would incorporate synthetic amino acids into its essential

proteins (28). Such amino acids could not be synthesized by the strain itself, thus making it dependant on exogenously delivered component. In this study, researchers have created a modified strain of laboratory *E. coli* having of its natural TAG stop codons replaced by either TGA or TAA, which act as alternative stop codons. Such a complex genome engineering procedure was possible thanks to the Multiplex Automated Genome Engineering (MAGE) approach (29), which allowed mutagenesis of multiple residues simultaneously. The changes in the codon pattern were followed by deletion of a release factor protein (RF1) that specifically recognizes the TAG codon. Normally, when the RF1 protein binds to its stop codon it causes the release of the newly synthesized protein and disassembly of the active ribosome. Since both stop codon and RF1 are gone from the genome of modified strain the TAG is no longer utilized by the host organism. However, once a heterologous tRNA molecule, capable of recognizing the TAG codon, is placed in the genome of the modified organism, the sequence can be used to encode amino acids. The introduced tRNA has to be paired with its own aminoacyl-tRNA synthetase that can load selected amino acid on the tRNA molecule. To test the efficiency of this approach residues within several essential *E. coli* genes have been replaced with TAG codons. Subsequently the codons became recognized by tRNA linked to p-acetyl-L-phenylalanine, p-iodo-L-phenylalanine or p-azido-L-phenylalanine. These non-biogenic amino acids are being incorporated into proteins crucial for survival of the modified strain. However since the strain cannot synthesize them on its own, it becomes an auxotroph for these compounds. A similar project has also successfully incorporated L-4,4’-biphenylalanine into several essential genes in the altered strain of *E. coli* (30). Such an approach proved to be highly efficient as the modified strain is very unlikely to mutate

into forms independent of the synthetic amino acid supplementation. However it requires far going changes in the genome of the host and therefore is unlikely to be implemented into a wide range of different organisms.

Essential genes

The idea of using essential genes as a part of genetic safety mechanism has gained interest in the last several years (31–33). In order to ensure the functionality of a kill switch in a particular host it is necessary to identify its own set of essential genes – the minimal genome. The initial search for essential genes can be helped by comparing the analyzed genome to the sequence of known essential genes. However ultimately it is necessary to experimentally validate the nature of such gene in the context of the host genome. Many research projects have been focused on establishing the minimal genome of representatives from different groups of living organisms.

Definition

Essential genes (EGs) are recognized as those that are necessary for survival of an organism. This means that loss of any essential gene would result in death or inability to reproduce. While some genes might be required for survival in a specific environment (e.g. an antibiotic resistance gene for culture containing that specific antibiotic), they are not considered essential. Such genes are not mandatory for an organism to prevail in the laboratory culture, where all crucial growth factors are provided. Loss of an EG might result in an immediate halt of important cellular functions e.g. loss of elements involved in transcription or translation. Alternatively the effect might be observed later if the gene was involved in cyclical processes like cell division (34). However the common effect of

EG loss is inability to produce viable offspring, whether it belongs to a uni- or multicellular organism.

Essential genes are responsible for maintaining basic metabolism. This includes such processes as DNA replication, DNA transcription, translation and biosynthesis of specific cofactors. Almost all proteins, whether working individually or as elements of multi-protein complexes, that facilitate these functions are encoded by essential genes. A defined set of essential genes within a particular organism is often referred to as a minimal genome. Finding minimal genomes of model organisms has become a focus of multiple research projects in the last decades. However since different groups of living organisms display varying level of complexity, the number of their EGs can be significantly different.

Minimal genome of prokaryotes

Many bacterial species have been investigated to evaluate the number of genes in their minimal genomes. It is estimated that approximately 300 genes in the genome of *E.coli* – a model prokaryotic, unicellular organism, are necessary for survival (35, 36). This number of EGs seems to be similar to that of other known prokaryotes, like *Bacillus subtilis* (37), *Helicobacter pylori* (38) or *Mycobacterium tuberculosis* (39). Interestingly that number tends to be very similar even between prokaryotic genomes that significantly vary in size, as the estimation of ~300 EGs seems to be true for both bacterium with smallest genome (0.58 Mbp) among free living organisms - *Mycoplasma genitalium* (40), and bacterium with one of the largest genomes (6.3 Mbp) – *Pseudomonas aeruginosa* (41). Some EG screening experiments suggest that the number of essential genes for

some of the bacterial strains might be around 600 – 700 (41, 42). However these research projects usually do not investigate each identified ORF deletion in details. Currently the most detailed analysis of essential genes has been performed in *E.coli*. The possible explanation of differences between observed results will be discussed in next chapter.

There are reports of prokaryotic organisms with less than 300 predicted genes in their genome, such as *Tremblaya princeps* (43), *Carsonella ruddii* (44) and many others (Table 1). However these organisms are obligatory symbionts and cannot live independently. Most of their essential metabolic functions are being fulfilled by a host organism or another co-symbiont. They also lack many known essential genes that are involved in replication, translation and biosynthetic pathways. It is often noticed that the dependency on the host factors makes these bacteria more similar to semi-autonomous cellular organelles, like mitochondria or chloroplasts, than free living bacteria. Therefore current understanding of prokaryote physiology strongly suggests that the minimal genome of typical bacterial species consists of approximately 300 genes.

Table 1 Characteristics of selected bacterial genomes.

Organism	Genome size (bp)	Number of genes	GC(%)	Ref.
Candidatus Tremblaya princeps*	138931	110	58.8	(43)
Candidatus Carsonella ruddii*	159662	182	16.5	(44)
Candidatus Zinderia insecticola*	208564	202	13.4	(45)
Candidatus Sulcia muelleri*	276511	246	21.1	(45)
*Candidatus Moranella endobia**	538294	452	44	(43)
*Buchnera aphidicola**	638852	593	25.6	(46)
Mycoplasma genitalium	580070	482	31	(40)
Mycobacterium tuberculosis	4411529	3924	65.6	(47)
Bacillus subtilis	4214810	4100	43.5	(48)
Escherichia coli	4639221	4288	50.8	(49)

The list represents some of the known species with the smallest genomes, as well as typical model bacterial organisms. Organisms marked with (*) are obligatory endosymbionts and cannot be cultured independently in laboratory conditions.

While multiple bacterial species were investigated to identify their essential genes, only a few members of *Archaea* have become a subject of similar analysis. In 2013, a research paper describing transposon mutagenesis based investigation of the genome of the methanogenic archaeon *Methanococcus maripaludis* was published (50). The results suggested that approximately 500 genes are essential for its growth. Earlier comparative analysis of 16 complete archaeal genomes revealed a conserved core of over 300 genes that were represented in all investigated species (51). These findings would suggest that the typical archaeal minimal genome is comparable in size to a bacterial one.

Minimal genome of eukaryotes

The number of EGs is significantly higher among eukaryotes. The model organism - *Saccharomyces cerevisiae*, which is a unicellular eukaryote, requires 1100 genes to maintain its basic functions (52). A similar number (~1200 EGs) has been proposed for another yeast organism – *Schizosaccharomyces pombe* (53). The significant difference between unicellular prokaryotes and eukaryotes is most likely related to the more complex construction of eukaryotic cell and cellular mechanisms that are unique to eukaryotic gene expression, e.g. splicing. Following this reasoning we would expect that the number of essential genes would be even higher among multicellular eukaryotes. Studies of *Caenorhabditis elegans* that focused on 86% of its 19,427 genes suggested that over 1700 of them are important for organism's fitness (54). In details, inhibition of approximately 270 of these genes was associated with defects in post-embryonic development (regarding e.g. movement or body shape). Inhibition of another 270 genes seemed to cause slowed or arrested development at post-embryonic stage. The loss of

functionality from the remaining genes caused death or sterility. This would imply that approximately 1500 genes are crucial for the survival of this species. Establishing the minimal genome for other multicellular eukaryotes is an ongoing project that could help with understanding the complexity of these organisms.

Identification

Identifying the essential genes in a particular genome requires a method of high-throughput analysis. Usually it relies on selecting mutants with individual genes inactivated. If one cannot isolate a clone lacking a functional copy of a certain gene, it strongly suggests that particular sequence might be essential. There are several high throughput methods of detecting essential genes that utilize this approach. The most commonly used ones (especially among simple organisms) are transposon mutagenesis, single ORF deletion and trapping lethal insertion mutants. Transposon mutagenesis seems to be the most commonly utilized approach. It is predominantly due to its relative simplicity as well as the fact that it can be used in various different organisms. The basic principle of this technique is to induce transposition of mobile genetic element containing selective cassette into the genome of investigated organism. This allows isolation of mutants with insertions in random genes. Usually at least 10,000 mutants are analyzed for the insertion site of the mobile sequence to ensure a proper coverage of the entire genome. The drawback of this system is the fact that certain transposable elements might have a preference for insertion in specific regions (55), which reduces the ratio of transposition to other parts of the genome. Subsequently this imposes a requirement to increase the amount of analysed clones in order to avoid missing non-essential genes.

Certain regions might not contain the transposon cassette not because they are crucial for the host, but because of random nature of transposition.

An alternative method, also relying on gene inactivation, is called trapping of lethal insertion mutants (56). It uses a population of sequences obtained from a known genome by cloning its fragments into a conditionally replicating vector. The collection of such sequences is introduced into the analysed organism, which is subsequently incubated at conditions non-permissive for vector's replication. This allows selection for clones that have the constructs integrated into their chromosome via a homologous recombination. This event is facilitated by the presence of a DNA fragment cloned from the genome. Integration of the vector should cause an inactivation of any gene in that region, just like transposition of selectable cassette. The advantage of this approach is the possibility of guiding the constructs into designated regions of the genome, thus ensuring that each gene can be targeted. However both of these methods share one common flaw. Many genomes are redundant and single function can be performed by multiple different genes. Therefore a loss of one gene can be complimented by the presence of another. On the other hand an insertion in an operon containing essential gene might have lethal effect even if the inactivated gene is not essential itself. This makes it harder to define all the essential genes within a single genome.

Utilizing essential genes in biosafety control mechanisms

Because of the crucial role of essential genes in maintaining the basic metabolism, they can be utilized as elements of genetic circuits focused on controlling the fitness and survival of an organism. Because of the development of genetically modified organisms

there is a desire to create mechanisms that provide control over life and death of generated strains. This would prevent unwanted spreading of such organisms, act as a safety switch or simply allow inactivation of cells as a step in an industrial process. Such an effect could be achieved by either inducing the expression of lethal genes or preventing the expression of a necessary one. Utilizing lethal or conditionally lethal genes would require insertion of a heterologous sequence into the genome of a modified organism and putting it under tight control to ensure expression in a response to specific stimulus. However this approach carries a risk of losing the introduced genes. Since they do not provide any advantage and often might be a burden, it would be favorable for the host to remove them from its genome. Therefore an organism that has lost inserted genetic elements would increase its fitness within the population. Because of this risk the control over essential genes seems to be a more reliable solution. Simple loss of essential gene itself would be lethal for an organism and therefore would be selected against. However it would still require a specific system for regulation of the target gene expression. This would usually involve systems that utilize regulation at the various levels of transcription or translation of chosen genes. Yet these systems are also likely to mutate within a large population. Additionally tight and precise control might be hard to implement without sufficient knowledge of the host's native regulatory mechanisms.

Designing an essential gene encoding protein that itself has feature allowing conditional inactivation would grant control over organism's growth without the need to create a complex expression system. It would mean that any special genetic circuits for regulation of expression are obsolete. Additionally, mutations in the coding sequence are much more likely to create a disadvantage for an organism and therefore be under

negative selection pressure. It would mean that the whole system is much more stable. Attempts to create such protein have been made with the recent development of modified *E.coli* strains that can utilize non-biogenic amino acids. Such strain can incorporate additional amino acids, ones that are usually not utilized during translation, into its proteins (30). This however means that an adequate synthesis of modified protein requires presence of these amino acids in the growth medium. If that modified protein is providing an essential function in the cell, then survival of the entire organism is dependent on the presence of externally provided amino acids. The downside of this method is the necessity of creating a specific strain of *E. coli* that has been adapted to utilize non -biogenic amino acids. On top of that it only allows control over its growth in response to a very specific chemical compound. A much more interesting approach would be to adapt a protein to react independently, without interactions with other cellular components, to a specific environmental signal and change its own features. A good example of this solution is the use of temperature-sensitive (TS) essential proteins.

Advantage of TS proteins is the possibility of using them in multiple different hosts. Since essential genes are conserved among many different, even distantly related, species it is possible to utilize the host's native expression regulators to control the production of introduced protein. The only necessity is to place the coding sequence of TS sequence under the control of regulatory elements responsible for expression of the wild type allele. Therefore there is no necessity to develop separate mechanisms allowing the precise control over protein biosynthesis, like promoters or ribosome binding sites (RBS). This means that even organisms with poorly understood regulatory elements can be modified to become temperature-sensitive. The only restriction in creating strains

relaying on TS essential genes is the efficiency of recombination in the organism of interest. This presents a much easier method or implementing biosafety mechanisms and it can be used in a broad range of host species.

The TS genes have been already successfully utilized as a mean of controlling the bacterial growth. One such example could be the creation of strains of pathogenic bacteria unable to grow at the temperature of human body (31). These modified bacteria can be used as live vaccine strains. In similar fashion, creating the temperature-sensitive strains of human viruses allows generation of much safer means of vaccination (57).

Examples of essential genes

Essential genes govern many metabolic pathways. This creates a variety of enzymes that can be utilized in genetic circuits that control the survival of a modified organism. In this work several essential genes have been successfully adapted in a mechanism regulating the fitness of altered prokaryotes and eukaryotes. The following chapters will briefly discuss the basic information regarding these genes. Emphasis has been placed on explaining the role that the genes' products play within the cellular metabolism. Additionally the presence of gene's homologues in different groups of living organism and structure of their protein products was mentioned. These chapters serve to help with placing the investigated elements within the complexity of metabolism of living organisms.

DNA ligase

DNA ligase is the enzyme that we have placed most focus on during the course of this work. Our observations suggest that temperature dependant inactivation of this enzyme can cause a rapid halt to the growth of a genetically modified organism. The results of investigating temperature sensitive alleles of DNA ligase are described in the Chapters 3 and 4.

Role

DNA ligase is an essential enzyme commonly found in all cellular organisms. The role of this protein is focused on forming the covalent bonds between the 5' phosphoryl and the 3' hydroxyl groups in breaks of the DNA backbone. Because of this activity, it plays an important part during DNA replication, recombination and repair. Loss of a DNA ligase activity can quickly cause formation of double stranded breaks in the DNA followed by an enzymatic degradation of the chromosome (58).

There are two types of DNA ligases that can be distinguished based on the cofactor utilized for reaction catalysis (59, 60). The first group consists of ATP-dependent ligases, mostly found among eukaryotes and archeons, as well as encoded by viruses and phages. The second group is classified as NAD-dependent ligases, which are predominantly found in bacteria. Research suggests that original ligases were utilizing ATP as a cofactor and the NAD-dependant variants appeared only in bacteria. A remarkable exception is DNA ligase from thermophilic archeon *Thermococcus fumicolans*, which can utilize both ATP and NAD as a cofactor (61). Additionally, while there are multiple cases of ATP-dependent ligases encoded by bacterial genomes, such

enzymes are non-essential and cannot complement lack of the NAD-dependent ligase (62, 63).

The most crucial role of DNA ligase in an organism is to create covalent links between short DNA fragments (Okazaki fragments) generated on the lagging strand during the replication. Mutations that cause loss or significant down regulation of this function are usually lethal to the host organism. It was proven that a single round of DNA replication in *E.coli* cells lacking sufficient DNA ligase activity is causing a lethal effect (58). This is caused by the generation of double stranded breaks in the DNA backbone, which arises from the DNA replication that uses nicked DNA as a template. The damaged DNA is then quickly degraded by exonucleases resulting in breakdown of the entire chromosome. While DNA repair systems can temporarily inhibit the lethal effect of DNA ligase deficiency, they cannot prevent it in the long run. This proves that DNA ligase plays crucial role in maintaining the integrity of the chromosome.

Organisation

The essential bacterial ligase contains several domains. The most N-terminal fragment is the Ia subdomain, unique for NAD-dependent ligases. It is involved in interactions with the cofactor. The next element is the adenylation domain, localized near the N-terminal end of the protein, just next to the Ia fragment. It contains residues responsible for recognizing the cofactor, forming the enzyme-cofactor complex and catalyzing the reaction. This domain is also highly similar to the corresponding region from ATP-dependent ligases. Next is the OB domain, which is partially involved in the catalytic activity, but mostly contributes as a DNA binding surface. Further elements are

zinc binding domain and HHH (helix-hairpin-helix) motif. These elements are believed to be responsible uniquely for binding DNA. The last part of the sequence encodes the BRCT domain. While this domain is often found among proteins responsible for the DNA repair and cell cycle checkpoints, its role in DNA ligase activity is unconfirmed. So far it has been shown that deletion of this domain does not stop the *in vitro* activity of the enzyme. It does however inhibit it significantly (64). This domain composition is highly conserved among bacteria and can be found in all essential NAD-dependant DNA ligases. All known NAD-dependent ligases are acting as monomeric proteins, 70-80 kDa in size. In summary the N-terminal region of NAD-dependant ligase is mostly responsible for catalytic activity and interactions with the cofactor, while the C-terminal part binds DNA.

The essential ATP-dependant DNA ligases are significantly more variable than their bacterial counterparts. The domain composition of these enzymes is not a highly conserved as the NAD-dependant ones. This might be caused by the fact that ATP-dependant ligases are found among greater variety of organisms and therefore have adapted to different conditions. Additionally ATP-dependant ligases are the only DNA ligases found among multicellular eukaryotes. The complexity of these organisms caused them to rely on several specialized types of DNA ligases (65). Among them ATP-dependant DNA ligase I is the one fulfilling the role most similar to the NAD-dependant ligase in bacteria. It is conserved in all eukaryotes and is involved in nuclear DNA replication, repair and recombination. In budding yeasts, like *Saccharomyces cerevisiae*, it is also involved in mitochondrial DNA replication and repair. The latter role is fulfilled by DNA ligase III in vertebrates, which also helps in nuclear DNA repair. The last type

of enzyme – DNA ligase IV, is conserved among all eukaryotes and plays role in nuclear DNA repair. Because of these specialized roles ATP-dependant ligases tend to be more variable.

The domain composition of ATP-dependant ligases is much less conserved than that of essential bacterial ligases (Fig. 2). Their size can vary between 30 and 100 kDa. All of them possess a catalytic domain and non-catalytic regions. The smallest ATP-dependant DNA ligases, like T7 DNA ligase, consist of only catalytic domain and OB domain. Other ligases from this group might possess additional domains, similar to those found in NAD-dependant DNA ligases. Additionally they often have cellular localization signals, like nuclear localization signal (NLS).

Despite the differences in composition all DNA ligases possess a set of conserved motifs involved in their catalytic activity. These motifs are named I-VI and are found in the catalytic region and the OB domain. Same motifs can be found among related nucleotidyl transferases like RNA ligases and GTP-dependant capping enzymes. While some of the motifs are poorly conserved (e.g. motif VI), especially between bacterial and eukaryotic enzymes, certain residues that are crucial for catalytic activity are preserved among all functional enzymes. Motif I is the most crucial one as it contains conserved lysine residue. That lysine is responsible for formation of the covalent bond with the cofactor, which is a necessary step in the early stage of reaction. Any mutations of this amino acid result in complete loss of activity and when introduce into essential DNA ligases, have lethal effect.

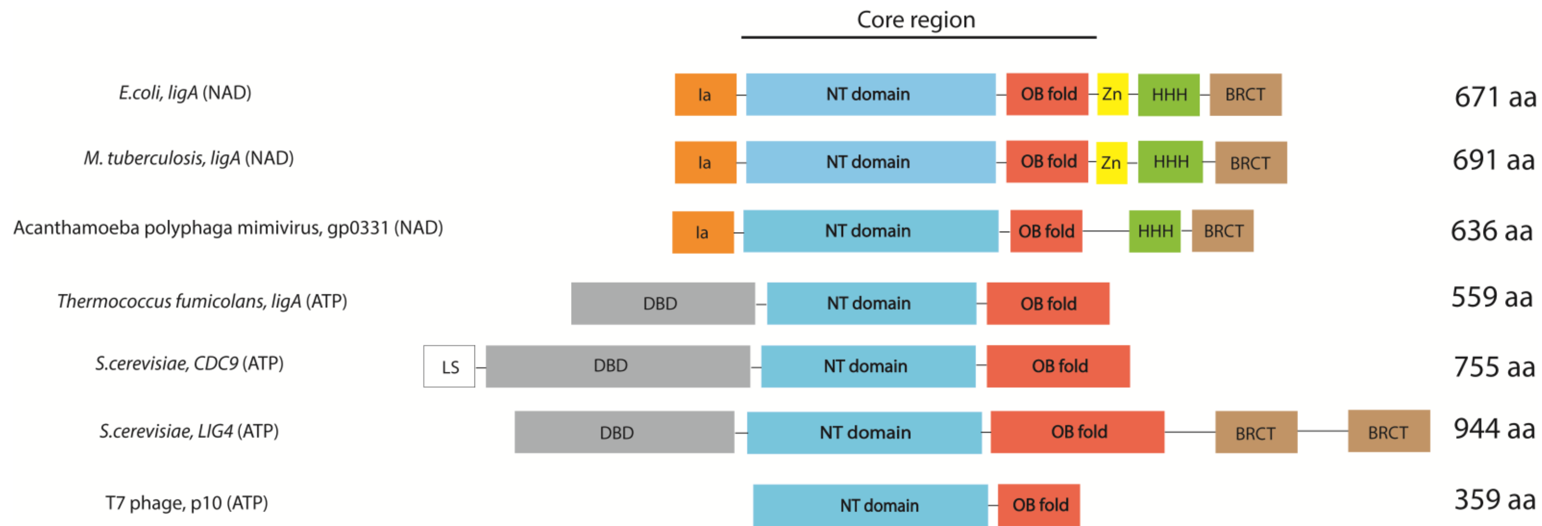

Figure 2 Organisation of DNA ligases in representatives of different domains of life.

The graph shows schematic organisation of DNA ligase regions found in viruses, bacteria, archeons and yeasts. The enzymes were differentiated based on their cofactor specificity (NAD or ATP). Individual domains have been labeled: Ia – NAD binding domain, NT – catalytic domain, OB – oligonucleotide binding domain, Zn – zinc binding domain, HHH – helix-hairpin-helix domain, BRCT – BRCT domain, DBD – DNA binding domain conserved among the ATP-dependant DNA ligases, LS – localization signal (short sequences that can direct protein to mitochondrion or nucleus). The conserved fragment present in each enzyme has been underlined. The size of each protein has been placed next to the graph.

Catalysis

The reaction of DNA ligation consists of three steps. The first one involves formation of enzyme-AMP intermediate. This event is facilitated by creation of a phosphoramidate bond between cofactor and the ε amino group of the conserved lysine from the motif I in the catalytic domain. Among NAD-dependant ligases this step involves cleaving NAD, which results in creation of two mononucleotides – AMP and nicotinamide. The ATP-dependant ligases break an ATP molecule into AMP and pyrophosphate. During the second step AMP particle is transferred from the enzyme onto the DNA backbone at the site of a single strand break. It ends with formation of bond between the free 5’-phosphate group of the DNA strand and the 3’-hydroxyl group from the cofactor. In the last step the AMP-phosphate bond is attacked by the 3’-hydroxyl group of the other DNA fragment, which results in connection of two DNA strands and release of free AMP.

In addition to the organic cofactor DNA ligases require presence of metal ions for proper activity. This includes divalent and monovalent ions. The divalent metals are bound close to the active site of the enzyme. Two such metal binding sites have been found in human DNA ligase I (66). Both of them are located in the catalytic site, close to the covalently linked AMP. Two glutamate residues seem to be responsible for interacting with the ions. It is believed that the presence of these metal particles plays a role in positioning the bound cofactor to allow the reaction with DNA backbone. Usually magnesium is used to fulfill that role in the *in vitro* conditions. However other divalent ions have also been proven to facilitate the enzymatic activity of DNA ligase. This includes ions of cobalt or manganese (67). On the other hand ions of heavy metals, like

zinc and cadmium had a strong inhibitory effect on the enzymatic activity (68). Additionally some monovalent cations, like Na^+ and K^+, have a inhibitory effect on DNA ligase (69, 70). Others, like NH_4^+ and Li^+, were proven to raise the reaction efficiency of DNA ligation when present at low concentrations (71). Therefore most of ligation reactions are supplemented with selected set of metal ions. This kind of ion dependency is common among different types of DNA ligases. The one known exception is the ligase from acidophilic archaeon *Ferroplasma acidiphilum* which requires two trivalent ions of iron per active enzyme and does not need magnesium or potassium (72). It is believed to be an adaptation to the extreme conditions of its environment. However despite some differences in preferred element the necessity of metal ions for enzyme activity is a shared feature of all DNA ligases.

Methionyl-tRNA formyltransferase

Investigation of methionyl-tRNA formyltransferase was a part of large scale screening for temperature sensitive alleles of essential genes that is described in the Chapter 6.

Translation initiation

Translation of information from mRNA occurs on the ribosomes. In bacteria the 70S ribosome consists of large 50S and small 30S subunit. The assembly of translation complex begins with association of the 30S subunit with mRNA and fMet-tRNA. These elements are then stabilized together and the interaction between the start codon on mRNA and anticodon on tRNA are formed. That complex then binds to the 50S subunit.

At the end of the process, thanks to the presence of protein initiation factors and changes in the conformation, the complete 70S ribosome is formed. During the translation acylo-tRNAs can interact with the three binding pockets located between the ribosome subunits. These pockets are named aminoacyl (A), peptidyl (P) and exit (E). Initially the tRNA molecule binds in the A site. During the translation tRNA changes its position, moving to another pocket region. The elongating peptide chain is attached to the bound acyl-tRNA while it moves to the P site. Later peptide chain is transferred to another tRNA and the used RNA particle is transferred to E site where it dissociates from the ribosome. Every acylo-tRNA involved in the protein synthesis goes through all three sites in that order. The only exception is the initiation tRNA that starts bound directly to the P site. The ribosome assembly is a multistage process that involves many components, both protein and RNA, and relies on specific interactions between these elements (73).

Initiation tRNA binds with the ribosome in a unique way. Positioning the initiation tRNA on the P site occurs in three steps called codon-independent binding, codon dependent binding and tRNA adjustment. Interactions between fMet-tRNA and the ribosome are facilitated by the initiation factor 2 (IF2). First IF2-GTP complex helps with binding initiation tRNA to the 30S subunit. The binding is then tightened by the interactions between the start codon on the mRNA and the anticodon in the loop of tRNA. At this stage the IF3 ensures proofreading of the intiation tRNA. What follows is assembly of the complete ribosome through interaction with 50S subunit. Finally fMet-tRNA adjustment is facilitated by the IF2 and hydrolysis of GTP. This causes changes in the position that tRNA takes on the P site. The ability of IF2 to specifically interact with fMet-tRNA is facilitated by the presence of the formyl group on the loaded amino acid

(74). The selectivity of interactions between IF2 and fMet-tRNA are an important factor for the correct initiation of translation.

Role of formyltransferase

Methionyl-tRNA formyltransferase (FMT) is an enzyme involved in modification of the amino acid methionine. During the catalyzed reaction a formyl group from N^{10}-formyltetrahydrofolate is transferred to the α-amino group of methionine. When the process of protein synthesis is complete the formyl group is usually removed from the amino acid chain through the activity of deformylase. The biosynthesis of fMet is a phenomenon limited to bacteria and eukaryotic organelles. There are no enzymes with similar functions in the cells of organisms from groups Eukarya or Archaea. Additionally only the methionine on tRNA involved in initiation of translation gets modified. The isoacceptor tRNAs utilized during the elongation stage carry non-modified amino acids. Therefore the role of FMT is limited to preparing the elements necessary at the beginning of the translation process.

Formyltransferase is important for the proper growth of bacteria. Early studies suggested that deletion of the *fmt* gene significantly reduces *E.coli* growth at 37°C and completely prevents growth at 42°C (75). This would suggest that while important the gene is not essential for the survival. Studies in other bacterial species suggested that disruption of *fmt* might be even less damaging than it was in *E.coli* (76). Furthermore several types of bacteria were capable of growing even in the absence of folate, which is necessary for the synthesis of cofactor delivering the formyl group for the reaction.

However more recent studies of essential genes in *E.coli* and related bacteria showed that *fmt* is essential at all conditions (31, 35, 77). Additionally genes encoding functional formyltransferase can be found among many different members of Bacteria group, suggesting important role in the cellular metabolism. This suggests that while in selected species the essentiality of *fmt* might be debatable it does provide indispensable function in the model organisms.

Tyrosyl-tRNA synthetase

Investigation of tyrosyl-tRNA synthetase was a part of large scale screening for temperature sensitive alleles of essential genes that is described in the Chapter 6.

Role

Aminoacyl-tRNA synthetases (aaRS) are a group of enzymes responsible for linking amino acids with their partner tRNAs. This reaction is necessary for the progression of translation. It allows recognition of each individual codon as carrying information about the selected amino acid. Since the translation mechanism is highly conserved in all domains of life, similar tRNA synthetases are found in all organisms. Majority of the living organisms possess 20 different aaRS, one for each biogenic amino acid (78). Each enzyme has to recognize specific tRNA molecules that carry codons for its amino acid and distinguish them from the other tRNAs. If a single amino acid is encoded by multiple codons then its aaRS will utilize all of the appropriate isoacceptor tRNAs. Since all tRNA types have similar secondary and tertiary structure the enzymes have to rely on less obvious differences to distinguish between them. This might be based

on a specific base pairs in the acceptor stem, different length of loops or the interactions with anticodon sequence (79). Additionally many tRNAs contain modifications that prevent recognition by a wrong aaRS e.g. lysidine (modified cytosine) in the $tRNA^{Ile}$ of *E.coli* (80, 81). The final mechanism ensuring accurate discrimination of amino acids is the hydrolytic activity that can remove misplaced residues. Many aaRSs possess a second active site that can facilitate such reaction. While some bacterial ssRSs do not have correcting activity they are complemented by separate enzymes with editing capabilities. Overall the aminoacyl-tRNA synthetases play a crucial role in facilitating an adequate progression of each translation.

Aminoacyl-tRNA synthetases catalyse the aminoacylation in two distinct steps (82). First the enzyme binds amino acid and ATP molecule. The amino acid is activated by adenylation, which requires hydrolysing ATP into AMP and pyrophosphate. Subsequently AMP molecule is being transferred onto the amino acid. In second step the amino acid part of created aminoacyl adenylate is moved to the 2' or 3' hydroxyl group of 3'-end of bound tRNA. The reaction involves release of pyrophosphate, which is later hydrolysed by an enzyme called inorganic pyrophosphatase. Since the hydrolysis is highly exoergic the whole process becomes energetically favourable. The initial adenylation step also requires presence of magnesium ions, like many similar reactions. Therefore the specific reaction of aminoacylation in addition to its substrates requires presence of specific cofactors and metal ions for its completion.

Structure

The aminoacyl-tRNA synthetases can be divided into two basic classes (83). The distinction is based on their structure and sequence similarity. First class consists mostly of monomeric enzymes. During the reaction they transfer the amino acid onto the 2' hydroxyl group of tRNA. They possess two conserved sequences in their active site: "KMSKS" (Lys-Met-Ser-Lys-Ser) and "HIGH" (His-Ile-Gly-His). Both motifs are found in the loop regions of their enzymes. The members of the second class are usually dimeric or tetrameric proteins. They are known to catalyse transfer of amino acid onto the 3' hydroxyl group of target tRNA. These enzymes have three motifs characterized by a weak consensus. Other differences between two classes are in the way they bind ATP and tRNA. Therefore both classes have very distinct features and structure. The tyrosyl-tRNA synthetase is a class one aaRS and operates as a monomer.

Porphobilinogen deaminase

The prophobilinogen deaminase has been analysed to investigate the capability of alleles of essential genes from the genome of *Colwellia* sp. C1 to generate a temperature sensitive phenotype (described in the Chapter 5). We specifically focused on this enzyme because of its involvement in the biosynthesis of biologically important cofactors. As such we expected that the inactivation of this enzyme would cause much slower cell death than expected from the products of other investigated essential genes.

Role

Tetrapyrroles are a class of chemical compounds that perform multiple important functions in living organisms. Among those molecules are porphyrins, chlorins and phycobilins, which have key role in metabolism. Their common feature is a structure of four pyrrole rings arranged in linear or circular fashion. One of the essential features of circular tetrapyrroles is their ability to chelate divalent metal cations such as Fe^{2+} or Mg^{2+}. Hemes with ability to chelate iron ions play crucial role as cofactors in electron transport complexes e.g. cytochromes. The cytochrome molecules contain different heme compounds as prostetic groups necessary for their activity. Therefore the presence of enzymes that are responsible for synthesising tetrapyrroles is required for many essential metabolic pathways.

Biosynthesis of tetrapyrroles occurs in multiple steps and involves several enzymes (84). The single pyrrole molecule, which forms one fourth of tetrapyrrol, is a heterocyclic pentamer consisting of four carbon atoms and one nitrogen atom. It also has two methine groups. However pyrrole is not an intermediate in the synthesis pathway. The precursor compound is 5-aminolevulini acid (ALA). In bacteria ALA can be synthesized in one of two alternative pathways. The Alphaproteobacteria do it by condensating succinylo-CoA and glycin in the Shemin pathway also known as C4 pathway. Other bacteria, such as *E.coli*, utilize the C5 pathway that starts from glutamate. In the first step of tetrapyrrole synthesis two ALA molecules are condensed to form a porphobilinogen. This reaction is catalyzed by porphobilinogen synthetase, encoded by *hemB* gene. Then four molecules of porphobilinogen are linked into a linear tetrapyrrole preuroporphyrinogen. At this point the created product contains four modified pyrrole

rings. The condensation of four porphobilinognes is facilitated by porphobilinogne deaminase (PBGD), encoded by *hemC* gene. Finally preuroporphyrinogen is converted to uroporphyrinogen III when the ends of linear molecule are ligated together into cyclic structure. This last step is helped by the uroporphyrinogen III synthetase, encoded by *hemD* gene. After that the pathway splits into two possible directions. One leads to synthesis of vitamin B_{12}, coenzyme F_{430} and heme d_1, while the other follows with formation of hemes and chlorphylls. The first three reactions and the enzymes that catalyze them – *hemB*, *hemC* and *hemD* – are common for all tetrapyrrole synthesis pathways.

Porphobilinogen deaminase requires the presence of its cofactor to catalyse polymerisation of porphobilinogen into preuroporphyrinogen. The specific cofactor for PBGD is a molecule of dipyrromethane covalently attached to protein through a tioether bond. Interestingly, this element is synthesized by its apoenzyme from two molecules of prophobilinogen. The cofactor serves as an attachment site for the polymerizing prophobilinogens. The substrate prophobilinogen is deaminated forming methylene pyrroline, which subsequently forms covalent bond with the cofactor. Next substrate ring is processed in exactly the same way, but if forms link with previously bound modified prophobilinogen. Finally the complete preuroporphyrinogen is released, which regenerates the original cofactor. Because of the unique nature of dipyrromethane cofactor the presence of substrate can directly affect the activity of enzyme.

Structure

Porphobilinogen deaminase acts as a monomeric protein with three distinct domains. All three domains have relatively similar size. The domains are connected with each other by hinge regions that allow high degree of flexibility. The dipyrromethane interacts with the cavity between domain 1 and domain 2, but is covalently bound to cysteine in domain 3 (85). There are three sites for binding pyrrole rings, two of them seem to be responsible for binding the cofactor while the third one interacts with the substrate. The porphobilinogen deaminase has a similar structure and organisation, even between distantly related organisms. It is believed that enzyme undergoes significant conformational changes during the catalysis of reaction. Additionally the presence of multiple cofactor binding regions helps stabilize the apoenzyme.

Peptidylt tRNA-hydrolase

Peptidyl tRNA-hydrolase is an essential gene used during the investigation of a plasmid based allele exchange system (described in the Chapter 7). This gene has been selected for the experiments because of its small size and universal presence among all of the groups of living organisms.

Role

Peptidyl tRNA-hydrolase is responsible for cleaving the peptidyl-tRNA bond on a stalled ribosome. This function is crucial, as during translation some ribosomes happened to halt the process resulting in inhibition of protein synthesis. Usually it occurs as a result of defective mRNA, amino acid starvation or insufficient level of tRNA. This process can possibly be lethal for the cells and therefore factors that can release and recycle the stuck

elements of translational mechanism are required. These factors include tmRNA, RF and EF proteins and enzymes such as peptidyl tRNA-hydrolases (Pth). The Pth enzyme catalyzes the hydrolysis of ester bond between the C-terminal end of peptide chain and the 2' or 3' residue of ribose in tRNA. This allows the release of tRNA, which can be later reused in another translation reaction (86). Therefore Pth provides important function that is essential for survival of an organism.

Currently there are two distinct types of Pth proteins known to be utilized by living organisms. The first group (Pth1) has been found in Bacteria and Eukarya. Those enzymes were the first discovered peptidyl-tRNA hydrolases. At the same time no elements with similar sequence were localized in the archaeal genomes. The Archaea domain possesses its own class of peptidyl tRNA-hydrolases (Pth2), which is distinct from the previously characterized enzymes in both sequence and structure. While it is believed that Pth1 class is specific to Bacteria and Pth2 is specific to Archaea, the Eukarya domain can possess representatives of both of these groups. Additionally eukaryotes might utilize other set of enzymes that fulfill similar role in their cells. The presence of Pth1 enzymes is essential to the survival of bacteria, however it has been proven that despite their differences the Pth2 enzymes can complement its loss (87). At the same time none of these two enzyme groups seem to be essential for the survival of eukaryotes (88, 89). Therefore it looks like there are at least two groups of protein with the activity of peptidyl tRNA-hydrolase, which are responsible for crucial functions in the cells of living organisms.

The loss of Pth activity in *E.coli* results in the cell death (90). The lethal effect is associated with accumulation of peptydyl-tRNA (91). Furthermore, the antibiotics that

cause rise in the cellular level of peptidyl-tRNA can increase the severity of observed phenotype. The most likely explanation for the observed lethality is the significant decrease of free tRNA level, which directly reduces the rate of protein synthesis. Specifically certain types of tRNA are more prone to accumulate in form of peptidyl-tRNA than others. Several fold increased expression of $tRNA^{Lys}$ is sufficient to supress the temperature-sensitive mutation of *pth* gen (92). At the same time higher level of other types of tRNA does not provide such effect. In line with these results the decrease in tRNA expression is causing, just like certain antibiotics, more severe effect of Pth depletion. Therefore it seems very likely that functional Pth is essential to maintaining a minimal level of free tRNA that is crucial for efficient protein synthesis and, as a result, the survival of the cell.

Organisation and catalysis

Two classes of identified Pth enzymes are known to have significant differences in terms of their structure. Members of the first class (Pth1) are usually bigger (~190 aa), while the proteins from the second class (Pth2) are on average smaller (~120 aa). While Pth1 enzymes are monomeric, at least some of the Pth2 proteins act as homodimers (87). Overall the Pth2 hydrolases seem to be more compact than the Pth1 ones. Both classes display α/β fold. The Pth1 fold consists of six α-helices and seven β-strands, while the smaller Pth2 has four α-helices and four β-strands (93, 94). The active site of Pth1 enzymes is believed to consist of histidine in α1 helix, aspartic acid in the loop between β5 - β6 strands and asparagine in α3 helix. At the same time the suggested catalytic site of Pth2 class consists of lysine in α1 helix and aspartic acid and threonine, both of them

in the loop between β3 – β4 strands. Two substrate recognition sites have been found in the Pth1 structure. They can bind the 5' and 3' ends of the tRNA respectively. The first one consists of conserved lysine and arginine residues in β6 and β7 strands, while the second one is a cluster of asparagine residues localized mostly in the loop regions. The alignment of Pth2 sequences does not reveal similar conserved elements. Despite both being involved in the same highly essential process, and their ubiquity among living organisms, both Pth classes have a distinct features.

Temperature-sensitive enzymes and psychrophiles.

In order to obtain alleles of essential genes that can be efficiently inactivated at elevated temperatures we decided to utilize the sequences found in the genomes of psychrophilic bacteria. The advantage of this approach is the possibility of generating a large number of unique alleles that have been adapting to low-temperature conditions for millions of years. Because of the cold adaptation the psychrophilic bacteria exhibit unique growth preferences as well as mechanisms allowing them to cope with their harsh habitat. Another advantage of using psychrophiles is the fact that they can be easily isolated from environments that constantly or periodically are exposed to cold (95, 96). On the other hand, the creation of temperature sensitive strains requires placing a heterologous gene in the mesophilic host. Therefore it is important to understand the genetic context from which this allele originates to ensure its proper expression.

Psychrophilic organisms

Psychrophiles are a specific environmental group of organisms that have adapted to grow at low temperatures. For that reason they are called cold-loving, or psychrophilic. The growth temperature of these organisms is usually in a range of approximately -20^{0}C to 20^{0}C. The bacteria capable of growing at the lowest temperature that would still allow their survival have been found in permafrost soil and sea ice. This kind of specific environmental adaptation can be found among various life forms, including bacteria, fungi and algae. However the greatest number of psychrophiles has been characterized in the domain of bacteria. Some of the well-known groups of psychrophilic bacteria include the genera *Shewanella*, *Pseudoalteromonas*, *Collwelia* and *Psychrobacter*.

Psychrophilic bacteria are represented by a diverse set of species. The most commonly found in low-temperature environment bacteria are gram negative, predominantly Proteobacteria (*Alpha*-, *Beta*- and *Gammaproteobacteria*), as well as members of Bacteroidetes (*Cytophaga, Flavobacterium and Bacteriodes*) phylum (96). The members of *Proteobacteria* group have been mostly found in deep waters, while *Bacteroidetes* seem to dominate in surface waters and sediments. The recognized psychrophilic Gram positive bacteria are mostly coryneforms, *Arthrobacter* sp. and *Micrococcus* sp. members (97). Other noteworthy groups are the various psychrophilic cyanobacteria – *Oscillatoria sp.*, *Phormidium sp.* and *Nostoc sp* (98). Psychrophilic microorganisms are represented by a variety of different groups of bacteria, which most likely cause a significant diversity of their specific features.

Psychrophilic adaptation

The psychrophilic prokaryotes have a number of specific traits that allow them to succeed in their specific environment. This includes synthesis of cryoprotecting agents, antifreeze proteins, more labile membranes and altered amino acid composition of their proteins (99–101). Combination of these features helps them to avoid lethal damage to the cell or inactivation of essential metabolic functions that would be normally caused by a low temperature encountered in their habitats.

The antifreeze proteins can bind to growing ice particles and slow the formation of crystal. This creates a phenomenon known as thermal hysteresis – situation when physical property of body depends not only on its current temperature but also on the previous observed temperature. In the case of antifreeze proteins it means that the freezing point of water is several degrees lower (102). This sort of cryoprotecting agents can be found in various different groups of organisms, from bacteria to vertebrates. Specific antifreeze proteins offer varying level of protection. Production of specific saccharides is an alternative way of protecting the intracellular environment from effects of low temperature (103, 104). Specifically high level of polysaccharides has been detected in some of the Antarctic marine bacteria, which could act as a possible low-temperature adaptation (104).

The cellular membranes of psychrophilic bacteria are characterized by a distinct composition of fatty acids. It has a higher content of unsaturated, polyunsaturated and methyl-branched fatty acids. It also possesses cis-isomeric acids, rather than trans-isomeric ones. Additionally the fatty acid chains tend to be shorter in the membranes of psychrophiles (97). This allows membrane to keep its fluidity even at low temperatures.

Otherwise the low permeability would hinder the transport of nutrients inside the cell. Another strategy developed by these organisms is the synthesis of carotenoid pigments and wax esters which help maintaining proper homeoviscosity (105). Some of these components can make up a significant part of the whole cell membrane.

The enzymes of psychrophilic bacteria are adapted to remain functional despite the decreased processivity at low temperatures. Because of the low temperatures chemical reactions occur at a slower pace. This can have a highly negative effect on the metabolism where multiple reactions occur at the same time and the ratio of reaction velocities between linked processes is essential. To counter that the psychrophiles utilize enzymes characterized by a less rigid conformation. With their structure being more flexible, they can efficiently catalyze their specific reactions. In return, however, they have become more unstable and heat labile. This suggests that while they can perform superbly at low temperatures, their activity at elevated temperatures would drastically drop (106).

Psychrophilic enzymes

Several characterized psychrophilic enzymes have been found to be temperature-sensitive (Fig. 3). This temperature-sensitive phenotype can be very useful for research. Possibility of inactivating specific proteins within the cell allows detailed investigation of its function. This is especially true for essential genes, which cannot be deleted without killing the organism. For a long time the temperature-sensitive mutants were obtained with mutagenesis, utilizing UV radiation and chemical agents (107–109). Instead of introducing cold-adapted variants of particular protein it relayed on randomly altering the

sequence of gene encoding it, in hope that the changed alleles would be less stable. Multiple such mutants have been characterized among eukaryotes, bacteria and viruses/phages. However this type of mutation is not easy to find, which limits the number of known temperature-sensitive alleles. Additionally, since most of them possess only a few mutations at most the chance for a mutation reversing the effect is quite high. Another significant downside of mutants obtained in such a manner is the fact that often they are characterized by an overall decrease in enzymatic activity at all temperatures. Psychrophilic enzymes, on the other hand, have evolved to display high efficiency at lower temperatures and should not suffer from inhibition under the permissive conditions.

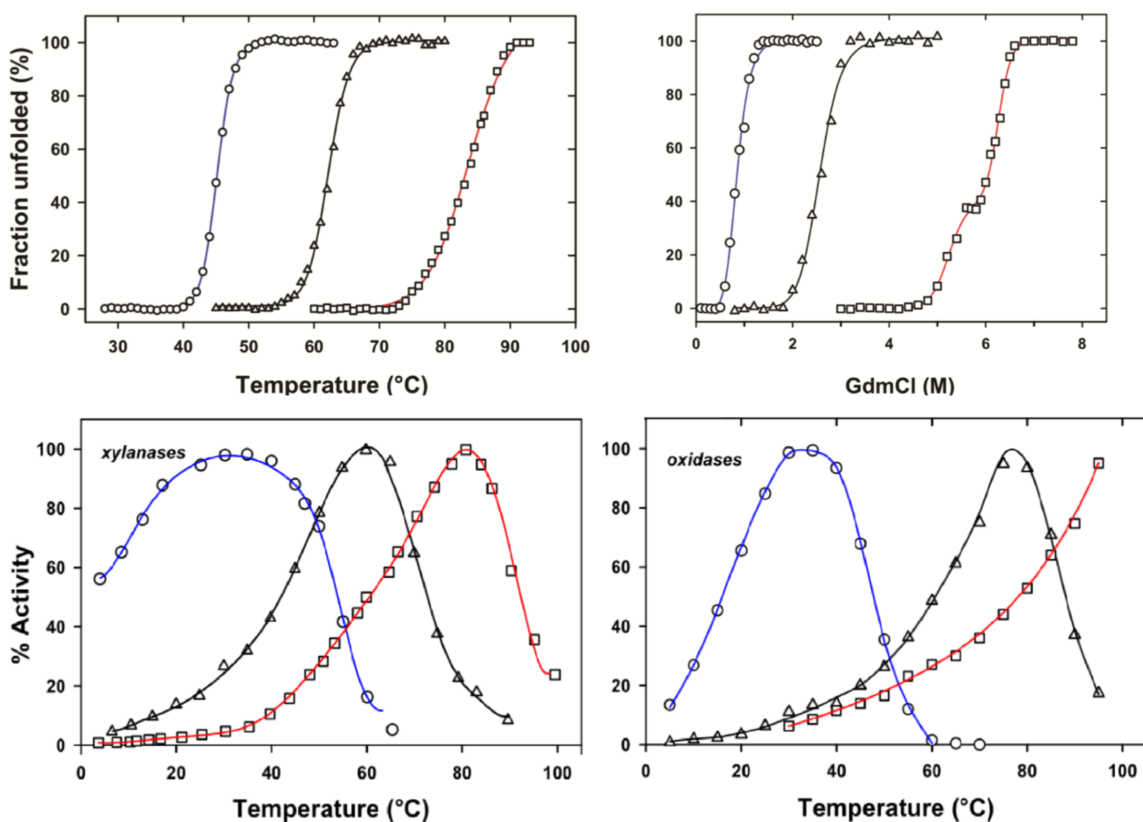

Figure 3 Properties of enzymes from different tropic groups of microorganisms.

(**Upper panels**) Measured sensitivity to temperature (left) and chemical agents (right) of psychrophilic (blue), mesophilic (black) and thermophilic (red) homologues of α-amylases. Data was obtained by fluorescence spectroscopy.

(**Lower panels**) Different enzymatic activity of psychrophilic (blue), mesophilic (black) and thermophilic (red) xylanases (left) and oxidases (right) across temperature spectrum. The maximum activity of cold-adapted enzymes is usually shifted towards lower values. Figures were modified from manuscript by Georges Feller (101).

Psychrophilic genomes

Psychrophilic bacteria have a few unique features of their genomes that distinguish them from other environmental groups of organisms. Among those elements are low GC content of their sequences, preferences for specific amino acids in the coding regions, multiple copies of tRNA genes and greater variety of encoded chaperon proteins (110). It is disputable if all of these factors are directly involved in adapting to cold environment. Additionally they are not strictly present among all of the cold adapted organisms (111). However these traits are commonly associated with the psychrophilic bacteria.

It is often recognized that psychrophilic bacteria have significantly lower than average GC content of their genome (112). One of the explanations for such state is the necessity to have less hydrogen bond in the DNA sequence. The adenine-thymine pair forms only two hydrogen bonds, while the guanine-cytosine pair has three hydrogen bonds. The weaker bonding of nucleotides would facilitate melting of double stranded DNA, which is necessary for many essential processes e.g. transcription and replication. Since psychrophiles live at lower temperatures it would make the strands of DNA less likely to dissociate. Therefore the number of GC bonds in psychrophilic genome is lesser than that in the genome of mesophilic organism. This idea is linked to the observation that genomes of thermophilic and hyperthermophilic bacteria tend to have higher GC content then the mesophiles. This would increase the stability of dsDNA at higher temperatures and prevent unwanted melting of double stranded structure.

Allele exchange

Overview

Allelic exchange is a method of swapping one form of a gene for another within the genome of a selected organism. This way it is possible to evaluate how changes in single coding sequence can affect the phenotype and molecular processes of the host. Specifically it could be used to search for an allele with certain properties from a pool of variable clones. This can be particularly interesting when looking for temperature sensitive genes or trying to exchange conserved genes between distantly related organisms. The known methods of allele exchange include swapping allele in its original locus through homologous recombination or replacing the original copy by one provided *in trans* on an autonomous vector.

BAC vectors and plasmid shuffling

Bacterial artificial chromosomes (BACs) are a specific class of vectors known for their stability and capability of supporting large DNA sequences. Such vectors are based on minimal forms of very specific plasmids such as F or P1 (113, 114). The unique aspects of these plasmids are low copy number (1-2 copies per cell) and presence of multiple plasmid maintenance systems. The low number of plasmid copies is maintained through a complex replication system involving iteron sequences (115). This system helps plasmid to stably replicate even when carrying a large genetic load, which has been used in molecular biology. Because of the low copy number plasmids cannot rely on random diffusion to both daughter cells during the division like some of the high copy

plasmids e.g. ColE1. Without a way to ensure stable transmission from generation to generation the plasmids would quickly be diluted in population. Therefore they utilize a mechanism that ensures a placement of at least one copy of plasmid in each cell. This crucial element was named partitioning system. It consists of two protein encoding genes (e.g. *parA* and *parB* on P1 plasmid) and a centromere-like sequence. During cell division the two proteins form a polymeric structure that interacts with the centromeric region on the plasmids. The polymer forms a pair of the plasmids and then places each individual plasmid molecule at one of the cell poles. Due to this process when the dividing cell separates, each offspring cell is ensured to have one copy of the plasmid (116). These systems have proven to be extremely efficient and allow stable propagation of plasmids for many generations even without any selective pressure. Usually replication system and partitioning system are placed together on synthetic vectors to ensure proper regulation of replication and stable maintenance.

Both replication and partitioning systems are strong incompatibility factors for the BAC vectors. Plasmid incompatibility is a phenomenon observed among closely related plasmids that share highly similar elements controlling their propagation. If two plasmids are incompatible they cannot be stably maintained in one strain without strong selective pressure. Without that the population splits into two strains each carrying one of the vectors. The molecular basis of incompatibility is caused by crosstalk between the plasmid systems of different vectors. Plasmids are often segregated into incompatibility groups, where plasmids that cannot be maintained together are placed in one group. There are approximately 30 groups of incompatibility among plasmids of gram negative bacteria and 7 groups among gram positive bacteria (117). For some of the low copy

plasmid types (e.g. F plasmid) the presence of incompatible vectors in one strain should cause separation into two kinds of clones after just a few generations.

Plasmid incompatibility can be used as a mean of exchanging one vector for another via plasmid shuffling. This process relies on two vectors with different selection markers but identical, and therefore incompatible, replication/partitioning systems. One vector is present in host strain and carries a specific gene of interest. Then the incompatible vector, with different allele, is introduced to the same host. Selection for clones carrying the introduced vector but not the old one can be made based on different selection cassettes. Low copy plasmids are ideal for this system since it is much easier for one of them to be lost. This method has been proposed as a way of analyzing sequences generated through PCR mutagenesis of essential genes (118). Similar solutions have been also applied to yeast genetics (119).

Chromosomal recombination

Recombination between a heterologous fragment of DNA and chromosome allows for the changing of native sequences *in situ*. Thanks to this a selected fragment of a DNA sequence can be inserted or deleted. It is also possible to exchange a specific piece of sequence for a foreign one. This includes swapping different alleles of the same gene. The introduced fragment of DNA is placed in the context of host genome. Therefore the native sequences can influence the expression of new element. This means that a coding region could be placed directly under the control of promoter and other regulatory sequences specific for a certain gene. The homologous recombination gives a

possibility of introducing new genetic components that become incorporated into host regulatory network.

Sequence manipulation requires presence of either functional host recombination system or introduction of a foreign one. Most of organisms possess their own mechanism facilitating recombination of DNA fragments. It is easy to find organisms among both prokaryotes and eukaryotes that are known for their capabilities of incorporating foreign DNA via homologous recombination. Typical examples include *Saccharomyces cerevisiae* yeast as an example of eukaryotic recombination and *Francisella novicida* or *Bacillus subtilis* as prokaryotic examples. However some of the laboratory strains have been rendered recombination deficient to increase the stability of genetic constructs that they carry. Additionally some of the organisms, which are particularly interesting from the scientific point of view, e.g. *Mycobacteria*, are known for a high frequency of nonspecific illegitimate recombination (120). However there is a possibility of introducing a heterologous recombination system into a selected host, which should generate a high recombination potential. Such systems are usually expressed from plasmids under a control of inducible promoters. This means that their activity can be directly regulated and when they are no longer necessary curing host from the plasmid can render the host recombination inefficient. The lambda recombination system is a commonly used element of that type. It allows high efficiency recombination with linear DNA in *E.coli* and related organisms (121, 122). Similar system has been developed for *Mycobacteria* (123, 124). Major limitation of these elements is the fact that their activity is usually limited to a narrow range of hosts. Ultimately the choice between different

approaches to *in vivo* recombination depends on the type of host that is used in experiments.

System for selecting recombinants is usually necessary to screen for the desired clones. The selection is simplified if the integrated DNA fragment encodes elements that are necessary for survival under specific conditions e.g. genes for biosynthesis can be selected for on minimal media. Otherwise the introduced sequence has to be combined with the selective cassette. Usually this role is fulfilled by an antibiotic marker. In the case when excision of selected DNA region has to follow the integrative recombination, a counter-selective marker is also added. This cassette is usually a conditionally lethal gene that can be easily selected against. It ensures selection for clones that underwent deletion of the specific sequence. However the presence of this selective cassette can be an obstacle since it can affect the gene expression of flanking sequences. Recently there has been a substantial effort into developing recombination methods that allow selection without antibiotic marker cassettes (121). The recombination methods allow selection of clones that possess the modified sequence in their genome as they were designed.

Chapter 2

Materials and methods

Strains and growth conditions.

Escherichia coli DH10B (Invitrogen) was used as the host for DNA manipulation experiments. The different *ligA* alleles were tested for the phenotype that they confer in a *Francisella novicida* restriction negative strain (125). To test the full length *ligA*-C1 allele we used the *S. enterica* ser. Typhimurium strain LT2 TT18389 (126) (gift from John Roth) as the host for recombinant *ligA*-C1 carried by a chloramphenicol resistant plasmid pBC SK+ (Agilent Technologies); the TT18389 strain has a chromosomal *ligA* partial deletion. The *Saccharomyces cerevisiae* YBSΔL1 strain with a deleted CDC9 gene, as well as plasmids expressing CDC9 were the kind gift of Dr. Steward Shuman (127).

The *ligA* alleles C2, P6 and S1 and the corresponding *F. novicida* strains carrying these alleles have been described previously (31), and were isolated from the ocean psychrophilic strains *Colwellia psychrerythraea* 34H (“C2”), *Pseudoalteromonas haloplanktis* TAC125 (“P6”), and our isolate of *Shewanella frigidimarina* (“S1”)(31). The other *ligA* alleles were isolated from ocean waters collected at the following GPS coordinates: C1, 82.32.04N, 62.45.86W; P3, 82°32N, 62°46.4W; S2, 48°20.26N, 123°36.24W; S3, 45°55,24’N, 129°59,48W; P4, 35°52.16 N, 114°39.73W; P5, 82°32N, 62°46.43W; P7, 64°47.9N, 168° 36W; S4, 48°20.26N, 123°36.24W. P1 and P2 were collected from the North Pacific Ocean at an unknown location (Fig. 4). The known

maximal growth temperatures for the psychrophiles that served as the source of the *ligA* alleles are C1 (12.5°C); C2 (18°C); P4/P5 (~30°C); P6 (18°C); S1 (27°C); S2 (22°C); S4 (24°C);

E. coli recombinants were grown at 37°C in Luria broth (128) supplemented with chloramphenicol (10 µg/ml), kanamycin (50 µg/ml) or tetracycline (10 µg/ml) as needed. *F. novicida* strains were grown in tryptic soy broth supplemented with 0.1% (w/v) L-cysteine (TSBC) and kanamycin (15 µg/ml) or sucrose (10% w/v) as needed. *F. novicida* strains harboring psychrophilic alleles were routinely grown at 30°C. *Saccharomyces cerevisiae* VL648 and YBSΔL1 were grown in YPAD (129) medium at 30°C. For selection of recombinants *S. cerevisiae* strains were plated on synthetic complete drop out (-Ura) medium or synthetic complete drop out (-Trp) medium (129).

Figure 4 Locations where psychrophilic strains were isolated.
Map of the Canadian Shield and surrounding regions. The markers point to the approximate locations where the strains of psychrophilic bacteria were isolated, (1) location where strains C1, P3 and P5 were isolated, (2) location where strains S2, S3, S4 and P4 were isolated, (3) location where strain P7 was isolated.

Cloning and insertion of psychrophilic *ligA* alleles into the *F. novicida* chromosome.

Partial genomic sequence was available for several unidentified psychrophilic bacteria, and this allowed the identification of some of the *ligA* alleles. For strains with no available genomic sequence data we assumed that many of our psychrophilic isolates were in the *Colwellia*, *Pseudoalteromonas* and *Shewanella* genera and we designed primers to amplify the *ligA* alleles from these bacteria. The PCR products of the appropriate size were partially sequenced to identify those that contained the *ligA* ORF. To prepare *ligA* alleles for insertion into *F. novicida* we needed to join them to DNA that corresponded to regions of the *F. novicida* chromosome that flank *ligA*. Each *ligA* allele was amplified with primers that had overlapping regions to the *F. novicida* flanking regions, and PCR products corresponding to the *ligA* allele, the flanking regions, and a yeast cloning vector were seamlessly assembled using transformation assisted recombination in *S. cerevisiae* VL648 according to the procedure described by Geitz and Schiestl (130). The vector used for this assembly was a derivative of pRS426 (131) that had the ampicillin marker replaced with genes encoding kanamycin resistance and sensitivity to sucrose (*sacB*). This plasmid can replicate in *S. cerevisiae* and *E. coli* but not in *F. novicida* (Fig. 5).

Once the *ligA* alleles were joined to the *F. novicida* flanking regions and sequence verified, the recombinant plasmids were amplified in *E. coli*, isolated, and used to transform chemically competent *F. novicida* (132). The exchange of the psychrophilic *ligA* for the *F. novicida* homologue was accomplished via a Campbell-like integration followed by excision as previously described (31). In essence the transformation step

was used to select for the integration of the plasmid; after co-integrants were identified the merodiploids were cultured in the presence of 10% sucrose and plated on agar medium with sucrose (see schematic in Fig. 6). Colonies were screened for sensitivity to Km and temperature-sensitivity. The region including approximately 0.5 kb upstream and downstream to the substituted *ligA* allele was PCR amplified and sequenced.

Cloning and insertion of other psychrophilic alleles into the *F. novicida* chromosome.

Alleles of *hemC*, *fmt* and *tyrS* genes were identified in partial genomic sequences of psychrophilic bacteria. Individual sequences were PCR amplified and assembled with respective flanking regions from the genome of *F. novicida* in the same manner as it was done for the alleles of DNA ligase. In order to facilitate allele exchange in the host chromosome the same vector pRS426KmSacB was used. Selection on kanamycin followed by counter-selection with 10% sucrose helped with discovering clones characterized by a temperature sensitive phenotype.

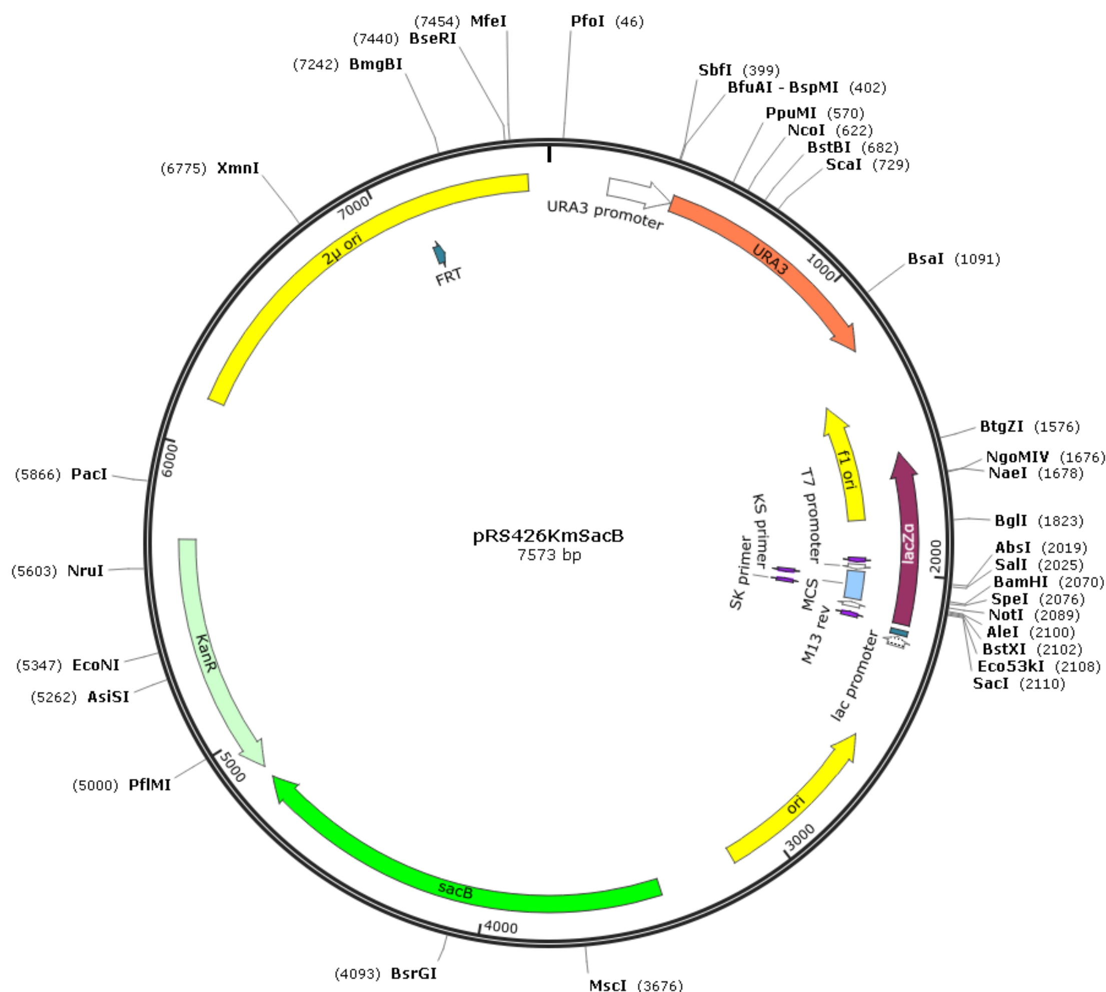

Figure 5 Map of vector pRS426KmSacB

Modified *S. cerevisiae* shuttle plasmid pRS426 designed for integration-excision experiments in *F. novicida*. The pRS426KmSacB plasmid was used to assemble DNA fragments in *S. cerevisiae*. When transformed into *F. novicida* the kanamycin resistant marker was used for selection of integration of the plasmid. The *sacB* cassette was subsequently used for counter-selection when enriching for excision events.

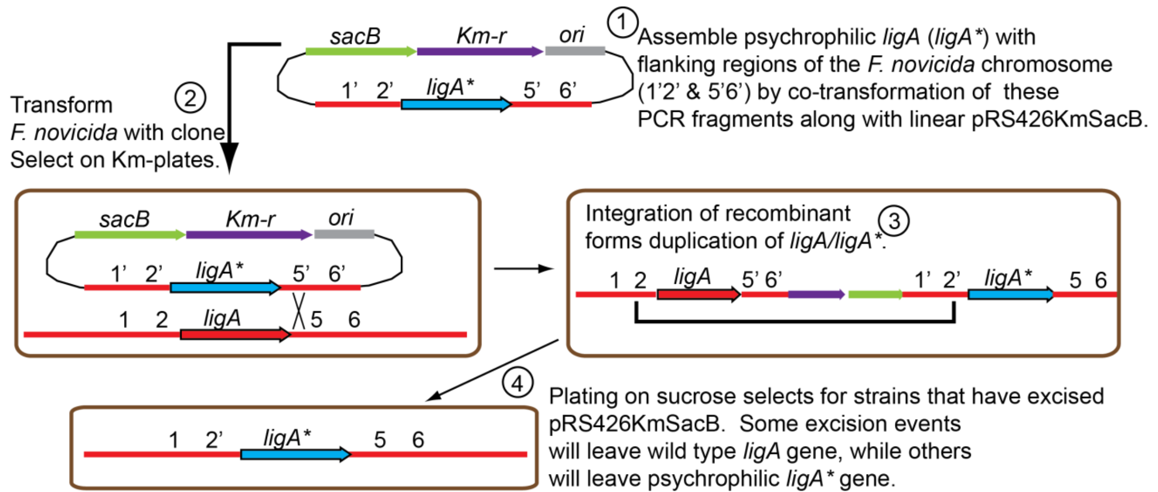

Figure 6 Creation of constructs with psychrophilic alleles.

Transformation assisted recombination-based assembly (step 1) of psychrophilic *ligA* with *F. novicida* flanking regions in *Saccharomyces cerevisiae*, and integration-excision based gene replacement in *F. novicida* (steps 3 and 4).

Determination of restrictive temperatures of *F. novicida*-TS strains.

We defined the restrictive temperature as the lowest temperature that failed to support the formation of isolated colonies on agar medium. To determine this temperature we spread ("streaked") each culture on agar plates, which were buried in aluminum beads in order to buffer the temperature fluctuations of the incubator. The variation of the temperature of the beads was monitored with a Fluke 53II thermometer which is accurate to 0.2°C. Even with all of our efforts to control the temperature of the petri plates, our stated restrictive temperatures should be considered accurate to about ±0.5°C.

Reversion frequency determination.

TS *F. novicida* strains were grown in TSBC at 30°C until the A_{600} was greater than 2. The cultures were plated without dilution on agar medium which was incubated for 48h at temperature 3°C above the restrictive temperature of the strain being studied. In parallel the culture was diluted to determine the number of CFU of the culture at its permissive temperature. The reversion frequency was calculated from the number of colonies that appeared at the restrictive temperature divided by the number of CFU in the cultures. These determinations were done at least three times with independent cultures.

Temperature shift experiments.

F. novicida strains were grown at 25°C in TSBC overnight, diluted 1:50 (4 µL in 196µl of medium) and inoculated in quadruplicate into 96-well plates. A BioTek ELx808IU™ microplate reader was used to incubate and agitate the plates at different temperatures, and automatically record changes in the optical density. Cultures were initially incubated at 26°C for 2 h after which the temperature was raised to a target restrictive temperature.

Yeast growth experiments.

The yeast strains carrying different alleles of DNA ligase (CDC9, P2, P7, *F. novicida*) were grown in synthetic complete drop out (-Trp) medium overnight. Cultures were then diluted in fresh medium in order to generate an A_{600} of 0.2, and incubated with shaking at selected temperatures for 12 hours. Every 2 hours the A_{600} of the cultures was measured. Each time the cultures were diluted to an A_{600} value 0.2 – 0.8 for accurate measurements and the culture density was calculated based on the dilution factor. For each strain three independent cultures were analyzed at each temperature.

Colwellia sp. C1 DNA sequencing.

The *Colwellia* sp. C1 genomic DNA was isolated using Qiagen's Spin Columns and Genomic DNA Buffer Set. We have prepared sample at a concentration of 100ng/µl. The quality of sample was confirmed on agarose gel and with Qubit DNA Assay Kit. The sequencing has been performed using MiSeq sequencer at Illumina. Obtained fragments were assembled into contigs using ABySS 1.3.1. software.

Colwellia sp. C1 sequence annotation and submission.

The obtained *Colwellia* sp. C1 readings were annotated using Rapid Annotation of microbial genomes using Subsystems Technology (RAST) method (133). The sequences have been submitted to GenBank with the accession numbers KT428294 and KT428295.

GenBank Accession numbers.

The sequences of the psychrophilic *ligA* alleles as they exist in the *F. novicida* chromosome have been submitted to GenBank with the accession numbers C1 [KR154478]; C2 [HM003389]; P1 [KR154476]; P2 [KR154477]; P3 [KR154481]; P4

[KR154482]; P5 [KR154483]; P6 [HM003396]; P7 [KR154484]; S1 [HM003397]; S2 [KR154479]; S3 [KR154480]; S4 [KR006256]; S5 [KR864903]. The C1 gene formed a hybrid when inserted into the *F. novicida* chromosome, thus the native C1 complete allele was submitted separately and assigned [KR818907].

Generation of *E. coli* ∆pth strain.

The *pth* gene deletion was performed in *E. coli* strain GB05. Deletion cassette consists of chloramphenicol cassette amplified form vector pBC SK+ and 500bp region upstream and downstream of the target gene. Fragments were cloned in pBluescript vector using Gibson assembly. Amplified linear cassette was transformed into competent *E. coli* cells that were expressing λ recombination proteins for recombination. The recombination strain carried a vector expressing functional *pth* allele to complement the chromosomal deletion.

Assembly of vector shuffling system.

Both vectors for the shuffling system were based on pBeloBAC11 plasmid (New England Biolabs). We created two alternative forms of this construct, with different antibiotic markers, one with kanamycin cassette – pBACK, and one with tetracycline cassette – pBACT. The vectors were made by PCR amplifying the whole plasmid with the exception of its native chloramphenicol resistance cassette. We also amplified Tet resistance marker from plasmid pJB86 and Km resistance marker from plasmid pMP814. The primers used for the marker amplification has an overhang regions homologous to the ends of pBeloBAC11 backbone. The fragments were cloned using method of Gibson Assembly (134).

Cloning the pth sequence into pBeloBAC11 derived vectors.

We have PCR amplified the pBACT and pBACK vectors using primers binding close to the MCS region. The cassette containing the *pth* coding sequence from *Salmonella enterica* under the control of lac promoter was amplified with primers containing regions of homology to the vectors. The fragments were put together using Gibson Assembly (134). The correct constructs were confirmed by PCR with primers flanking the cloning site.

Primers.

Oligonucleotide primers used in this work are described in the table below.

Table 2 Primers used in this study.

Name	Sequence	Role
Primers used to generate pRS426KmSacB integration vector		
KmSacAMPF	TTCAAATATGTATCCGCTCATGAGATCGTGAAGAA GGTGTTGCTGACTCATACCAGG	forward primer for amplification of KmSac cassette
KmSacAMPR	AAAGGATCTTCACCTAGATCCTTTTAATACTTGAAA AGATTATAAATATGCCCATCTAGTAAGTAGTC	reverse primer for amplification of KmSac cassette
primers used to create allele exchange constructs for *Francisella novicida*		
pRS426RevF	TCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTG	forward primer for pRS426 amplification
pRS426RevR	AGTGAGTCGTATTACGCGCGCTCACTG	reverse primer for pRS426 amplification
U112ligAFR1	TGGAGTCATTAAAGAATATTTTTAAATAATTATTT TG	reverse primer for upstream *ligA* flank amplification
U112ligAFF1	GTGAGCGCGCGTAATACGACTCACTCTCGAGACTA TCGCCAACGAGGATAC	forward primer for upstream *ligA* flank amplification
U112ligAFF2	TGAATATCAATAACTTAAAAAAGGGCGATGTAG	forward primer for downstream *ligA* flank amplification
U112ligAFR2	CTCACTAAAGGGAACAAAAGCTGGACTCGAGATCC CGACATCAAAGTACTAAGAAC	reverse primer for downstream *ligA* flank amplification
ligAC1F	TTAAAAATATTCTTTAATGACTCCAGCCTCATCAA TAAGCGCTTTACAAG	forward primer for *ligA* allele C1 amplification
ligAC1R	CCCTTTTTTAAGTTATTGATATTCAGTTATATTTT TCCAATAAAGCAACTAAGTCATCTTC	reverse primer for *ligA* allele C1 amplification
ligAS2F	TTAAAAATATTCTTTAATGACTCCAATTGAACTTG AAATTACTGAACTCAAAAATGAACTTAAC	forward primer for *ligA* allele S2 amplification
ligAS2R	CCCTTTTTTAAGTTATTGATATTCATGACAAAATG GCAATCAAGCCCTCTTC	reverse primer for *ligA* allele S2 amplification

ligAS3F	TTAAAAATATTCTTTAATGACTCCAATTCAAACTG AAATGGATCAACTTACTCACACCATTAATC	forward primer for *ligA* allele S3 amplification
ligAS3R	CCCTTTTTTAAGTTATTGATATTCAGTTAGCCGCA TTCAATAGATCCATCAG	reverse primer for *ligA* allele S3 amplification
ligAS4F	TTAAAAATATTCTTTAATGACTCCAATAGAAGAGA TCAAACAACTAACCGACATAC	forward primer for *ligA* allele S4 amplification
ligAS4R	CCCTTTTTTAAGTTATTGATATTCATTCTGAAGCT GAAAGTAGTTGTAAC	reverse primer for *ligA* allele S4 amplification
ligAP1F	GCAAATTAATCATCTTCGTACTACATTAGAGCAGC ACAATTACAATTATTATG	forward primer for *ligA* alleles P1 and P2 amplification
ligAP1R	CCGTGAGTATGTCGATACCTAAATCTTGAGCCTTT GTGAGTTTAG	reverse primer for *ligA* alleles P1 and P2 amplification
ligAP3F	TTAAAAATATTCTTTAATGACTCCAAGCATTAGTG AGCAAATTAATTATCTTCGTACTAC	forward primer for *ligA* allele P3 amplification
ligAP3R	CCCTTTTTTAAGTTATTGATATTCAACCATTATGC TTTTGAAGCAGCTTAATTAAC	reverse primer for *ligA* allele P3 amplification
ligAP4F	TTAAAAATATTCTTTAATGACTCCAAGCATTAGTG AGCAAATTAATCATTTTC	forward primer for *ligA* allele P4 amplification
ligAP4R	CCCTTTTTTAAGTTATTGATATTCAACCATTATGC TTTTGAAGCAGTTCTATTAATTC	reverse primer for *ligA* allele P4 amplification
ligAP5F	TAATGACTCCAAGCATTAGTGCAATTAATTATCTT CGTACTACATTAG	forward primer for *ligA* allele P5 amplification
ligAP5R	TTTTGAAGCAGCTCAATTAACTCATCTTCCGTGAG TATCTC	reverse primer for *ligA* allele P5 amplification
ligAP7F	TTAAAAATATTCTTTAATGACTCCAAGCATTAGTG AGCAAATTAATCATCTTCGTACTAC	forward primer for *ligA* allele P7 amplification
ligAP7R	CCCTTTTTTAAGTTATTGATATTCAACCATTATGC TTTTGAAGCAGCTCAATTAATTC	reverse primer for *ligA* allele P7 amplification
ligARev1	ACTAATGCTTGGAGTCATTAAAG	flank amplification primer for *ligA* P5 allele
ligARev2	TTAATTGAGCTGCTTCAAAAG	flank amplification primer for *ligA* P5 allele
ligARev3	TGTAGTACGAAGATGATTAATTTGCTCACTAATG	flank amplification primer for *ligA* P1 and P2 alleles

ligARev4	GATTTAGGTATCGACATACTCACGGAAGATGAATTAATTG	flank amplification primer for *ligA* P1 and P2 alleles
hemCLFL-1	GTGAGCGCGCGTAATACGACTCACTCTCGAGGTGTTGCAGTAGGACAGTTTTG	forward primer for *hemC* flank amplification
hemCLFR-1	GACTAGCTATAGTTATTTGTTTCATATGGGGATGCAGATGCG	reverse primer for *hemC* flank amplification
hemCRFL-1	TGCATATAAAATATTGGAGAGTTAAGGTATGATGGCTGCATTCTTG	forward primer for *hemC* flank amplification
hemCRFR-1	CTCACTAAAGGGAACAAAAGCTGGACTCGAGACAGGAGCATTAGTTGGAGC	reverse primer for *hemC* flank amplification
U112tyrSFR1	GCTCGACTTATATTTATAATATTAAAAATTTC	reverse primer for upstream *tyrS* flank amplification
U112tyrSFF1	GTGAGCGCGCGTAATACGACTCACTGTCGACCTCAGTTACCATCCGGAAATATTTTAG	forward primer for upstream *tyrS* flank amplification
U112tyrSFF2	ATAAAATAAGGGGTTGTAAGATGGATAAG	forward primer for downstream *tyrS* flank amplification
U112tyrSFR2	CTCACTAAAGGGAACAAAAGCTGGACTCGAGTTGCATGATAACCATCCGGAAATATTTTAG	forward primer for downstream *tyrS* flank amplification
1011tyrSF	TAATATTATAAATATAATGTCGAGCTTAGAGCAAGCATTAGCAGAAATTAAAC	forward primer for *tyrS* allele amplification
1011tyrSR	TCCATCTTACAACCCCTTATTTTATGGTGATTTTAGCGAATTTACGCTTAC	reverse primer for *tyrS* allele amplification
U112fmtFR1	TTTTTTCATTTACCTAATCTCTAAAAGCGTATCCACGTAATTG	reverse primer for upstream *fmt* flank amplification
U112fmtFF1	ACGACGGCCAGTGAGCGCGCGTAATACGACTCACTGTCGACCAACCTTAACTACTTGAAATGGCAAGCTAATT	forward primer for upstream *fmt* flank amplification
U112fmtFF2	CTAGGATAAATAATGAAAGTTGGATTTATAATAG	forward primer for downstream *fmt* flank amplification
U112fmtFR2	GCAATTAACCCTCACTAAAGGGAACAAAAGCTGGACTCGAGGCTGCAACAAGAAACTCATC	reverse primer for downstream *fmt* flank amplification

911fmtF	TTTAGAGATTAGGTAAATGAAAAAATTAAACGTTA TTTTTGCTGGTACACCGGATTTTG	forward primer for *fmt* allele amplification
911fmtR	ATCCAACTTTCATTATTTATCCTAGGCGTGTGCCA GGCGTAAACC	reverse primer for *fmt* allele amplification
primers used to clone the DNA ligase allele into yeast shuttle vector pYX132		
YXP2F	AAAACACATACAGGAATTCAATGACTCCAAGCATT AGTGAGCAAATTAATC	forward primer for *ligA* alleles P2 and P7 amplification
YXP2R	GCGTAAGCTTGTGGGCCCTATCAACCATTATGCTT TTGAAGCAGCTCAATTAATTC	reverse primer for *ligA* alleles P2 and P7 amplification
YXU112F	AAAACACATACAGGAATTCAATGACTCCAAACGAA TTTTTTTCTATAAAATATC	forward primer for Francisella *ligA* allele amplification
YXU112R	GCGTAAGCTTGTGGGCCCTATCATAATAAATCTTT CAGATTATCTTCAAGAATAAC	reverse primer for Francisella *ligA* allele amplification
YXcdc9F	AAAACACATACAGGAATTCAATGCGCAGATTACTG ACCGGTTG	forward primer for yeast CDC9 allele amplification
YXcdc9R	GCGTAAGCTTGTGGGCCCTACTAATTTTGCATGTG GGATTGGTTTTCATAC	reverse primer for yeast CDC9 allele amplification
primers used to amplify the *Francisella novicida* DNA ligase to complement the TS mutants		
ligAcomplF	GCGAAATCCCGACATCAAAG	forward primer
ligAcomplR	ACTATCGCCAACGAGGATAC	reverse primer
primers used to amplify the *Francisella novicida hemC* allele to complement the TS mutant		
hemC_recF	TGGGTACATATCGCAGATCC	forward primer
hemC_recR	CAGCACCAACTGCTGTAAAG	reverse primer
Primers used to create BAC constructs for allele exchange		
pBelo_backF	CCAGCTGAACGGTCTGGTTATAGG	forward primer for pBeloBAC11 backbone amplification
pBelo_backR	TGCCCTTAAACGCCTGGTTG	reverse primer for pBeloBAC11 backbone amplification

tet_BAC_F	ACCTATAACCAGACCGTTCAGCTGGCATAATATGT CCACCAACTTATC	forward primer for tetracycline cassette amplification
tet_BAC_R	CGTAGCAACCAGGCGTTTAAGGGCATGCCGGGAAG CTAGAGTAAG	reverse primer for tetracycline cassette amplification
km_BAC_F	CGTAGCAACCAGGCGTTTAAGGGCAAACAAATAAA AACGCAAAAGAAAATG	forward primer for kanamycin cassette amplification
km_BAC_R	ACCTATAACCAGACCGTTCAGCTGGTGCAAGAGCT TGGAACTTTG	reverse primer for kanamycin cassette amplification
pthF	CAGTGAGCGCAACGCAATTACAGGTTTCCCGACTG GAAAG	forward primer for *pth* gene amplification
pthR	CGATTAAGTTGGGTAACGCCTCATCGCAGTCGGCC TATTG	reverse primer for *pth* gene amplification
BAC-F	TAATTGCGTTGCGCTCACTG	forward primer for reverse amplification of pBACK and pBACT
BAC-R	GGCGTTACCCAACTTAATCG	reverse primer for reverse amplification of pBACK and pBACT

The underlined region represents the 5’ overhang sequence.

Chapter 3

Collection of temperature-sensitive alleles of DNA ligase

The temperature sensitive strain of *Salmonella enterica* described in this chapter has been generated by Stephanie Puckett.

Introduction

One aspect of synthetic biology is the development of genetic elements that can be used for genome engineering (5). These elements include promoters (6, 7), transcriptional enhancers (135), transcriptional stop elements (8), riboswitches (136) or site-specific recombinases (137). Some include reporter genes that encode fluorescent proteins, pigments or odors (10). Others (28, 30) have proposed that essential genes could be a useful class of genetic elements that might be widely used to engineer the limits of viability of a variety of microbes under a specified restrictive condition, such as high temperature. One application of temperature restriction of growth could be in creating bacterial pathogens that are wild type in every trait except for growth above a defined temperature, such as 35°C. Such pathogens could be used in research, teaching and diagnostic antigen preparation with minimal chance of causing invasive disease in humans. Another application is in creating attenuated vaccines, where temperature-sensitivity is already a well-established approach for attenuation.

Essential genes are defined as those that are required for the viability of an organism under all growth conditions (35, 36, 138). Biologists try to determine an organism's complement of essential genes for the inherent interest in knowing which

minimal set of genes are needed for the functioning of a cell, or for the practical goal of knowing which gene products are good targets for lethal chemicals, such as antibiotics.

Because essential genes are required for growth it is difficult to prove any one gene is essential, since proof of essentiality requires establishing the impossibility of deleting a particular gene. High throughput approaches, such as saturation mutagenesis (41, 42) have eased the task of identifying essential genes: those genes that fail to tolerate an insertion are suspected of being essential. As well, comparative genomics help identify highly conserved essential genes through their evolutionary retention index (ERI) (77, 139). Finding that a gene is present in all bacteria (ERI of 1), or in a large proportion of bacteria, is considered evidence that a gene is essential. Using this approach biologists have identified about 600 genes with high ERI values that are good candidates for essential genes (77). Further experiments with single open reading frame (ORF) deletions have refined this to 300 putative essential genes (35, 36). These genes are thought to be essential in nearly all bacteria.

One of the essential genes found in every living organism is *ligA*, which encodes DNA ligase (60, 140). This enzyme is responsible for catalyzing the formation of a phosphodiester bond between the 5' phosphate of one nucleotide and the 3' hydroxyl group of another in the backbone of a DNA strand. The loss of DNA ligase function causes quick formation of double strand breaks in the DNA structure, followed by exonucleolytic degradation of the fragmented DNA (58). For this reason DNA ligase has a crucial role in DNA replication, recombination and repair.

There are two general classes of DNA ligases, the ATP-dependent and the NAD-dependent enzymes (59, 60). The ATP-dependent ligases are a more diverse group that

usually are the essential ligases of Eukarya and Archea. They are also commonly found in the genomes of viruses and bacteriophages. On the other hand the NAD-dependent ligases are much more conserved, and are almost uniquely found among Bacteria. Usually a bacterial species possesses a single NAD-dependent DNA ligase, but some bacteria also have ATP-dependent ligases (62). However these are considered to be non-essential, since experiments have shown that the native bacterial ATP-dependent ligases cannot substitute for the loss of the NAD-dependent ligase (62, 63).

The complex eukaryotic cell requires several specialized types of DNA ligases (65). Of these the ATP-dependant DNA ligase I is the one that has a role most similar to NAD-dependant ligase in bacteria. It is conserved in all eukaryotes and is involved in nuclear DNA replication, repair and recombination. In *Saccharomyces cerevisiae* and other simple eukaryotes DNA ligase I is also involved in mitochondrial DNA replication and repair. In vertebrates, DNA ligase III aids in nuclear DNA repair but has its primary function in the mitochondria. The enzyme, DNA ligase IV, is conserved among all eukaryotes and plays a role in nuclear DNA repair. Because of these specialized roles ATP-dependant ligases are more diverse than NAD-dependent ligases.

The enzymes found in psychrophilic (cold-loving) bacteria have adapted to function at cold temperatures, and one consequence of this adaptation is that many psychrophilic enzymes are more heat-labile than their mesophilic counterparts (141). Predictably, this cold adaptation and heat lability properties extend to essential proteins. In our previous work we have found that psychrophilic forms of DNA ligase are inactivated in a temperature range appropriate for engineering mesophilic bacterial to temperature-sensitive (TS) forms that can be used for bio-containment or live vaccines

(31). In this work new psychrophilic alleles of *ligA* were discovered and shown to impart a range of TS-phenotypes in bacteria. Two alleles encoding NAD-dependent DNA ligase were also proven to confer temperature sensitivity in yeast.

Results and discussion

Isolation of *ligA* alleles from psychrophilic bacteria.

We created several DNA constructs designed to replace the native DNA ligase in the chromosome of the mesophilic bacterium *Francisella novicida* with *ligA* alleles from psychrophilic bacteria. We had previously observed that *F. novicida* is a good host for the expression of psychrophilic alleles, and this is likely due to its low G+C content (33%) which is close to the G+C content of chromosomes of the psychrophilic organisms that served as the source of the *ligA* alleles (31). All of our *ligA* gene substitutions were designed to use the native *F. novicida* promoter and ribosome binding site (Fig. 7). To minimize a change of the mRNA structure in the region of the RBS while substituting the maximal amount of the psychrophile-derived *ligA* gene we utilized the first three codons of the *F. novicida ligA* ORF in place of the codons from the psychrophile-derived *ligA* gene.

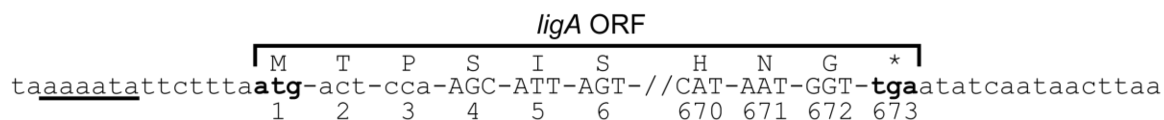

Figure 7 Structure of typical psychrophilic gene substitution into the *F. novicida* chromosome.

Example here shows the P1 *ligA* allele. Lower case letters represent *F. novicida* DNA sequence, and upper case letters represent psychrophilic *ligA* gene sequence. The start and stop codons are shown in bold letters. The putative ribosome binding site is underlined. Numbers indicate codon position in the *ligA* ORF.

All of the *ligA* alleles derived from the genera *Colwellia*, *Pseudoalteromonas* or *Shewanella* functioned in *F. novicida*, and their substitutions for the native *ligA* gene resulted in temperature-sensitive (TS) *F. novicida* strains with restrictive temperatures ranging from around 33 to 39°C (Table 3 and Fig. 8). When we subjected the *ligA* region to sequence analysis we found that all but one of the strains (see discussion of C1-*ligA* allele below) incorporated the complete psychrophile-derived allele, including the intentionally altered first three codons. In addition, to rule out the possibility that the TS phenotype was due to a misfolded psychrophilic DNA ligase interfering with the native enzyme expressed from a cryptic chromosomal location, we reintroduced the *F. novicida ligA* gene to form merodiploids of strains P1, P2, C1, S2, S3, P3, P4, P5, P7 and S4 that harbored both the native and psychrophilic *ligA* genes. In all cases the merodiploids grew well above the restrictive temperature of the cognate TS strain; figure 9 shows these results for three merodiploids.

Table 3 Bacterial origin of psychrophilic *ligA* alleles, restrictive temperature and reversion frequencies in F. novicida strains harboring these alleles.

Allele	Origin	Restrictive Temperature	Reversion Frequency
P1	*Pseudoalteromonas sp.*	33.0°C	1.8×10^{-8}
P2	*Pseudoalteromonas sp.*	33.0°C	4.1×10^{-9}
C1[¥]	*Colwellia sp.*	33.0°C	4.0×10^{-8}
S1	*Shewanella frigidimarina*	33.0°C	4.0×10^{-6}
S2	*Shewanella sp.*	33.0°C	3.7×10^{-8}
S3	*Shewanella sp.*	33.5°C	2.8×10^{-4}
P3	*Pseudoalteromonas sp.*	34.5°C	6.7×10^{-9}
C2	*Collwelia psychrerythraea 34H*	35.0°C	$<1.2 \times 10^{-10}$
P4	*Pseudoalteromonas sp.*	35.0°C	$<4.2 \times 10^{-10}$
P5	*Pseudoalteromonas sp.*	36.0°C	$<2.3 \times 10^{-7}$
P6	*Pseudoalteromonas haloplanktis* TAC125	37.0°C	$<1.5 \times 10^{-10}$
P7	*Pseudoalteromonas sp.*	38.0°C	$<3.0 \times 10^{-10}$
S4	*Shewanella sp.*	38.5°C	6.8×10^{-10}
S5	*Shewanella oneidensis* MR1	>44.0°C	ND

*All of these *ligA* alleles were PCR amplified from chromosomal DNA, except the P6 allele, which was made synthetically as a codon-optimized version suitable for expression in *F. novicida*.

[¥]When the full C1 allele supported the growth of a *S. enterica* Δ*ligA* strain, the restrictive temperature was 27°C and the reversion frequency was $1.5 \text{ X } 10^{-8}$/cell.

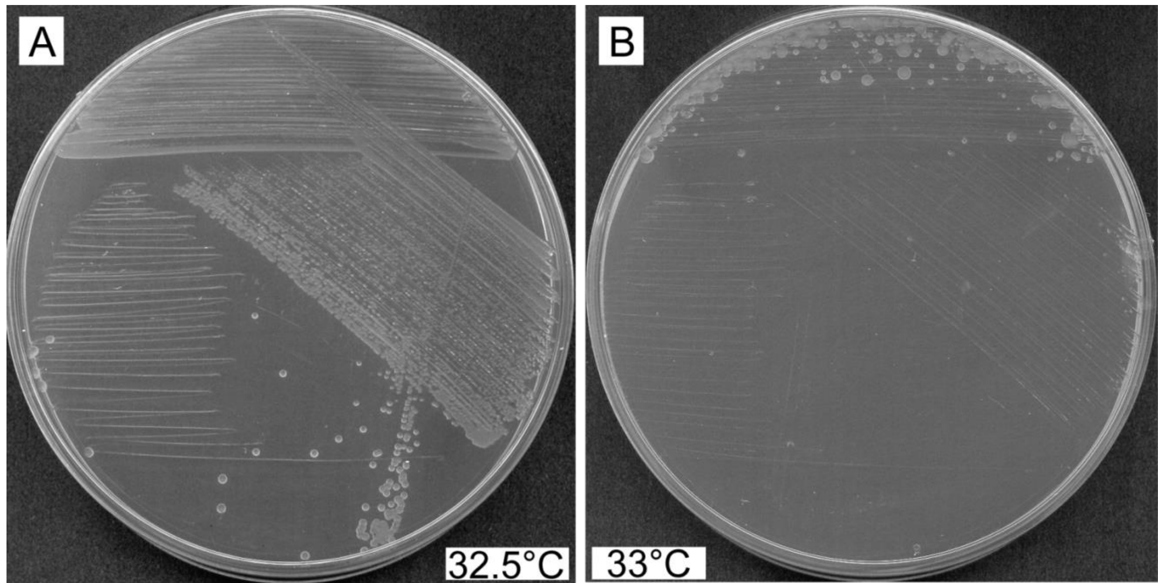

Figure 8 Temperature sensitive phenotype of *Francisella novicida*

Inability to form isolated colonies defines restrictive temperature of *F. novicida*-P2. Panel A, Inoculated agar plate incubated at the maximal permissive temperature of 32.5°C. Panel B, inoculated agar plate incubated at restrictive temperature of 33°C.

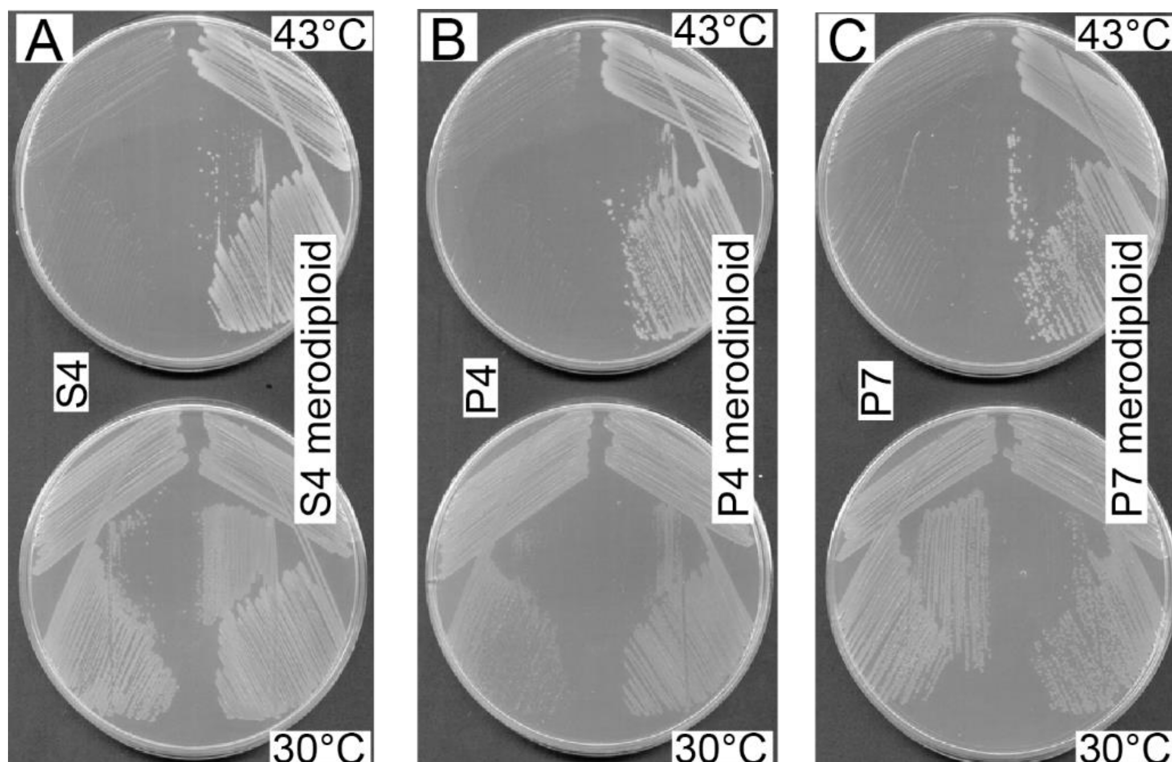

Figure 9 Phenotype of merodiploid strains

The pRS426KmSac constructs with the following psychrophilic alleles of DNA ligase were used to transform mesophilic *Francisella novicida*: P1, P2, C1, S2, S3, P3, P4, P5, P7 and S4 (basically everything except the S1, C2 and P6). The recombination mutants were selected on kanamycin at 30^{0}C. Then the isolated colonies were streaked on TSBC plates at 43^{0}C, next to *Francisella novicida* strains with corresponding psychrophilic alleles (TS strains). As a control identical plates were grown at 30^{0}C. Temperature sensitive strains were capable of growing only at 30^{0}C, while the merodiploid strains could survive at both 30^{0}C and 43^{0}C.

To confirm the presence of psychrophilic alleles in the merodiploid strains we have isolated genomic DNA from P2, P7, S3, P4, P3, C1, S4 merodiploids. We used it as a template for PCR with primers designed to bind inside the psychrophilic alleles. As a control we run PCR with the same primers against the genomic DNA of wild type *Francisella novicida*. We obtained the expected product from the genomic DNA of merodiploids but not from the wild type *Francisella novicida*. Additionally the PCR products obtained from merodiploid strains S3 and P4 were sequenced. The sequencing results confirmed the presence of introduced psychrophilic alleles.

To test the possibility that the TS phenotype of the strains with *ligA* gene substitutions were due to mutations outside of the *ligA* alleles we returned the *F. novicida ligA* homologue to each of the TS strains. The *F. novicida* strain used in these studies allows high level transformation with linear DNA destined for integration into the chromosome. We PCR-amplified the wild type *F. novicida ligA* allele along with approximately 1 kb of flanking DNA regions. This amplicon was used to transform each of the TS *F. novicida* strains carrying psychrophilic *ligA* alleles. After plating the transformation mixtures at restrictive temperatures a lawn of putative transformants were found for each transformation (Fig. 10). Parallel experiments without added DNA yielded no growth at the restrictive temperatures. DNA from isolates from three of the transformation experiments were used as templates for PCR amplification of the *ligA* regions, and these amplicons were subjected to DNA sequence analysis. All of these showed the presence of the *F. novicida ligA* gene and the absence of any of the psychrophilic genes.

Another possibility is that the introduction of a foreign gene or gene product induces a TS phenotype through an unknown mechanism. While this possibility is difficult to test directly we approached this indirectly by introducing a foreign mesophilic *ligA* gene. The mesophilic bacterium *Shewanella oneidensis* MR-1 is closely related to the three psychrophilic genera used as the source of the *ligA* alleles. Indeed the DNA ligase of *S. oneidensis* has 74.5%, 73.5%, 74.6% and 73.5% amino acid identity with the products of the S1, S2, S3 and S4 *ligA* alleles respectively. When the *S. oneidensis ligA* allele was substituted into the *F. novicida* chromosome the resulting strain had a temperature profile indistinguishable from that of the parental *F. novicida* (Fig. 11). This

demonstrated that the mere introduction of a foreign *ligA* allele does not induce a TS phenotype.

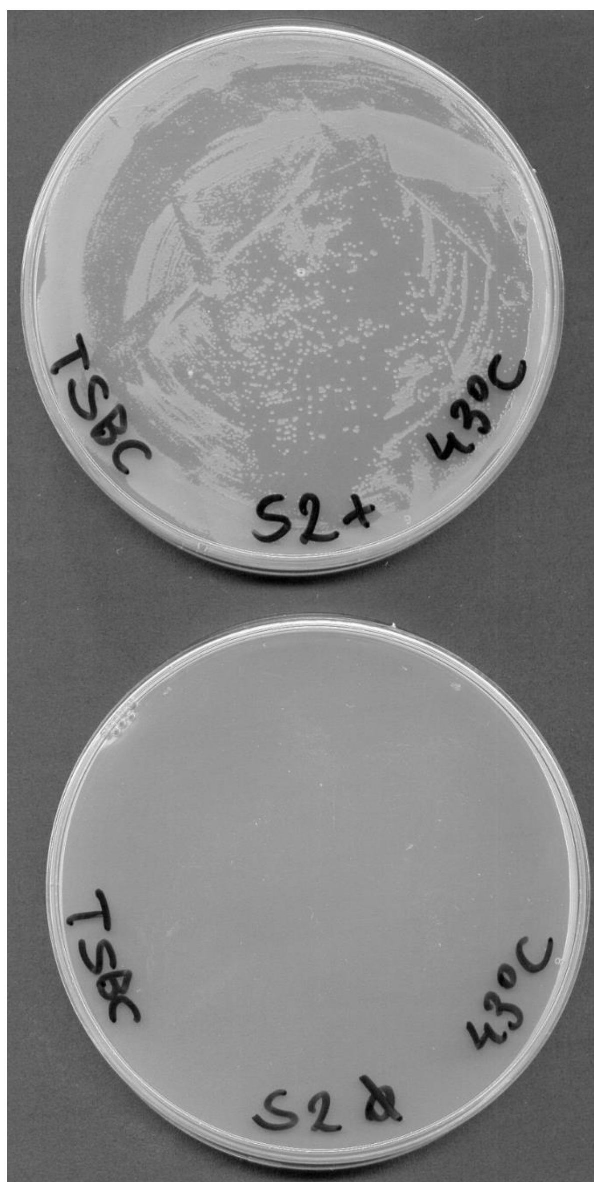

Figure 10 Complementation of temperature sensitive phenotype

Transformation of *F. novicida-S2* with PCR amplicon encoding the wild type *F. novicida ligA* and surrounding chromosomal regions and selection at 43°C (upper panel) or with no DNA (lower panel). Transformation-competent *F. novicida* readily take up and integrate linear DNA. The *F. novicida*-S2 strain has a restrictive temperature of 33°C. Similar results were found when all other temperature sensitive *F. novicida* strains (P1. P2, C1, S1, S3, P3, C2, P4, P5 P6 and P7) were transformed with the PCR amplicon encoding the wild type *F. novicida ligA* chromosomal region.

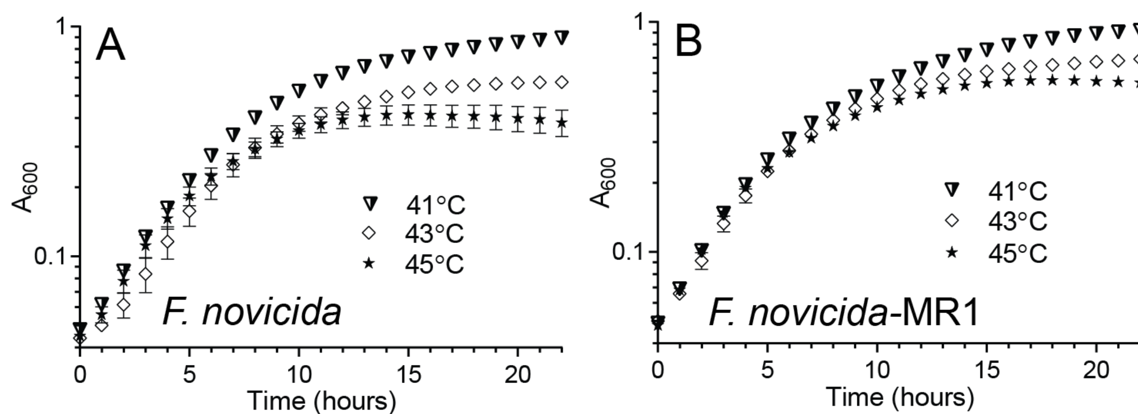

Figure 11 Growth pattern of mesophilic *F. novicida* strains.

Growth characteristics of wild type *F. novicida* (Panel A) and *F. novicida*-MR-1 (S5) that harbors a *ligA* gene from *S. oneidensis* MR-1 (Panel B). The maximal growth temperature for *F. novicida* is about 45°C. Error bars (SEM) are presented but are usually too small to be discerned.

The gene substitution of the C1-*ligA* allele yielded unusual results. The *F. novicida*-C1 strain carried a hybrid *ligA* gene, with the first 457 codons from the *F. novicida ligA* gene and the last 223 codons from the *Colwellia ligA* gene (Fig. 12). Since the source bacterium of the *Colwellia* sp. *ligA* gene has a maximal growth temperature of 12.5°C we suspected that the *ligA* gene product may not function at the 30°C that was used to select recombinants. Further experiments to isolate a *F. novicida* carrying the *ligA* C1 allele at lower temperature failed, and thus we do not know the reason why a complete C1 allele substitution was not found.

To further test the properties of the C1-*ligA* allele we introduced it on plasmid pBC SK+ (Cm^R) into an *S. enterica* strain with an inactivated chromosomal copy of *ligA* that was supported by a pBR313-borne (Ap^R) copy of the *S. enterica ligA*. After growth in broth with choloramphenicol but without ampicillin, we found isolates that had lost the Ap^R plasmid. These isolates had a restrictive temperature of 27°C (Fig. 13), and DNA sequencing of the plasmid present in the cell showed that it carried the C1-*ligA* allele. These findings suggest that the C1-*ligA* gene product is inactivated at about 27°C.

```
   F. novicida DNA            Colwellia DNA          F. novicida DNA
 F   V   S   R   K    A   H   D    K   Y   N    *
ttt-gtt-tca-cgt-aaa-GCC-CAT-GAT//AAA-TAT-AAC-tgaatatcaataacttaa
453 454 455 456 457 458 459 460  678 679 680 681
```

Figure 12 Hybrid form of C1 allele of DNA ligase.

Integration of the C1 *ligA* allele resulted in a hybrid gene with 456 codons from the *F. novicida ligA* gene that encodes the N-terminal region of the 680 amino acid hybrid protein.

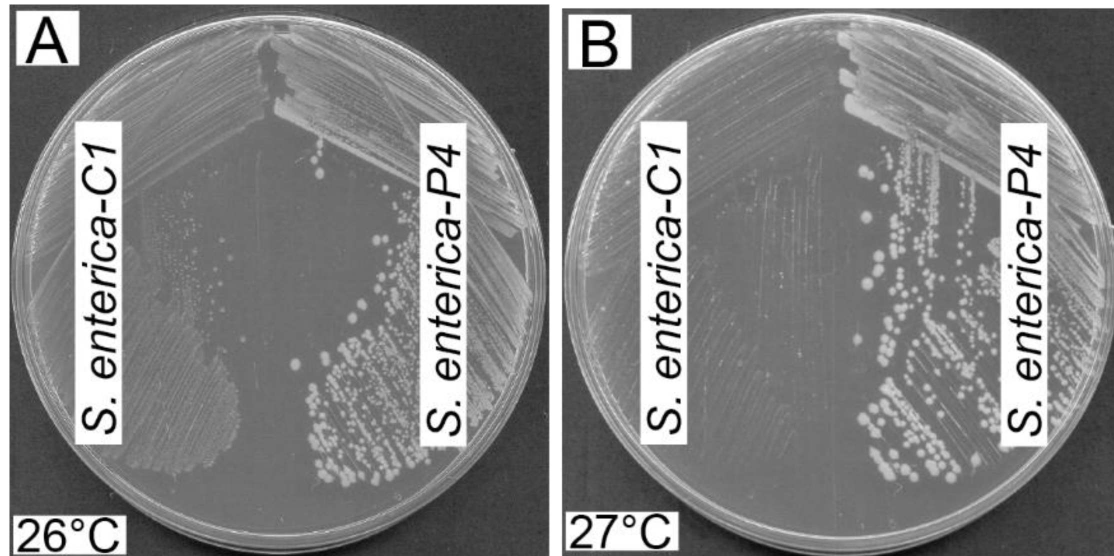

Figure 13 Temperature sensitive phenotype of *Salmonella enterica*

Low restrictive temperature of *S. enterica* strain depends on *ligA*-C1 for growth. *S. enterica* dependent on *ligA*-C1 or *ligA*-P4 both can form isolated colonies at 26°C (Panel A) but the strain carrying *lig*A-C1 allele cannot form isolated colonies at 27°C

Comparative analysis of the *ligA* deduced amino acid sequences

In order to ascertain if there are discernible amino acid differences that correlate with the restrictive temperatures induced by the *ligA* alleles we analyzed the sequences using multiple alignments (Fig. 14). We found that large segments of the deduced amino acid sequences are highly similar even between *ligA* alleles of psychrophilic and mesophilic origin (over 70% similarity). This is especially true in some of the conserved motifs such as motif I and motif III. Yet, while most *ligA* deduced amino acid sequences with high identity (over 95% identity) have similar restrictive temperature, there are cases of closely similar alleles with different phenotypes, for example the S2 and S4 alleles (Fig. 15). Thus, a simple analysis of any amino acid sequence differences does not allow a prediction of the inactivation temperature of the *ligA* gene product.

```
P1    33.0°C  MTP------------------SISEQINHLRTTLEQHNYNYYVLDTPSIPDAEYDRLLRELSALETQNPEFLSADSPTQK  62
P2    33.0°C  MTP------------------SISEQINHLRTTLEQHNYNYYVLDTTSIPDAEYDRLLRELSALETQNPEFLSADSPTQK  62
C1    33.0°C  MTPNEFFSI----KYHILAKAELKAYIDKLADYLSQQSYLYHTLDKPIISDSDYDKLFRLLQDLVNDNPQFKPINSVLDR  76
S1    33.0°C  MTP------------------IQTEMDQLTHTINQHNIRYYVDDAPSIPDAEYDRLIKRLTELERDYPQFKSVDSPTQR  61
S2    33.0°C  MTP------------------IELEITELKNELNQHNYHYYVDDNPTIPDAEYDRLLRKLDELEVAHPEFASADSPTQR  61
S3    33.5°C  MTP------------------IQTEMDQLTHTINQHNIRYYVDDAPSIPDAEYDRLIKRLTELEGDYPQFKSVDSPTQR  61
P3    34.5°C  MTP------------------SISEQINYLRTTLEQHNYNYYVLDTPSIPDSEYDRLLRELSALETEHPEFLTADSPTQK  62
C2    35.0°C  MTP------------------VEKKISQLQQQLNQYNHEYYVLDQPSVPDAEYDRLMTALIDLEKTNPELKTIDSPSQK  61
P4    35.0°C  MTP------------------SISEQINHFRTILEQHNYNYYVLDTPSIPDSEYDRLLRELSALENEHPEFLTADSPTQK  62
P5    36.0°C  MTP------------------SIS-AINYLRTTLEQHNYNYYVLDTPSIPDSEYDRLLRELSALETEHPEFLTADSPTQK  61
P6    37.0°C  MTP------------------SISEQINHLRSTLEQHSYNYYVLDTPSIPDAEYDRLLQQLSALETQHPELITADSPTQK  62
P7    38.0°C  MTP------------------SISEQINHLRTTLEQHNYNYYVLDTPSIPDSEYDRLLRELSALEMQHPEFLTADSPTQK  62
S4    38.5°C  MTP------------------IE-EIKQLTDILNEHNHRYYVEDSPSVPDAEYDRLINRLKALEAEHPELCLATSPTQR  60
S5   >44.0°C  MTPIQLDKRLSELLSQAVTPQTAQPLMQALSQSLNEHNIRYYVDDAPSITDSEYDRLMQQLKKLEAEYPQFVLADSPTQR  80
Fnov >44.0°C  MTPNEFFSI----KYHILAKAELKAYIDKLADYLSQQSYLYHTLDKPIISDSDYDKLFRLLQDLVNDNPQFKPINSVLDR  76
Ecol >44.0°C  M------------------ESIEQQLTELRTTLRHHEYLYHVMDAPEIPDAEYDRLMRELRELETKHPELITPDSPTQR  61
Bsub >44.0°C  M---------------DKETAKQRAEELRRTINKYSYEYYTLDEPSVPDAEYDRLMQELIAIEEEHPDLRTPDSPTQR  63
Mtb  >44.0°C  MSSPD----------ADQTAPEVLRQWQALAEEVREHQFRYYVRDAPIISDAEFDELLRRLEALEEQHPELRTPDSPTQL  70

                                                                EhKhDG                   TRGDG
P1    33.0°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEDEFTGFNRRIKERLMSTDELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGF  141
P2    33.0°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEDEFTGFNRRIKDRLMSTDELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGF  141
C1    33.0°C  VGGEVLA-EFETIKHKKKMTSLANVFSLEELRDFYDKIEY------DIELECEPKMDGLAISIFYKNGKFDYAVTRGDGI  149
S1    33.0°C  VGGIALQ-KFAQITHLKPMLSLDNAFEQADFAAFNKRITDKVD--S-VDYVCEPKLDGLAVSITYRFGVLERAATRGDGS  137
S2    33.0°C  VGGEALA-KFEQITHLKPMLSLDNVFDEPEFTAFHSRISDKVG--TELSYCCEPKLDGLAVSIVYRDGVFERAATRGDGQ  138
S3    33.5°C  VGGIALQ-KFAQITHLKPMLSLDNAFEQADFAAFNKRITDKVD--N-VDYVCEPKLDGLAVSITYRLGVLERAATRGDGS  137
P3    34.5°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEEEFTAFNRRIKERLMSTDELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGF  141
C2    35.0°C  VGGQALK-SFTQVTHQLPMLSLDNVFSLDDFHAFVKRVKDRLNDNQAIVFCAEPKLDGLAVSLRYEHGQLIQAATRGDGS  140
P4    35.0°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEEEFTAFNRRIKERLMSTDELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGF  141
P5    36.0°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEEEFTAFNRRIKERLMSTDELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGF  140
P6    37.0°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEEEFIAFNRRIKERLMSTEELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGL  141
P7    38.0°C  VGGAALS-KFEQVAHQVPMLSLDNAFSEEEFTAFNRRIKERLMSTDELTFCCEPKLDGLAVSIIYRDGVLVQAATRGDGF  141
S4    38.5°C  VGGVALA-KFEQITHLKPMLSLDNVFSEEEFEAFYKRISGKTS--EAPHFCCEPKLDGLAVSIIYRDGVYERAATRGDGT  137
S5   >44.0°C  VGGLALA-KFEQITHLKPMLSLDNAFDEADFTAFHKRVSDRVG--Q-VSFCCEPKLDGLAVSILYRNGVLERAATRGDGT  156
Fnov >44.0°C  VGGEVLA-EFETIKHKKKMTSLANVFSLEELRDFYDKIEY------DIELECEPKMDGLAISIFYKNGKFDYAVTRGDGI  149
Ecol >44.0°C  VGAAPLA-AFSQIRHEVPMLSLDNVFDEESFLAFNKRVQDRLKNNEKVTWCCELKLDGLAVSILYENGVLVSAATRGDGT  140
Bsub >44.0°C  VGGAVLE-AFQKVTHGTPMLSLGNAFNADDLRDFDRRVRQSVG--DDVAYNVELKIDGLAVSLRYEDGYFVRGATRGDGT  140
Mtb  >44.0°C  VGGAGFATDFEPVDHLERMLSLDNAFTADELAAWAGRIHAEVG--DAAHYLCELKIDGVALSLVYREGRLTRASTRGDGR  148

                                        hEVRGE                        NPRNAAAGSLRQ
P1    33.0°C  TGENITQNVKTIRNVPLKLR---G-DFPKELEVRGEVFMDSAGFEKLNSEAQKRGEKVFVNPRNAAAGSLRQLDSKITAK  217
P2    33.0°C  TGENITQNVKTIRNVPLKLR---G-DFPKELEVRGEVFMDSAGFEKLNSEAQKRGEKVFVNPRNAAAGSLRQLDSKITAK  217
C1    33.0°C  QGEKVSENVKTIRNVPLKLN---TSNPPEELEVRGEIILDKQSFLSLNEYMQTHENKTFANPRNAAAGSIRMLDSKVVAK  226
S1    33.0°C  VGEDITANVRTIRSIPLKLR---GEGFPDLVEVRGEVFMPKAAFEALNQRQISKGDKVFVNPRNAAAGSLRQLDSKITAS  214
S2    33.0°C  TGENITENVRTIKSIPLTLR---GDNFPPLVEVRGEVIMPHQAFDDLNKRALAKGDKLFVNPRNAAAGSLRQLDSKITAS  215
S3    33.5°C  VGEDITANVRTIRSIPLKLR---GEGFPDLVEVRGEVFMPKAAFEALNQRQISKGDKVFVNPRNAAAGSLRQLDSKITAS  214
P3    34.5°C  TGENITQNVKTIRNVPLKLR---G-DYPKELEVRGEVFMDSAGFDKLNTEAQKRGEKVFVNPRNAAAGSLRQLDSKITAK  217
C2    35.0°C  VGENITTNIRTIKSIPLKLMGTPGKDFPDIVEVRGEVFMPKASFDALNTLAKKRGEKGFANPRNAAAGSLRQLDSKITAK  220
P4    35.0°C  TGENITQNVKTIRNVPLKLR---G-DYPKELEVRGEVFMDSAGFDKLNTEAEKRGDKVFVNPRNAAAGSLRQLDSKITAK  217
P5    36.0°C  TGENITQNVKTIRNVPLKLR---G-DYPKELEVRGEVFMDSAGFDKLNTEAQKRGEKVFVNPRNAAAGSLRQLDSKITAK  216
P6    37.0°C  TGENVTQNVKTIRNVPLKLR---GSDYPAELEVRGEVFMDNAGFEKFNIEAEKRGEKVFVNPRNAAAGSLRQLDSKITAK  218
P7    38.0°C  TGENITQNVKTIRNVPLKLR---G-DYPKELEVRGEVFMDSAGFDKLNTEAQKRGEKVFVNPRNAAAGSLRQLDSKITAK  217
S4    38.5°C  TGENITENVRTILSIPLTLR---GDNFPPLLEVRGEVIMPKKAFDALNDRARKKGDKLFVNPRNAAAGSLRQLDSKITAS  214
S5   >44.0°C  VGEDITENVKTIKSIPLRLR---GNNYPELVEVRGEAFMPKAAFEALNERARLKDDKQFVNPRNAAAGSLRQLDSKITAS  233
Fnov >44.0°C  QGEKVSENVKTIRNVPLKLN---TSNPPEELEVRGEIILDKQSFLSLNEYMQTHENKTFANPRNAAAGSIRMLDSKVVAK  226
Ecol >44.0°C  TGEDITSNVRTIRAIPLKLH---GENIPARLEVRGEVFLPQAGFEKINEDARRTGGKVFANPRNAAAGSLRQLDPRITAK  217
Bsub >44.0°C  TGEDITENLKTIRNIPLKMN------RELSIEVRGEAYMPKRSFEALNEERIKNEEEPFANPRNAAAGSLRQLDPKIAAK  214
Mtb  >44.0°C  TGEDVTLNARTIADVPERLTPGDDYPVPEVLEVRGEVFFRLDDFQALNASLVEEGKAPFANPRNSAAGSLRQKDPAVTAR  228

                 hIsYsxGhh
P1    33.0°C  RPLMFYAYSTGLVADGS-------------IPDDHFQQLEKLTDWGLPLCPETKLVEGPQAALDYYGDILTRRSELKYEI  284
P2    33.0°C  RPLMFYAYSTGLVADGS-------------IPDDHFQQLEKLTDWGLPLCPETKLVEGPQAALDYYGDILTRRSELKYEI  284
C1    33.0°C  RPLKLYSYGIGYFSKDF------------VHPETQFELMQLLQSFGFTISDNMFLAKNFSEVEEYHHKMSHQRADLAYDI  294
S1    33.0°C  RALGFYAYALGVVEGESQ-----------PMQTSHYGQLTQLQQWGIPVSSEVKVTDLLEKVYAYYADIMARRSALEYEI  283
S2    33.0°C  RSLGFYAYALGVVEPETW-----------PLADSHFGQLEQLRSWGVPVSQEVKVCDSVAKVMEYYNDIQQRRSDLAFEI  284
S3    33.5°C  RALGFYAYALGVVEGESQ-----------PMQTSHYGQLTQLQQWGVPVSSEVKVTDSLEKVYAYYADIMARRSALEYEI  283
P3    34.5°C  RPLMFYAYSTGLVANGN-------------IPEDHYQQLEKLTDWGLPLCPETKLVEGPKAALEYYSDILTRRGELKYEI  284
C2    35.0°C  RNLAFYAYSLGFVGKLSDGGAESTDLTNDFFANSHERLCQLKRLGLPMCPEVRLLESEQACDAFYQDILAKRSALSYEI  297
P4    35.0°C  RPLMFYAYSTGLVADGN-------------IPQDHYQQLEKLTDWGLPLCPETKLVEGPKAALEYYSDILTRRSSLKYEI  284
P5    36.0°C  RPLMFYAYSTGLVANGN-------------IPEDHYQQLEKLTDWGLPLCPETKLVEGPKAALEYYSDILTRRGELKYEI  283
P6    37.0°C  RPLMFYAYSTGLVADGS-------------IAEDHYQQLEKLTDWGLPLCPETKLVEGPQAALAYYTDILTRRGELKYEI  285
P7    38.0°C  RPLMFYAYSTGLVADGS-------------IPEDHYQQLEKLTDWGLPLCPETKLVEGPKAALDYYSDILTRRGELKYEI  284
S4    38.5°C  RALGFYAYALGVVEPESW-----------PLADTHYGQLMQLKSWGFPVSAEVKQCHDVSSVVAYYTDIMTRRDSLDYEI  283
S5   >44.0°C  RALSFYAYALGVVEPSSH-----------ELAKTHYEQLQQLKSWGLPVSSEIKVCDELNQVFAYYKDILTRRSDLPFEI  312
Fnov >44.0°C  RPLKLYSYGIGYFSKDF------------VHPETQFELMQLLQSFGFTISDNMFLAKNFSEVEEYHHKMSHQRADLAYDI  294
Ecol >44.0°C  RPLTFFCYGVGVLEGGE-------------LPDTHLGRLLQFKKWGLPVSDRVTLCESAEEVLAFYHKVEEDRPTLGFDI  284
Bsub >44.0°C  RNLDIFVYSIAELDEMG-------------VETQSQGLDFLDELGFKTNQERKKCGSIEEVITLIDELQAKRADLPYEI  280
Mtb  >44.0°C  RRLRMICHGLGHVEGFR-------------PATLHQAYLALRAWGLPVSEHTTLATDLAGVRERIDYWGEHRHEVDHEI  294

              DGhVhK                      AhAYKFPA
P1    33.0°C  DGVVIKVNNKALQERLGFVARAPRWAIAYKFPAQEEVTQLLDVEFQVGRTGAITPVARLEPVFVGGVTVSNATLHNGDEI  364
P2    33.0°C  DGVVIKVNNKALQERLGFVARAPRWAIAYKFPAQEEVTQLLDVEFQVGRTGAITPVARLEPVFVGGVTVSNATLHNGDEI  364
C1    33.0°C  DGLVFKVNNIKLQDIIGYTARGPKWAVAYKFPAEEVESEVLNVEFQVGRTGAITPVARLKPVAVGGVIVSNATLHNINEI  374
S1    33.0°C  DGVVIKVNDIAKQQTLGFVAKAPRWAIAYKFPAQEEMTLLESVDFQVGRTGAVTPVARLKPIFVGGVTVSNATLHNADEI  363
S2    33.0°C  DGVVIKVNQIAHQLNLGFVAKAPRWATAFKFPAQEEMTLLEGVDFQVGRTGAVTPVARLKPVFVGGVTVSNATLHNADEI  364
S3    33.5°C  DGVVIKVNDIAKQQTLGFVAKAPRWAIAYKFPAQEEMTLLESVDFQVGRTGAVTPVARLKPIFVGGVTVSNATLHNADEI  363
P3    34.5°C  DGVVIKINQKALQERLGFVARAPRWAIAYKFPAQEEITQLLDVEFQVGRTGAITPVARLEPVFVGGVTVSNATLHNGDEI  364
C2    35.0°C  DGTVLKVDEISLQKRLGFVARAPRWAIAYKFPAEEELTCVEDVEFQVGRTGAITPVARLKPVFVGGVTVSNATLHNQDEI  364
P4    35.0°C  DGVVIKINEKALQERLGFVARAPRWAIAYKFPAQEEITQLLDVEFQVGRTGAITPVARLEPVFVGGVTVSNATLHNGDEI  364
P5    36.0°C  DGVVIKINQKALQERLGFVARAPRWAIAYKFPAQEEITQLLDVEFQVGRTGAITPVARLEPVFVGGVTVSNATLHNGDEI  363
P6    37.0°C  DGVVIKINQKALQERLGFVARAPRWAIAYKFPAQEEITKLLDVEFQVGRTGAITPVARLEPVFVGGVTVSNATLHNGDEI  365
P7    38.0°C  DGVVIKINQKALQERLGFVARAPRWAIAYKFPAQEEITQLLDVDFQVGRTGAITPVARLKPVFVGGVTVSNATLHNSDEV  364
S4    38.5°C  DGVVIKVDSIEHQGQLGFVAKAPRWATAEKFPAQEEMTLLEGVDFQVGRTGAVTPVARLKPIFVGGVTVSNATLHNADEI  363
S5   >44.0°C  DGVVMKVNDIAQQQTLGFVAKSPRWAIAYKFPAQEEMTLLEGVDFQVGRTGAVTPVARLKPVFVGGVTVSNATLHNADEI  392
Fnov >44.0°C  DGLVFKVNNIKLQDIIGYTARGPKWAVAYKFPAEEVESEVLNVEFQVGRTGAITPVARLKPVAVGGVIVSNATLHNINEI  374
Ecol >44.0°C  DGVVIKVNSLAQQEQLGFVARAPRWAVAFKFPAQEQMTFVRDVEFQVGRTGRITPVARLEPVHVAGVLVSNATLHNADEI  364
Bsub >44.0°C  DGIVIKVDSLDQQEELGFTAKSPRWAIAYKFPAEEVVTKLLDIELNVGRTGVITPTAILEPVKVAGTTVSRASLHNEDLI  360
Mtb  >44.0°C  DGVVVKVDEVALQRRLGSTSRAPRWAIAYKYPPEEAQTKLLDIRVNVGRTGRITPFAFMTPVKVAGSTVGQATLHNASEI  374
```

```
                    PphxxhhhccR                      C   C                C     C
P1    33.0°C ARLGVKVGDTVIIRRAGDVIPQITQVVLDRRPADARDIVFPSACPICDSHVE-KVEGEAVARCTGGLVCPAQRKQAIKHF 443
P2    33.0°C ARLGVKVGDTVIIRRAGDVIPQITQVVLDRRPADARDIVFPSACPICDSHVE-KVEGEAVARCTGGLVCPAQRKQAIKHF 443
C1    33.0°C KRKDIRVGDRVIVRRAGDVIPEVVKSLPQYRKSDAQMVEMPTNCPVCDSAIE-NVNDQAIYRCTGGWHCQAQTTERLKHF 453
S1    33.0°C ARLGVKIGDTVIIRRAGDVIPQIVAIVPEKRPDDAQDIIFPLHCPVCQSIVE-RLEGEAVARCSGGLFCEAQRKEAIKHF 442
S2    33.0°C ARLGVKVGDTIIIRRAGDVIPQIVAVVADKRLENATDITFPAQCPVCQSLVE-RIEGEATARCTGGLFCEAQRKEAIKHF 443
S3    33.5°C ARLGVKIGDTVIIRRAGDVIPQIVAIVPEKRPDDAQDIIFPQHCPVCQSIVE-RLEGEAVARCSGGLFCEAQRKEAIKHF 442
P3    34.5°C ARLGVKVGDTVIIRRAGDVIPQITQVVLERRSDDARDIEFPTTCPICDSHVE-KVAGEAVARCTGGLVCPAQRKQAIKHF 443
C2    35.0°C TRLGLKVNDFVVIRRAGDVIPQIVSVVLDKRPDNAVDIVFPTSCPVCDSAVA-KPEGEAVLRCTAGLFCAAQRKEAIKHF 443
P4    35.0°C ARLGVKVGDTVIIRRAGDVIPQITQVVLERRPDDARDIEFPVTCPICDSHVE-KVEGEAVARCTGGLVCPAQRKQAIKHF 443
P5    36.0°C ARLGVKVGDTVIIRRAGDVIPQITQVVLERRPDDARDIEFPTTCPICDSHVE-KVAGEAVARCTGGLVCPAQRKQAIKHF 442
P6    37.0°C ARLGVKVGDTVIIRRAGDVIPQITQVVLERRPDDARDIEFPVTCPICDSHVE-KVEGEAVARCTGGLVCPAQRKQAIKHF 444
P7    38.0°C ARLGVKVGDTVIIRRAGDVIPQITQVVLERRPDDASDIEFPVTCPICDSHVE-KVEGEAVARCTGGLVCPAQRKQAIKHF 443
S4    38.5°C ARLGVKVGDTVIVRRAGDVIPQIVAIVADKRPDDAADIEFPKTCPVCDSMVE-RIEGEAVARCTGGLFCEAQRKEAIKHF 442
S5   >44.0°C ERLGVMVGDTVIIRRAGDVIPQIVAIVPERRSEDAQAIVFPQHCPVCGSLVE-RLEGEAVARCSGGLFCEAQRKEAIKHF 471
Fnov >44.0°C KRKDIRVGDRVIVRRAGDVIPEVVKSLPQYRKSDAQMVEMPTNCPVCDSAIE-NVNDQAIYRCTGGWHCQAQTTERLKHF 453
Ecol >44.0°C ERLGLRIGDKVVIRRAGDVIPQVVNVVLSERPEDTREVVFPTHCPVCGSDVE-RVEGEAVARCTGGLICGAQRKESLKHF 443
Bsub >44.0°C KEKDIRILDKVVVKKAGDIIPEVVNVLVDQRTGEEKEFSMPTECPECGSELV-RIEGEVALRCIN-PECPAQIREGLIHF 438
Mtb  >44.0°C KRKGVLIGDTVVIRKAGDVIPEVLGPVVELRDGSEREFIMPTTCPECGSPLAPEKEGDADIRCPNARGCPGQLRERVFHV 454

P1    33.0°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALNEAKATTLAKFLYSLGI 517
P2    33.0°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALNEAKATTLAKFLYSLGI 517
C1    33.0°C VSRKAHDIDGLGDKLVEQLVDENLITTPADLFKLTEIDVSTMERMGKK------SAVNLMVALEKAKETTLAKFIYGLGI 527
S1    33.0°C ASRKALNIDGMGDKIVEQLIDKELVKTPADLFSLTASSITMLDRMAMK------SATNIVAAIKHAKATTLARFLYSLGI 516
S2    33.0°C ASRKALDIDGMGDKVVEQLIDKELVESPADLFRLTASAMTMLERMGMK------SATKLVASIEVAKRTTFSRFLYALGI 517
S3    33.5°C ASRKALNIDGMGDKIVEQLIDKELVKTPADLFSLTASSVTMLERMAMK------SATNIVAAIKDAKTTTLARFLYSLGI 516
P3    34.5°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALEEAKGTTLAKFLYSLGI 517
C2    35.0°C ASRKAHDVDGLGDKLVEQLVDEKLINTPADLFKLTEIQVSTIDRMGKK------SATNLINGLEQAKSTTLAKFIYGLGI 517
P4    35.0°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALEEAKGTTLAKFLYSLGI 517
P5    36.0°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALEEAKGTTLAKFLYSLGI 516
P6    37.0°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALQDAKATTLAKFLYSLGI 518
P7    38.0°C ASRKALDIDGLGDKIVDQLVDRELIKTPADLFILKQGHFESLERMGPK------SAKNLVTALEEAKGTTLAKFLYSLGI 517
S4    38.5°C ASRKALDIDGMGDKVVEQLIDKELVESPADLFKLTASAITMLDRMGMK------SATNLVAAIDVAKTTTFNRFLYGLGI 516
S5   >44.0°C ASRKALDIDGMGDKIVEQLIDKELVQSPADLFKLTASMMTMLDRMGMK------SATNLALAIEAAKTTTLPRFLYALGI 545
Fnov >44.0°C VSRKAMDIDKLGAKLIEQLVAANLIKYPADIYKLNFDQLTGLERMGAK------SSQNVLDSIKKSKTPSLARFIFAIGI 527
Ecol >44.0°C VSRRAMDVDGMGDKIIDQLVEKEYVHTPADLFKLTAGKLTGLERMGPK------SAQNVVNALEKAKETTFARFLYALGI 517
Bsub >44.0°C VSRNAMNIDGLGERVITQLFEENLVRNVADLYKLTKERVIQLERMGEK------STENLISSIQKSKENSLERLLFGLGI 525
Mtb  >44.0°C ASRNGLDIEVLGYEAGVALLQAKVIADEGELFALTERDLLRTDLFRTKAGELSANGKRLLVNLDKAKAAPLWRVLVALSI 534

P1    33.0°C REAGEATAQNLANHFLTLENIFSASVDSLTQVSDVGDIVASHVRGFFDEEHNLAVVNALIEQGIHW--PALSAPSE--DD 593
P2    33.0°C REAGEATAQNLANHFLTLENIISASVDSLTQVSDVGDIVASHVRGFFDEEHNLAVVNALIEQGIHW--PALSAPSE--DD 593
C1    33.0°C REVGEATAANLANHFITLAAIQSANFEALKAVSDVGDVVAKNIVNFFKESHNIAVVAALDELM-TW--PDIIKKSA--DK 602
S1    33.0°C REVGEATAANLAQHFAEFERIRTASVEQLLEVADVGDIVAKHIRQFFAQPHNIEVIEQLLEAGITW--PVIEQADE--SQ 592
S2    33.0°C REVGEATAANLAAYFKTLEALKASNAEEFIKVDDVGTIVAAHLQHFFEQPHNLEVIDKLIEAGITW--PVIEAVSE--ED 593
S3    33.5°C REVGEATAANLAQHFAEFERIRTASVEQLLEVADVGDIVAKHIRQFFSQPHNIEVIDQLLEAGITW--PVIEQADE--SQ 592
P3    34.5°C REAGEATAQNLANHFLTLENIINASIDSLTQVNDVGEIVATHVCGFFDEEHNLAVVNALIEQGVNW--PALSAPSE--EE 593
C2    35.0°C REVGEATAANLANHFYTLAAIESASLEDLQNVSDVGEVVAKNIINFFKEEHNLAIVSGLS-EVMHW--PTIEIKSA--EE 592
P4    35.0°C REAGEATAQNLANHFLTLENVINASIDSLTQVSDVGEIVAAHVRGFFDEEHNLAVVNALVDQGVNW--PQLSAPSE--EE 593
P5    36.0°C REAGEATAQNLANHFLTLENIINASIDSLTQVSDVGEIVATHVRGFFDEEHNLAVVNALIEQGVNW--PALSAPSE--EE 592
P6    37.0°C REAGEATTQNLANHFLTLENVINASIDSLTQVSDVGEIVATHVRSFFAEQHNLDVVNALVEQGINW--PELTPPSA--QE 584
P7    38.0°C REAGEATAQNLANHFLTLENIINASIDSLTQVSDVGEIVATHVRGFFDEEHNLAVVNALIEQGVNW--PALSAPSE--EE 583
S4    38.5°C REVGEATAANLANYFKTLDKLKAADAETFIKVDDVGAIVAQHLTHFFAQPHNLEVVDSLVEAGVNW--PDIEEVAQ--ES 582
S5   >44.0°C REVGEATAANLAAHFGSLDALRIATIEQLIEVEDIGEVVAQHVAHFFAQPHNLEVIDALIAAGVNW--PAIEAPSA--DE 621
Fnov >44.0°C KDIGEVSSDVLANHFGSLESFRDAKFEELIEINDIGEIMANNIVSFWHDSLNIKIVEELLAIG-IKIQNPVKVEHA--YN 604
Ecol >44.0°C REVGEATAAGLAAYFGTLEALEAASIEELQKVPDVGIVVASHVHNFFAEESNRNVISELLAEGVHW--PAPIVINAEEID 595
Bsub >44.0°C RFIGSKAAKTLAMHFESLENLKKASKEELLAVDEIGEKMADAVITYFHKEEMLELLNELQELGVNTLYKGPKKVKAEDSD 605
Mtb  >44.0°C RHVGPTAARALATEFGSLDAIAAASTDQLAAVEGVGPTIAAAVTEWFAVDWHREIVDKWRAAGVRMVDE-----RDESVP 619

P1    33.0°C QPLAGLTYVLTGTLNTLNRNDAKARLQQLGVKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIDILTEDELIELLQKHNG- 672
P2    33.0°C QPLAGLTYVLTGTLNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIDILTEDELIELLQKHNG- 672
C1    33.0°C QPLAEQTFVLTGTLNQMGRSEAKAALQSLGAKVSGSISAKTHFLVAGEKSGSKLTKAQDLGVSILTEDDLVALLEKYN-- 680
S1    33.0°C LSLKGQTWVLTGTLTQLNRNDAKAQLQALGAKVAGSVSKNTDCLVAGEAAGSKLAKAEELGVKVIDEQALMDLLNAAN-- 670
S2    33.0°C LSLKGQTWVLTGTLTQLNRNDAKAQLQALGAKVAGSVSKNTDCLVAGAAAGSKLTKAQELGVKVINEEGLIAILS----- 668
S3    33.5°C LSLKGQTWVLTGTLTQLNRNDAKAQLQALGAKVAGSVSKNTDCLVAGEAAGSKLAKAEELGVKVIDEQALMDLLNAAN-- 670
P3    34.5°C QPLAGLTYVLTGTLNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIEILTEDELIKLLQKHNG- 672
C2    35.0°C LPLAEQIFVLTGTLTQMGRTEAKTALQSLGAKVSGSVSKNTHFVVAGDKAGSKLTKAQDLGISVLTEDGLVALLAEHGIT 672
P4    35.0°C QPLAGLTYVLTGTLNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIDILTEDELIELLQKHNG- 672
P5    36.0°C QPLAGLTYVLTGTLNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIEILTEDELIELLQKHNG- 671
P6    37.0°C QPLAGLVYVLTGTLNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIDVLTEEDLINLLEQHNG- 663
P7    38.0°C QPLAGLTYVLTGTLNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAGSKLTKAQDLGIDILTEDELIELLQKHNG- 662
S4    38.5°C LSLKGQTWVLTGTLVQLNRNDAKARLQALGAKVAGSVSKNTDCLVAGEAAGSKLTKAQDLGVKVIDEEALLQLLSASE-- 660
S5   >44.0°C QPLKGQTWVLTGTLNQLNRNDAKAQLQVLGAKVAGSVSKNTDCLVAGEAAGSKLAKAQELGVKVIGEDELLALLAANR-- 699
Fnov >44.0°C ESFTGKTVVITGSFENYGRTELTQLLKSIGAKVTSSVSKKTDMVICGDNAGSKLTKAQELGVEVILEDNLKDLL------ 678
Ecol >44.0°C SPFAGKTVVLTGSLSQMSRDDAKARLVELGAKVAGSVSKKTDLVIAGEAAGSKLAKAQELGIEVIDEAEMLRLLGS---- 671
Bsub >44.0°C SYFAGKTIVLTGKLEELSRNEAKAQIEALGGKLTGSVSKNTDLVIAGEAAGSKLTKAQELNIEVWNEEQLMGELKK---- 681
Mtb  >44.0°C RTLAGLTIVVTGSLTGFSRDDAKEAIVARGGKAAGSVSKKTNYVVAGDSPGSKYDKAVELGVPILDEDGFRRLLADGPAS 699

P1    33.0°C --   672
P2    33.0°C --   672
C1    33.0°C --   680
S1    33.0°C --   670
S2    33.0°C --   668
S3    33.5°C --   670
P3    34.5°C --   670
C2    35.0°C I-   673
P4    35.0°C --   672
P5    36.0°C --   671
P6    37.0°C --   663
P7    38.0°C --   662
S4    38.5°C --   660
S5   >44.0°C --   699
Fnov >44.0°C --   678
Ecol >44.0°C --   671
Bsub >44.0°C --   681
Mtb  >44.0°C RT   701
```

Figure 14 Sequence alignment of LigA protein sequences

Multiple sequence alignment of the deduced amino acid sequence of psychrophilic *ligA* alleles that were successfully substituted into the *F. novicida* chromosome as well as selected sequences from mesophilic forms. Sequences are arranged, from top to bottom, in order of restrictive temperatures. Domains have been labelled under the alignment (black – cofactor binding domain, grey – nucleotransferase domain, green – OB fold domain, brown – zinc binding domain, yellow – HHH domain, red – BRCT domain). Black letters above the alignments indicate conserved nucleotransferase motifs (upper case letters indicate strong conservation of amino acid in that residue, c – charged amino acid, h – hydrophobic amino acid, x- not conserved). Motif shown in red text forms an α-helix that interacts with DNA. Green letter indicates lysine critical for the activity. Orange letters indicate positions of conserved cysteine residues in zinc binding domain.

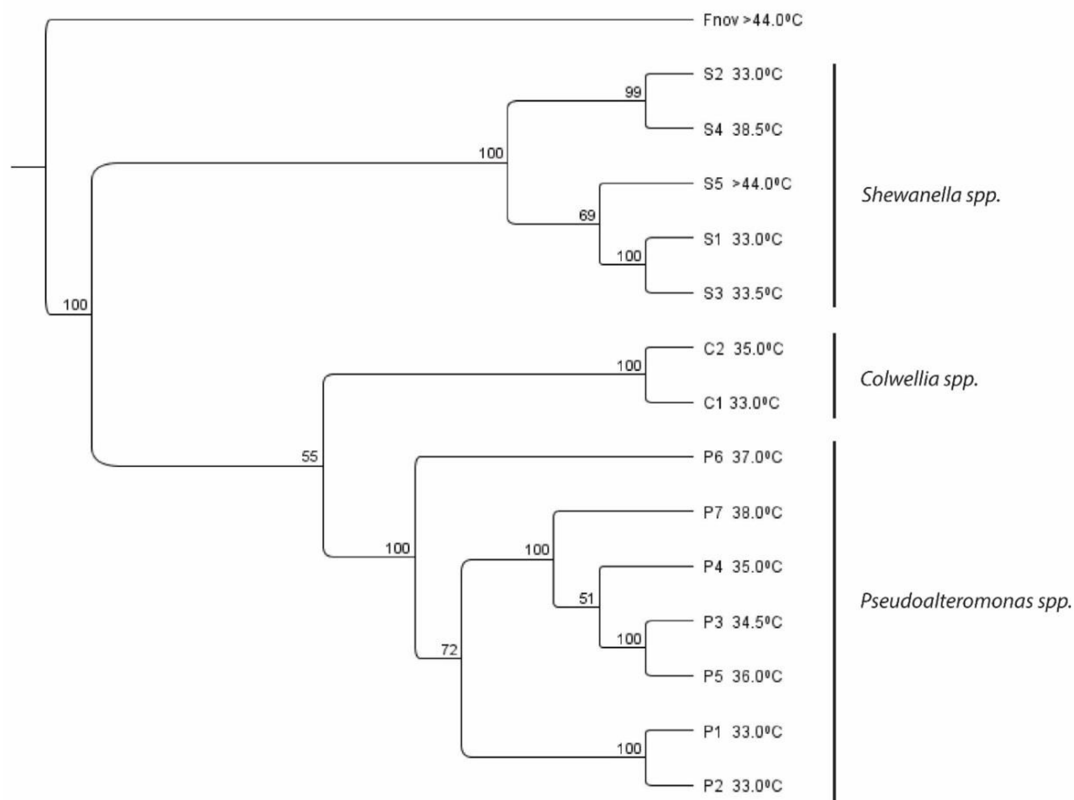

Figure 15 Evolutionary relation of DNA ligase alleles.

Phylogram of the amino acid sequences of the deduced product of *ligA* alleles postulating their evolutionary relatedness. The temperature of inactivation of the *F. novicida* strain harboring the *ligA* alleles is listed next to each allele. S5 is from a mesophilic *Shewanella* species. The phylogram was generated using the neighbor joining method in Geneious version 6.0.6 version. The wild type *ligA* sequence from *F. novicida* was used as an outgroup.

Growth properties of *F. novicida* harboring psychrophilic *ligA* alleles.

We wanted to assess if the substitution of a foreign *ligA* gene affected the growth rate of the hybrid strains of *F. novicida*. Thus, we analyzed their growth in liquid media over a range of temperatures below and above the restriction temperatures of the different strains (Fig. 16). In all of these experiments 30°C was used as the lowest incubation temperature and served as a baseline for the other growth temperature conditions. For the wild type strain the generation time did not change appreciably with incubation temperatures up to 42°C. However each of the TS variants of *F. novicida* grew more slowly at incubation temperatures that were 2°C–3°C below their restrictive temperatures. Progressively higher incubation temperatures correlated with earlier time points when growth slowed or stopped.

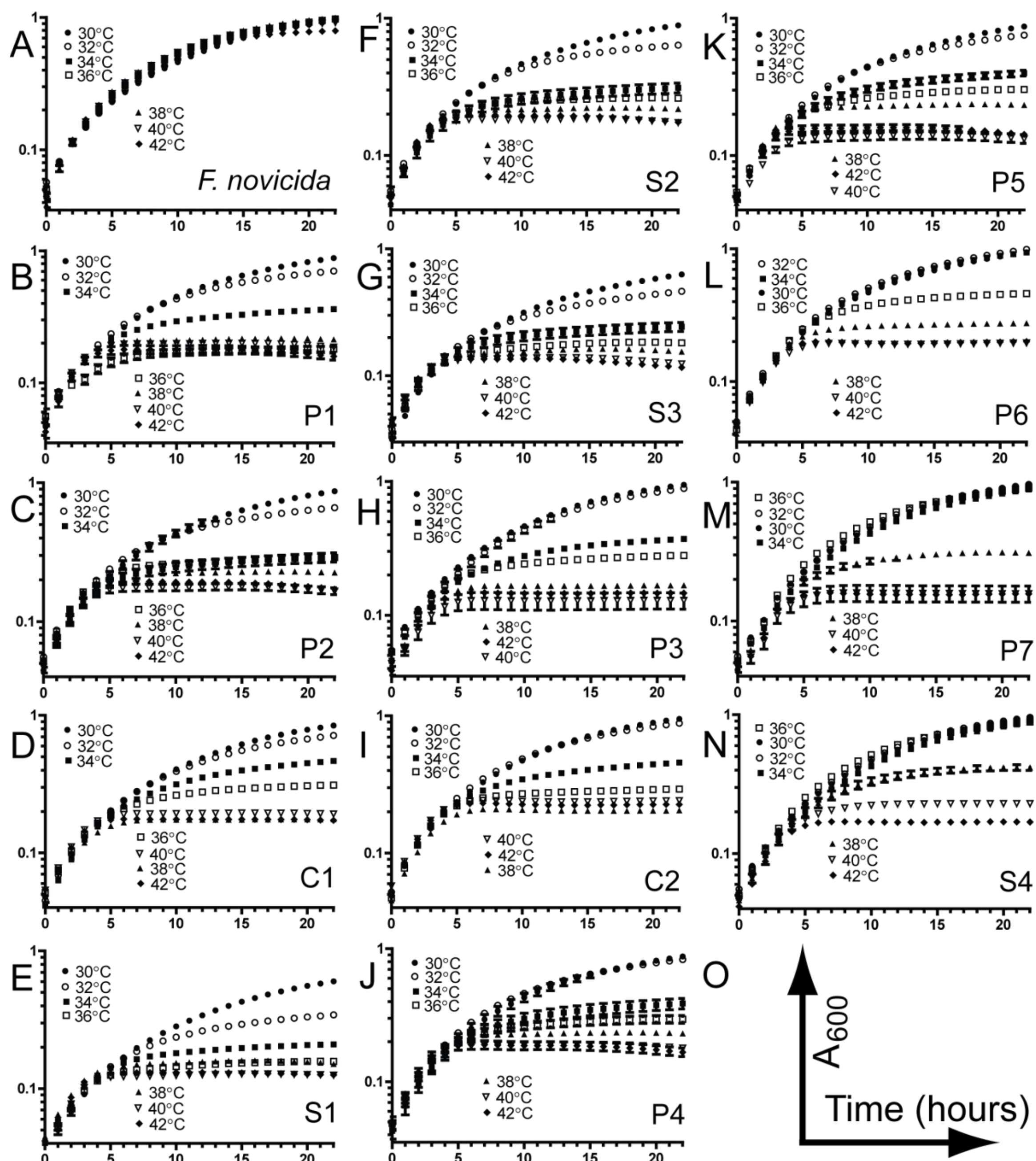

Figure 16 Growth characteristics of *F. novicida* strains harboring psychrophilic *ligA* alleles The strains were cultured at 25°C for 16h, diluted and then grown at the elevated temperatures. The incubation temperatures are listed in each panel from top to bottom to correspond to the order of the growth curves; this allows the reader to discern the position of each growth curve even when two or more curves overlap. Error bars (SEM) are presented but are usually too small to be discerned. The wild type strain displays changes in the growth rate at different temperatures that might be too small to notice.

Mutation frequencies to temperature resistance of *F. novicida* strains carrying *ligA* alleles.

One of the potential advantages of using naturally occurring essential gene from psychrophiles is that millions of years of evolutionary adaptation may have generated some proteins that cannot be converted to a temperature-stable form through a single amino acid change resulting from a single nucleotide change. Of the thirteen TS *F. novicida* strains, four (C2, P4, P6 and P7) failed to generate colonies when greater than 10^{10} CFU were plated on agar medium and incubated 3°C above their restrictive temperatures (Table 3). For the P5 strain we did not find temperature-resistant colonies, but this strain generated only low CFU (~10^{7}) at the permissive temperature, so it is impossible to predict a meaningful mutation frequency to temperature resistance of the P5 *ligA* allele. The *F. novicida* strains harboring the S3 *ligA* allele had an unusually high frequency of mutation to temperature-resistance, but the rest of the alleles that generated temperature-resistant strains mutated at frequencies typical for point mutations. In every case we found the temperature-resistant phenotypes to be stable when re-inoculated onto agar plates.

Psychrophilic *ligA* alleles that support *S. cerevisiae* viability.

The essential genes of different biological kingdoms are represented by widely divergent genes and by different sets of genes. Thus it was surprising that the *ligA* gene of *E. coli* could support the viability of *S. cerevisiae* that had a knock out in its essential *CDC9* DNA ligase encoding gene (DNA ligase type I) (127) even though the bacterial enzyme is NAD-dependent and the yeast enzyme is ATP-dependent. Given these results

we speculated that some psychrophilic *ligA* genes could support *S. cerevisiae* growth and wondered if a TS phenotype would result in such a hybrid strain.

To test our hypothesis we cloned two *Pseduoalteromonas ligA* alleles, P2 and P7 into the centromere vector, pYX132, which is based on tryptophan auxotrophy for selection and utilizes the strong constitutive TPI1 promoter to drive the expression of recombinant genes. As a representative of a bacterial mesophilic gene we cloned the *F. novicida ligA* gene into this vector; and as a positive control we also cloned the yeast *CDC9* gene. These recombinant plasmids were transformed into the *S. cerevisiae* YBSΔL1 haploid strain which has a chromosomal knock out of *CDC9* and carries *CDC9* on a plasmid that is based on uracil selection. After transformation and selection for the recombinant pYX132 (trp-based plasmid) colonies were spread on agar medium containing 5-fluoroortic acid to counter-select against the uracil selection-based plasmid, which removes the only copy of *CDC9*. The surviving colonies were tested for the presence of the original *CDC9*-bearing plasmid, and those lacking the plasmid were studied further. We found recombinants that lacked any copy of *CDC9* and that bore psychrophilic *ligA* alleles P2 and P7, and we found that all of these isolates were temperature sensitive (Fig. 17). However the restrictive temperatures for the *S. cerevisiae* harboring *ligA* alleles P2 and P7 were 34.5, and 36.5°C as compared to the restrictive temperatures for *F. novicida* strains with these alleles which were 33 and 38°C. *S. cerevisiae* strains harboring the *F. novicida ligA* allele or strain in which we introduced the *CDC9* gene on the pYX132 plasmid, grew in the same temperature range as the parent *S. cerevisiae* strain (Fig. 18).

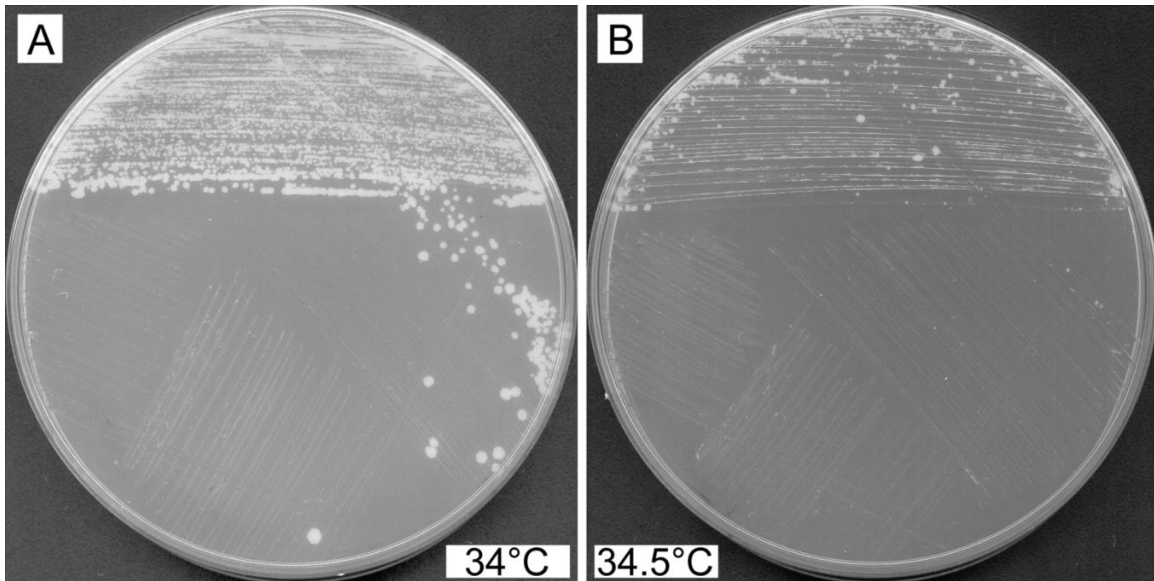

Figure 17 Temperature sensitive phenotype of *Saccharomyces cerevisiae*

S. cerevisiae strain supported by the P2 *ligA* allele. Panel A. Inoculated agar plate incubated at maximal permissive temperature of 34°C. Panel B. Inoculated agar plate incubated at restrictive temperature of 34.5°C.

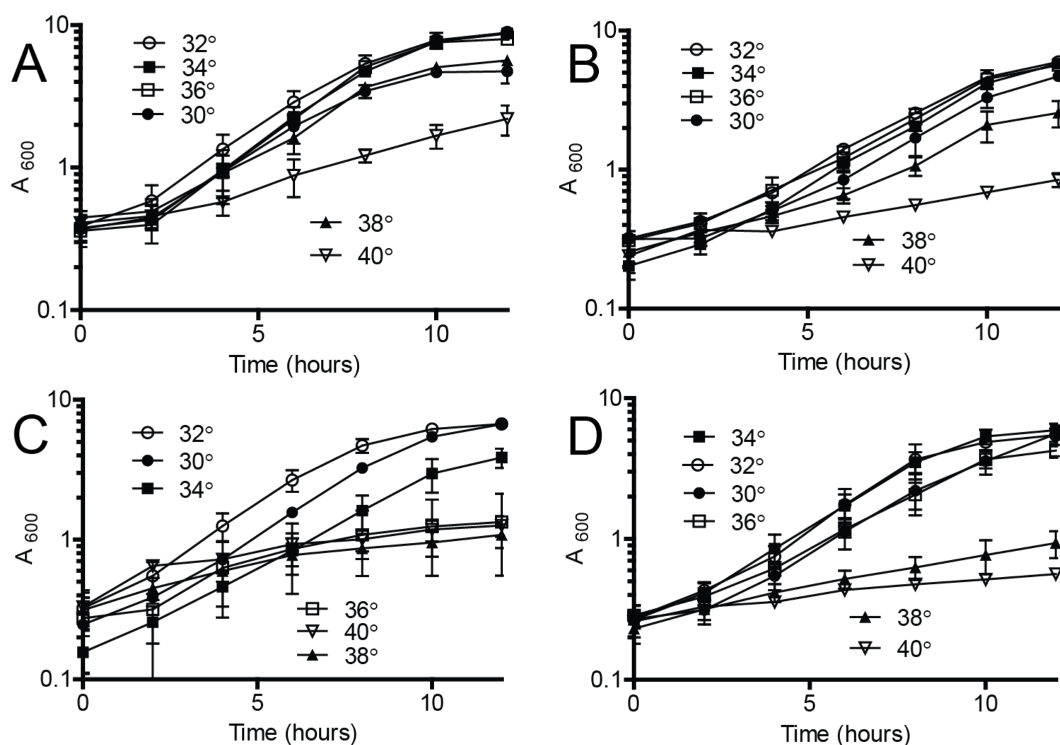

Figure 18 Growth pattern of yeast strains.

Growth of *S. cerevisiae* strains dependent on bacterial psychrophilic *ligA* alleles. The *S. cerevisiae* strains carrying *ligA* alleles CDC9 (Panel A), wild type *F. novicida ligA* (Panel B), *ligA*-P2, (Panel C), and *ligA*-P7 (Panel D) were grown for 16 h at 30°C and diluted to an A_{600} of 0.3 to generate new cultures. Triplicate cultures were grown at each temperature and A_{600} values were measured every two hours. Error bars represent standard errors of the means.

In this work we have identified several psychrophilic *ligA* alleles that can be used as BioBricks to engineer TS variants of prokaryotes and, for some alleles, a eukaryote microbe. The range of temperature-sensitivity induced by the alleles, from 28°C to 39°C, is appropriate for bio-containment while allowing robust growth of the TS microbe at temperatures suitable for bio-process applications and for live vaccines. Since every psychrophilic allele of bacterial DNA ligase proved to generate TS phenotype it shows that these genes can be efficiently used to generate attenuated strains. Furthermore, selected alleles were able to alter the phenotype of yeast strain, meaning they can act as universal control mechanism. Despite the high level of similarity between the sequences, strains that carry different alleles are often characterized by unique phenotypes. Therefore they can be further modified, by altering individual amino acid residues, to fine tune the level of temperature sensitivity in the mesophilic organisms. Previous analysis of multiple TS alleles of essential gene (31) focused on several different essential genes. While establishing the role of psychrophilic genes as potential regulatory mechanism it did not allow for comparative analysis of many alleles of the same gene. With the new collection those residues that affect the protein stability could be identified. This offers greater insight into the nature of TS alleles.

Chapter 4

Analysis of temperature resistant mutants

Introduction

Living organisms have evolved to live at various different environments. This includes extreme conditions, such as Antarctic permafrost or volcanic pools. However adaptation to such harsh conditions does not require completely remodeling the known structure of basic cellular compounds. In fact many of the enzymes found in organisms living at drastically different environments have very similar structure (Fig. 19). At the same time enzyme from a particular extremophile might share only as little as 40% identity of its amino acid sequence with the corresponding enzyme adapted to the opposite type of environment (e.g. proteins from hyperthermophile and psychrophile) (141). Such level of similarity is sufficient to maintain comparable protein architecture. However it also makes the enzymes too distinct to allow easy identification of the crucial changes that affect stability and activity. In addition to that the differences are at least partially affected by the adaptations to other aspects of environment, like high pressure or highly acidic/basic pH. On top of that level of similarity will changed based on evolutionary relatedness of the source organism. Finding a way to detect the changes that could render a given enzyme low- or high-temperature adapted would have a significant impact on biotechnology where necessity to catalyse reactions at specific conditions is crucial. Therefore a significant effort has been placed on localizing amino acid substitutions that affect enzymatic activity in the context of temperature.

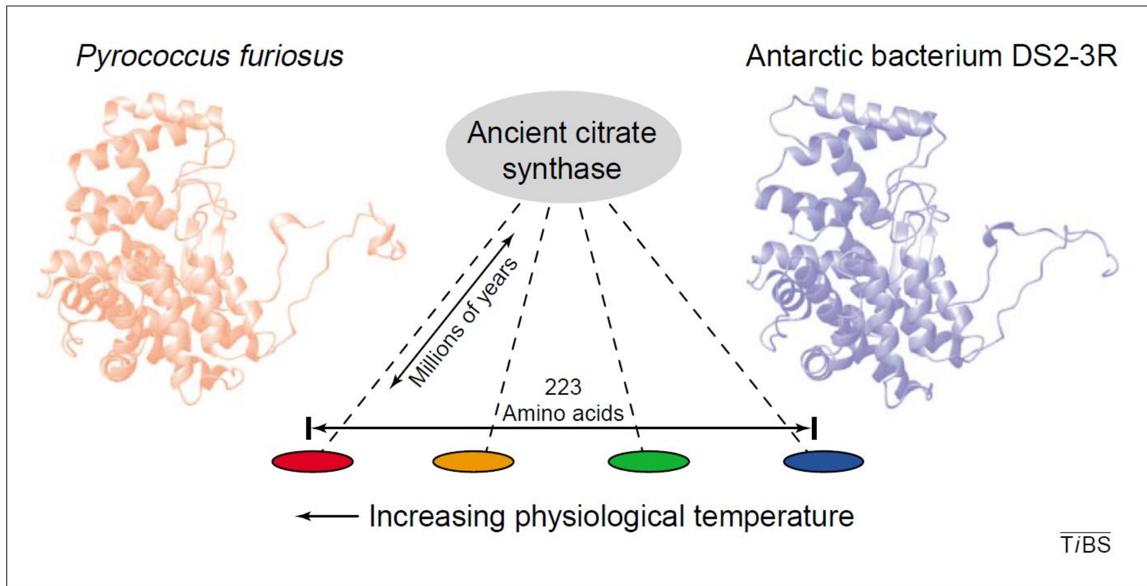

Figure 19 Common ancestry of enzymes adapted to extreme conditions

The illustration presents subunit of citrate synthetase from thermophilic (left) and psychrophilic (right) organism. Despite functioning at drastically different environments both enzymes have retained very similar structure. Figure was adapted from Frances H. Arnold et al (142).

Analysis of multiple psychrophilic and thermophilic enzymes revealed features that might be associated with adaptation to extreme environment. This includes preference for specific amino acids (143, 144). In general it has been noticed that proteins from thermophilic organisms tend to contain more hydrophobic and charged amino acids and less non-charged polar residues than the proteins from mesophiles. At the same time their psychrophilic counterparts possess less aromatic and charged residues and more small amino acids. Such analysis suggests the kind of changes that allows adaptation to different temperatures; however it is insufficient to facilitate rational engineering of proteins by altering individual residues. At this moment it is necessary to establish individual pattern of specific substitutions for groups of proteins that would impact their activity at different conditions.

In order to establish amino acid changes necessary for temperature adaptation researchers have tried to recreate the protein evolution in laboratory conditions. The most common way to achieve this is by utilizing direct evolution techniques. This includes mutagenesis of a particular sequence by error-prone PCR and *in vitro* recombination through DNA shuffling (145). Obtaining functional enzyme with new properties requires analyzing multiple mutants and distinguishing the sequence alterations that contribute to the desired phenotype from the neutral or deleterious ones. Such task is extremely difficult when employing only *in vitro* methods. To ease the process of screening an earlier step of *in vivo* selection can be performed. Mutated sequences are placed in the context of the host organism and clones capable of surviving at desired conditions are further analysed. However for this method to yield effect the investigated enzyme has to perform a biological function that directly impacts the host's fitness. Because of this

requirement the proteins encoded by the essential genes seem to be best suited for such analysis. Any alterations to their activity would have a direct influence on the survival capabilities of the organism. Therefore essential genes make good models for investigating types of amino acid alterations that have impact on protein adaptation to different temperatures.

It could be beneficiary to investigate the mutations that are autonomously selected by organisms carrying temperature-sensitive alleles of their essential genes. In the previous chapter we have described a collection of *F. novicida* clones that depend on DNA ligase originating from psychrophilic organisms, resulting in those clones being incapable of surviving at elevated temperatures. We observed that several of our strains were capable of generating spontaneous temperature resistant mutants, which were able to grow at temperatures that were previously restrictive for them. We decided to investigate the changes in psychrophilic DNA ligases that would affect the observed phenotype of the analyzed strains.

Results and discussion

Mutations that convert TS *ligA* gene products into temperature-resistant forms.

To determine the nature of the mutations leading to temperature-resistance we analyzed the DNA sequence of the *ligA* region from at least two temperature-resistant isolates from each strain that generated such mutants. Seven out of eight strains had a single nucleotide alteration that caused a change in the encoded amino acid in the *ligA* gene product (Table 4). However, all of the three analyzed *F. novicida* strains carrying *ligA*-P2 had no changes in the *ligA* coding sequence nor in the nearest upstream and downstream region. Additionally while we found three mutated forms of the *ligA-S3* allele, we have also isolated four temperature resistant *F. novicida*-S3 strains with no changes in their *ligA* ORF (Table 4), indicating that extragenic suppressors altered their maximal growth temperatures.

Since it appeared that some of the temperature-resistant isolates derived their phenotypes from extragenic suppressor mutations we wanted to test if the intragenic *ligA* mutations found in most of the isolates truly generated their temperature-resistant phenotypes. To do this we PCR amplified the *ligA* regions from five temperature-resistant isolates derived from temperature-sensitive strains of *F. novicida* carrying psychrophilic *ligA* substitutions and used the PCR products to transform *F. novicida*-C2 (max 35°C) and selected for growth at 37°C. In all cases the temperature-resistant transformants carried the *ligA* mutations associated with the *ligA* from the temperature-resistant donor strains. The results from one such experiment are shown in Figure 20.

Table 4 Examples of mutations found in temperature-resistant mutants

	Amino Acid Change	Codon Change	Protein Region
P1	A520V	GCA→GTA	HHH domain
P2*	None found	None found	--
C1	E475K	GAA→AAA	HHH domain
S2	E433K	GAA→AAA	zinc binding domain
S3[¥]	M8I	ATG→ATA	N-terminal end
S3[¥]	D9N	GAT→AAT	N-terminal end
P3	A520V	GCT→GTT	HHH domain
S4	D550G	GAT→GGT	HHH domain
S4	E518G	GAA→GGA	HHH domain

*Presumably mutations extragenic to *ligA* altered the restrictive temperature.

[¥]Some temperature-resistant strains carrying the S3 allele had no mutations in the *ligA* gene, and extragenic mutations were presumably responsible for the altered phenotype.

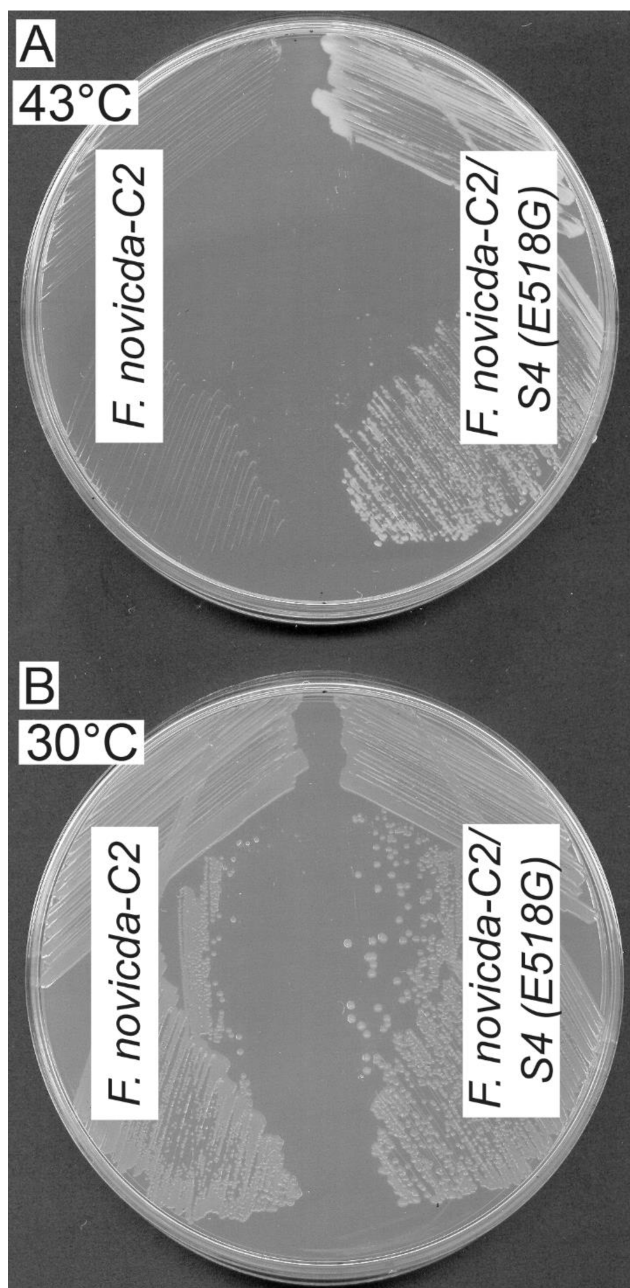

Figure 20 Relation between temperature resistant phenotype and observed mutations

Introduction of a temperature-resistant variant of a psychrophilic allele (S4-E518G) into a temperature-sensitive *F. novicida* strain results in temperature-resistant isolates. The TS *F. novicida*-C2 strain was transformed with a PCR amplicon of the S4 (E518G) *ligA* allele and surrounding DNA. The one temperature-resistant colony that was examined carried the S4 (E518G) *ligA* allele. Sequencing confirmed presence of S4 (E518G) allele and lack of the original psychrophilic one

The majority of intragenic temperature resistant mutations occurred in the C-terminal region of DNA ligase protein, including the zinc binding domain and the HHH region (Fig. 21) both of which are involved in binding DNA. A Phyre2 structure prediction of temperature-resistant S4 LigA placed both the E518G and the D550G changes in loop regions between two α-helices. The A520V changes found in temperature-resistant variants of LigA-P1 and LigA-P3 were predicted to lie in a portion of the enzyme that has no predicted secondary structure. The E433K mutation in LigA-S2 occurred at the beginning of a predicted α-helix in the zinc binding domain adjacent to one of the cysteine that are involved in binding zinc ion (Fig. 22).

Figure 21 Location of mutations that converted TS psychrophilic alleles into ones that produced temperature-resistant products

Blocks signify the protein domains of bacterial NAD-dependent DNA ligases. Ia, cofactor binding domain; NT, nucleotransferase domain; OB, OB fold domain; Zn, zinc binding domain; HHH, helix-hairpin-helix domain; BRCT, BRCT domain. The double headed arrow indicates the mutations that occurred at the very N-terminal end of the protein and did not have a consistent character other than to increase the strength of the RBS; the circular arrows indicates the mutations changing glutamic acid into lysine; the single arrow indicates the mutation changing alanine into valine; and the square arrows indicate mutations changing acidic residue into glycine.

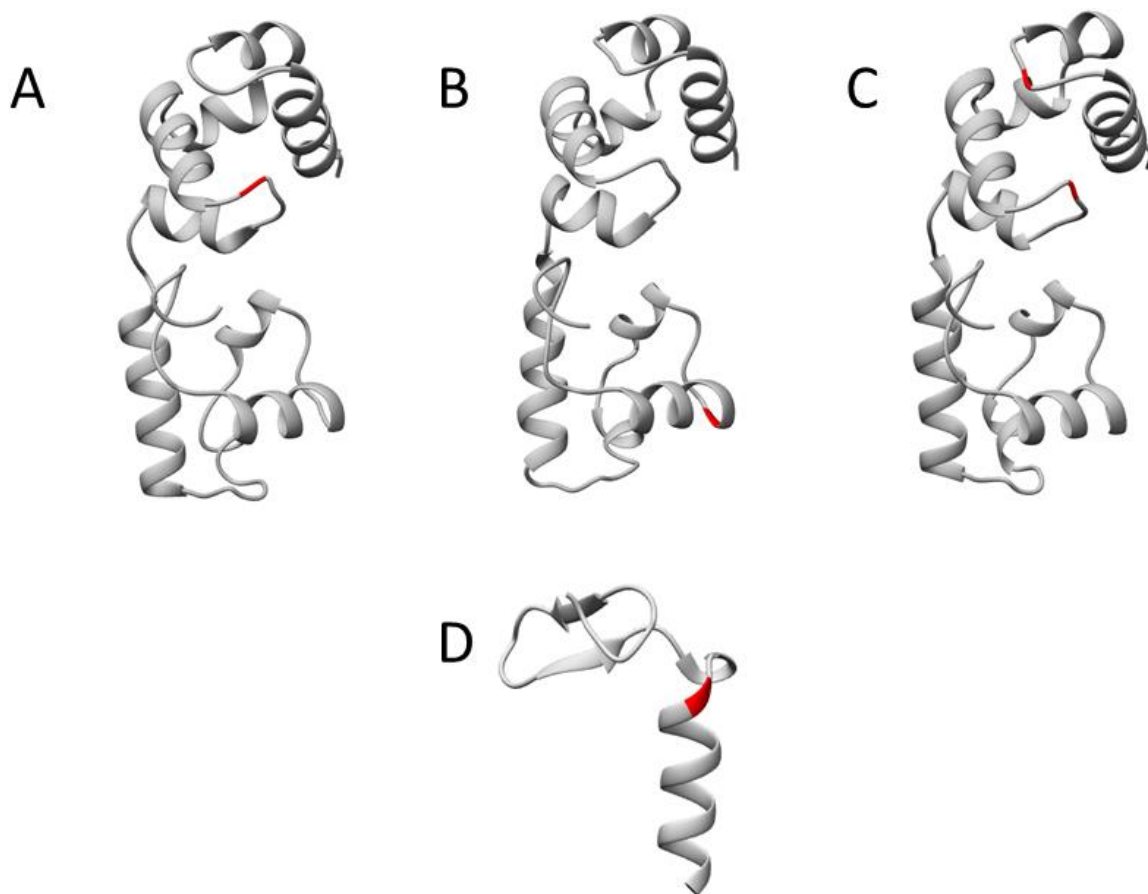

Figure 22 Location of mutations that converted TS psychrophilic alleles into temperature-resistant products, within the local structures of protein

The complete structures of the proteins were modeled based on their sequences. The local structures where mutations have occurred are presented on the illustration. The mutated residues were labelled in red within the HHH domain of allele P1 (Panel A), C1 (Panel B) and S4 (Panel C). The mutation identified within the zinc binding domain of S2 (Panel D) was annotated in the same way.

Temperature-resistant variants of ligA-S3 had mutations that introduced changes at the N-terminus of the LigA product (Table 5), and we were surprised to find that the mutations were identical to those previously found in temperature-resistant mutants of *ligA*-S1 (31). We noted that both of the S1 and S3 alleles had the lowest predicted RBS strengths (86 au) of all of the alleles, and wondered if the mutations to temperature-resistance affected the strengths of the RBSs. We found that indeed both the S1 and S3 predicted strengths were increased at least three-fold (317-1,531 au, Table 6) by the mutations found in the temperature-resistant variants. The mRNA of the S1 and S3 alleles is predicted to form secondary structure that encompasses the beginning of coding sequence and the upstream region that includes the RBS, which could affect the gene translation (Fig. 23). At the same time all identified mutations have visibly negative effect on its stability (Table 6). If the temperature-resistance phenotype of the S3 variants is due to a change in mRNA structure rather than a change in protein structure then this type of mutation should be regarded as a form of "RBS strength suppressor" mutation analogous to multicopy suppressor mutations that result from having many copies of a particular gene.

Table 5 Effect of mutations on predicted RBS strength.

Allele	Amino Acid/nucleotide change	Predicted Translation Initiation Rate (au)
F. novicida	N/A	748.4
S1 (parental)	N/A	85.9
S1†	P3S/C7T	316.8
S1†	Q5K/C13A	496.9
S1†	M8I/G24A	1530.7
S1†	D9N/G25A	1117.1
S3 (parental)	N/A	85.9
S3	M8I/G24A	1530.7
S3	D9N/G25A	1117.1

Calculated strengths* of Ribosome Binding Sites for temperature-resistant psychrophilic *ligA* alleles with changes near the 5'-end of the *ligA* ORF.

†These amino acid changes were previously reported by Duplantis et al., Proc. Natl. Acad. Sci. U. S. A. 107:13456–60.

*Calculations for RBS strengths performed by algorithm available at https://salislab.net/software/forward. For each calculation of the RBS strength 50 bases upstream and downstream of the ATG start codon were entered into the "reverse engineering" version 2.0 algorithm. The anti-Shine-Delgarno sequence of *F. novicida* is identical to that of *E. coli* K12.

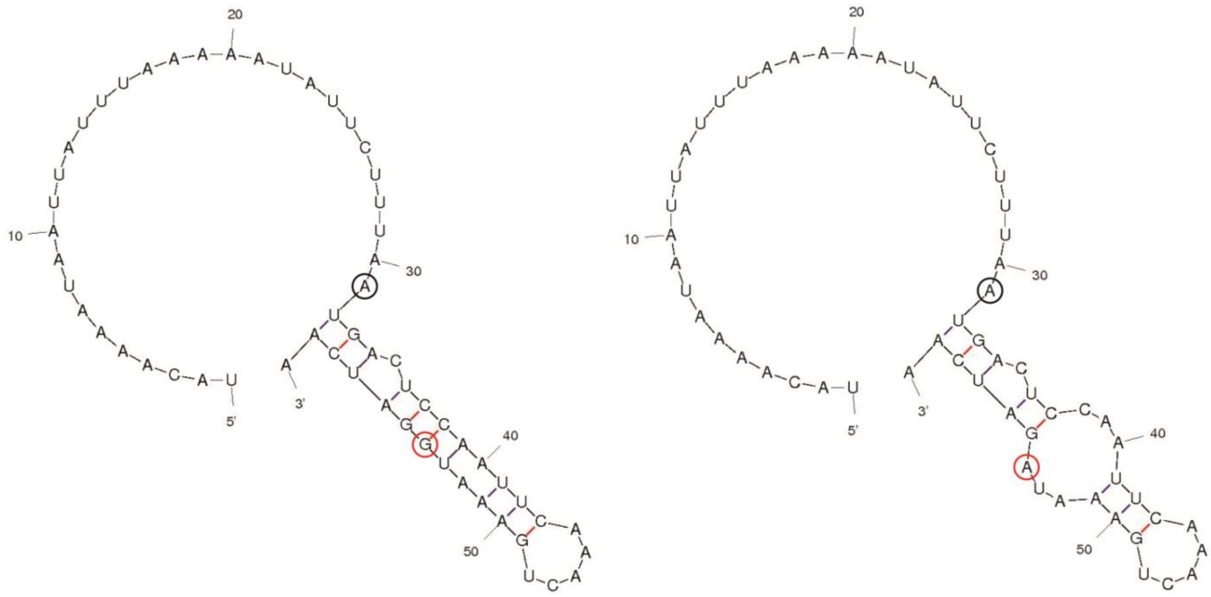

Figure 23 Secondary structure of mRNA from *ligA* alleles.

Predicted structure of mRNA of temperature sensitive allele S3 (left) and its temperature resistant mutant M8I (right). The adenine in the first ATG codon was circled in black. The mutated nucleotide has been circled in red. The structures were prepared by using mfold program (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form).

Table 6 Stability of *ligA* alleles mRNA

Allele	Amino Acid/nucleotide change	ΔG (kcal/mol)
S1	N/A	-8.0
S3	N/A	-8.0
S1	P3S/C7T	-5.1
S1	Q5K/C13A	-4.1
S1	M8I/G24A	-1.6
S1	D9N/G25A	-2.3
S3	M8I/G24A	-1.6
S3	D9N/G25A	-2.3

Calculated stability of secondary structure of mRNA of alleles S1 and S3, as well as their temperature resistant mutants. The structures were prepared by using mfold software (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form).

Growth properties of *F. novicida* harboring psychrophilic *ligA* alleles.

When substituting the *ligA* alleles into *F. novicida* we preserved the *F. novicida* transcriptional control, and we tried to minimize disruption of mRNA structure surrounding the ribosome binding site (RBS) regions. However the fusion of the *F. novicida* region distal to the *ligA* promoter to the psychrophilic *ligA* ORF inevitably resulted in a new mRNA and a new RBS-environment. Our calculations (Table 7) of the predicted strengths of the RBSs for the psychrophilic *ligA* alleles showed that most of them were within a factor of two from the original, *F. novicida*, *ligA* RBS. The substitution of the S1, S3 and C2 alleles substantially lowered the strength of the RBSs but, in general, there is not an apparent relationship between the predicted strength of the RBSs lying in front of the psychrophilic *ligA* alleles and the growth rates of the strains harboring those alleles (Fig. 24).

Table 7 Analysis of bacterial growth and RBS strength.

	Calculated translation Initiation Rate (arbitrary units) of allele fused to *F. novicida* translation start regions	Generation time at 30°C of *F. novicida* strains harboring psychrophilic *ligA* alleles.
P1	496.9	170 min
P2	496.9	169 min
C1	748.4	166 min
S1	85.9	170 min
S2	1752.0	167 min
S3	85.9	163 min
P3	386.2	175 min
C2	176.5	170 min
P4	650.9	170 min
P5	954.3	175 min
P6	475.0	169 min
P7	496.9	156 min
S4	680.9	175 min
S5	1917.0	ND
F. novicida	748.4	176 min

Calculated strengths* of Ribosome Binding Sites for psyhrophilic *ligA* alleles substituted into the *F. novicida* chromosome and the growth generation time for cognate strains.

*Calculations for RBS strengths performed by algorithm available at https://salislab.net/software/forward. For each calculation of the RBS strength 50 bases upstream and downstream of the ATG start codon were entered into the "reverse engineering" version 2.0 algorithm. The anti-Shine-Delgarno sequence of *F. novicida* is identical to that of *E. coli* K12.

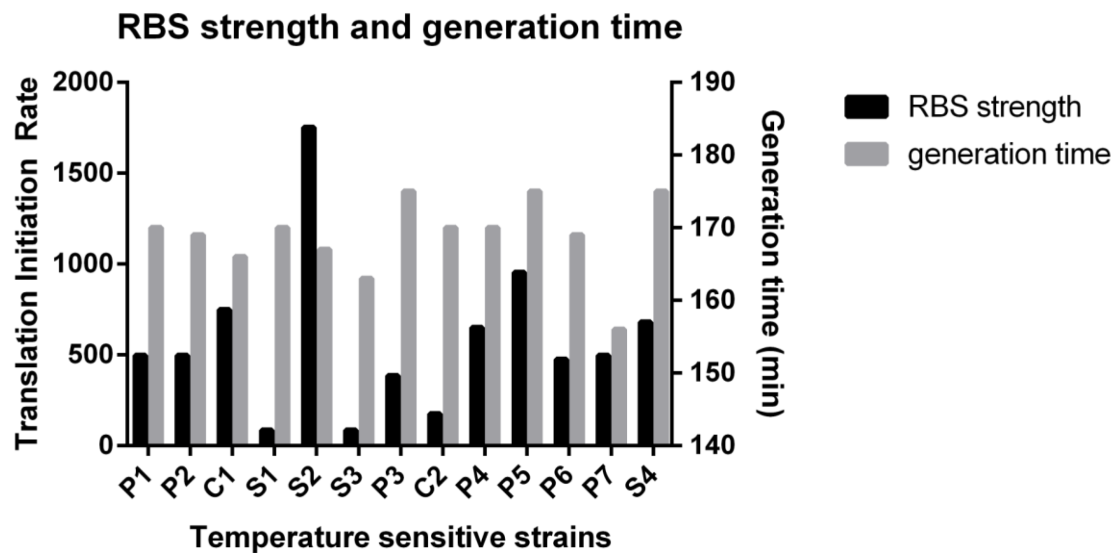

Figure 24 RBS strength and generation time

We observed lack of relationship between the predicted strength of the ribosome binding site (RBS) and the growth rate of *F. novicida* strains carrying substituted psychrophilic *ligA* alleles.

Temperature resistant mutations in the context of *ligA* sequence similarity.

The mutations localized in temperature resistant mutants have occurred in specific locations within the coding sequence. All of the isolated clones of *F. novicida* carrying the DNA ligase amplified from *Pseudoalteromonas* sp. had A520V substitution (Table 4). Interestingly all of the analyzed alleles from that group of psychrophiles have alanine in position 520. Despite a high level of similarity between these sequences only some of them managed to generate stable clones with this substitution. At the same time valine is found among all the other psychrophilic sequences in the same position. The region where mutation occurred is otherwise highly conserved even between psychrophiles and mesophiles. This would suggest that the mutations require presence of other residues that can be found only in certain variants of *ligA* gene.

Just like the previously described example other C-terminal mutations also occurred in mostly well conserved regions. Moreover the sequence alterations that were found in temperature resistant mutants do not overlap with the differences between TS alleles that have low and high restrictive temperature. Thus it seems that mechanism of generating temperature resistant variants found a different means of stabilizing the protein structure at elevated temperatures.

Discussion

The investigation of TS strains of *F. novicida* carrying the psychrophilic alleles of DNA ligase revealed the stability of an artificially generated phenotype. There are several types of changes that could occur within the mesophilic host that would allow escape from the lethal effect of the safety mechanism. These have to be considered when using the alleles in real-life situations. The mutations observed within the C-terminal region of

the proteins seem to be typical alterations that prevent the loss of activity of the enzyme at suboptimal conditions. They follow a specific pattern of amino acid substitutions that depends on the residue position e.g. acidic residues changing to glycine in the terminal part of HHH domain (Fig. 21). It can be presumed that these mutations are equally likely to occur in different hosts organisms and therefore the frequency of their occurrence would be similar to the already observed one.

A different type of alteration that generates temperature stable phenotype has been observed in the N-terminal region of protein. These mutations do not follow a clear pattern of amino acid substitutions. Instead they are likely affecting the gene expression at the mRNA level, by impacting the strength of the RBS site. Therefore these mutations should depend on the genetic context of the allele and will not necessary appear in different host genome. On the other hand the alleles that previously displayed no tendency to mutate could spawn similar clones when placed in a different context. The reason why this type of mutation is generated with the highest frequency might come from the advantage given at the permissive conditions. The clones that obtain the mutation affecting the RBS strength can outperform the clones carrying the native allele. Therefore later they form a significant portion of the final population. Since these mutations tend to appear at very high frequency it is important to analyse any new TS organism for its tendency to undergo such changes.

The last type of the observed temperature resistant clone had no mutations within the *ligA* coding sequence or the flanking DNA sequences. This suggests that an extragenic mutation was responsible. This kind of change might be specific only for certain hosts and therefore it is impossible to know how likely it is to occur in a different

genome. These mutations occurred at a very low level in the strain carrying allele P2 (Table 4). They also occurred in the strain S3 where they constitute a majority of temperature resistant mutants generated by that strain. Therefore the frequency of extragenic mutations might be also allele specific. Individual investigation would be necessary to determine whether a new host is prone to such alterations.

Chapter 5

Analysis of genome from extreme psychrophilic *Colwellia* sp. C1

Introduction

Psychrophilic bacteria are an environmental group that has been adapted to thrive at low temperatures. Typically the psychrophilic bacteria are capable of growing in a range of temperatures from approximately -20^{0}C to 20^{0}C. The most extreme psychrophiles, which can survive at lowest temperature known to still support growth, have been found in permafrost soil and sea ice. Adaptation to low temperature environment has been characterized among members of various different groups of eukaryotic organisms such as fungi and algae. However the greatest number of psychrophiles has been characterized in the domain of bacteria. Psychrophilic bacteria can be found among both gram-negative and Gram positive phyla. Many psychrophilies are also adapted to low concentration of nutrients and high baric pressure.

In order to survive at low temperatures the psychrophilic bacteria have developed a number of specific traits. Some of these are synthesis of cryoprotecting agents, antifreeze proteins and more labile membranes or altered amino acid composition of their proteins. Altogether these elements allow psychrophiles to avoid cell damage caused by the growing ice shards or inactivation of essential metabolic functions because of the low temperature. These conditions would be lethal for regular mesophilic organisms that have not been adapted to such habitats.

The enzymes synthesized by the psychrophilic bacteria have evolved to maintain their catalytic activity despite the extreme conditions. The major obstacle that they are facing is caused by the slower pace of chemical reactions at the low temperatures. Such changes might have a devastating impact on the cell homeostasis when many processes have to occur simultaneously and the correct ratio of reaction velocity between linked metabolic pathways is essential. To suppress the negative effect of low temperature the psychrophiles utilize enzymes characterized by a less rigid conformation (146). Since their proteins have more flexible structure, they can more efficiently catalyze their specific reactions. However, because of this kind of adaptation they have become more unstable and prone to heat denaturation. This suggests that while they can perform superbly at low temperatures, their activity at elevated temperatures could be completely lost.

Various different enzymes produced by psychrophilic organisms were found to be inactivated at moderate temperatures (101). Therefore an organism that would have to rely on such labile protein would risk losing its activity if growing at conditions of increased temperature. Such temperature-sensitive phenotype can be extremely useful for investigating the specific function of genes. This is especially true for essential genes, which cannot be deleted without killing the organism. A default method of obtaining genes encoding temperature-sensitive proteins was to generate altered sequences through mutagenesis, utilizing UV radiation and chemical agents. The random changes in their sequence could create alleles encoding proteins that are still functional, yet less stable. As a result they would lose their activity at increased temperatures. This approach allowed generation of multiple mutated versions of genes from eukaryotes, bacteria and

viruses/phages. However temperature-sensitive mutations are not easy to select from a pool of randomly altered sequences. This causes a limitation to how easy new temperature-sensitive alleles can be created. On top of that the few amino acid changes that were introduced to make protein temperature-sensitive can be reverted by naturally occurring mutations. Finally the obtained mutants usually are not being screened for high processivity at the permissive conditions. As a result they are often characterized by an overall decrease in enzymatic activity at all temperatures. Alternative method of generating temperature-sensitive mutants involves proteins that have evolved in the environment where high stability is not selected for, yet high efficiency at low temperatures is necessary. Because of this protein adaptation, the modified strains that utilize psychrophilic genes should not suffer from an inhibition under the permissive conditions.

The genomes of psychrophilic bacteria are characterized by several specific features that distinguish them from other groups of living organisms. One of such elements is the low GC content of their sequences. Usually psychrophilic genome has 30% - 40% GC pairs, which is significantly lower than the typical genome of mesophilic bacteria. Psychrophiles also have preferences for specific amino acids in their coding sequences e.g. preferring lower content of charged amino acids. It is not certain if all of these elements are a result of adaptation to low temperature environment. Many of them are not strictly present among all of the cold adapted organisms. These traits are also not unique to psychrophiles. However they are commonly associated with the psychrophilic organisms.

Essential genes are the genes that provide crucial role in the organism and are indispensable. Therefore loss of any of them results in cell death. This is because they control the processes that lie at the base of cellular metabolism. The number of essential genes in a specific organism may vary depending on its complexity. Singe cell organisms will have fewer genes than multicellular ones. Also prokaryotes tend to have fewer genes than eukaryotic organisms. Based on detailed analysis of *E. coli* genome it is currently believed that a minimal bacterial genome contains ~300 essential genes (35, 36). This gene set contains sequences responsible for such reactions as DNA replication, transcription, translation or biosynthesis of important metabolites. Because of the essential role of these elements at least one functional copy of each gene has to be present in the host's genome in order to sustain growth.

The stability of psychrophilic proteins usually would decrease at temperatures typical for the mesophiles. The more flexible structure gives them less resistance to denaturing factors such as heat or chemical agents (101). Since the rate of chemical reactions is lower in the cold environment the bacteria have to find means to compensate for that. Therefore their enzymes are usually characterized by a more flexible structure that increases their processive abilities. As a downside they will usually loose most of their activity when exposed to temperature above 25^{0}C - 30^{0}C. It is disputed whether the decreased stability is a direct result of increased flexibility or is it caused by evolutionary drift allowing accumulation of temperature-sensitive mutations.

Results and discussion

Characterization of partial genomic sequence of *Colwellia* sp. C1.

We have obtained approximately 0.4 Mbp of a total sequence from the analysis of the genome of psychrophilic *Colwellia* sp. C1. The annotated fragments possess 354 predicted open reading frames (ORFs) (Fig. 25). Identified genes displayed a high level of similarity to known sequences of psychrophilic bacteria, especially to *Colwellia piezophila* (147). This organism is a described psychrophile with optimum growth temperature of 10°C. Based on the sequence similarity we suspect that many identified ORFs are involved in the saccharide metabolism, including polysaccharide biosynthesis, and assembly of outer cell layer. This would suggest that the investigated strain has adapted to the low temperature environment by developing features typical for the psychrophilic bacteria. In similar fashion all genomic fragments were characterized by low GC content, which is another feature typical for psychrophilic microbes (Table 8).

In several cases we found ORF transcribed in the same direction, with similar predicted metabolic functions. It would suggest that these genes are organized into an operon and are being expressed as a single mRNA molecule. Based on the analysis some of the predicted operons would be involved in biosynthesis of important organic compounds, such as histidine and heme. Additionally some of the putative glycotrasferases and sugar epimerases were organized together with surface proteins in a similar fashion.

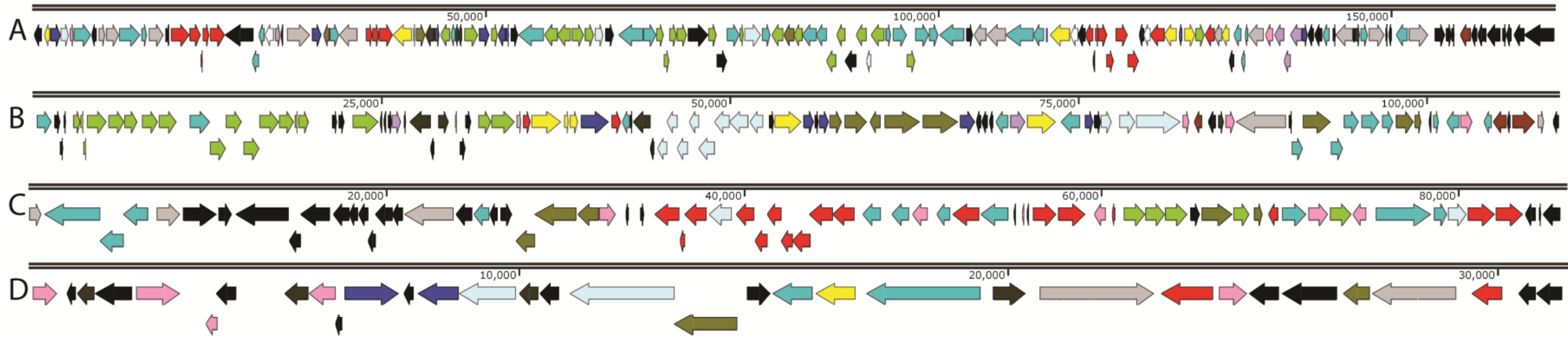

Figure 25 Predicted ORFs found in the sequenced fragments of *Colwellia* sp. C1 genome

(A) contig #1, (B) contig #2, (C) contig #3, (D) contig #4. Genes were colour coded based on their predicted function. Red ORFs indicate genes that are predicted to be involved in biosynthesis of cofactors and nucleotides. Yellow ORFs indicate genes involved in DNA replication and recombination as well as oligonucleotide processing. Dark yellow ORFs indicate genes that are predicted to be involved in biosynthesis of lipids. Light green ORFs indicate genes that are predicted to be involved in biosynthesis, modification and degradation of saccharides (including polysaccharides). Dark green ORFs indicate genes involved in cell persistence and detoxification. Light blue ORFs indicate genes involved in amino acid biosynthesis. Sea blue ORFs indicate genes that encode membrane bound and surface proteins, like transporters. Dark blue ORFs indicate genes involved in translation, protein modification and degradation. Pink ORFs indicate genes that encode transcription regulators and elements of transcription machinery. Violet ORFs indicate genes involved in cell division. Brown ORFs indicate genes that encode transposases. Gray ORFs indicate genes that encode proteins involved in signal transduction, like two-component systems. Black ORFs indicate hypothetical proteins with no predicted function.

Table 8 Characteristic of analyzed fragments of genomic DNA from *Colwellia* sp. C1

	contig #1	contig #2	contig #3	contig #4
size (bp)	168027	109592	86007	31443
GC%	37%	37%	39%	36%
genes	146	104	71	33
essential genes	14	9	0	1
tRNA genes	1	0	0	0
amino acid synthesis	4	11	2	2
carbohydrate metabolism	21	20	5	0
DNA metabolism	6	5	0	1
cell surface proteins	28	13	11	2
nucleotide and cofactor synthesis	16	2	17	2
protein synthesis and folding	9	5	0	2
persistence mechanism	4	4	0	4
cell division	4	2	0	0
lipid metabolism	3	8	5	2
transcription factors	2	3	6	5
signal transduction	12	3	3	2
transposases	1	3	0	0
energy transfer	3	0	0	0
hypotetical protein genes	33	25	22	11

The identified genes were segregated based on their predicted function. Altogether 14 different groups of ORFs were recognized among the characterized sequences.

We aimed to identify alleles of the essential genes among the predicted ORFs. Such genes have been used to generate temperature sensitive (TS) mutants of mesophilic organisms (31), which can act as a biosafety mechanism for laboratory strains. We assumed a gene to be essential based on the analysis performed in *E. coli* mutants from the Keio collection and PEC database (35, 36, 148). This way the identified protein could be used to generate TS mutant of known Gram negative bacteria. Searching the sequence revealed 24 genes that meet our criteria (Table 9). The identified sequences represent genes involved in the crucial processes that control cellular metabolism. Among those are DNA replication, transcription, translation, cell division and heme biosynthesis. We have also found two homologues of phosphoglucosamine mutase (*glmM*), which is an enzyme involved in the UDP-N-acetylglucosoamine biosynthesis pathway. Therefore it is an important element responsible for the assembly of bacterial cell wall. It is possible that one of these copies is inactive or the cell requires activity of both of them in order to sustain its requirements.

One of the identified alleles has already been proven to generate temperature-sensitive phenotype when placed in the genome of mesophilic bacteria. The allele of DNA ligase (*ligA*) amplified from the *Colwellia* sp. C1 genome would support the growth of *Salmonella enterica* strain with deletion of its native *ligA* sequence (149). However the resulting strain would not be capable of surviving at a temperature of 27°C or above. This is one of the lowest restrictive temperatures imposed on bacterial strain by simple allele exchange. It can be speculated that other essential genes from the genome of *Colwellia* sp. C1 would be capable of generating TS phenotype when integrated into the genome of mesophilic bacteria.

Table 9 Possible essential genes identified in the genome of *Colwellia* sp. C1.

gene name	contig	strand	gene start	gene end
orn	#1	—	1909	1364
rsgA	#1	+	2090	3139
psd	#1	+	3166	4092
yihA	#1	+	26856	27491
gpsA	#1	—	46108	45092
glmM	#1	—	57994	56600
hemC	#1	+	117755	118702
hemD	#1	+	118686	119474
rho	#1	—	126303	125044
hemB	#1	—	130510	129512
rpoH	#1	—	136980	136126
ftsX	#1	—	138172	137198
ftsE	#1	—	138893	138165
ftsY	#1	—	140164	138932
glmM	#2	+	3926	5368
dnaX	#2	+	35845	37932
adk	#2	+	41592	42236
yeaZ	#2	+	55317	56039
fabG	#2	—	60785	60048
zipA	#2	+	70139	71206
ligA	#2	+	71342	73414
dapB	#2	+	76661	77467
lptB	#2	+	90349	91167
dapE	#4	—	9957	8794

Each gene was named after the known essential genes that it was most similar to.

Generation of *Francisella novicida hemC* TS strain.

As a proof of concept we wanted to show that other identified essential genes could be used to generate TS strains of mesophilic organisms. For this purpose we have selected *hemC* gene encoding prophobilinogen deaminase, a crucial enzyme in the heme biosynthesis pathway. While depletion of an essential DNA ligase results in almost immediate cell death (58), the prophobilinogen deaminase participates in a synthesis of chemical compound. Therefore it is the lack of heme that would impact the cell fitness. Because of this the cell death should be delayed, giving the organism a chance to adapt its metabolism. We expect that the observed phenotype would be different from that of *ligA* mutant. In order to test that hypothesis, we prepared genetic construct designed to replace the native *hemC* sequence in the chromosome of the mesophilic bacterium *Francisella novicida* with our psychrophilic allele. We had previously utilized *F. novicida* as a host for the psychrophilic genes since its low G+C content (33%) would closely resemble that of a psychrophilic organism. The allele substitution construct was prepared in a way that would allow the introduced *hemC* gene to remain under the control of the native *F. novicida* promoter and ribosome binding site. To prevent any significant changes of the mRNA structure in the RBS region after the gene substitution, we placed the first three codons of the *F. novicida hemC* ORF in place of the corresponding codons from the psychrophilc gene.

The introduced *hemC* was successfully substituted into *F. novicida*, resulting in temperature-sensitive (TS) form of *F. novicida* strain with restrictive temperature of 36°C. The *hemC* gene from generated mutant was sequenced to confirm the presence of the psychrophilic sequence. In order to confirm that the change in phenotype was caused

by the exchange of alleles and not due to mutations outside of the *hemC* gene we returned the native *F. novicida hemC* homologue to the TS strains. The *F. novicida* strain used as a host in our studies can be easily transformed with linear DNA. Therefore we PCR amplified the wild type *F. novicida hemC* allele along with approximately 1 kb of flanking DNA regions. This amplicon was used to transform the TS strain of *F. novicida.* The transformed cells were plated above the restrictive temperature of the TS strain. After overnight incubation we observed a lawn of viable clones on the plate. Parallel experiments without added DNA yielded no growth at the same temperature. To further confirm the impact of our allele on the observed phenotype we isolated DNA from three of the temperature resistant clones and used it as templates for PCR amplification of the *hemC* region. The resulting amplicon was subjected to DNA sequence analysis. All of the investigated clones showed the presence of the *F. novicida hemC* gene and the absence of the psychrophilic gene.

Frequency of mutation to temperature resistant phenotype in *F. novicida hemC* TS strain

As mentioned before by using the naturally occurring essential gene from the psychrophilic organism that has adapted to a low temperature environment during millions of years of evolution we limit the possibility of mutations that could converted the TS strain to a temperature-stable form through a single nucleotide alteration that causes change of a single amino acid residue. We tested our temperature sensitive strain of *F. novicida* for its ability to generate temperature resistant forms. We discovered that, even when incubated above its restrictive temperature, plated strain forms patches of

solid growth on the plate. It might be concluded that this is the effect of slower death caused by the lack of heme. This effect is different from the death caused by TS DNA ligase, where plating even large number of cells resulted in clear plates. Since we could not obtain individual colonies to calculate the reversion rate we decided to modify the original method of analysis. As before the cells were grown at permissive temperature to a large number. Then a small amount of culture was spun down and suspended in ten times greater volume of PBS buffer. The suspended cells were incubated above their permissive temperature for 1 hour. Then they were plated on pre-warmed TSBC plates and incubated at the same conditions, approximately 4°C above their restrictive temperatures. This time we were able to observe individual colonies. While there were remains of a hazy growth on some regions of plates, it did not obstruct the process of counting colonies. We calculated that the strain is capable of reverting to temperature stable form at a high rate of ~1 cell in 10^5. We selected several colonies and streaked them again above the restrictive temperature. They were capable of forming regular growth, which confirms their viability. At the same time the hazy growth observed on some plates could not form any visible colonies when re-streaked at the same temperature. This result proves that the hazy regions did not contain any living cells. It, most likely, consists of cells that were able to divide several times before dying as a result of temperature inactivation of an essential gene. This is the expected result of delayed lethality. The cells are not directly affected by the lack of functional prophobilinogen deaminase, only the lack of product of its biosynthesis pathway.

The observed reversion rate is much higher than that the ones seen for most of the TS mutants of DNA ligase. However it is comparable to the reversion capabilities of

Francisella novicida ligA S3 strain described before. The mentioned temperature sensitive clone was characterized by the ability to generate two types of mutations that would grant it resistance to the elevated temperature. One type of mutation occurred outside of the coding sequence of the TS allele. Its location could not be determined as it did not occur anywhere in the nearest vicinity of the gene. This would suggest that certain extragenic mutations would affect the expression or stability of the *hemC* product. The second class of mutations happened at the very beginning of the coding sequences, within the several first nucleotides. Since, other than the location, the amino acid changes seemed to be completely random it is possible that the mutations might actually affected the structure of mRNA of the gene. This would mean that temperature resistant clones were characterized by a higher expression level of their DNA ligase. Specifically the modification would have impact at the level of translation. Multiple temperature resistant clones of *F. novicida* were sequenced to confirm the cause of the altered phenotype. Among twelve investigated clones none had any changes within the *hemC* gene. This strongly suggests the presence of an extragenic mutation as a cause for the changes in temperature sensitive phenotype.

Conclusions

The partial sequence of *Colwellia* sp. C1 genome reveals possible adaptations to life in the low temperature environment. We also found homologues of genes that are essential for the bacterial metabolism. We proved that such genes can be used to replace an allele of a homologous gene in a mesophilic organism and render it temperature sensitive. The large number of identified essential genes gives a possibility of creating multiple different TS mutants. Additionally, applying the temperature inactivation

mechanism to different essential genes can result in varying pace of cell death. There is a visible difference between the lethal effect of inactivating the replication genes and inactivation of genes involved in the cofactor biosynthesis. The further investigation of one of such mutants showed capability of easily mutating into temperature resistant forms. This means that a single psychrophilic allele might not be sufficient for generating a safe TS strain for future work. Therefore just like we did with the investigation of DNA ligase alleles there is a need to compare multiple different forms of the same gene in order to find ones that could meet a number of criteria, including a low reversion frequency.

Chapter 6

Screening of psychrophilic alleles of multiple essential genes

Introduction

The set of genes necessary for maintenance of the basic metabolism within an organism, the minimal genome, is crucial for survival. Obtaining control over the activity of these genes would give the possibility of regulating the survival of genetically modified organism at specified conditions. The number of essential genes varies depending on the complexity of organism. Currently it is estimated that the prokaryotic genomes possess several hundred essential genes. Latest research suggests that a basic bacterial genome contains ~300 essential genes. Within this group we have elements responsible for such processes as DNA replication, transcription, translation or biosynthesis of important cofactors. Since each genome usually contains a single copy of each essential gene, its inactivation would be lethal for the organism. Therefore the control over the single essential gene should be sufficient to inhibit the growth of altered bacterial strain.

We have established that the genomes of psychrophilic bacteria can serve as a source of temperature sensitive essential genes. The proteins encoded by these sequences are adapted to function at low temperatures. Among the features typical for enzymes adapted to such environment is the decrease in their stability. The reaction speed is significantly lowered in the cold environment. In order to compensate for that the enzymes are usually characterized by a more flexible structure that increases their processive abilities. However, at the same time the psychrophilic enzymes are often more

sensitive to heat denaturation. They usually loose significant portion of their activity when exposed to the temperature above 25°C - 30°C. It has been postulated that the instability of psychrophilic proteins is a result of their increased flexibility. However alternative explanation suggested that the cause lies in the evolutionary drift that allows accumulation of mutations, which prevent the enzymes from functioning at higher temperatures.

Our previous work shows that using psychrophilic alleles of essential genes can be an efficient method of generating temperature sensitive strains of mesophilic bacteria. The characterized collection of *ligA* alleles contained sequences that would generate TS phenotype at a range of temperatures. In order to identify similar alleles of different essential genes we have investigated genome of the extreme psychrophilic *Colwellia* sp. strain. The investigated allele of *hemC* gene managed to successfully generate a temperature sensitive phenotype in *F. novicida*. However the obtained strain was prone to mutate into a temperature resistant form. Therefore we decided that it is necessary to analyze multiple alleles for each essential gene, before finding elements that would guarantee stable phenotype. In this study we have attempted to characterize several alleles of selected essential genes that were obtained from the genomes of extreme psychrophiles.

Results and discussion

Generation of TS strains.

The aim of this study was to generate a collection of temperature sensitive alleles of essential genes similar to that of *ligA* alleles. For this purpose we have selected the

alleles encoding several genes identified as essential for the growth of bacteria. For some of them, like prophobilinogen deaminase (*hemC*), methionyl-tRNA formyltransferase (*fmt*) and tyrosyl-tRNA synthetase (*tyrS*), a number of temperature sensitive variants has been identified in earlier research concerning the psychrophilic genes (31). Other ones have not been previously investigated in such way (Table 10). While individual temperature sensitive mutant of mesophilic alleles were reported, there are no collections of alleles that would support a stable temperature sensitive phenotype for these genes (150). We have used *F. novicida* as a default acceptor for the psychrophilic alleles, as the G+C content of its genome (33%) closely resembles that of the chromosomes of the psychrophilic organisms. As described previously all of the allele substitutions were designed to utilize the native *F. novicida* regulatory elements e.g. promoter and ribosome binding site of the corresponding gene. To avoid causing significant changes in the structure of mRNA we placed the first three codons of the native *F. novicida* coding sequence in place of the codons of psychrophile gene. We tested genes amplified from the strains belonging to groups such as *Shewanella*, *Colwellia* and *Pseudoalteromonas*.

Table 10 Essential genes investigated in search for temperature sensitive alleles

Gene	Role	Gene size (bp)	Tested alleles
fmt	N-formyltransferase	942	3
tyrS	aminoacyl-tRNA synthetase	1191	3
hemC	propohbilinogen deaminase	903	3
dnaB	replicative DNA helicase	1398	4
dnaK	protein chaperone	1929	3
dnaA	replication initiation	1476	3

The number of tested alleles does not include the previously characterized sequences. Gene size is that of corresponding sequence found in the genome of *F. novicida* U112.

Only two of the 19 tested alleles managed to substitute the native gene and generated a temperature-sensitive (TS) *F. novicida* strain. They encoded a *Shewanella* type N-formyltransferase and aminoacyl-tRNA synthetase. The restrictive temperature of the mutants was 40°C and 43°C respectively. The sequencing confirmed the presence of the complete psychrophilic *fmt* allele. However the *tyrS* allele possessed only the last 36 amino acids from the introduced psychrophilic sequence. The vast majority of the coding regions remained unchanged (Fig. 26). This situation is very similar to the hybrid allele of DNA ligase that we previously observed in *F. novicida* (see Chapter 3). We suspect that the hybrid allele was created by an intergenic recombination. The C-terminal region of the enzyme, which is now encoded by the psychrophilic sequence, contains the S4 domain. This domain is unique for the tyrosyl-tRNA synthetase and participates in binding of the RNA molecule.

Several other investigated constructs have generated TS clones. However sequencing showed that they possess only the native form of respective essential gene in its locus. Since we never observed a spontaneous formation of TS clones in *F. novicida* population it would suggest that a nonspecific recombination between introduced construct and the chromosome caused the changes in the phenotype. It could cause insertion within a vital part of the genome, which resulted in impaired growth of the strain.

```
     F. novicida DNA              Shewanella DNA          F. novicida DNA
 I   Q   Q   G   A     A   K   I    T   I   K     *
att-cag-caa-ggt-gct-GCG-AAG-ATC//ACC-ATA-AAA-taaggggttgtaagatgg
356 357 358 359 360 361 362 363  394 395 396 397
```

Figure 26 Hybrid form of aminoacyl-tRNA synthetase.

Integration of the psychrophilic allele of aminoacyl-tRNA synthetase resulted in a hybrid gene with 360 codons from the native *F. novicida* gene that encodes the N-terminal region of the 396 amino acid hybrid protein.

Frequency of mutation to temperature resistance of *F. novicida* strains carrying temperature sensitive alleles.

We investigate the genetic stability as part of the characterization of the TS strains. We do this by observing if the generated clones could mutate into temperature resistant forms. Usually we test for the spontaneous formation of temperature resistance forms at temperature approximately 3°C higher than the restrictive temperature of each strain. However the high restrictive temperature of *tyrS* and *fmt* mutants would force us to analyze them at temperatures very close to or above the maximum growth temperature of *F. novicida*. In such case the elevated temperature would impact the entire cellular metabolism, not just the product of altered essential gene, resulting in either decreased fitness or cell death. Such experiments would not give the accurate picture of the TS phenotype of new strain. Therefore we decided to abandon this aspect of analysis of the *fmt* and *tyrS* strains.

Conclusions

In this work we aimed to generate a number of new collections of TS alleles representing selected essential genes. Previous attempts yielded a large number of TS strain with varying restrictive temperature. However in this approach only two alleles, each of a different gene, were capable of generating temperature sensitive phenotype. This proves that not every type of psychrophilic gene, when placed in mesophilic genome, is going to have the expected impact on the fitness of its host. It is possible that only certain genes, like *ligA* encoding DNA ligase, can efficiently establish a temperature sensitive effect in mesophilic organisms. This could be caused by several different factors. The products of specific genes might be stable even at temperatures above the

maximum growth temperature of their host organism. Alternatively the introduced psychrophilic allele might be incapable of performing its functions in the mesophilic host. Therefore clones carrying only that allele would be deleterious and could not be selected for. The screening for TS clones often requires analyzing the phenotype of hundreds of clones and the allele swapping rate can be as low as 1%. Because of this it is often hard to establish if any of the obtained clones actually possesses the introduced gene. In order to facilitate more efficient analysis of psychrophilic essential genes a more efficient system of exchanging alleles has to be established.

Chapter 7

Plasmid based system for efficient allele exchange

The *E.coli* Δpth strain described in this chapter has been generated by Daniel Kemp.

Introduction

The essential genes are sequences encoding elements that are crucial for maintaining the metabolism of a living organism. The set of all essential genes in given organism is often called the minimal genome. These elements provide backbone for all of the secondary cellular processes. Therefore by manipulating the essential genes we can provide better support for synthesis of valuable secondary metabolites. Using essential genes also allows development of biosafety mechanisms that prevent the spread of genetically modified organisms into the environment. Therefore the investigation of essential genes plays an important role in modern molecular and synthetic biology. Any genetic alterations in the sequence of essential genes can have a unique effect on the host organism. Since simple organisms usually have only one copy of each essential gene, any alterations to that sequence would have immediate effect of cell fitness. However investigation of essential genes is hindered by the requirement of possessing at least one active copy at a time to ensure the survival of an organism. Deleting essential genes would cause, by definition, a lethal effect.

There are systems that allow replacing a native copy of a selected essential gene with an alternative sequence. They usually rely on a double recombination with the chromosome, resulting in insertion of introduced alleles and deletion of the native sequence. This approach has a few major drawbacks that limit its efficiency. It usually

utilizes the native recombination system of the host organism to facilitate the gene insertion. Therefore it can be performed only in organisms characterized by high recombination proficiency. In case of *E.coli* strain that lacks active form of RecA protein (e.g. as a result of a gene deletion) its functional form has to be provided (151). This is usually done by introducing conditionally replicating vector that facilitates expression of *recA* gene. Furthermore the exchange of alleles has to be linked with marker that allows selection of desired clones. Recently developed system for recombination combining λ phage genes and the CRISPR system allows for efficiency high enough to skip the selection step (152). However this method requires the use of several plasmids, which have to be maintained in the strain at the same time. Finally, for most of the allele exchange protocols the efficiency of the process is limited by two factors – the transformation rate and the recombination rate. While the transformation rate is usually the same for a specific organism, the recombination rate can vary greatly depending on which system is used.

The system for allele exchange that we propose here focuses on decoupling the gene of interest from the rest of the genome by transferring it onto an autonomously replicating plasmid. Such vector can be small enough to introduce the whole construct, including allele of an essential gene, into the organism via a standard transformation. Since it can persist independently in the cell, allowing expression of its genes, there is no need for recombination with the chromosome. In order to sustain the growth of strain deleterious for selected essential gene at least one such construct has to be constantly maintained in the host organism. To remove the former vector we took advantage of the phenomenon known as plasmid incompatibility. This effect causes plasmids with similar

replication or partitioning systems to be incapable of remaining in the same strain for a longer period of time. Eventually after several cell divisions two sub-strains are created, each one with different plasmid. The strain separation tends to be faster when vectors are characterized by a low copy number. This would allow introduction of a new allele of essential gene into the modified organism and selection against the presence of the old one.

Results and discussion

We aimed to create a simple system that would allow rapid exchange of alleles of essential genes in the host strain. For this purpose we have created two vectors based on the F plasmid maintenance systems. Both constructs were based on the same plasmid backbone from pBeloBAC11 vector. To facilitate selection for a specific vector they were given different antibiotic markers. One of the plasmids –pBACK- has a kanamycin cassette, while the second one – pBACT – has a tetracycline one. Since they share the same replication and partitioning system they are being incompatible with each other. Therefore after a few rounds of division most of the cells should carry only one of the designed constructs. The process of losing additional vector should be fastened by the low copy of these constructs. To perform its function as a system for exchange of essential genes it has to allow expression of selected ORF at a level high enough to support the growth of its host. To assess that we have generated an *E. coli* strain with deletion of an essential gene. We have chosen *pth* encoding the peptidyl tRNA hydrolase. The product of this gene is responsible for cleaving the bond between peptide chain and the tRNA molecule on a halted ribosome. Therefore it is responsible for recycling the

tRNAs in the cells and a proper progression of translation. Studies have shown that *pth* gene is indispensable in bacterial cells (90). We generated the Δpth strain with the help of inducible λ recombination system placed on the chromosome of *E.coli* GB05 strain. The loss of the essential gene was complemented by introduction of a modified pSEVA plasmid carrying *pth* gene from *Salmonella enterica*. The plasmid was designed to have its replication and expression of the essential gene dependent on the presence of anhydrotetracycline (aTc). This compound de-represses the promoters driving expression of genes responsible for plasmid replication as well as the native *pth* gene. We introduced that vector to our GB05 strain. Upon the transformation with linear DNA we have obtained *E. coli* clones with chloramphenicol resistance suggesting that they have integrated the disruption cassette into their chromosome. We confirmed the presence of insertion in the native *pth* gene through sequencing.

The *E. coli* Δpth strain would require the presence of episomaly encoded copy of peptidyl tRNA hydrolase in order to survive. We wanted to confirm that the performed recombination did render the clone dependant on the presence of pSEVA vector. To achieve this we have grown the strain in the absence of aTc; the compound is required for the replication of the modified pSEVA vector. The strain was eventually unable to survive. However the default high copy number of pSEVA vector and possibly leaky expression allowed the strain to go through multiple rounds of divisions before that happened. We have then transformed the Δpth clone with pBACK vector carrying the same *pth* gene as pSEVA. However the sequence carried on the introduced plasmid was placed under the control of a lac promoter. The transformed cells were recovered overnight and then diluted for subsequent cultures. Finally we have selected individual

clones. We confirmed the presence of pBACK vector and the absence of pSEVA through agarose gel analysis of purified plasmid DNA. Initially the new clone was grown in the presence of ITPG in order to allow the expression of the essential gene. However later we discovered that even in the absence of this inducer the expression of *pth* was sufficient to sustain the growth.

We tested whether the exchange of incompatible vectors carrying a copy of the essential gene could still allow survival of the host strain. To do that we introduced pBACT vector with and without *pth* gene to the pBACK clone. The transformed cells were placed under the tetracycline selection, meaning they could only survive if they acquired the pBACT plasmid. After an overnight incubation we observed regular colonies from the pBACT-pth transformation. The control transformation with empty vector resulted in a few barely visible colonies. We replica plated both types of colonies on medium with either kanamycin or tetracycline. We observed that all of the clones with pBACT-pth construct would grow on tetracycline and majority of them would not be able to grow on kanamycin. Many of the clones that could grow on the kanamycin would form only a few individual colonies instead of a continuous growth. This suggests that analyzed colony was a heterologous population consisting predominantly of pBACT carrying cells with a minority of pBACK cells. The small colonies that were generated from empty pBACT vector transformation could all grow fine on kanamycin, but only formed single colonies when replicated on tetracycline. It would suggest that the pBACT vector, which was the only source of functional peptidyl tRNA hydrolase, remained in the cells. However due to the high incompatibility of vectors it had limited possibility of

persisting in population, resulting in severely crippled growth of these clones. This also further proves that characterized strain requires external *pth* gene in order to survive.

The presented system allows for a quick and efficient swapping of copies of *pth* essential genes. It is possible to adopt it for investigation of other essential genes by simply generating strains with appropriate deletions. The advantage of this approach stems from the possibility of avoiding the chromosomal recombination step. This additional stage lowers efficiency of the whole process and increases the risk of obtaining clones with non-specific insertions. With the growing interest in investigating the essential genes this type of method could find multiple applications. Additionally it is an improvement the previous system of allelic exchange that relay on homologous recombination with the chromosome.

Chapter 8

Summary, conclusions and future directions

Summary

Genetic devices designed to act as a biosafety mechanisms have been developed as means of control over organisms created artificially in the laboratory conditions. The primary motive behind their creation was always to prevent unplanned and accidental leakage of potentially dangerous strains of microbes. Such strains, depending on their nature, could pose a threat to environment or people’s health. Usually if there is a risk of contamination there are the conventional methods of preventing its spread, like physical barriers or safety protocols. The genetic safety switches act as a last resort, when every other attempt at preventing microbe release has failed. Just like many other genetic mechanisms they were created to activate upon exposure to a specific signal. When engaged they aim to damage the vital components of their host. This results in preventing organism from replicating, destroying the integrity of its cells and allowing disintegration of its genetic material. As a result the modified organism cannot spread further nor cause any harm. Even most basic safety features, such as auxotrophy, have been following the concept of damaging the released strain when specific conditions are met. With the development of molecular biology multiple different biochemical pathways were targeted by the genetic devices at the same time. Each one of them would essential for microbial survival. Disabling multiple targets allows additional security as it is less likely for each individual pathway to escape the triggered effect.

Attacking multiple crucial metabolic pathways is currently the most common way of increasing the efficiency of safety devices. This requires development of features that would allow direct interaction with these elements. By having control over several essential genes, or their products, it is possible to simultaneously affect several biochemical pathways. This approach has been utilized with the recently created strains of *E. coli* that require non-biogenic amino acids for their survival (30). However in this case high efficiency of mechanism comes at the price of very narrow specialization. Such elaborate method can be applied only to heavily engineered strains that underwent multiple changes in their chromosome. The biosafety system would become more useful if it could be easily transferred between different organisms. Such versatility might become a necessity, since a single group of engineered strains is unlikely to become a standard for every molecular research in the future. Therefore safety switch would have to be adaptable to the various desired hosts. This feature has to be combined with a mechanism of inactivating selected essential genes.

I utilized temperature sensitive alleles of essential genes to generate a biosafety system that is useful in protecting humans and other mammals from potentially dangerous microbes. With this system the input signal that activates the safety mechanism is the increase in environmental temperature. Since the triggering element is a physical factor and not a chemical one it gives a greater independence with using modified strain. There is no need to consider cellular transport across the membranes or degradation of the inducer. Additionally the modified strains should meet the most important requirement expected from genetically altered organisms – they are safe for humans and animals. Almost all of the strains that were created during our research are

incapable of surviving at the temperature of human body. Furthermore since essential genes are mostly conserved among all domains of life it is possible to move them between even distantly related strains. Because these genes are involved in all the crucial metabolic processes they can be used to inactivate different pathways simultaneously. Therefore the system using temperature sensitive essential genes has the capacity to be universal, cheap and efficient.

We used the psychrophilic genomes as a source of temperature sensitive essential genes. These alleles have naturally evolved to function at low temperatures. This kind of adaptation results in decreased stability of protein, making it more susceptible to heat or detergent caused denaturation. Screening psychrophilic genes in order to find temperature sensitive alleles seems to be much more efficient method than regular *in vitro* mutagenesis. With the mutagenesis approach usually few mutations per gene are introduced causing random changes in the amino acid composition of the protein. Such changes can be easily reverted by naturally occurring single mutations. Furthermore the sequence alterations might have a negative effect on the overall catalytic activity of the enzyme. This would highly impact the fitness of the host organism, possibly rendering it useless. To find the best possible mutant from the randomly altered pool it would be necessary to perform multiple rounds of mutagenesis, each time using generated products as a template for another series. This would impose the necessity to analyse exponentially growing number of clones, which would make the entire process extremely difficult. While introducing purely random changes into the sequence of investigated gene seems to be unfeasible, the possibilities for rational design of protein are still very limited making it unlikely to design a properly functioning temperature sensitive protein.

Therefore the use of psychrophilic genes allows us to skip a major obstacle, which is obtaining the temperature sensitive alleles. The sequences obtained from the genomes of low temperature adapted organisms have been changing under selective pressure for millions of years, giving them a significant advantage over sequences created in laboratory conditions.

Conclusions

Our research confirmed that psychrophilic genes are a good source of temperature sensitive alleles. I have generated a collection of 13 alleles of DNA ligase that are capable of generating temperature sensitivity in mesophilic organisms. Additionally these genes can function in both prokaryotes and simple eukaryotes like yeasts. This collection presents the fulfilment of our initial concept – having universal and simple mechanism of biocontainment. Almost all of the created strains had their restrictive temperature below or equal to that of the human body. As such they could be used to alter pathogenic strains making them safer, for people, to work with. The temperature sensitive pathogens could also be used as vaccines. On top of that, majority of the alleles display none or very little capability for mutating into temperature stable forms. This makes them reliable as safety devices. Finally the inactivation of *ligA* product causes very quick degradation of host genetic material and cell death. Overall the created collection of temperature sensitive DNA ligase alleles can be applied into multiple processes as a biosafety mechanism.

Investigation of some of the psychrophilic alleles revealed the limitations to our method. Among the analysed *ligA* alleles as well as psychrophilic alleles of the other genes we have found specimen that were highly likely to convert into temperature stable

forms. This seemed to occur mostly due to mutation outside of the coding region or just at the beginning of the coding sequence. The mutations in the other parts of the chromosome have not been identified. Therefore we cannot make any reasonable prediction regarding the nature of their influence on the psychrophilic protein. On the other hand the mutations occurring at the beginning of the gene can be analysed more deeply. The sequence changes localized at the N-terminal end of protein do not follow any specific pattern. I suspect that the changes actually affect the level of transcription of the gene. This was partially confirmed with software analysis of the sequence near RBS region in the native psychrophilic and mutated (i.e. temperature resistant) sequences. Based on this we can attempt to alter the beginning of the coding sequence so that it would not change into unwanted mutant as easily. Countering of the other type of mutations, which occurs outside of the coding region, is less feasible. It would have to involve more detailed analysis and possibly engineering of the host genome, which goes against the concept of easily applicable safety mechanism. On the other hand these mutations might not occur in a different genetic background. While the majority of investigated alleles presented sufficient level of stability, it would be beneficial to find means of predicting if a gene would be likely to convert into temperature resistant form. However currently we do not have sufficient tools. Therefore each used psychrophilic gene has to be analysed for its stability in a mesophilic host.

We found that not all psychrophilic genes are going to induce the temperature sensitive phenotype. Large number of alleles from different groups of psychrophilic bacteria failed to generate an altered strain with temperature based safety mechanism. Despite the success with DNA ligase we could not generate similar collection for genes

involved in translation or cofactor biosynthesis. Only individual alleles proved to be temperature sensitive. It is possible that some of the genes could not function properly in their new host. Therefore recombinant strain would either not be viable or had severely reduced fitness. This suggests the need to screen a much greater number of samples in order to find any specimen that meet the desired features. Our current system based on double homologues recombination with host chromosome is limiting us from high throughput analysis of many samples. Therefore we created an alternative method that utilizes vector shuffling as a mean of allele exchange. This approach offers a much easier way of introducing a new allele to the mesophilic host and immediately observing the changes in its phenotype. Because of the high efficiency of allele exchange we can recognize when new allele has a negative impact on modified organism, since we would observe a decrease in the exchange rate. This way we should be able to generate multiple temperature sensitive variants for more essential genes.

Future directions

Psychrophilic genes could be scavenged for domain sequences to generate custom alleles. With a large enough set of various temperatures sensitive sequences, e.g. like our *ligA* collection, we can use individual domains of these genes to combine them with elements from different alleles. This way we would create a hybrid forms that would display phenotype different from their parent sequences. With the techniques of seamless DNA assembly and vector based allele exchange we could generate a greater number of new versions of essential genes. Such approach would make us partially independent from the influx of new genomic samples. Since all of the domains would come from

homologous sequences it should not obstruct the activity of the created enzymes. Additionally we could use domains obtained from mesophilic organisms and possibly increase restrictive temperature of particular allele. This would allow us to tune the phenotype of a modified strain to our desires. The only limitation for proposed method would be the necessity of having several temperature sensitive alleles of one gene with different restrictive temperature. Currently we possess sufficient number of DNA ligase clones to carry on with such approach.

Genes from different groups of psychrophiles can be used to screen for temperature sensitive alleles. So far we have utilized the samples of bacterial genomic DNA as a source of temperature sensitive sequences. However a large proportion of essential genes are being well conserved between all three domains of life. Additionally some genes involved in replication are being carried by mobile genetic elements such as phages and plasmids. We have already demonstrated the possibility of using psychrophilic DNA ligase obtained from bacteria in the yeast cells. Among all the groups of living organisms we can find specimens that have adapted to thrive at low temperatures. This would imply that their enzymes underwent a similar evolution as their bacterial equivalents. Especially psychrophilic bacteriophages could carry a great variety of useful genes. It has been previously demonstrated that a DNA ligase from T7 phage can rescue *E. coli* clone lacking its own active enzyme (153). Similar approach could be successful with psychrophilic genes. On top of that isolation of a large number of phage particles from environmental samples is not a limiting factor. This way we can generate a new collection of temperature sensitive alleles for at least some of the essential genes.

Bibliography

1. **Szybalski W**. 1974. In vivo and in vitro initiation of transcription. Adv. Exp. Med. Biol. **44**:23–4.
2. **Kim H**, **Han H**, **Ahn J**, **Lee J**, **Cho N**, **Jang H**, **Kim H**, **Kwon S**, **Bang D**. 2012. "Shotgun DNA synthesis" for the high-throughput construction of large DNA molecules. Nucleic Acids Res. **40**:e140.
3. **Carlson R**. 2009. The changing economics of DNA synthesis. Nat. Biotechnol. **27**:1091–4.
4. **Ma S**, **Tang N**, **Tian J**. 2012. DNA synthesis, assembly and applications in synthetic biology. Curr. Opin. Chem. Biol. **16**:260–7.
5. **Nandagopal N**, **Elowitz MB**. 2011. Synthetic Biology: Integrated Gene Circuits. Science (80-.). **333**:1244–1248.
6. **Davis JH**, **Rubin AJ**, **Sauer RT**. 2010. Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res. **39**:1131–1141.
7. **McWhinnie RL**, **Nano FE**. 2014. Synthetic promoters functional in Francisella novicida and Escherichia coli. Appl. Environ. Microbiol. **80**:226–34.
8. **Chen Y-J**, **Liu P**, **Nielsen AAK**, **Brophy JAN**, **Clancy K**, **Peterson T**, **Voigt CA**. 2013. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat. Methods **10**:659–64.
9. **Salis HM**, **Mirsky EA**, **Voigt CA**. 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. **27**:946–50.
10. **French CE**, **Mora K de**, **Joshi N**, **Elfick A**, **Haseloff J**, **Ajioka J**. 2011. Synthetic biology and the art of biosensor design. National Academies Press (US).
11. **Wang Y-H**, **Wei KY**, **Smolke CD**. 2013. Synthetic biology: advancing the design of diverse genetic systems. Annu. Rev. Chem. Biomol. Eng. **4**:69–102.
12. **Kelwick R**, **Bowater L**, **Yeoman KH**, **Bowater RP**. 2015. Promoting microbiology education through the iGEM synthetic biology competition. FEMS Microbiol. Lett. **362**:fnv129.
13. **Røkke G**, **Korvald E**, **Pahr J**, **Oyås O**, **Lale R**. 2014. BioBrick assembly standards and techniques and associated software tools. Methods Mol. Biol. **1116**:1–24.
14. **Temme K**, **Zhao D**, **Voigt CA**. 2012. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. Proc. Natl. Acad. Sci. U. S. A. **109**:7085–90.
15. **Stricker J**, **Cookson S**, **Bennett MR**, **Mather WH**, **Tsimring LS**, **Hasty J**. 2008. A fast, robust and tunable synthetic gene oscillator. Nature **456**:516–9.
16. **Perry N**, **Ninfa AJ**. 2012. Synthetic networks: oscillators and toggle switches for Escherichia coli. Methods Mol. Biol. **813**:287–300.

17. **Claesen J**, **Fischbach MA**. 2015. Synthetic microbes as drug delivery systems. ACS Synth. Biol. **4**:358–64.

18. **Wen M**, **Bond-Watts BB**, **Chang MCY**. 2013. Production of advanced biofuels in engineered E. coli. Curr. Opin. Chem. Biol. **17**:472–9.

19. **Moon TS**, **Lou C**, **Tamsir A**, **Stanton BC**, **Voigt CA**. 2012. Genetic programs constructed from layered logic gates in single cells. Nature **491**:249–53.

20. **Jackson DA**, **Symons RH**, **Berg P**. 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. **69**:2904–9.

21. **Berg P**, **Baltimore D**, **Brenner S**, **Roblin RO**, **Singer MF**. 1975. Summary statement of the Asilomar conference on recombinant DNA molecules. Proc. Natl. Acad. Sci. U. S. A. **72**:1981–4.

22. **Gibson DG**, **Benders GA**, **Andrews-Pfannkoch C**, **Denisova EA**, **Baden-Tillson H**, **Zaveri J**, **Stockwell TB**, **Brownley A**, **Thomas DW**, **Algire MA**, **Merryman C**, **Young L**, **Noskov VN**, **Glass JI**, **Venter JC**, **Hutchison CA**, **Smith HO**. 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science **319**:1215–20.

23. **Kvitko BH**, **Bruckbauer S**, **Prucha J**, **McMillan I**, **Breland EJ**, **Lehman S**, **Mladinich K**, **Choi K-H**, **Karkhoff-Schweizer R**, **Schweizer HP**. 2012. A simple method for construction of pir+ Enterobacterial hosts for maintenance of R6K replicon plasmids. BMC Res. Notes **5**:157.

24. **Wright O**, **Stan G-B**, **Ellis T**. 2013. Building-in biosafety for synthetic biology. Microbiology **159**:1221–35.

25. **Torres B**, **Jaenecke S**, **Timmis KN**, **García JL**, **Díaz E**. 2003. A dual lethal system to enhance containment of recombinant micro-organisms. Microbiology **149**:3595–601.

26. **Bej AK**, **Molin S**, **Perlin M**, **Atlas RM**. 1992. Maintenance and killing efficiency of conditional lethal constructs in Pseudomonas putida. J. Ind. Microbiol. **10**:79–85.

27. **Curtiss R**, **Nakayama K**, **Kelly SM**. Recombinant avirulent Salmonella vaccine strains with stable maintenance and high level expression of cloned genes in vivo. Immunol. Invest. **18**:583–96.

28. **Rovner AJ**, **Haimovich AD**, **Katz SR**, **Li Z**, **Grome MW**, **Gassaway BM**, **Amiram M**, **Patel JR**, **Gallagher RR**, **Rinehart J**, **Isaacs FJ**. 2015. Recoded organisms engineered to depend on synthetic amino acids. Nature **518**:89–93.

29. **Wang HH**, **Kim H**, **Cong L**, **Jeong J**, **Bang D**, **Church GM**. 2012. Genome-scale promoter engineering by coselection MAGE. Nat. Methods **9**:591–3.

30. **Mandell DJ**, **Lajoie MJ**, **Mee MT**, **Takeuchi R**, **Kuznetsov G**, **Norville JE**, **Gregg CJ**, **Stoddard BL**, **Church GM**. 2015. Biocontainment of genetically modified organisms by synthetic protein design. Nature **518**:55–60.

31. **Duplantis BN**, **Osusky M**, **Schmerk CL**, **Ross DR**, **Bosio CM**, **Nano FE**. 2010. Essential genes from Arctic bacteria used to construct stable, temperature-sensitive bacterial vaccines. Proc. Natl. Acad. Sci. U. S. A. **107**:13456–60.

32. **Chan CTY**, **Lee JW**, **Cameron DE**, **Bashor CJ**, **Collins JJ**. 2015. "Deadman" and "Passcode" microbial kill switches for bacterial containment. Nat. Chem. Biol. **12**:82–86.

33. **Rovner AJ**, **Haimovich AD**, **Katz SR**, **Li Z**, **Grome MW**, **Gassaway BM**, **Amiram M**, **Patel JR**, **Gallagher RR**, **Rinehart J**, **Isaacs FJ**. 2015. Recoded organisms engineered to depend on synthetic amino acids. Nature **518**:89–93.

34. **Dai K**, **Lutkenhaus J**. 1991. ftsZ is an essential cell division gene in Escherichia coli. J. Bacteriol. **173**:3500–6.

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

36. **Yamamoto N**, **Nakahigashi K**, **Nakamichi T**, **Yoshino M**, **Takai Y**, **Touda Y**, **Furubayashi A**, **Kinjyo S**, **Dose H**, **Hasegawa M**, **Datsenko KA**, **Nakayashiki T**, **Tomita M**, **Wanner BL**, **Mori H**. 2009. Update on the Keio collection of Escherichia coli single-gene deletion mutants. Mol. Syst. Biol. **5**:335.

37. **Commichau FM**, **Pietack N**, **Stülke J**. 2013. Essential genes in Bacillus subtilis: a re-evaluation after ten years. Mol. Biosyst. **9**:1068–75.

38. **Salama NR**, **Shepherd B**, **Falkow S**. 2004. Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. J. Bacteriol. **186**:7926–35.

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

40. **Glass JI**, **Assad-Garcia N**, **Alperovich N**, **Yooseph S**, **Lewis MR**, **Maruf M**, **Hutchison CA**, **Smith HO**, **Venter JC**. 2006. Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. U. S. A. **103**:425–30.

41. **Jacobs MA**, **Alwood A**, **Thaipisuttikul I**, **Spencer D**, **Haugen E**, **Ernst S**, **Will O**, **Kaul R**, **Raymond C**, **Levy R**, **Chun-Rong L**, **Guenthner D**, **Bovee D**, **Olson M V**, **Manoil C**. 2003. Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. U. S. A. **100**:14339–44.

42. **Akerley BJ**, **Rubin EJ**, **Novick VL**, **Amaya K**, **Judson N**, **Mekalanos JJ**. 2002. A genome-scale analysis for identification of genes required for growth or survival of Haemophilus influenzae. Proc. Natl. Acad. Sci. U. S. A. **99**:966–71.

43. **McCutcheon JP**, **von Dohlen CD**. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. **21**:1366–72.

44. **Nakabachi A**, **Yamashita A**, **Toh H**, **Ishikawa H**, **Dunbar HE**, **Moran NA**, **Hattori M**. 2006. The 160-kilobase genome of the bacterial endosymbiont Carsonella. Science **314**:267.

45. **McCutcheon JP**, **Moran NA**. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol. Evol. **2**:708–18.

46. **Cassone BJ**, **Wenger JA**, **Michel AP**. 2015. Whole Genome Sequence of the Soybean Aphid Endosymbiont Buchnera aphidicola and Genetic Differentiation among Biotype-Specific Strains. J. genomics **3**:85–94.

47. **Cole ST**, **Brosch R**, **Parkhill J**, **Garnier T**, **Churcher C**, **Harris D**, **Gordon S V**, **Eiglmeier K**, **Gas S**, **Barry CE**, **Tekaia F**, **Badcock K**, **Basham D**, **Brown D**, **Chillingworth T**, **Connor R**, **Davies R**, **Devlin K**, **Feltwell T**, **Gentles S**, **Hamlin N**, **Holroyd S**, **Hornsby T**, **Jagels K**, **Krogh A**, **McLean J**, **Moule S**, **Murphy L**, **Oliver K**, **Osborne J**, **Quail MA**, **Rajandream MA**, **Rogers J**, **Rutter S**, **Seeger K**, **Skelton J**, **Squares R**, **Squares S**, **Sulston JE**, **Taylor K**, **Whitehead S**, **Barrell BG**. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393**:537–44.

48. **Kunst F**, **Ogasawara N**, **Moszer I**, **Albertini AM**, **Alloni G**, **Azevedo V**, **Bertero MG**, **Bessières P**, **Bolotin A**, **Borchert S**, **Borriss R**, **Boursier L**, **Brans A**, **Braun M**, **Brignell SC**, **Bron S**, **Brouillet S**, **Bruschi C V**, **Caldwell B**, **Capuano V**, **Carter NM**, **Choi SK**, **Cordani JJ**, **Connerton IF**, **Cummings NJ**, **Daniel RA**, **Denziot F**, **Devine KM**, **Düsterhöft A**, **Ehrlich SD**, **Emmerson PT**, **Entian KD**, **Errington J**, **Fabret C**, **Ferrari E**, **Foulger D**, **Fritz C**, **Fujita M**, **Fujita Y**, **Fuma S**, **Galizzi A**, **Galleron N**, **Ghim SY**, **Glaser P**, **Goffeau A**, **Golightly EJ**, **Grandi G**, **Guiseppi G**, **Guy BJ**, **Haga K**, **Haiech J**, **Harwood CR**, **Hènaut A**, **Hilbert H**, **Holsappel S**, **Hosono S**, **Hullo MF**, **Itaya M**, **Jones L**, **Joris B**, **Karamata D**, **Kasahara Y**, **Klaerr-Blanchard M**, **Klein C**, **Kobayashi Y**, **Koetter P**, **Koningstein G**, **Krogh S**, **Kumano M**, **Kurita K**, **Lapidus A**, **Lardinois S**, **Lauber J**, **Lazarevic V**, **Lee SM**, **Levine A**, **Liu H**, **Masuda S**, **Mauël C**, **Médigue C**, **Medina N**, **Mellado RP**, **Mizuno M**, **Moestl D**, **Nakai S**, **Noback M**, **Noone D**, **O'Reilly M**, **Ogawa K**, **Ogiwara A**, **Oudega B**, **Park SH**, **Parro V**, **Pohl TM**, **Portelle D**, **Porwollik S**, **Prescott AM**, **Presecan E**, **Pujic P**, **Purnelle B**, **Rapoport G**, **Rey M**, **Reynolds S**, **Rieger M**, **Rivolta C**, **Rocha E**, **Roche B**, **Rose M**, **Sadaie Y**, **Sato T**, **Scanlan E**, **Schleich S**, **Schroeter R**, **Scoffone F**, **Sekiguchi J**, **Sekowska A**, **Seror SJ**, **Serror P**, **Shin BS**, **Soldo B**, **Sorokin A**, **Tacconi E**, **Takagi T**, **Takahashi H**, **Takemaru K**, **Takeuchi M**, **Tamakoshi A**, **Tanaka T**, **Terpstra P**, **Togoni A**, **Tosato V**, **Uchiyama S**, **Vandebol M**, **Vannier F**, **Vassarotti A**, **Viari A**, **Wambutt R**, **Wedler H**, **Weitzenegger T**, **Winters P**, **Wipat A**, **Yamamoto H**, **Yamane K**, **Yasumoto K**, **Yata K**, **Yoshida K**, **Yoshikawa HF**, **Zumstein E**, **Yoshikawa H**, **Danchin A**. 1997. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature **390**:249–56.

49. **Blattner FR**. 1997. The Complete Genome Sequence of Escherichia coli K-12. Science (80-.). **277**:1453–1462.

50. **Sarmiento F**, **Mrázek J**, **Whitman WB**. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. Proc. Natl. Acad. Sci. U. S. A. **110**:4726–31.

51. **Makarova KS**, **Koonin E V**. 2003. Comparative genomics of Archaea: how much have we learned in six years, and what's next? Genome Biol. **4**:115.

52. **Giaever G**, **Chu AM**, **Ni L**, **Connelly C**, **Riles L**, **Véronneau S**, **Dow S**, **Lucau-Danila A**, **Anderson K**, **André B**, **Arkin AP**, **Astromoff A**, **El-Bakkoury M**, **Bangham R**, **Benito R**, **Brachat S**, **Campanaro S**, **Curtiss M**, **Davis K**, **Deutschbauer A**, **Entian K-D**, **Flaherty P**, **Foury F**, **Garfinkel DJ**, **Gerstein M**, **Gotte D**, **Güldener U**, **Hegemann JH**, **Hempel S**, **Herman Z**, **Jaramillo DF**, **Kelly DE**, **Kelly SL**, **Kötter P**, **LaBonte D**, **Lamb DC**, **Lan N**, **Liang H**, **Liao H**, **Liu L**, **Luo C**, **Lussier M**, **Mao R**, **Menard P**, **Ooi SL**, **Revuelta JL**, **Roberts CJ**, **Rose M**, **Ross-Macdonald P**, **Scherens B**, **Schimmack G**, **Shafer B**, **Shoemaker DD**, **Sookhai-Mahadeo S**, **Storms RK**, **Strathern JN**, **Valle G**, **Voet M**, **Volckaert G**, **Wang C**, **Ward TR**, **Wilhelmy J**, **Winzeler EA**, **Yang Y**, **Yen G**, **Youngman E**, **Yu K**, **Bussey H**, **Boeke JD**, **Snyder M**, **Philippsen P**, **Davis RW**, **Johnston M**. 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature **418**:387–91.

53. **Kim D-U**, **Hayles J**, **Kim D**, **Wood V**, **Park H-O**, **Won M**, **Yoo H-S**, **Duhig T**, **Nam M**, **Palmer G**, **Han S**, **Jeffery L**, **Baek S-T**, **Lee H**, **Shim YS**, **Lee M**, **Kim L**, **Heo K-S**, **Noh EJ**, **Lee A-R**, **Jang Y-J**, **Chung K-S**, **Choi S-J**, **Park J-Y**, **Park Y**, **Kim HM**, **Park S-K**, **Park H-J**, **Kang E-J**, **Kim HB**, **Kang H-S**, **Park H-M**, **Kim K**, **Song K**, **Song K Bin**, **Nurse P**, **Hoe K-L**. 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. **28**:617–23.

54. **Kamath RS**, **Fraser AG**, **Dong Y**, **Poulin G**, **Durbin R**, **Gotta M**, **Kanapin A**, **Le Bot N**, **Moreno S**, **Sohrmann M**, **Welchman DP**, **Zipperlen P**, **Ahringer J**. 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature **421**:231–7.

55. **Lodge JK**, **Weston-Hafer K**, **Berg DE**. 1988. Transposon Tn5 target specificity: preference for insertion at G/C pairs. Genetics **120**:645–50.

56. **Knuth K**, **Niesalla H**, **Hueck CJ**, **Fuchs TM**. 2004. Large-scale identification of essential Salmonella genes by trapping lethal insertions. Mol. Microbiol. **51**:1729–1744.

57. **Jin H**, **Lu B**, **Zhou H**, **Ma C**, **Zhao J**, **Yang C**, **Kemble G**, **Greenberg H**. 2003. Multiple amino acid residues confer temperature sensitivity to human influenza virus vaccine strains (FluMist) derived from cold-adapted A/Ann Arbor/6/60. Virology **306**:18–24.

58. **Kouzminova EA**, **Kuzminov A**. 2012. Chromosome demise in the wake of ligase-deficient replication. Mol. Microbiol. **84**:1079–96.

59. **Shuman S**. 2009. DNA ligases: progress and prospects. J. Biol. Chem. **284**:17365–9.

60. **Pascal JM**. 2008. DNA and RNA ligases: structural variations and shared mechanisms. Curr. Opin. Struct. Biol. **18**:96–105.

61. **Nakatani M**, **Ezaki S**, **Atomi H**, **Imanaka T**. 2000. A DNA ligase from a hyperthermophilic archaeon with unique cofactor specificity. J. Bacteriol.

182:6424–33.

62. **Wilkinson A**, **Day J**, **Bowater R**. 2001. Bacterial DNA ligases. Mol. Microbiol. **40**:1241–1248.

63. **Petit MA**, **Ehrlich SD**. 2000. The NAD-dependent ligase encoded by yerG is an essential gene of Bacillus subtilis. Nucleic Acids Res. **28**:4642–8.

64. **Feng H**, **Parker JM**, **Lu J**, **Cao W**. 2004. Effects of deletion and site-directed mutations on ligation steps of NAD+-dependent DNA ligase: a biochemical analysis of BRCA1 C-terminal domain. Biochemistry **43**:12648–59.

65. **Ellenberger T**, **Tomkinson AE**. 2008. Eukaryotic DNA ligases: structural and functional insights. Annu. Rev. Biochem. **77**:313–38.

66. **Taylor MR**, **Conrad JA**, **Wahl D**, **O'Brien PJ**. 2011. Kinetic mechanism of human DNA ligase I reveals magnesium-dependent changes in the rate-limiting step that compromise ligation efficiency. J. Biol. Chem. **286**:23054–62.

67. **Sriskanda V**, **Kelman Z**, **Hurwitz J**, **Shuman S**. 2000. Characterization of an ATP-dependent DNA ligase from the thermophilic archaeon Methanobacterium thermoautotrophicum. Nucleic Acids Res. **28**:2221–8.

68. **Yang SW**, **Becker FF**, **Chan JY**. 1996. Inhibition of human DNA ligase I activity by zinc and cadmium and the fidelity of ligation. Environ. Mol. Mutagen. **28**:19–25.

69. **Tang Z**. 2003. Real-time monitoring of nucleic acid ligation in homogenous solutions using molecular beacons. Nucleic Acids Res. **31**:148e–148.

70. **Hayashi K**, **Nakazawa M**, **Ishizaki Y**, **Obayashi A**. 1985. Influence of monovalent cations on the activity of T4 DNA ligase in the presence of polyethylene glycol. Nucleic Acids Res. **13**:3261–3271.

71. **Liu L**, **Tang Z**, **Wang K**, **Tan W**, **Li J**, **Guo Q**, **Meng X**, **Ma C**. 2005. Using molecular beacon to monitor activity of E. coli DNA ligase. Analyst **130**:350–7.

72. **Ferrer M**, **Golyshina O V**, **Beloqui A**, **Böttger LH**, **Andreu JM**, **Polaina J**, **De Lacey AL**, **Trautwein AX**, **Timmis KN**, **Golyshin PN**. 2008. A purple acidophilic di-ferric DNA ligase from Ferroplasma. Proc. Natl. Acad. Sci. U. S. A. **105**:8878–83.

73. **Leung EKY**, **Suslov N**, **Tuttle N**, **Sengupta R**, **Piccirilli JA**. 2011. The Mechanism of Peptidyl Transfer Catalysis by the Ribosome.

74. **Wakao H**, **Romby P**, **Westhof E**, **Laalami S**, **Grunberg-Manago M**, **Ebel J**, **Ehresmann C**, **Ehresmann B**. 1989. The solution structure of the Escherichia coli initiator tRNA and its interactions with initiation factor 2 and the ribosomal 30 S subunit. J. Biol. Chem. **264**:20363–20371.

75. **Guillon JM**, **Mechulam Y**, **Schmitter JM**, **Blanquet S**, **Fayat G**. 1992. Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli. J. Bacteriol. **174**:4294–301.

76. **Newton DT**, **Creuzenet C**, **Mangroo D**. 1999. Formylation Is Not Essential for Initiation of Protein Synthesis in All Eubacteria. J. Biol. Chem. **274**:22143–22146.

77. **Gerdes SY**, **Scholle MD**, **Campbell JW**, **Balazsi G**, **Ravasz E**, **Daugherty MD**, **Somera AL**, **Kyrpides NC**, **Anderson I**, **Gelfand MS**, **Bhattacharya A**, **Kapatral V**, **D'Souza M**, **Baev M V.**, **Grechkin Y**, **Mseeh F**, **Fonstein MY**, **Overbeek R**, **Barabasi A-L**, **Oltvai ZN**, **Osterman AL**. 2003. Experimental Determination and System Level Analysis of Essential Genes in Escherichia coli MG1655. J. Bacteriol. **185**:5673–5684.

78. **RajBhandary UL**. 1997. Once there were twenty. Proc. Natl. Acad. Sci. U. S. A. **94**:11761–3.

79. **Fukai S**, **Nureki O**, **Sekine S-I**, **Shimada A**, **Vassylyev DG**, **Yokoyama S**. 2003. Mechanism of molecular interactions for tRNA(Val) recognition by valyl-tRNA synthetase. RNA **9**:100–11.

80. **Suzuki T**, **Miyauchi K**. 2010. Discovery and characterization of tRNAIle lysidine synthetase (TilS). FEBS Lett. **584**:272–7.

81. **Muramatsu T**, **Nishikawa K**, **Nemoto F**, **Kuchino Y**, **Nishimura S**, **Miyazawa T**, **Yokoyama S**. 1988. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature **336**:179–81.

82. **Fersht AR**, **Kaethner MM**. 1976. Mechanism of aminoacylation of tRNA. Proof of the aminoacyl adenylate pathway for the isoleucyl- and tyrosyl-tRNA synthetases from Escherichia coli K12. Biochemistry **15**:818–23.

83. **Eriani G**, **Delarue M**, **Poch O**, **Gangloff J**, **Moras D**. 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature **347**:203–206.

84. **Frankenberg N**, **Moser J**, **Jahn D**. 2003. Bacterial heme biosynthesis and its biotechnological application. Appl. Microbiol. Biotechnol. **63**:115–27.

85. **Louie G V**, **Brownlie PD**, **Lambert R**, **Cooper JB**, **Blundell TL**, **Wood SP**, **Warren MJ**, **Woodcock SC**, **Jordan PM**. 1992. Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. Nature **359**:33–9.

86. **Das G**, **Varshney U**. 2006. Peptidyl-tRNA hydrolase and its critical role in protein biosynthesis. Microbiology **152**:2191–5.

87. **Fromant M**, **Ferri-Fioni M-L**, **Plateau P**, **Blanquet S**. 2003. Peptidyl-tRNA hydrolase from Sulfolobus solfataricus. Nucleic Acids Res. **31**:3227–35.

88. **Menez J**, **Buckingham RH**, **Zamaroczy M de**, **Campelli CK**. 2002. Peptidyl-tRNA hydrolase in Bacillus subtilis, encoded by spoVC, is essential to vegetative growth, whereas the homologous enzyme in Saccharomyces cerevisiae is dispensable. Mol. Microbiol. **45**:123–129.

89. **Rosas-Sandoval G**, **Ambrogelly A**, **Rinehart J**, **Wei D**, **Cruz-Vera LR**, **Graham DE**, **Stetter KO**, **Guarneros G**, **Söll D**. 2002. Orthologs of a novel archaeal and of the bacterial peptidyl-tRNA hydrolase are nonessential in yeast. Proc. Natl. Acad. Sci. U. S. A. **99**:16707–12.

90. **Menninger JR**. 1979. Accumulation of peptidyl tRNA is lethal to Escherichia coli. J. Bacteriol. **137**:694–6.

91. **Menninger JR**. 1978. The accumulation as peptidyl-transfer RNA of isoaccepting transfer RNA families in Escherichia coli with temperature-sensitive peptidyl-transfer RNA hydrolase. J. Biol. Chem. **253**:6808–13.

92. **Heurgué-Hamard V**, **Mora L**, **Guarneros G**, **Buckingham RH**. 1996. The growth defect in Escherichia coli deficient in peptidyl-tRNA hydrolase is due to starvation for Lys-tRNA(Lys). EMBO J. **15**:2826–33.

93. **Singh A**, **Kumar A**, **Gautam L**, **Sharma P**, **Sinha M**, **Bhushan A**, **Kaur P**, **Sharma S**, **Arora A**, **Singh TP**. 2014. Structural and binding studies of peptidyl-tRNA hydrolase from Pseudomonas aeruginosa provide a platform for the structure-based inhibitor design against peptidyl-tRNA hydrolase. Biochem. J. **463**:329–37.

94. **Powers R**, **Mirkovic N**, **Goldsmith-Fischman S**, **Acton TB**, **Chiang Y**, **Huang YJ**, **Ma L**, **Rajan PK**, **Cort JR**, **Kennedy MA**, **Liu J**, **Rost B**, **Honig B**, **Murray D**, **Montelione GT**. 2005. Solution structure of Archaeglobus fulgidis peptidyl-tRNA hydrolase (Pth2) provides evidence for an extensive conserved family of Pth2 enzymes in archea, bacteria, and eukaryotes. Protein Sci. **14**:2849–61.

95. **Zhang D-C**, **Brouchkov A**, **Griva G**, **Schinner F**, **Margesin R**. 2013. Isolation and characterization of bacteria from ancient siberian permafrost sediment. Biology (Basel). **2**:85–106.

96. **Bowman J**, **McCammon S**, **Brown M**, **Nichols D**, **McMeekin T**. 1997. Diversity and association of psychrophilic bacteria in Antarctic sea ice. Appl. Envir. Microbiol. **63**:3068–3078.

97. **D'Amico S**, **Collins T**, **Marx J-C**, **Feller G**, **Gerday C**. 2006. Psychrophilic microorganisms: challenges for life. EMBO Rep. **7**:385–9.

98. **Pandey KD**, **Shukla SP**, **Shukla PN**, **Giri DD**, **Singh JS**, **Singh P**, **Kashyap AK**. 2004. Cyanobacteria in Antarctica: ecology, physiology and cold adaptation. Cell. Mol. Biol. (Noisy-le-grand). **50**:575–84.

99. **Gilbert JA**, **Hill PJ**, **Dodd CER**, **Laybourn-Parry J**. 2004. Demonstration of antifreeze protein activity in Antarctic lake bacteria. Microbiology **150**:171–80.

100. **Chintalapati S**, **Kiran MD**, **Shivaji S**. 2004. Role of membrane lipid fatty acids in cold adaptation. Cell. Mol. Biol. (Noisy-le-grand). **50**:631–642.

101. **Feller G**. 2010. Protein stability and enzyme activity at extreme biological temperatures. J. Phys. Condens. Matter **22**:323101.

102. **Jia Z**, **Davies PL**. 2002. Antifreeze proteins: an unusual receptor-ligand interaction. Trends Biochem. Sci. **27**:101–6.

103. **Phadtare S**. 2004. Recent developments in bacterial cold-shock response. Curr. Issues Mol. Biol. **6**:125–36.

104. **Marx JG**, **Carpenter SD**, **Deming JW**. 2009. Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium Colwellia psychrerythraea strain 34H under extreme conditions. Can. J. Microbiol. **55**:63–72.

105. **Fong NJ**, **Burgess ML**, **Barrow KD**, **Glenn DR**. 2001. Carotenoid accumulation in the psychrotrophic bacterium Arthrobacter agilis in response to thermal and salt stress. Appl. Microbiol. Biotechnol. **56**:750–6.

106. **Feller G**. 2013. Psychrophilic enzymes: from folding to function and biotechnology. Scientifica (Cairo). **2013**:512840.

107. **Harper JDI**, **Wu L**, **Sakuanrungsirikul S**, **John PCL**. 1995. Isolation and partial characterization of conditional cell division cycle mutants inChlamydomonas. Protoplasma **186**:149–162.

108. **Flett F**, **Platt J**, **Oliver SG**. 1992. Isolation and characterization of temperature-sensitive mutants of Streptomyces coelicolor A3(2) blocked in macromolecular synthesis. J. Gen. Microbiol. **138**:579–85.

109. **Patterson D**, **Waldren C**, **Walker C**. 1976. Isolation and characterization of temperature-sensitive mutants of Chinese hamster ovary cells after treatment with UV and X-irradiation. Somatic Cell Genet. **2**:113–123.

110. **Dutta A**, **Chaudhuri K**. 2010. Analysis of tRNA composition and folding in psychrophilic, mesophilic and thermophilic genomes: indications for thermal adaptation. FEMS Microbiol. Lett. **305**:100–8.

111. **Marashi S-A**, **Ghalanbor Z**. 2004. Correlations between genomic GC levels and optimal growth temperatures are not “robust”. Biochem. Biophys. Res. Commun. **325**:381–3.

112. **Musto H**, **Naya H**, **Zavala A**, **Romero H**, **Alvarez-Valín F**, **Bernardi G**. 2004. Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. FEBS Lett. **573**:73–7.

113. **Shizuya H**, **Birren B**, **Kim UJ**, **Mancino V**, **Slepak T**, **Tachiiri Y**, **Simon M**. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc. Natl. Acad. Sci. **89**:8794–8797.

114. **Sternberg N**. 1990. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci. **87**:103–107.

115. **Uga H**, **Matsunaga F**, **Wada C**. 1999. Regulation of DNA replication by iterons: an interaction between the ori2 and incC regions mediated by RepE-bound iterons inhibits DNA replication of mini-F plasmid in Escherichia coli. EMBO J. **18**:3856–67.

116. **Bouet J-Y**, **Nordström K**, **Lane D**. 2007. Plasmid partition and incompatibility--the focus shifts. Mol. Microbiol. **65**:1405–14.

117. **Couturier M**, **Bex F**, **Bergquist PL**, **Maas WK**. 1988. Identification and classification of bacterial plasmids. Microbiol. Rev. **52**:375–95.

118. **Kato J**, **Ikeda H**. 1996. Construction of mini-F plasmid vectors for plasmid shuffling in Escherichia coli. Gene **170**:141–142.

119. **Widlund PO**, **Davis TN**. 2005. A high-efficiency method to replace essential

genes with mutant alleles in yeast. Yeast **22**:769–74.

120. **Muttucumaru DGN**, **Parish T**. 2004. The molecular biology of recombination in Mycobacteria: what do we know and how can we use it? Curr. Issues Mol. Biol. **6**:145–57.

121. **Yu BJ**, **Kang KH**, **Lee JH**, **Sung BH**, **Kim MS**, **Kim SC**. 2008. Rapid and efficient construction of markerless deletions in the Escherichia coli genome. Nucleic Acids Res. **36**:e84.

122. **Sawitzke JA**, **Thomason LC**, **Costantino N**, **Bubunenko M**, **Datta S**, **Court DL**. 2007. Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic EngineeringMethods in Enzymology. Elsevier.

123. **van Kessel JC**, **Marinelli LJ**, **Hatfull GF**. 2008. Recombineering mycobacteria and their phages. Nat. Rev. Microbiol. **6**:851–7.

124. **van Kessel JC**, **Hatfull GF**. 2007. Recombineering in Mycobacterium tuberculosis. Nat. Methods **4**:147–52.

125. **Gallagher LA**, **McKevitt M**, **Ramage ER**, **Manoil C**. 2008. Genetic dissection of the Francisella novicida restriction barrier. J. Bacteriol. **190**:7830–7.

126. **Park UE**, **Olivera BM**, **Hughes KT**, **Roth JR**, **Hillyard DR**. 1989. DNA ligase and the pyridine nucleotide cycle in Salmonella typhimurium. J. Bacteriol. **171**:2173–80.

127. **Sriskanda V**, **Schwer B**, **Ho CK**, **Shuman S**. 1999. Mutational analysis of Escherichia coli DNA ligase identifies amino acids required for nick-ligation in vitro and for in vivo complementation of the growth of yeast cells deleted for CDC9 and LIG4. Nucleic Acids Res. **27**:3953–63.

128. **Bertani G**. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J. Bacteriol. **62**:293–300.

129. **Murthy MSS**, **Rao BS**, **Reddy NMS**, **Subrahmanyam P**, **Madhvanath U**. 1975. Non-equivalence of YEPD and synthetic complete media in yeast reversion studies. Mutat. Res. Mol. Mech. Mutagen. **27**:219–223.

130. **Gietz RD**, **Schiestl RH**. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. **2**:31–4.

131. **Christianson TW**, **Sikorski RS**, **Dante M**, **Shero JH**, **Hieter P**. 1992. Multifunctional yeast high-copy-number shuttle vectors. Gene **110**:119–122.

132. **Anthony LS**, **Gu MZ**, **Cowley SC**, **Leung WW**, **Nano FE**. 1991. Transformation and allelic replacement in Francisella spp. J. Gen. Microbiol. **137**:2697–703.

133. **Overbeek R**, **Olson R**, **Pusch GD**, **Olsen GJ**, **Davis JJ**, **Disz T**, **Edwards RA**, **Gerdes S**, **Parrello B**, **Shukla M**, **Vonstein V**, **Wattam AR**, **Xia F**, **Stevens R**. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. **42**:D206–14.

134. **Gibson DG**, **Young L**, **Chuang R-Y**, **Venter JC**, **Hutchison CA**, **Smith HO**. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods **6**:343–5.

135. **Amit R**, **Garcia HG**, **Phillips R**, **Fraser SE**. 2011. Building enhancers from the ground up: a synthetic biology approach. Cell **146**:105–18.

136. **Ceres P**, **Trausch JJ**, **Batey RT**. 2013. Engineering modular “ON” RNA switches using biological components. Nucleic Acids Res. **41**:10449–61.

137. **Yang L**, **Nielsen AAK**, **Fernandez-Rodriguez J**, **McClune CJ**, **Laub MT**, **Lu TK**, **Voigt CA**. 2014. Permanent genetic memory with >1-byte capacity. Nat. Methods **11**:1261–6.

138. **Zhang R**, **Lin Y**. 2009. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res. **37**:D455–8.

139. **Tzamali E**, **Poirazi P**, **Tollis IG**, **Reczko M**. 2011. A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities. BMC Syst. Biol. **5**:167.

140. **Lehman IR**. 1974. DNA ligase: structure, mechanism, and function. Science **186**:790–7.

141. **Georlette D**, **Damien B**, **Blaise V**, **Depiereux E**, **Uversky VN**, **Gerday C**, **Feller G**. 2003. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J. Biol. Chem. **278**:37015–23.

142. **Arnold FH**, **Wintrode PL**, **Miyazaki K**, **Gershenson A**. 2001. How enzymes adapt: lessons from directed evolution. Trends Biochem. Sci. **26**:100–6.

143. **Metpally RPR**, **Reddy BVB**. 2009. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. BMC Genomics **10**:11.

144. **Zhou X-X**, **Wang Y-B**, **Pan Y-J**, **Li W-F**. 2008. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. Amino Acids **34**:25–33.

145. **Stemmer W**. 1994. DNA Shuffling by Random Fragmentation and Reassembly: In vitro Recombination for Molecular Evolution. Proc. Natl. Acad. Sci. **91**:10747–10751.

146. **Struvay C**, **Feller G**. 2012. Optimization to low temperature activity in psychrophilic enzymes. Int. J. Mol. Sci. **13**:11643–11665.

147. **Nogi Y**, **Hosoya S**, **Kato C**, **Horikoshi K**. 2004. Colwellia piezophila sp. nov., a novel piezophilic species from deep-sea sediments of the Japan Trench. Int. J. Syst. Evol. Microbiol. **54**:1627–31.

148. **Hashimoto M**, **Ichimura T**, **Mizoguchi H**, **Tanaka K**, **Fujimitsu K**, **Keyamura K**, **Ote T**, **Yamakawa T**, **Yamazaki Y**, **Mori H**, **Katayama T**, **Kato J**. 2005. Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. Mol. Microbiol. **55**:137–49.

149. **Pankowski JA**, **Puckett SM**, **Nano FE**. 2016. Temperature Sensitivity Conferred by ligA Alleles from Psychrophilic Bacteria upon Substitution in Mesophilic Bacteria and a Yeast Species. Appl. Environ. Microbiol. **82**:1924–32.

150. **Chang SF**, **Ng D**, **Baird L**, **Georgopoulos C**. 1991. Analysis of an Escherichia coli dnaB temperature-sensitive insertion mutation and its cold-sensitive extragenic suppressor. J. Biol. Chem. **266**:3654–60.

151. **Kobayashi I**, **Ikeda H**. 1978. On the role of recA gene product in genetic recombination: An analysis by in vitro packaging of recombinant DNA molecules formed in the absence of protein synthesis. MGG Mol. Gen. Genet. **166**:25–29.

152. **Jiang W**, **Bikard D**, **Cox D**, **Zhang F**, **Marraffini LA**. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. **31**:233–9.

153. **Ashford SR**. 1996. Bacteriophage T7 DNA Ligase. J. Biol. Chem. **271**:11083–11089.