UNIVERSITY OF TARTU

FACULTY OF SCIENCE AND TECHOLOGY

INSTITUTE OF MOLECULAR AND CELL BIOLOGY

The possibilities of using plasmid ORF clone library for *ybeY* gene compensation studies in *Escherichia coli*

Bachelor thesis

12 EAP

Amata Žukova

Supervisor: Ismail Sarigül, MSc

Co-supervisor: Ülo Maiväli, PhD

TARTU 2020

The possibilities of using plasmid ORF clone library for *ybeY* gene compensation studies in *Escherichia coli*

The YbeY, encoded by the highly conserved *ybeY* gene, is a multi-functional protein in *Escherichia coli*. The processing of the 3′ terminus of 16S rRNA is its most important function, which has a direct effect on the biogenesis of ribosomes. The deletion of *ybeY* leads to phenotypic and translational defects. YbeY has an extensive network of interactions with other proteins, many of which are still not fully understood. ORF library can be used as a powerful biological tool to detect interactions between proteins. Two different selection strategies were applied using two pooled Gateway® entry clone libraries. Fully pooled library containing all genes represented in the library, and partially pooled library with the lack of some library genes, were electroporated into *ybeY* null mutants. According to the specificity of each pooled library, the growth selections and PCR methods for genes detecting were different. Results have shown that cultures growth improves during selection. Several genes, whose amount prevails over others in the last selection cycles, were sequenced.

KEYWORDS: YbeY, ORF library, selection, Gateway® entry clone library

CERCS: B230 Microbiology, bacteriology, virology, mycology

Geeniekspressiooni raamatukogu kasutamise võimalused *Escherichia coli` ybeY* geeni kompensatsiooni uuringutes

YbeY on multifunktsionaalne *Escherichia coli* valk, mida kodeerib kõrgelt konserveerunud *ybeY* geen. Tema peafunktsioon on 16S rRNA 3′ otsa töötlemine, millel on otsene mõju ribosoomide biogeneesile. *ybeY* deletsioon põhjustab fenotüübilisi ja translatsioonilisi defekte. YbeY valgul on laialdane interaktsioonide võrk teiste valkudega, millest siiani pole palju teada. Geenieksprессionni raamatukogu saab kasutada võimsa vahendina valkude vaheliste funktsionaalsete interaktsioonide tuvastamiseks. Antud töös rakendati kahte erinevat selektsiooni strateegiat kasutades kahte ühendatud Gateway® kloonide kogu. Täielikult ühendatud kloonikogu, mis sisaldas kõiki kloonikogus esindatud geene ja osaliselt ühendatud kloonikogu, kus puudusid mõned geenid elektroporeeriti *ybeY* nullmutantidesse. Kasvu selektsiooni ja PCR-meetodeid geenide tuvastamiseks kasutati vastavalt antud kloonikogu eripärale. Tulemused näitavad, et kultuuride kasv paraneb selektsiooni jooksul. Geenid, mille kogus prevaleeris teiste üle sekveneeriti.

MÄRKSÕNAD: YbeY, geenieksprессiooni raamatukogu, selektsioon, Gateway® kloonikogu

CERRCS: B230 Mikrobioloogia, bakterioloogia, viroloogia, mükoloogia

CONTENTS

LIST OF ABBREVIATIONS

Å - unit ångström

BAC - Bacterial Artificial Chromosome

bp - base pair

C-terminal - carboxyl-terminal

DNA - deoxyribonucleic acid

EHEC - Enterohemorrhagic *Escherichia coli*

kb - kilobase

LB - Luria-Bertani broth

nt – nucleotides

OD – optical density

ORF – open reading frame

IPTG - isopropyl β- d-1-thiogalactopyranoside

PCR - Polymerase Chain Reaction

RNA - ribonucleic acid

rRNA - ribosomal ribonucleic acid

sRNA - small ribonucleic acid

shRNA – short hairpin ribonucleic acid

SOB - super optimal broth

tRNA - transfer ribonucleic acid

UV - ultraviolet radiation

YAC - Yeast Artificial Chromosome

WT – wild-type

INTRODUCTION

Protein synthesis is a basic function of any living organism. Correct maturation of the ribosome is the key for normal functioning of the translation system. This is a complex process, in which a variety of proteins are involved, with different degrees of importance, for the survival of the organism. The YbeY, encoded by the highly conserved *ybeY* gene, is a good example of a multi-functional protein, which plays a crucial role in the processing of the 3′ terminus of 16S rRNA, and thus has a great effect on the ribosomal maturation in general (Sulthana and Deutscher, 2013). Although the main function of YbeY is connected with ribosomal activity, it participates in many other processes occurring in a bacterial cell: 70S ribosome quality control, transcription of the antitermination sequences, resistance to environmental stresses (high temperatures, oxidative, UV and antibiotic stresses) and sRNA-mediated regulation (Jacob et al., 2013); (Grinwald and Ron, 2013); (Vercruysse *et al*., 2016); (Rasouly et al., 2009); (Pandey *et al*., 2014). Despite the studies that are already performed on the YbeY, some of its functions and its mechanisms remain unknown. The multifunctionality of this protein generates a complex system of interactions with other proteins, what can be used to shed light on YbeY role in the cell.

The ORF libraries are a powerful tool to detect interactions between proteins, which can be complicated for multifunctional proteins. In this thesis, the Gateway® entry clone library was used, which has approximately 4000 different genes inserted into plasmids. The selection strategies presented in this study can significantly increase the number of proteins, whose functions and molecular mechanisms are potentially associated with YbeY, thereby providing the data for further research. The basis of our method is natural compensatory mechanisms. The ribosomal maturation is a vitally important set of processes, which have various alternative ways. The lack of *ybeY* makes the bacteria activate those pathways using library plasmids contained other genes for compensation. Thus, it indicates direct and indirect interactions of YbeY with other proteins.

The theoretical part of this thesis provides an overview of the ribosomal structure, its maturation process, focusing on rRNA processing, thereby shining light to the role of YbeY. The possibilities of using genomic libraries and features of the Gateway® cloning system are also discussed. The experimental part aims at designing a protocol for ORF library experiments, applying the growth selection for differently pooled libraries and analyzing the results.

1. LITERATURE REVIEW

1.1 Prokaryotic ribosomes - an overview

Ribosomes are essential cell organelles, whose main function is protein synthesis, also known as translation. Ribosome is a macromolecular ribonucleoprotein (RNP) complex, whose basic structure is conserved in all domains of life. In a broad sense, ribosome is a ribozyme and ribonucleic acid (RNA) based polymerase (Rodnina *et al*., 2007).

Ribosomes are found in both prokaryotes and eukaryotes, while the structural organization, location and the process of maturation of ribosomes have differences between prokaryotic and eukaryotic cells. Ribosomes are made up of ribosomal proteins and ribosomal RNA (rRNA). The approximate mass of bacterial ribosome is 2.6-2.8 MDa. It has a diameter of 200–250 Å and a sedimentation coefficient of 70S. The 70S ribosome includes two subunits: a large 50S subunit and a small 30S subunit. Each subunit consists of rRNA and proteins: a single 16S rRNA (~1500 nt) in the 30S subunit with 21 proteins, and a 5S (~120 nt) with 23S rRNA (~2900 nt) with 34 proteins in the large subunit. (Nierhaus and Wilson, 2004)

Ribosomal subunits have different functions in translation. The 50S subunit plays the main role in catalyzing peptide bond formation. Furthermore, it binds initiation, termination and elongation factors. (Ban et *al.*, 2000) The main functions of the small subunit are initiating the interaction with mRNA during translation initiation and decoding its message. (Kaczanowska and Rydén-Aulin, 2007).

Ribosomes have three tRNA-binding sites: A (aminoacyl), P (peptidyl) and E (exit). The decoding of mRNA into protein is based on codon-anticodon system. According to the codon of mRNA, the specific tRNA binds the ribosome at the A site. A covalent bond is created between amino acid from previous tRNA at the P site and the amino acid at the A site. E-site (the exit site) is needed for the deacylated tRNA to exit the ribosome (Rheinberger *et al.*, 1981).

1.2 Ribosomal ribonucleic acid (rRNA)

The ribosomal ribonucleic acid (rRNA) is an essential part of ribosome to perform its function-protein synthesis. The rRNA is present in all types of cells, and it is used to determine the taxonomic position of organisms (Smit *et al.*, 2007).

The rRNA can be represented in three structural levels: primary, secondary and tertiary (Brimacombe and Stiege, 1985).

The secondary structure of 23S rRNA includes four regions, which consist of 6 domains (Figure 1). A helix is formed by combining the 5’ and 3’ terminal ends. The loop of this helix gives a start to 11 stem-loop structures of differing degrees of complexity (Wilson *et al*., 2002). The 5S rRNA structure is firmly connected to 23S rRNA. There are 24 cross-link sites defining tertiary contacts within or between the 23S and 5S rRNAs (Petrov *et al*., 2013).

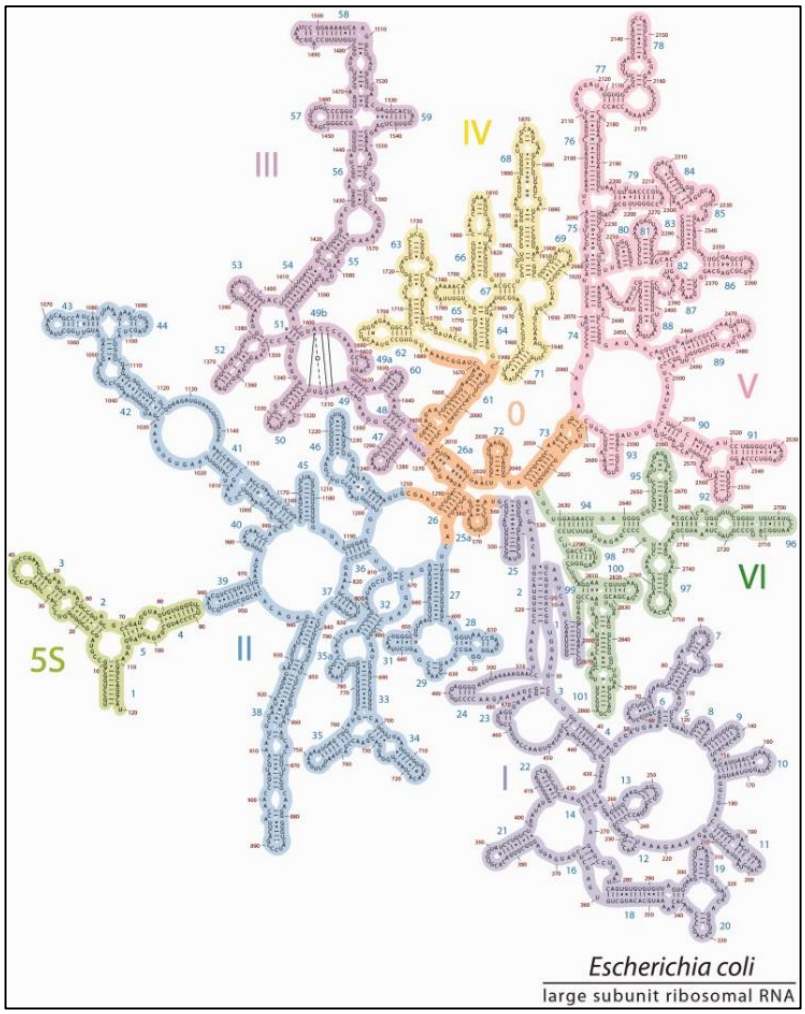


Figure 1 | The structures of 23S rRNA and 5S rRNA in *Escherichia coli*. The 23S rRNA has 6 domains: I, II, III, IV, V, VI. The figure also shows the structure of 5 S rRNA (Petrov *et al.*, 2013).

The secondary structure of 16S rRNA has 4 domains: 5’ domain, central domain, 3’ major and 3’ minor domains (Figure 2). The main difference comparing with a large subunit rRNA domains in the absence of the tertiary fold of 16S rRNA. Furthermore, the anti Shine-Dalgarno sequence is located in 3’ of 16S rRNA (Shine and Dalgarno, 1974). The binding between the anti-Shine-Dalgarno sequence and specific mRNA sequence (Shine-Dalgarno sequence) contributes translation initiation, and increases translation efficiency (Li *et al*., 2012).

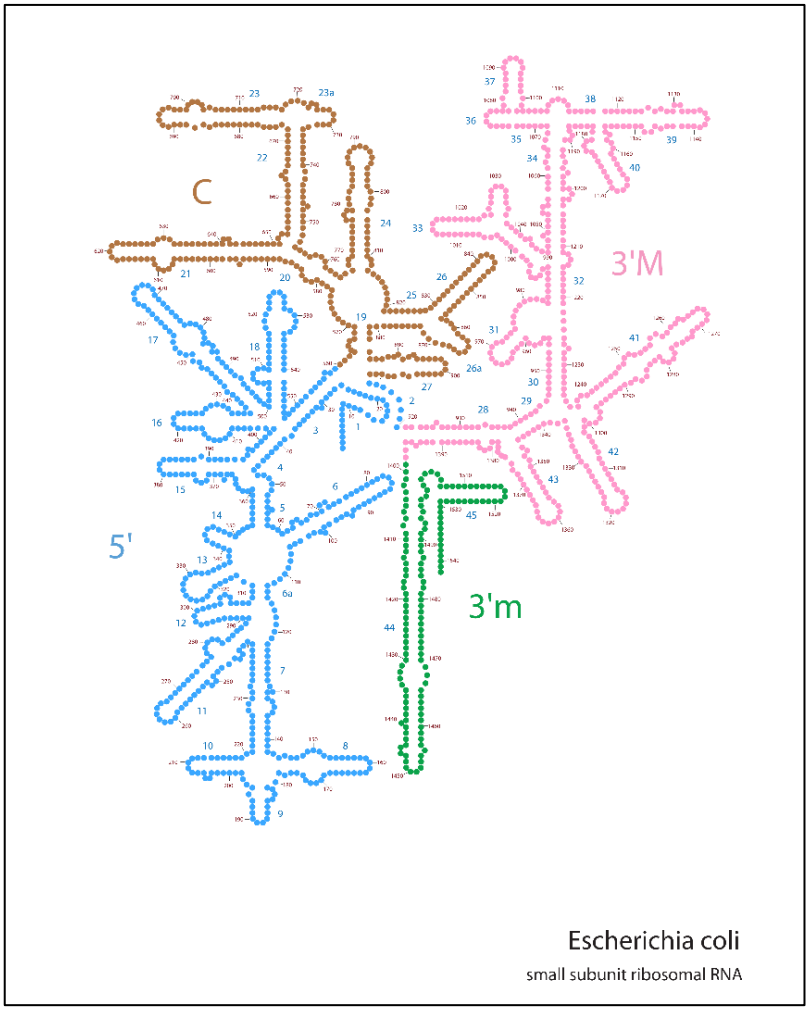


Figure 2 | The structures of 16S rRNA in *Escherichia coli*. The figure shows 5' domain, central domain, 3' major and 3' minor domains. (The scheme was taken from database of rRNA secondary structures http://apollo.chemistry.gatech.edu/RibosomeGallery/index.html).

1.2.1 rRNA processing in *Escherichia coli*

The transcription of the *Escherichia coli rrnB* operon containing 16S, 23S, and 5S rRNA coding sequences, result in a single primary transcript containing not only the above listed rRNAs, but can also include tRNAs. The maturation of the primary transcript begins before transcription is completed, which enables instantaneous formation of local secondary structures and the processing of rRNAs by several RNases (Williamson, 2003; Davis and Williamson, 2017). The process of complete maturation of rRNA consists of two stages. The double-strand-specific endoribonuclease RNase III cleaves the primary rRNA transcript by separating precursor rRNAs and tRNAs during the first stage. The RNase III recognizes double helical structures such as hairpin loops, which are formed by a transcript regions containing 16S and 23S rRNA. The presence of RNase III in a bacteria is critical to 23S rRNA full maturation, but not for full maturation of 16S rRNA (King and Schlessinger, 1983). The precursor 16S rRNA

(17S rRNA), precursor 23S rRNA, precursor 5S rRNA (9S rRNA), and few tRNA precursors are final points of the first stage.

The essence of the second stage is in giving the correct length to the. A 17S precursor of 16S rRNA contains an additional 115 nucleotides at its 5’ end and 33 nt at the 3’ end. Those nucleotides must be removed (Figure 3). RNase E and RNase G execute this removal process at the 5’ end of the 16S rRNA. First, RNase E forms the 16.3S precursor, which has extra 66 nt, which are eventually trimmed by RNase G (Li *et al.*, 1999a). The maturation process of the 3’ end of 16S rRNA have two possible pathways. The endoribonuclease YbeY plays a key role in one of the pathways. This protein is described in more details below. The second pathway involves such exoribonucleases as PNPase, RNase PH, RNase R and RNase II. Each of them processes the 3’ end with different efficiency in the absence of other exoribonucleases. Furthermore, it is important to mention that the processing of 3’ end of 16S rRNA has a big influence on 5’ end processing: in the absence of correct 3’ end maturation, 5’ processing occurs much less efficiently (Sulthana and Deutscher, 2013).

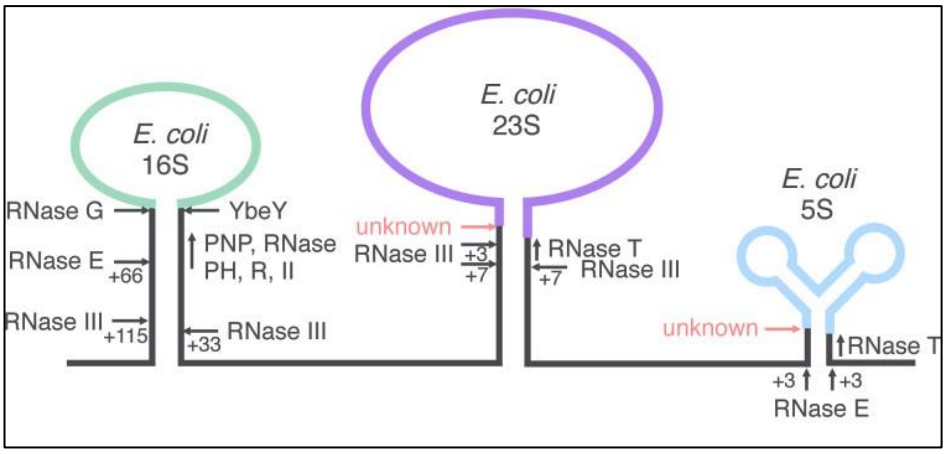


Figure 3 | Ribosomal RNA processing in *Escherichia coli.* The scheme shows cutting places of RNase E and RNase G at the 5’ end, and cutting places of YbeY, PNPase, RNase PH, RNase R and RNase II at the 3’ end of the 16S rRNA. Cutting sites of RNase III and RNase T are also noted on 23S and 5S rRNAs. Unknown enzymes are marked in red (Illustration is adapted from Baumgardt *et al*., 2018)

The final maturation of the 5’ end of the 23S rRNA is performed by an unknown enzyme, although for the 3’ end, it is well established that exoribonuclease RNase T is involved in its processing. In the absence of RNase T, other exoribonucleases can help to cut the 3’ end of 23S rRNA, but RNase T is required for removing the last few residues. In the absence of RNase T, 23S rRNA with extra nucleotides is incorporated into ribosomes, which ultimately has very little effect on bacterial growth (Li *et al.*, 1999b).

The final stage of 5S rRNA processing includes two steps. RNase E cleaves 9S rRNA precursor, as a result the 5' end is processed by an unknown enzyme, and the 3' end is trimmed by RNase T (Ghora and Apirion, 1978).

It is very important to take into account the fact that rRNA processing is only one link in the chain of ribosome biogenesis, and the accuracy of this process depends on the previous steps, just like this stage affects the subsequent steps. The sequence of biogenesis events is a complex system, which includes addition of ribosomal factors , cleavage and trimming of rRNA precursors, and rRNA modifications. Furthermore, substitutions and alternative pathways can be found in this system as well, giving it flexibility regarding mutation and environmental conditions (Deutscher, 2009).

1.3 *Escherichia coli* operon *ybeZYX-lnt*

Operon *ybeZYX-lnt* is an *Escherichia coli* operon that includes 4 genes: *ybeZ*, *ybeY*, *ybeX* and *lnt* (**Figure 4**). This operon is involved in various processes in the bacterial cell, however, its main function- the 16S rRNA processing- is determined by the *ybeY* gene (Baumgardt *et al*., 2018).

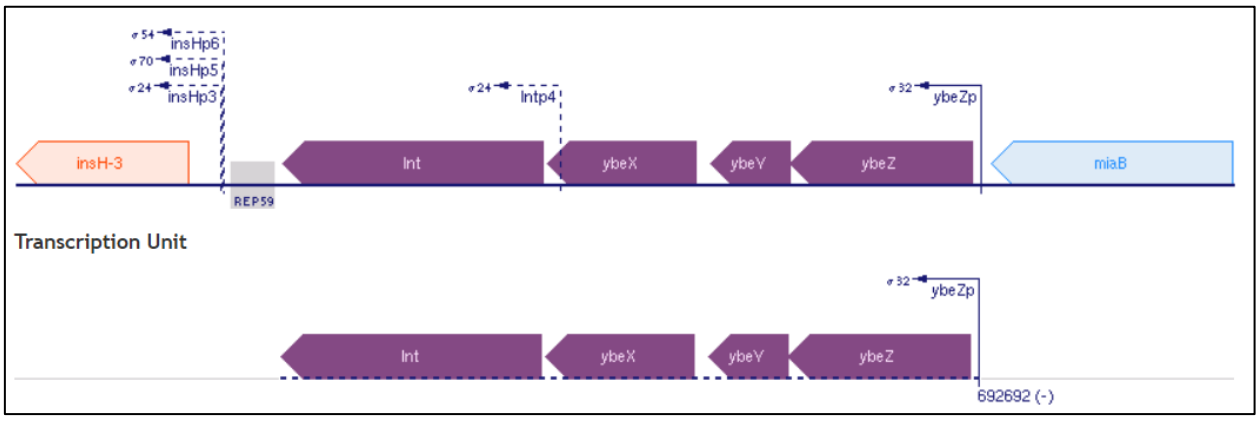


Figure 4 | Operon *ybeZYX-lnt*. The length of the genes: *ybeZ* – 1080 bp, *ybeY* – 468 bp, *ybeX* – 879 bp, *lnt* – 1539 bp. Genes are united under ybeZp promoter, which binds σ^{32} factor. The presence and location of Intp4 promoter is inferred computationally ("Summary of Escherichia coli K-12 substr. MG1655, version 19.5," n.d., p. 5)

The operon has two promoters: ybeZp and Intp4. The ybeZp promotor binds σ^{32} factor, which controls the heat shock response during log-phase growth. The σ^{32}- dependent heat shock response can be also induced by ethanol, alkaline pH changes and hyperosmotic shock (VanBogelen *et al*., 1987), (Bianchi and Baneyx, 1999), (Taglicht *et al*., 1987). The Intp4 promotor has similarity to the sequence of the σ factor promoters. According to its high homology index, its presence has been suggested computationally. The Intp4 promotor binds the minor σ^{E} factor, which is necessary for transcription of a number of genes, whose functions

are associated with heat shock reactions and misfolded proteins (Ades *et al*., 2003), (Mecsas and Rouviere, 1993).

The *ybeZ* encodes a putative ATPase YbeZ. This is a PhoH-like protein, which has a nucleoside triphosphate hydrolase domain, and can interact closely with metalloprotein YbeY. Its functions remain unknown (Vercruysse *et al.*, 2016).

The product of the conserved *ybeX* gene, the YbeX protein, shows a high homology to *Salmonella typhimurium* CorC, which is involved in Mg^{2+} and Co^{2+} efflux (Gibson *et al*., 1991).

The Mg^{2+} plays important role in regulation of ribosome biogenesis. Low level of Mg^{2+} in bacteria leads to reduction in the number of ribosomes, which is a genetically encoded response. Also, the growth of *Escherichia coli* with engineered tethered ribosomal subunits increases because of mutation in the *ybeX* (Orelle *et al*., 2015), (Pontes *et al*., 2016).

The *lnt* gene encodes enzyme apolipoprotein N-acyltransferase, which catalyzes the last step in lipoprotein maturation, and localizes to inner membrane (Gupta and Wu, 1991). The homologous gene *cutE* in *Salmonella typhimurium* is involved in the heat sensitivity and the copper sensitivity (Gupta *et al.*, 1993).

1.4 The *ybeY*

This highly conserved gene is one of 206 included in the minimal gene set of *Escherichia coli* (Gil *et al*., 2004). Homologues are found in gram-negative and in gram-positive bacteria, as well as in plants and mammals cells, although not in yeast or fungi. (Anantharaman and Aravind, 2003), (Ghosal *et al.*, 2017), (Liu *et al.*, 2015). The *ybeY* has strong genetic interactions with *rnc* (RNase III), *rnr* (RNase R) and *pnp* (PNPase) (Davies *et al*., 2010). The gene *ybeY* encodes metalloprotein YbeY (Zhan *et al*., 2005)

1.4.1 The structure of metalloprotein YbeY

The YbeY is a member of the poorly characterized highly conserved UPF0054 protein family. Its sequence similarity to metal-dependent hydrolases indicates a hydrolytic function. The YbeY structure consists of six α-helices and four β-strands in a β-α-α-β-α-β-β-α-α-α fold (**Annex 1**). This metalloprotein has a tetrahedral geometry with a metal ion in the center. The X-ray fluorescence analysis was used to recognize that this metal ion is probably a Ni^{2+} ion, which is surrounded by 3 highly conserved histidine residue (His114, His118 and His124). In addition to the histidines located around the metal-binding site, Arg59, Lys61, Asn66 and Ser69 are also present close by and are needed for the functional activity of the YbeY. According to

the structural similarity to a hypothetical monomeric protein AQ_1354 from *Aquifex aeolicus*, YbeY protein can be monomeric as well (Zhan *et al*., 2005).

1.4.2 The functions of metalloprotein YbeY

The YbeY is a multi-functional protein, whose tasks are closely related to ribosomal activity. Firstly, this protein is involved in the processing of the 3′ terminus of 16S rRNA. As mentioned earlier, this process has two possible pathways: the activity of endoribonuclease YbeY or interaction of multiple 3′-5′ exoribonucleases (PNPase, RNase PH, RNase R and RNase II) (Sulthana and Deutscher, 2013). The *ybeY* null mutants have more 17S rRNA than 16S rRNA, and substantial amounts of a shortened form of 16S rRNA, named as 16S* rRNA. Both 17S rRNA and 16S* rRNA are found in the 30S ribosomal subunits of mutant strains, although only 17S rRNA can be presented in the 70S. This could mean that 16S* rRNA is a non-functional product of ribosome subunits assembly. Furthermore, possibly due to interactions between *ybeY* and *rnc, rnr* and *pnp,* the YbeY has an effect on 23S and 5S rRNAs maturation as well. It is important to note that as only RNase III can also affect the processing of multiple rRNA termini, other RNases seem not have such strong effects on rRNA maturation as does the YbeY. Thus, the lack of this metalloprotein leads to defects in ribosome function, including decreased ribosome activity, decreased translational accuracy, and altered binding of translation initiation factors, which ultimately leads to bacteria phenotypic defects (Davies *et al*., 2010), (Rasouly *et al.*, 2010).

The process of the antitermination is an important regulatory element, which participates in the processing of the mature *rrn* transcript and in modulating the transcription elongation rate (Morgan, 1986). The YbeY is essential for transcription of those *rrn* regions that contain the antitermination sequences known as "*nut*-like sequences", which bind factors that are responsible for the formation of the antitermination complex (Grinwald and Ron, 2013).

The metalloprotein YbeY also participates in 70S ribosome quality control. The recognition and elimination of defective non-translating 70S ribosomes generated in a *ybeY* null mutant can be performed by cooperation of YbeY with RNase R. Together they can also degrade defective 70S ribosomes that are included into translating polysomes (Jacob *et al*., 2013). Interestingly the YbeY-RNase R complex activity is triggered by defective 30S subunit in the 70S ribosome, not the 50S ribosomal subunit, and it leads to the degradation of the entire 70S ribosome, including normal 50S subunits (Warner, 2013).

The YbeY is a heat shock protein. As temperature rises, functional activity of the YbeY becomes increasingly important for the survival of the bacteria. The *ybeY* deletion mutants have

growth defects at higher temperatures and no thermotolerance at lethal temperatures. Furthermore, null mutants have the translational defects, which can be detected already under normal growth conditions at 37 ° C, and their number increases with temperature (Rasouly *et al.*, 2009). In addition, YbeY protein functions are associated with the ability of bacteria to survive under various harmful conditions, such as oxidative, UV and antibiotic stresses (Vercruysse *et al*., 2016).

Some RNases, such as RNase E and RNase III, participate in sRNA-mediated regulation, and YbeY is one of them (Aiba, 2007), (Storz *et al*., 2011). The sRNA (small RNA) regulates gene expression most often by binding to an RNA chaperone or Hfq and using imperfect base-pairing interactions with mRNA target (Waters and Storz, 2009). The RNA-binding protein Hfq is necessary for the functioning of sRNAs in some bacteria species, other species have Hfq-independent sRNA-regulation pathway (Brennan and Link, 2007). The protein YbeY participate in both Hfq-dependent and independent sRNA regulation variants. The YbeY-dependent sRNAs play important role in adapting to oxidative and free radicals stresses. Although the exact molecular-biochemical mechanisms of interaction between YbeY and sRNA regulation system remain unknown (Pandey *et al.*, 2014).

Escherichia coli can be pathogenic for human, in this case it is called enterohemorrhagic *Escherichia coli* (EHEC). These strains can express specific effector proteins, which gives opportunity to colonize host cells. All these proteins are combined into type 3 secretion (T3S) system. Since T3S is activated indirectly through rRNA maturation, then YbeY plays important role in the formation of virulence in enterohemorrhagic *Escherichia coli.* Moreover, the absence of the YbeY leads to reduced motility cold sensitivity in EHEC (McAteer *et al*., 2018). Also in addition to the virulence issue, it is important to note that YbeY has a very similar functions in *Vibrio cholerae:* the lack of YbeY causes reduced biofilm formation, reduced intestinal colonization of mice, decreased production of cholera toxin (CT) and decreased amount of virulence-associated sRNAs. All this just emphasizes the importance of protein YbeY in the formation of virulence in bacteria (Vercruysse *et al*., 2014).

In conclusion to this chapter, it is important to pay attention to the multifunctionality of the YbeY, and to the related difficulties. The study of YbeY is complicated by its intricate system of interactions with other proteins, which can be involved in different functions or in overlapping functions. This provides a good basis for schemes of identifying YbeY partners and its interactions. For example, some work suggests that ribosomal protein S11 and GTPase Era are partners with the YbeY during 16S rRNA maturation. Other ribosomal factors can also be involved (Vercruysse *et al*., 2016). The production levels of initiation factors IF2 and IF3

was affected by deletion of *ybeY* gene (Davies *et al.*, 2010). The theoretical B2H (Bacterial two-hybrid analysis) data based partners of YbeY are proteins Der (ribosome biogenesis function), YbeZ (unknown function) and SpoT (stress response function) (Vercruysse *et al*., 2016).

1.4.3 The compensation of the absence of *ybeY* gene

The YbeY has putative non-specific endoribonuclease activity, which means that it could require the presence of proteins, that would limit and direct this activity to the appropriate region of the 16S rRNA precursor (Jacob *et al.*, 2013). The ribosome-associated highly conserved GTPase Era, whose expression can partially compensate for the lack of YbeY, is a candidate for this (Ghosal *et al*., 2018). Era is one of the ribosomal factors (in addition to RbfA, RimM and RsgA/YjeQ) that participate in the late assembly of the 30S ribosomal subunit, while being essential for efficient 16S rRNA processing. (Thurlow *et al*., 2016). Era has a unique KH domain, which recognizes specific RNA sequences. The GTP-hydrolyzing activity of Era is increasing RNA recognition (Tu *et al.*, 2009). Overexpression of Era improves the processing of 16S rRNA, the 70S ribosome assembly and culture growth in the strains with *ybeY* gene deletion, although it does not decrease sensitivity of *ybeY* null mutants to stresses such as heat shock, UV radiation, nitrofurazone and antibiotics. The Era activity in the rRNA processing depends on interactions of RNase II, RNase R, RNase PH and PNPase (Ghosal *et al*., 2018).

The overexpression of some transcription antitermination factors (NusA, NusB, NusE and NusG) can also partially compensate growth defects of *ybeY* deleted *Escherichia coli* strains (Grinwald and Ron, 2013).

1.5 Genomic library as a biological research tool

A genomic library represents a set of DNA clones, which make up the full or partial genome of a single organism. The different genes are isolated from the main organism's genome and inserted into the vectors. This stage consists of three processes: chromosomal DNA extraction, digestion with restriction enzyme and inserting the fragments into the identical vectors by using DNA ligase. Then each vector containing one insert gene is transformed into a host organism, *Escherichia coli* and yeast are the most commonly used hosts (James D. Watson *et al.*, 2004). The method of library preparation may vary depending on the type of library. For example, actively transcribed mRNA is the basis for creating cDNA libraries, which do not have nontranscribed sequences (Kooiker and Xue, 2014). The library can be screened to isolate clones with certain gene from library. Hybridization with labelled probes is a widely used

method for this purpose. Polymerase chain reaction (PCR) is also used for library screening (Green *et al*., 2012).

The types of vectors may vary, depending on the application for the library. Plasmids vectors, bacteriophage lambda and cosmids are used for single gene cloning, but they have a very limited cloning capacity (0,1-12 kb; 10-20 kb; 35-45 kb respectively). Bacteriophage P1, BAC (Bacterial Artificial Chromosome) and YAC (Yeast Artificial Chromosome) vectors are very common in genome mapping and sequencing. Moreover, BAC and YAC vectors have a high cloning capacity (30-300 kb and 100-1000 kb respectively) (Nierman and Feldblyum, 2001).

The possibility of using genetic libraries is important in studying the genomes of both eukaryotic and prokaryotic organisms. Characterizing of the genes, closing gaps in genomes, identifying haplotypes, aiding gene discovery – all this is only a part of the areas, where genetic libraries can be useful (Quail *et al*., 2012a), (Quail *et al.*, 2012b).

Pooled library is a simple and effective tool for studying various phenotypes and identifying related genes. Pooled library represents a mixture of all plasmids from a library in a single tube, which provides opportunities for various selection and screening experiments. This tool can be used with many different kinds of libraries: CRISPR libraries, virus libraries, including shRNA libraries, screening libraries, cDNA libraries, barcoding libraries (Kweon *et al*., 2018), (Zhang *et al*., 2017), (Coussens *et al.*, 2011), (Boettcher *et al*., 2019).

1.5.1 The Gateway® cloning system

The Gateway® cloning system by Invitrogen is a patented method to transfer DNA fragment into plasmid vector by using specific recombination sites - *att* sites. The system is widely used in creating ORF libraries. This technique is based on site-specific recombination system of phage λ. It integrates its DNA in the *E.coli* chromosome by using its own site called *att*P and *E.coli* site called *att*B. The process is catalyzed by the phage λ encoded enzyme Int (Integrase) and the *E. coli* enzyme IHF (Integration Host Factor). The *att*L and *att*R sites are generated after recombination between *att*P and *att*B sites (Hsu *et al.*, 1980).

The very specific sites - *att*1 and *att*2 for each recombination site – were developed in the Gateway® system (for instance, *att*B1 can interact only with *att*P1 resulting in *att*L1 and *att*R1, the same for *att*2 sites) as well as two enzymes "LR clonase" and "BP clonase". The system includes two steps: BP reaction and LR reaction. Firstly, the gene of interest containing *att*B sites interacts with donor vector containing *att*P sites by using BP clonase. The Gateway Entry clone, containing *att*L sites, flanking gene of interest is a result of this process. LR reaction consists of interaction between the Gateway Entry clone and destination vector containing

*att*R sites and regulatory sequences. The reaction is catalyzed by LR clonase. The result is an expression clone with gene of interest and *att*B sites. The Gateway® cloning system gives the opportunity to maintain integrity of open reading frame (ORF) during transferring. Moreover, it allows to clone multiple DNA fragments into one type vector in a very short time, what makes the Gateway® cloning system perfect for creating genetic libraries (Reece-Hoyes and Walhout, 2018).

2. EXPERIMENTAL PART

2.1 Objective of the study

The ORF libraries are powerful tools for identification and analysis of proteins and the genes encoding them. These kind of libraries are widely used to detect protein–protein interactions. My thesis is focused on finding genes, whose overexpression could phenotypically compensate for the absence of the *ybeY* gene. The pooled ORF library is the best option for this purpose.

Despite of the many advantages of using expression libraries in research, working with them requires long and thorough preparation. The methods of their use are highly dependent on the specificity of the studied protein and its encoding gene. As mentioned earlier, the *ybeY* gene is highly conserved and it is included in the minimal gene set of *Escherichia coli* (Gil *et al*., 2004). The absence of the *ybeY* gene has a strong negative effect on 16S rRNA processing, ribosomal maturation and resistance to various exogenous stresses, which leads to phenotypic and growth defects (Rasouly *et al*., 2010). This complicates the work with *ybeY* knockout strains.

The practical part can be divided into three parts: the optimization of the protocols to apply library selection experiment with *ybeY* null mutants, and two type of selections with fully and partially pooled library. After finding an effective way to transform pooled library plasmids into Δ*ybeY* strain, growth selection was applied. The final task of this thesis was sequencing of the found genes.

The objectives of this study are:

- To develop a suitable protocol for transformation of pooled library plasmids into Δ*ybeY* strain;
- To apply two selection strategies on Δ*ybeY* strains, containing fully or partially pooled library;
- Perform colony screening and total plasmids PCR to detect and sequence genes, whose products rescue *ybeY* phenotype(s).

2.2 Materials and methods

2.2.1 The Gateway® entry clone library

The basis for this library was another ORF library called ASKA library (Kitagawa *et al*., 2005). It was based on the genomic sequence data of *Escherichia coli* K-12 strain. 4333 protein-coding ORFs were cloned into the high copy-number plasmid pCA24N (Kitagawa *et al.*, 2005). Each clone encodes protein of one ORF with six histidines at the N-terminal end and GFP (green fluorescent protein) fusions at the C- terminal end. *Sfi*I restriction sites were generated at both sides of insert (Kitagawa *et al*., 2005). The main problem of ASKA library is that it is difficult to transfer ORFs into other expression vectors. Transferring the ASKA library into an Gateway® entry vector (pENTR/zeo) solves this problem by creating a new type of ORF library - Gateway® entry clone library. The plasmid ENTR/zeo was constructed using Gateway® recombination technology between pAZ677 and pDONOR. 3734 protein-coding ORFs from ASKA library were successfully transferred into pENTR/zeo. In addition 250 ORFs were cloned directly using Gateway® recombination. The library has *att*L1 and *att*L2 sites, what gives the transferring flexibility. Zeocin is used as a resistance marker (Rajagopala *et al*., 2010).

Schematic representation of creating the Gateway® entry clone library can be found in supplemental materials (**Annex 2**).

2.2.2 Pooled Gateway® entry clone libraries

Gateway® library was pooled into two versions by Ismail Sarigül:

- Pooled Gateway library called “Tenson”

This pooled library contains all clones from the Gateway® entry clone library.

- Pooled Gateway library called “ISM”

Some ribosome factors (*rbfA*, *rimM*, *yjeQ* and *era*), *ybeX* and *ybeY* were excluded from this pooled library.

2.2.3 Strains and plasmids

All used strains were from Keio collection. Keio collection is a *Escherichia coli* K-12 strain single-gene knockout collection (Baba *et al*., 2006). The ORF coding regions were replaced with a kanamycin cassette.

Strains	Characteristic	Source
Escherichia coli		
BW25113	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻rph-1, Δ(rhaD-rhaB)568, hsdR514*	(Baba et al., 2006)
BW25113Δ*ybeY*	BW25113 Δ*ybeY::Kan*	
Plasmids		
pENTR/zeo	It was created by cleaving of one clone from Gateway® library with restriction enzymes EcoRV and HpaI	Ismail Sarigül
pYbeY	Gateway® library clone containing *ybeY* gene	Gateway® library

2.2.3 Competent cells preparation

The competent cells were prepared using the Inoue method (Inoue *et al*., 1990). First of all, a stock from -80°C of Δ*ybeY::Kan* cells was plated on kanamycin containing LB agar and incubated overnight at 37°C. Then, one colony was picked and transferred into 3 ml of LB broth containing kanamycin (final concentration was 25 µg/ml) , incubated overnight at 37°C with vigorous shaking (300 rpm). After that, 25 ml of SOB containing kanamycin was inoculated with overnight culture so that diluted culture had OD_{600} = 0,1 in fresh SOB medium. In the next day, four flasks with 250 ml of fresh kanamycin containing SOB were inoculated with 4 ml and 6 ml of overnight culture (2×4 ml and 2×6 ml), and incubated 12-14 hours at 22°C with vigorous shaking (250 rpm). The point of this is to ensure that at least one from started cultures will be caught at the right OD_{600} value, which was monitored with spectrophotometer until it has reached the 0,55 value. This culture was transferred into five 50 ml-tubes, which were cooled in ice for 10 min. The Inoue buffer (250 mM KCI, 15 mM $CaCl_2$, 55 mM $MnCl_2$, 10 mM PIPES, pH 6,7) was put in ice to melt, and PIPES (pH 6,7) were added into it later. The cells culture was pelleted by centrifugation at 3900 rpm for 10 minutes at 4°C. The medium was poured off, and opened tubes were stored on a stack of paper towels for 2 minutes. The cells pellets were resuspended in 80 ml of ice-cold Inoue buffer (16 ml of buffer per one 50 ml-tube) and centrifuged at the previous settings. The supernatant was removed and cells pellets were resuspended in 20 ml of ice-cold Inoue buffer (4 ml of buffer per one 50 ml-tube). DMSO with final concentration 7% was added, the suspension was mixed and stored in ice for 10 min. The

mixture was divided between chilled, sterile microfuge tubes (one tube–250 µl of cells suspension), and snap-frozen using liquid nitrogen and stored at -80°C.

2.2.4 Heat shock transformation

The plasmids of ISM pooled library (pISM) and pYbeY were used for testing possibilities of using this method for library selection experiment. First of all, competent cells Δ*ybeY::Kan* were taken from -80°C and put in ice to melt. The concentration of plasmids was measured with NanoDrop™, and all concentrations were made 70 ng/µl using 1xTE. 50 µl of competent cells were pipetted into new microfuge tubes, and 1 µl of each plasmid (70 ng/µl) was added. The tubes with competent cells plus plasmids were stored in ice for 30 min. After that, heat shock was applied: 90 sec at 42°C, and put back to ice for 90 sec. 950 µl of pre-warmed SOC was added to the cells and plasmids mixtures, and incubated at 37°C for 1 hour. Finally, a part of recovery culture was used for plating on LB agar supplemented with 25 µg/ml of zeocin. 50 ml of LB containing zeocin in concentration 25 µg/ml was inoculated with 500 µl of recovery culture and incubated overnight at 37°C with vigorous shaking for use in subsequent experiments.

2.2.5 Electroporation

Electroporation requires electrocompetent cell preparation, whose main goal is to get rid of the salts in the medium in order to avoid arcing during electro pulse application. Stock cells Δ*ybeY::Kan* from -80°C were plated on LB agar supplemented with 25 µg/ml of kanamycin and incubated overnight at 37°C. One colony was picked, transferred into 3 ml of LB containing kanamycin, and incubated overnight at 37°C with vigorous shaking. The culture was then diluted into 50 ml of SOC with final OD_{600} = 0,1, and incubated at 37°C with vigorous shaking. OD_{600} was monitored until it reached 0,35-0,40 and then put on ice for 15 min. After that culture was transferred into 50 ml ice-cold tubes and centrifuged at 4000 rpm for 12 minutes at 4°C. The supernatant was poured off, and cells pellet was resuspended in 5 ml of ice-cold 10% glycerol, and centrifuged at 4000 rpm for 12 minutes at 4°C. This washing step was repeated two more times. After last centrifugation, cells pellet was resuspended in 830 µl of ice-cold 10% glycerol. This mixture was divided between chilled, sterile microfuge tubes (one tube–50 µl of cell suspension).

The plasmids of ISM pooled library (pISM) and pYbeY were used for method testing, and pYbeY, pENTR/zeo, pISM and the plasmids of Tenson pooled library (pTenson) were electroporated into Δ*ybeY::Kan* for growth selection experiment. Electroporation cuvettes (2

mm path length) and microcentrifuge tubes were placed on ice. The concentrations of plasmids were diluted to 70 ng/µl using 1xTE. 40 µl of Δ*ybeY::Kan* electrocompetent cells and 1 µl (70 ng/µl) of each plasmid were mixed together. The cell/DNA mix was transferred into cuvette, and electrical pulse - 25 uF capacitance, 2.5 kV and 200 ohm resistance- was applied. A time constant of 4-5 milliseconds with a field strength of 12.5 kV/cm was registered on the machine (The Gene Pulser Xcell system, Bio-Rad). 1 ml of pre-warmed SOC was immediately added. The recovery was incubated 1 hour at 37°C with vigorous shaking. In the end, some part of recovery was plated on LB agar supplemented with 25 µg/ml of zeocin. 50 ml of LB containing zeocin (25 µg/ml) was inoculated with 500 µl of recovery, and incubated overnight at 37°C with vigorous shaking for use in subsequent experiments.

2.2.6 Selection strategy for pTenson pooled library

Plasmids of fully pooled library (pTenson), pYbeY and pENTR/zeo as a controls were electroporated into Δ*ybeY::Kan.* After overnight growing in 50 ml of LB with 25 µg/ml of zeocin, cultures were diluted into fresh filtered 50 ml LB supplemented with 25 µg/ml of kanamycin in proportion 1:100. They grew 12 hours at 37°C with vigorous shaking. After this time, cultures were again diluted into fresh LB with kanamycin, and all steps were repeated in the same way. These growth intervals named as cycles. In total, five cycles were done (including recovery overnight growth). The OD_{600} of cultures was monitored between cycles, and after each cycle, 1 ml of cells were stored as 7% DMSO stock for further plating on zeocin containing plates and 24 hours kinetic measurements. (**Figure 6**)

2.2.7 Selection strategy for pISM pooled library

Plasmids of partially pooled library (pISM), pYbeY and pENTR/zeo as controls were electoporated into Δ*ybeY::Kan.* The overnight conditions, dilutions, cells storing method were made at the same way as in the pTenson selection experiment. The main difference was after the third cycle, when plasmid DNA was purified from 3 ml of each culture. These plasmids were re-electroporated into Δ*ybeY::Kan*, and put to grow overnight at 37°C with vigorous shaking. These 12-hour growth intervals were repeated three times. The dilutions were made as described earlier (1:100 dilution in LB with kanamycin, 12 hours at 37°C). After every third cycle plasmids were purified and re-electroporated into Δ*ybeY::Kan.* In total, there were nine cycles of growth including plasmids purification and re-electroporation after the third and the sixth cycles. The last plasmid purification was done after the ninth cycle. (**Figure 5**)

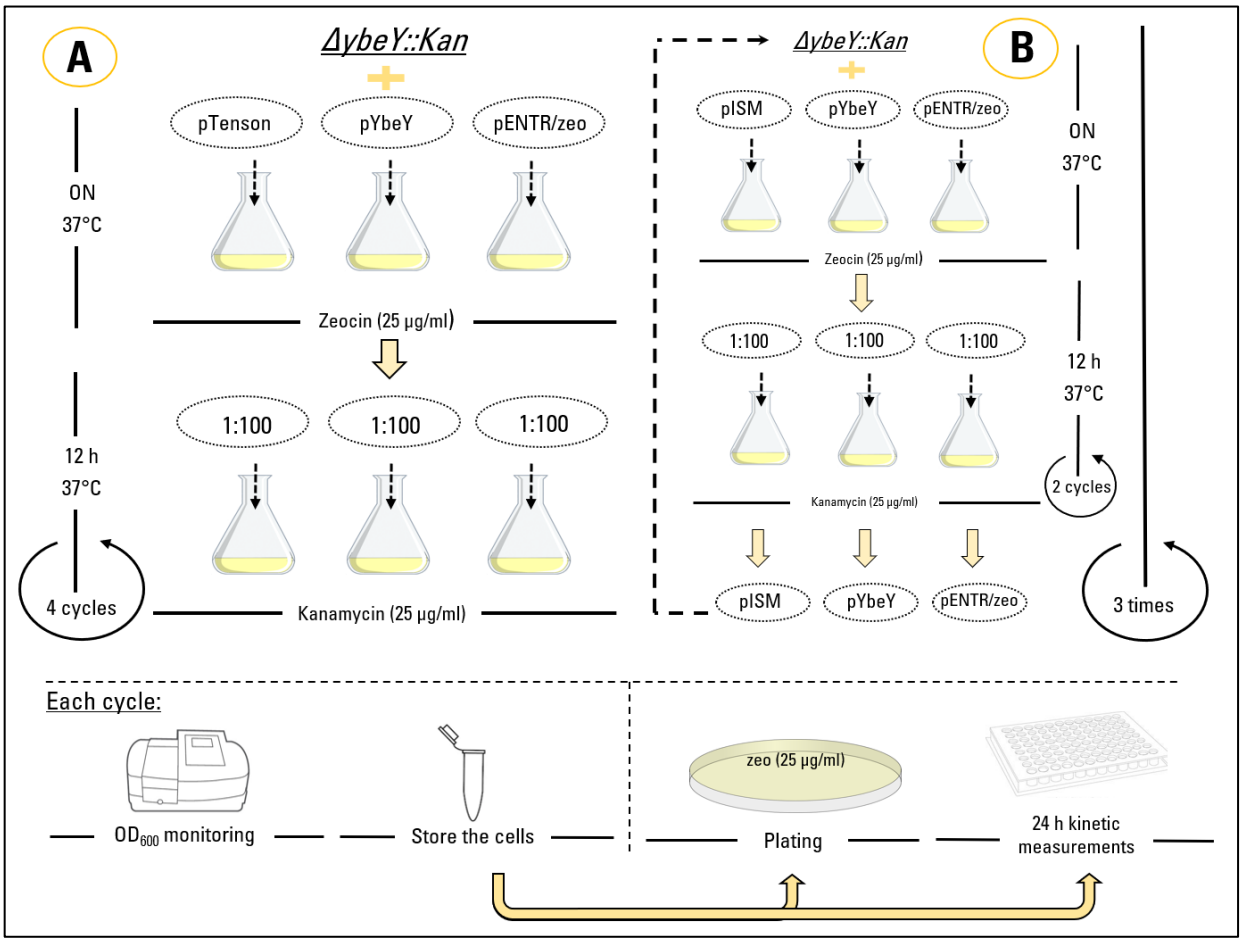


Figure 5 | The scheme to compare two different growth selection strategy for pISM and pTenson. A) Plasmids pTenson, pYbeY and pENTR/zeo were electroporated into Δ*ybeY::Kan*, all growth cycles were made in a row without additional electroporation. B) Plasmids pISM, pYbeY and pENTR/zeo were electroporated into Δ*ybeY::Kan*. After overnight growth and two 12 hours-growth cycles, plasmids were purified and re-electroporated into Δ*ybeY::Kan.* This was repeated three times.

2.2.8 Kinetic measurements

The 24 hours growth measurements were performed using BioTek Synergy Mx Microplate Reader with 15 min OD_{600} reading intervals. Cultures were incubated at 37°C with vigorous shaking for 24 hours. The microplate reader software Gen5 was used for data analysis. The cells from some selection cycles were used to create growth curves to show the main trends. The cultures were diluted to make OD_{600}=0,05, and for replica it was OD_{600}=0,1 as starting points. The dilutions were made into fresh LB supplemented with kanamycin on 96-well plates.

2.2.9 Purification of plasmid DNA

Plasmid DNA was isolated using FavorPrep™ Plasmid Extraction Mini Kit. 3 ml of cultures from each cycle (both pTenson and pISM) were used. The procedure was made according to

the kit protocol, which includes the cells lysing with kit`s buffers, binding plasmid DNA to the silica matrix of the FAPD column, washing it twice with special wash buffers and eluting plasmids in 50 µl of nuclease free MQ water. The plasmids concentrations were detected using NanoDrop™ spectrophotometer.

2.2.10 Restriction cleavage

Since experiments are associated with plasmids, which can be supercoiled, it is necessary to make them linear to ensure that it is the same plasmids. The restriction enzyme EcoO109I, whose cutting place is located near the *att*L1 site, was used to cut plasmids. Restriction reaction contained 16 µl of MQ water, 2 µl of 10X Buffer Tango, 1 µl of plasmids DNA (0,5-1 µg/µl) and 1 µl of EcoO109I enzyme; final volume was 20 µl. The mix was incubated at 37°C for 3-4 hours. The linear plasmids were run on 1,2% agarose gel containing ethidium bromide under electric voltage of 95 V for 45 min. GeneRuler 1 kb Plus DNA Ladder was used to control the length.

2.2.10 Colony screening for pTenson

This method of colony screening was applied for the fully pooled library (pTenson). Cells from the second cycle were plated on LB agar plates supplemented with 25 µg/ml of zeocin. Some colonies were picked and diluted into 150 µl of fresh LB with zeocin on 96-well plate. 50 µl from these cultures were again diluted into 100 µl of LB with 25 µg/ml of zeocin on 96-well plate. 17-hour growth measurements were done with the settings described above.

2.2.10.1 Colony PCR (Polymerase Chain Reaction)

According to the growth curves, some cells/LB mixture from 96-well plate were re-plated on LB agar plates supplemented with 25 µg/ml of zeocin for further colony PCR. Some colonies from this plate were randomly picked for screening. The sequence of the inserted into plasmid gene was a target to amplify it during PCR reaction. The used primers were designed by Ismail Sarigül:

GWlib_seq_attL1_for	5`GGCTTGGCCCTGAGGGCC
GWlib_seq_attL2_rev	5`GTGGCGGCCGCATAGGCC

PCR reaction was performed using 5x HOT FIREPol® Blend Master Mix containing all reagents required for PCR, except water, primers, and template (DNA). One colony was picked and dissolved in 50 µl of MQ water. PCR mix included:

Component	Amount
5x HOT FIREPol® Blend Master Mix	5 µl
Forward primer (10 µM)	0,6 µl
Reverse primer (10 µM)	0,6 µl
Water-cells suspension	5 µl
MQ water	Up to 20 µl

The PCR machine, Eppendorf mastercycler gradient thermal cycler™, was used to create the right conditions for the reaction:

	PCR settings		
Initial denaturation	95°C	15 min	
Denaturation	95°C	15 sec	30 cycles
Primers annealing	59°C	15 sec	
Extension	72°C	2 min	
Final extension	72°C	10 min	

The final step was directly loading 5 µl of samples on 1% agarose gel containing ethidium bromide. GeneRuler 1 kb Plus DNA Ladder was used as a reference.

2.2.10.2 Purification of PCR products from the gel and sequencing

Since it was not possible to accurately take a colony from plate containing only plasmids with gene that is different from *ybeY,* the unknown gene sequence had to be cut and purified from the agarose gel for subsequent sequencing. Zymoclean™ Gel DNA Recovery kit was used for this purpose. The kit protocol included: a melting a piece of agarose gel at 55°C, binding DNA to the silica of column, washing twice with a kit`s wash buffer and eluting DNA in 6 µl of nuclease-free water.

The purified genes have been sent to the Sanger sequencing in the Institute of Genomics Core Facility of University of Tartu.

2.2.11 Total plasmids PCR of pISM

Since pISM library has a differences from pTenson library in the composition and in the selection strategy, it was decided to use another method to detect the genes of interest. Purified plasmids were diluted to 4 ng/µl. The same primers and PCR mix were used as in a colony PCR. There was only one difference in reaction:

Component	Amount
5x HOT FIREPol® Blend Master Mix	5 µl
Forward primer (10 µM)	0,6 µl
Reverse primer (10 µM)	0,6 µl
Plasmids (4 ng/µl)	1 µl
MQ water	Up to 20 µl

The same Eppendorf mastercycler gradient thermal cycler™, was used with the same settings described earlier.

The samples were loading on 1,2% agarose gel containing ethidium bromide. GeneRuler 1 kb Plus DNA Ladder was used as a marker.

2.3 Results

2.3.1 Comparison of transformation methods

The first step in the selection experiments is transformation plasmids into Δ*ybeY::Kan* strain. One pooled library has approximately 4000 different genes, which are inserted into plasmids. The high transformation efficiency is extremely important to ensure that all presented plasmids with different genes will be successfully transformed into Δ*ybeY::Kan* strain, and will participate in further selection. Furthermore, the specificity of *ybeY* null mutants must be taken into account: phenotypic defects have an indirect effect on the success of the transformation. **Figure 6** shows that electroporation has higher transformation efficiency that heat shock transformation with competent cells prepared by the Inoue method.

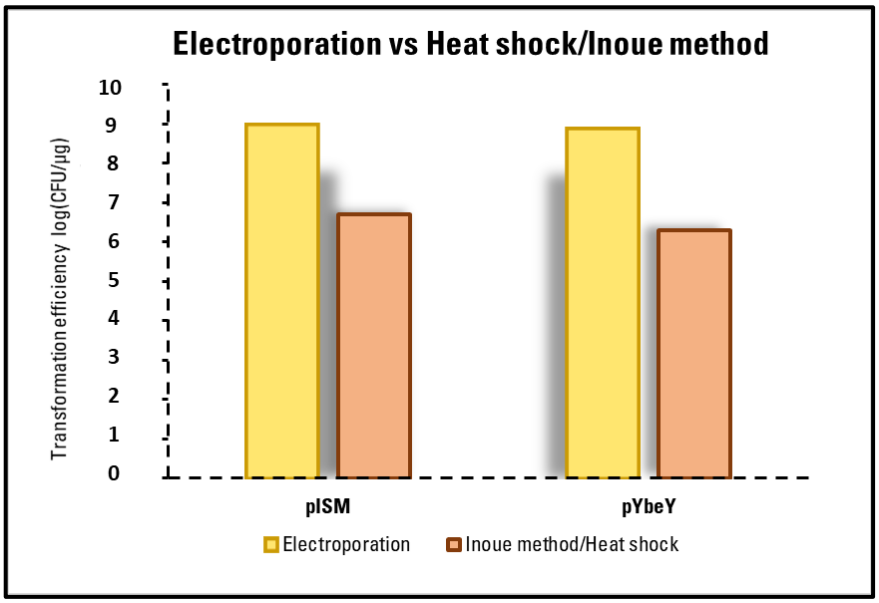


Figure 6 | Logarithmic graph shows the results of testing transformation methods. pISM and pYbeY were used as test plasmids. They were transformed into Δ*ybeY::Kan* strain by using two different methods - electroporation and heat shock. Cells were prepared in different ways – Inoue method for heat shock and washing with glycerol for electroporation. According to the results, electroporation gives a hundred times higher efficiency than the heat shock transformation. Electroporation was repeated in further experiments many times, heat shock was performed only in this testing.

2.3.2 The 24 hours kinetic measurements for pTenson selection

The selection experiment for fully pooled library can be seen as a positive control, but also as an opportunity to search for compensating genes. Every second growth cycle was used for the measurements.

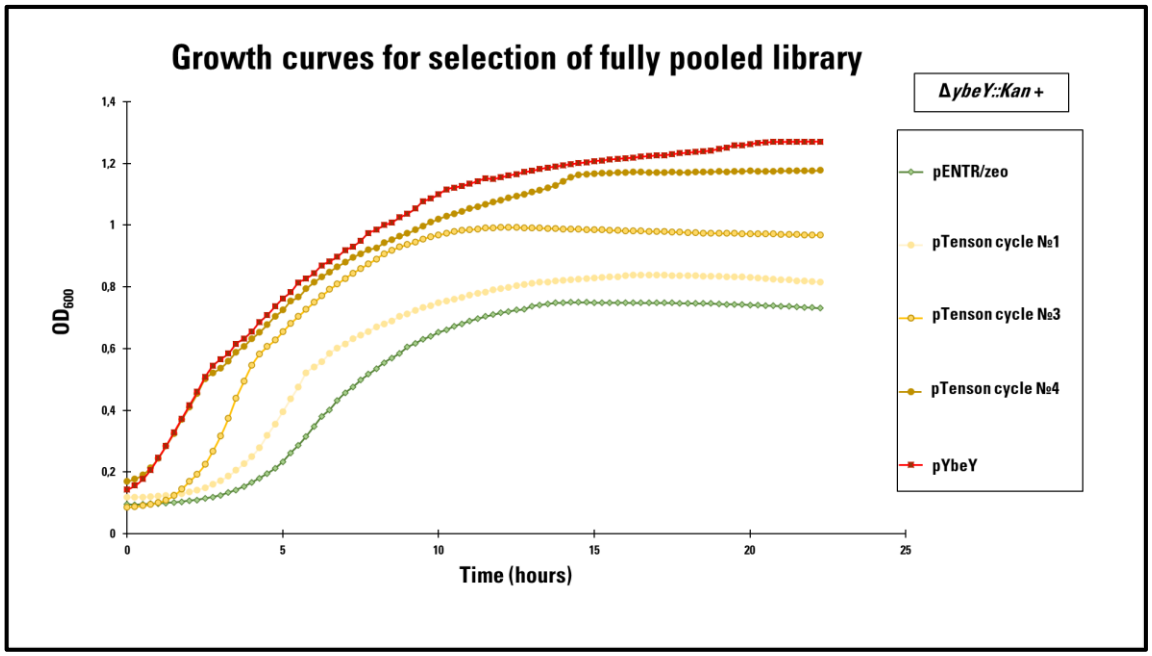


Figure 7 | The growth curves for selection experiment of fully pooled library (pTenson). Each culture had 3 technical replicates. The graph shows the growth trends for Δ*ybeY::Kan* + [pENTR/zeo, pYbeY], whose optical density fluctuated around the same level during all selection cycles. This graph also shows the selection process in fully pooled library (for cycles №1,3,4), where we can see how growth curve of pTenson with each new cycle is approaching closer to the growth curve of Δ*ybeY::Kan* pYbeY, which means successful selection for the missing *ybeY* gene.

According to the growth curves, it can be claimed that the optical density of fully pooled library is increasing during the selection experiment, approaching to the OD_{600} values of Δ*ybeY::Kan* pYbeY.

2.3.3 The restriction cleavage of pTenson

Although the pTenson library contains a plasmid with *ybeY* gene, and the selection for this gene becomes more apparent with each new growth cycle, it is important to do restriction analysis of total purified plasmids to assay the selection speed and to ascertain whether only plasmids containing *ybeY* are present in the last cycles. **Figure 8** illustrates the total plasmids of pTenson on different experiment stages in original and linear form. Only plasmids containing *ybeY* gene are present in the 4^{th} and 5^{th} cycles, what characterizes complete selection relative to the missing gene. This shows that this selection protocol is working properly. The speed of complete selection is high for fully pooled library. The interest in the search for compensatory genes may cause 2^{nd} cycle, where the amount of pYbeY is already quite high, though a little amount of other plasmids still represented in culture. The reason, why plasmids with some genes manage

to grow longer than others until complete pYbeY selection, can shed light on some YbeY protein interactions.

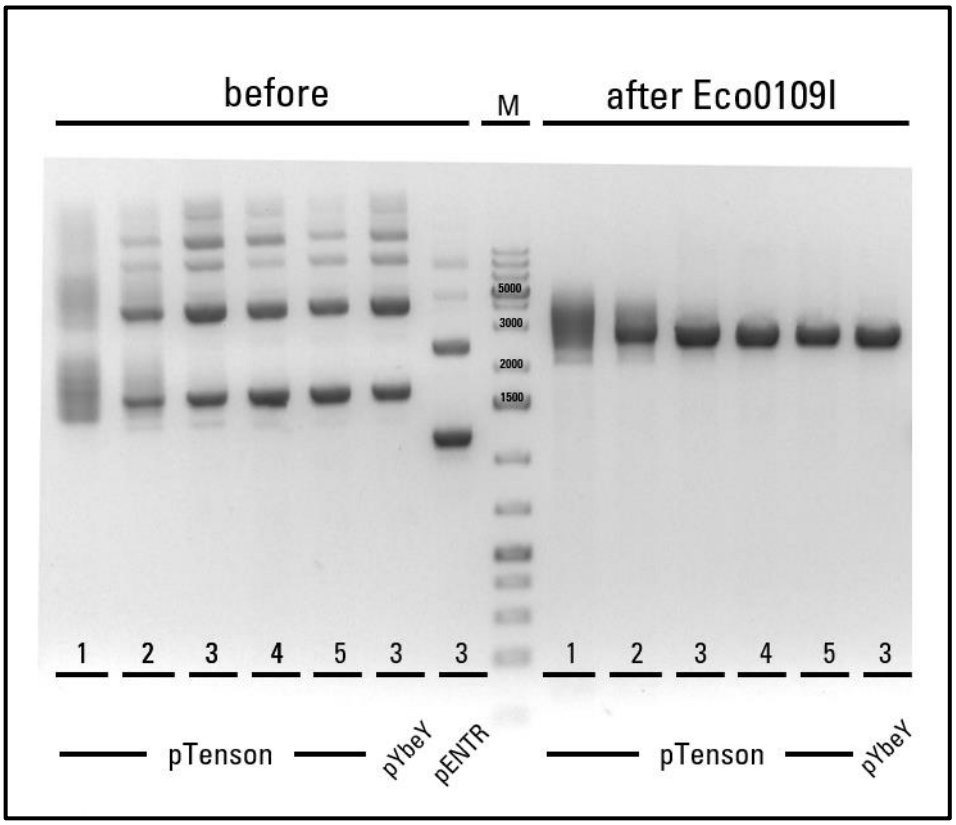


Figure 8 | The restriction cleavage of total plasmids of pTenson during the growth cycles. On the left side, plasmids without cutting, on the right side, plasmids linearized by Eco0109I. The numbers denote the number of the cycle. pYbeY from 3rd cycle was used as a positive control of reaction (since it has a target sequence of this restrictase). On the left side we can see that total plasmids can be present in different configurations, depending on the over- or under-winding of DNA. Cleavage can show the number of different plasmids in the culture. Starting from 2nd cycle the content of pYbeY in the culture is higher than other plasmids. Two last cycles have almost solely plasmids containing *ybeY* gene.

2.3.4 Colony screening of 2nd cycle of pTenson library

The amount of cells containing pYbeY is already quite high in the 2nd cycle of selection. Therefore, searching genes using total plasmids PCR would not give the results. Colony screening is more suitable in this case. When the pTenson culture was plated, the majority of colonies contained pYbeY, which shows the high speed of selection. The main idea was to find other genes, which are present along with pYbeY in the culture. The first step was monitoring the growth of randomly picked colonies. If growth curve of a single colony showed partial compensation (because of the high probability that a full rescue of growth would be caused by

pYbeY), it was used for re-plating and PCR. Two genes were found using this strategy of colony screening (**Figures 9,10**).

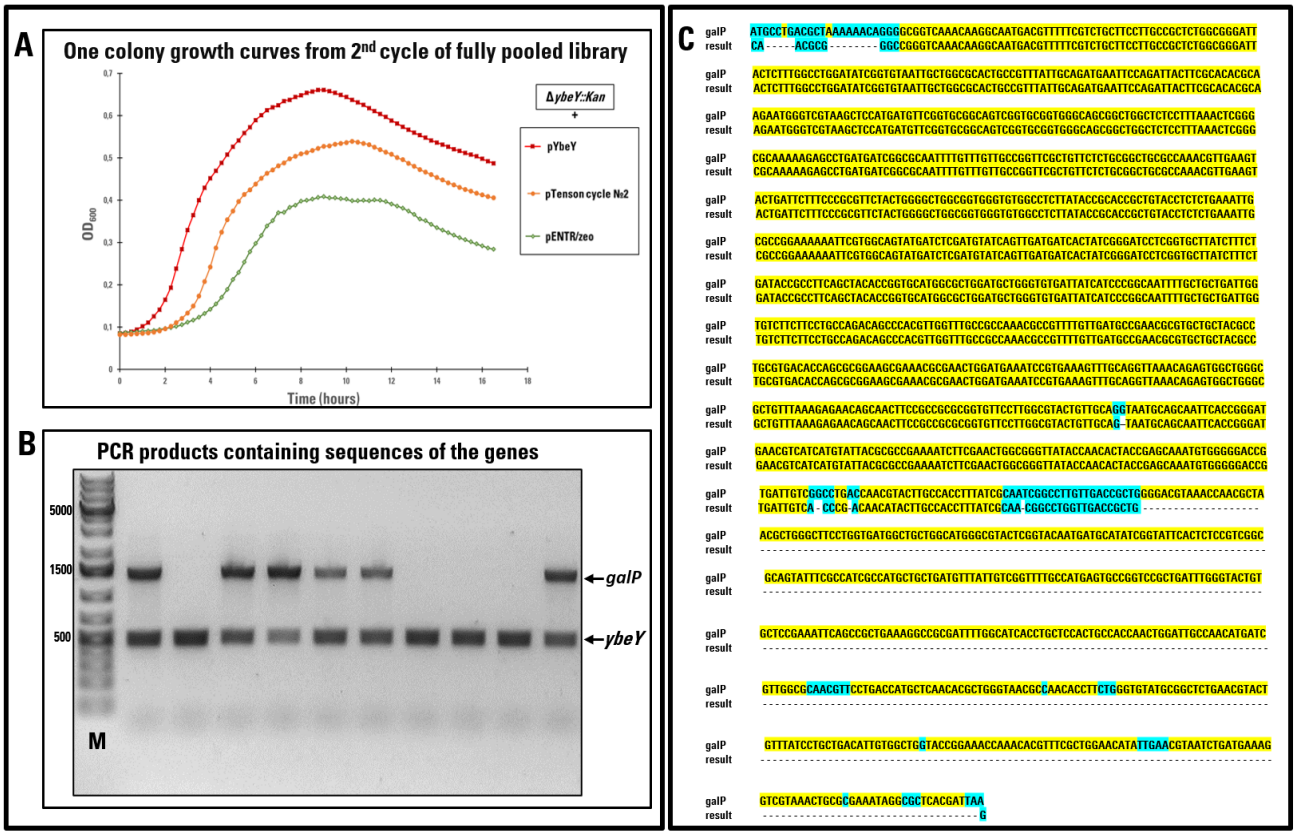


Figure 9 | A) The growth curves of single colonies of Δ*ybeY::Kan* + [pYbeY, pTenson 2nd cycle and pENTR/zeo]. Control colony containing pENTR/zeo has the lowest level of growth, and the colony with pYbeY has- the highest level. The library colony shows partial compensation. **B)** This colony was re-plated and colony PCR was performed. The target sequences for primers are located on both sides of the inserted gene. **C)** Sanger sequencing identified galactose-proton symporter *galP*. Yellow color shows good index of alignment, blue color- average. Percent of identity is 94%. The differences in the nucleotide sequences can be caused by sequencing defects. The alignment were made using T-Coffee server, version_11.00.d625267.

The presence of *ybeY* gene in all colonies gives cause for discussion. There can be several suppositions for explaining. For example, it can not be guaranteed that only cells with the same plasmids are present in one colony when the amount of cells with pYbeY in culture is very high.

Two found genes, *galP* and *rpIL*, are present in several colonies. Galactose-proton symporter is a transmembrane protein, which is involved in the transport of hexose, using proton gradient. This kind of transfer is quick, although GalP can transport sugars in a leaky fashion without proton gradient (Henderson *et al.*, 1977). The *rpIL* gene is coding 50S ribosomal protein

L7/L12. It is a binding site for several GTPase factors involves in translation. It plays an important role for accurate translation (Diaconu *et al*., 2005).

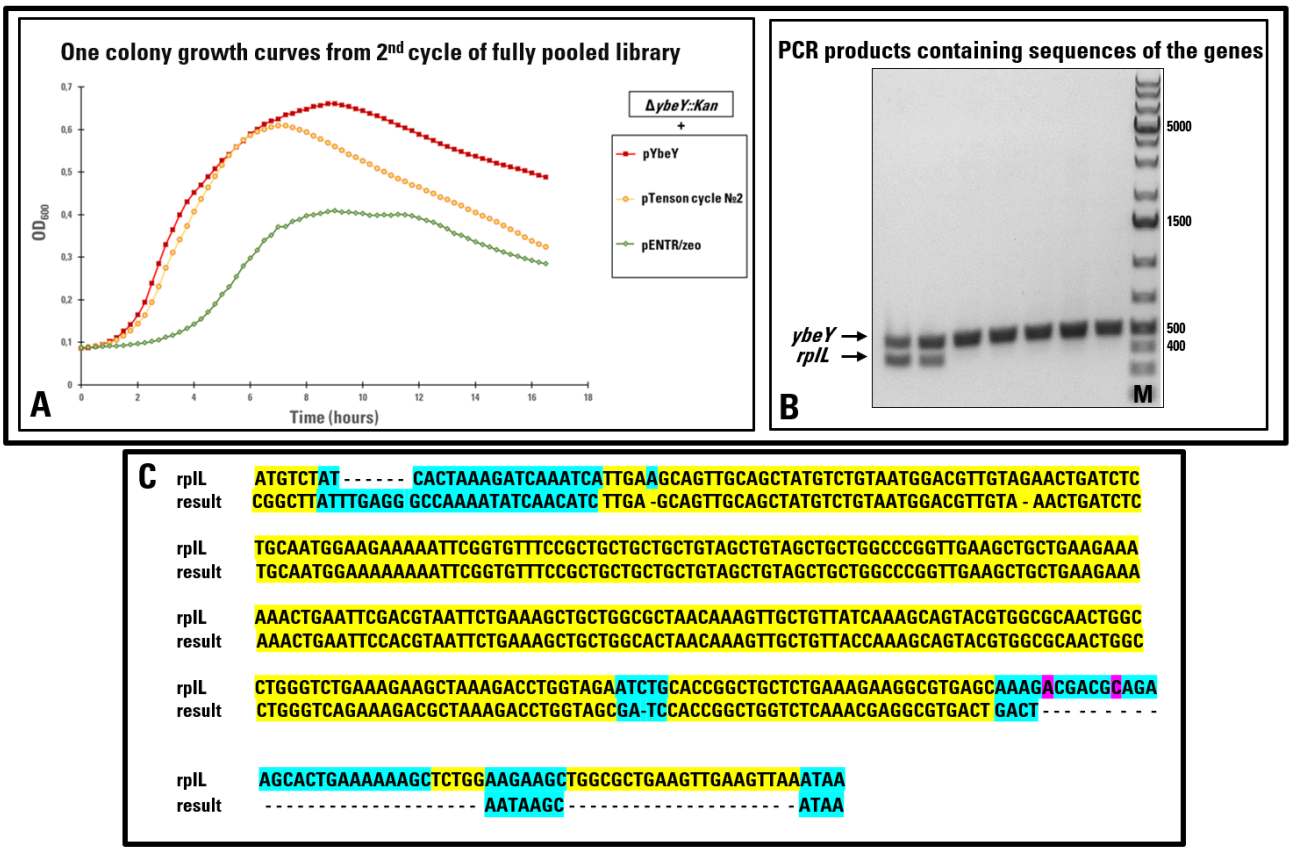


Figure 10 | A) The growth curves of single colonies of Δ*ybeY::Kan* + [pYbeY, pTenson 2nd cycle and pENTR/zeo]. **B)** PCR products on 1% agarose gel, where is another gene besides *ybeY* gene - *rpIL*. **C)** According to the Sanger sequencing results, the identity percent between original *rpIL* gene and sequences from material cut from agarose gel is 78%. Yellow color shows the good index of alignment, blue color- average, and purple – shows below average index. Alignments were made using T-Coffee server, version_11.00.d625267.

2.3.5 The 24 hours kinetic measurements for pISM selection

The ISM library is a partially pooled library. Some ribosome factors (rbfA, rimM, yjeQ and era), ybeX and ybeY were excluded from this pooled library, which means it could need a longer selection. Since continuous cell growth can lead to the unwanted chromosome mutations, it is important to purify plasmids from culture and re-electoporate it again to the Δ*ybeY::Kan* strain at regular intervals. The growth trend of ISM library culture is shown in the **Figure 11**. During the selection, the growth of culture containing pISM is approaching to the growth rates of culture containing pYbeY, although eventually ISM library does not reach the same level in nine cycles. Nonetheless, it is enough to see ongoing selection as a result of which

there is an improvement in library culture growth. It can be clearly seen in comparison with the culture containing the empty plasmid, which had the lowest level of growth.

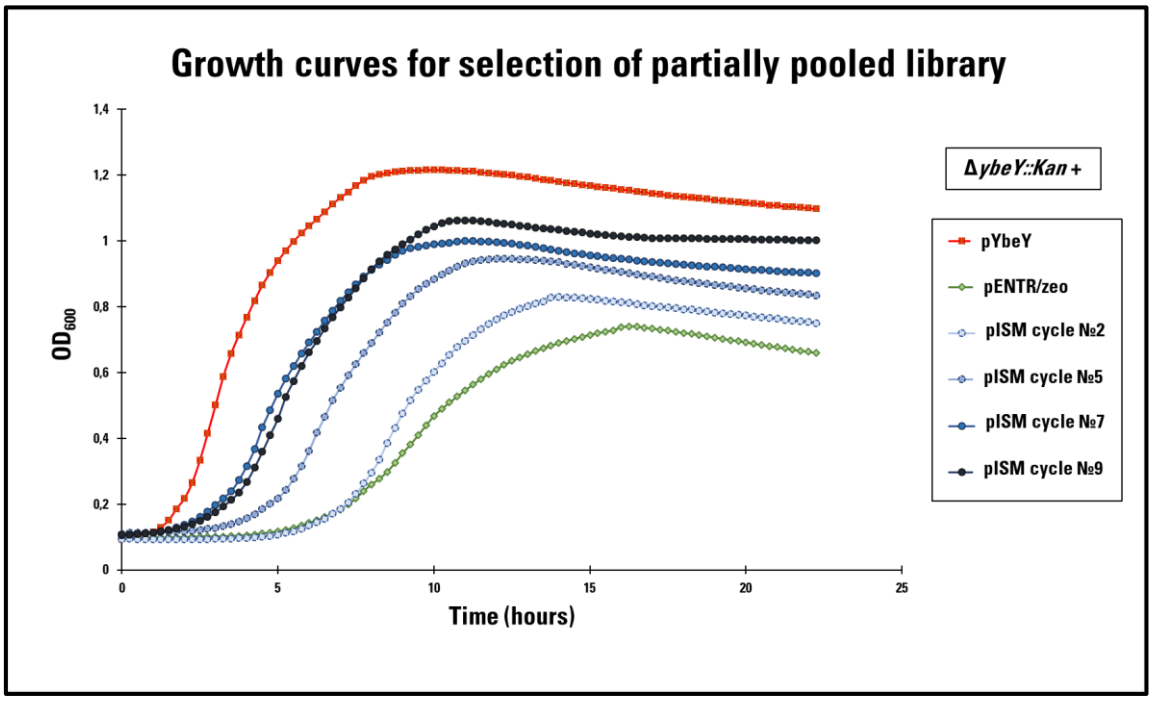


Figure 11 | The growth curves for selection of partially pooled library (pISM). Each culture had 3 technical replicates. The cultures of Δ*ybeY::Kan* strains containing pYbeY and pENTR/zeo were used as a controls during the selection experiment and for comparison in the 24 hours kinetic measurements. This graph shows the large difference between optical densities of library culture in the 2nd and 9th cycles. The intermediate growth of libraries 5-7 shows the gradualness of this process, which means that selection of plasmids with inserts is able to partially compensate for the lack of *ybeY* gene.

2.3.6 The restriction cleavage of pISM

The speed of ISM library selection is significantly slower than it is for the Tenson library. Restriction cleavage can give a general view of selective process in pISM culture. According to the gel picture of uncleaved samples, there are some plasmids whose amounts increase during the experiments. Although restriction cleavage does not give much information about how much plasmids are present in the 9th cycle, it shows that the number of plasmids with different inserts decreases in the culture with each next cycle. The uneven ratio of plasmids contents is clearly visible starting from 5th-6th cycles (**Figure 12**). In both panels (uncleaved and cleaved) the selection process is clearly visible when comparing the first and last cycles, which means that some plasmids borne genes can compensate for the lack of *ybeY*.

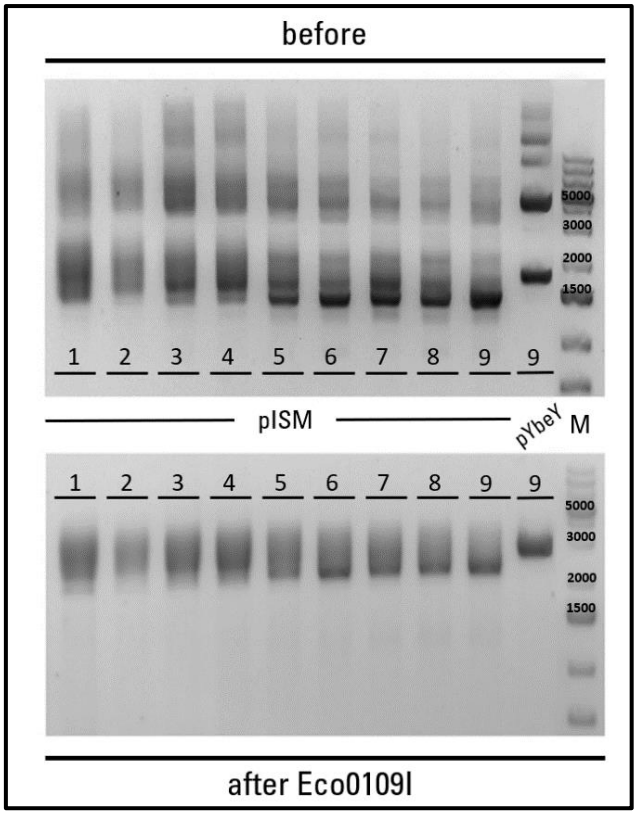


Figure 12 | Purified total plasmids of ISM partially pooled library before and after incubation with restriction enzyme Eco0109I. The numbers show the serial number of cycles. pYbeY from 9th cycle was used as a positive control. Both gel pictures show ongoing selection.

2.3.7 Total plasmids PCR of ISM pooled library

The PCR of total purified plasmids from a culture is a quick and sensitive method to detect specific genes. During the selection the number of certain plasmids increases, which means that after PCR the genes inserted into these plasmids should be brighter on the gel. In the **figure 13**, it can be observed how some genes prevail over others. The separate lines on the gel are visible, starting from the 8th cycle. The five brightest PCR products from 9th cycle were purified from the gel and sequenced (**Figure 14**).

The main problem of the total plasmids PCR in this experiment is gene purification from the gel. Although the five brightest genes are the most visible to our eyes, there can be many other genes, which could interfere during sequencing. Accordingly, the sequencing results were quite difficult to analyze. Instead of nucleotide alignment, it was possible to detect genes only with predictable amino acids alignment, which is more flexible in the search for identities.

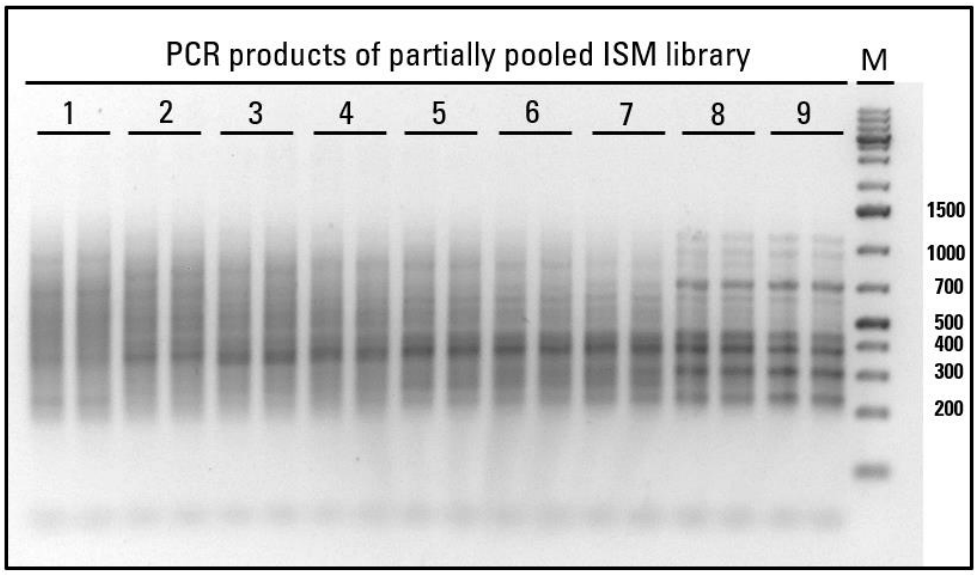


Figure 13 | PCR products of total purified plasmids from pISM culture on different selection stages. The numbers show the serial numbers of the cycles. The certain genes are becoming brighter on the gel with each subsequent cycle. There are at least five genes in the 9th cycle, which prevail over other genes.

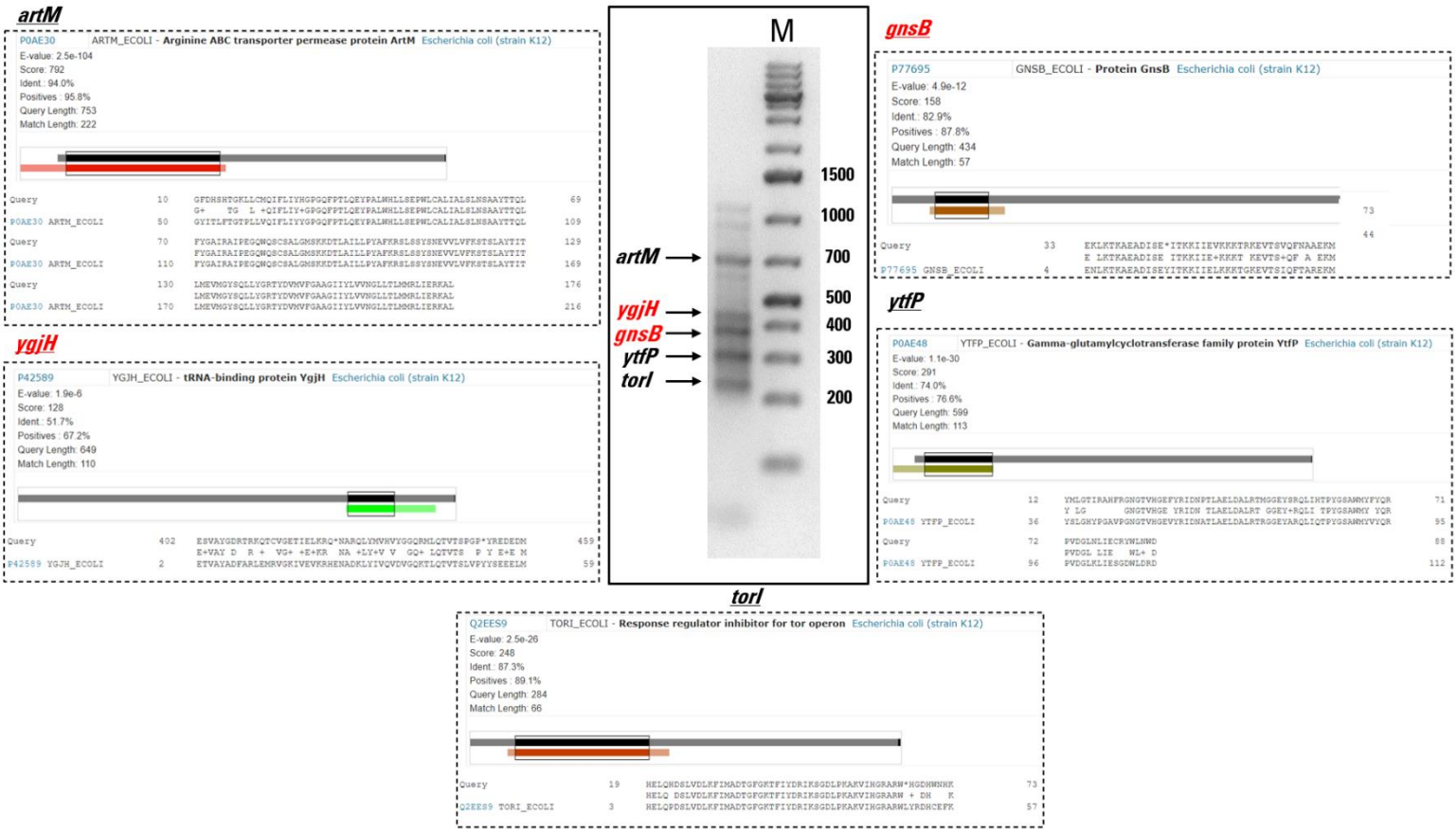


Figure 14 | The results of gene identification based on sequencing of five purified PCR products of 9th cycle ISM library. The genes marked in red have inaccurate identification. The alignment results are shown in dotted frames. The Basic Local Alignment Search Tool of UniProt was used as a program.

According to the alignments, five genes were detected: *artM, ygjH, gnsB, ytfP* and *torI*. Two of them, *ygjH* and *gnsB,* have low percentage of identity or inconsistencies in length.

The *artM* gene encodes arginine ABC transporter permease protein ArtM. It participates in binding-protein-dependent transport system of L-arginine, being connected to the neighboring operon- *artPIQJ*- responsible for arginine transport in *Escherichia coli*. The *artM* product is similar to the structure of the transmembrane proteins (Wissenbach et *al*., 1995).

The *ygjH* gene encodes putative uncharacterized tRNA-binding protein YgjH.

The *gnsB* gene encodes GnsB protein with not well understood functions. Overexpression of *gnsB* increases the amount of unsaturated fatty acids of phospholipids. Moreover, GnsB suppresses the temperature-sensitive *fabA6* mutation (inability in synthesis of unsaturated fatty acid) and cold-sensitive *secG* (the function of protein translocation across the cytoplasmic membrane) null mutation (Sugai *et al*., 2001).

The *ytfP* gene encodes gamma-glutamylcyclotransferase family protein YtfP. It is a member of the UPF0131 protein family of unknown function (Aramini *et al.*, 2005).

The *torI* gene encodes a negative regulator of the *torCAD* operon, which plays important role in the two-component regulatory system of the anaerobic utilization of trimethylamine-N-oxide (Ansaldi *et al*., 2004).

2.4 Discussion

ORF library is a multifaceted tool, which can be used in different ways depending on the purpose of the study. The methods described above can be very useful for studies of functional proteins interactions. Moreover, they can shed light not only on direct pathways of molecular interactions between proteins, but also on indirect ones, thereby replenishing knowledge about the complex, but very flexible physiology of a bacterial cell.

The first step was to decide on an effective transformation method. The absence of a conserved gene like *ybeY* affects the phenotype of the bacteria, complicating the process of transformation. In selection, it is very important that all plasmids presented in pooled libraries are successfully transformed into *ybeY* null mutants. Electroporation has shown higher transformational efficiency than heat shock in a test experiment (Figure 6). Even in the case of electroporation it can not be claimed that all plasmids with different genes are present in bacteria, or that they are present in equal amounts, therefore it was important to make a sufficient number of growth cycles at the selection stage. Thus, this error is compensated by the long growth of culture: if specific plasmids really compensate for the lack of *ybeY*, then even in the case of uneven transformation, their number will begin to prevail over a longer growth time.

The next step was the design of selection experiments, taking into account the features of two pooled libraries. The fully pooled library has a high selection speed. It did not need a big number of cycles, unlike the partially pooled library, where the selection process required many cycles. Based on the functions of YbeY, it can be assumed that changing the growth conditions of the culture can increase the selection speed, which opens up new horizons for further experiments. For example, growth defects at higher temperatures are more critical in *ybeY* deletion mutants, therefore an increase in temperature during selection can accelerate it (Rasouly *et al.*, 2009). According to the growth curves both selection strategies were working (Figures 8,12). Restriction cleavages confirm that specific plasmids cause this improvement in growth, and that their number is increasing relative to other plasmids (Figures 9,13).

The main question here is whether the selected genes really have functional connection with *ybeY* . The selection of a gene in these experiments does not necessarily mean that its encoded protein has a direct interaction with YbeY. The compensatory mechanism can trigger pathways in which proteins are involved that are not related to the functions of YbeY, but in its absence can help cells maintain an acceptable level of functioning. This raises another question: is compensation determined by the activation of alternative ways to fulfill the missing function,

or rather by eliminating negative consequences of YbeY`s absence? Of course, compensation can also be a combination of these two strategies.

A theoretical basis for action can be hypothesized for some of the compensating genes. For example, *rpIL* gene plays important role in an accuracy of translation, and *ybeY* null strains have inaccurate translation (Rasouly *et al*., 2009). Thus the overexpression of *rpIL* can help to improve it. On the other hand, this gene was found in experiment with fully pooled library in the context of pYbeY co-transformation, where it is possible that some non-pertinent plasmids manage to survive longer before complete pYbeY selection occurs. The presence of pYbeY in all selected colonies can be caused by its high content in the culture. Thus, it is more reliable to use partially pooled library to search for compensatory genes, and fully pooled library – as a positive control.

The data obtained in selection experiments should always be verified, ideally by independent methods. There are several options. First of all, genes can be checked using the same ORF library. For example, one can separately electroporate library plasmids with genes of interest into the original null strain and do growth assays. If a culture shows improvement in growth, then Northern blot can be done to check for aberrant forms of 16S rRNA. Ribosome profiling by sucrose gradient ultracentrifugation is the best option to check if compensation occurs at the ribosome assembly level. Based on this, it will be possible to design experiments for each gene individually, considering its functions and level of influence on bacteria processes. Another excellent verification option is repeating growth selection using another ORF library, which contains the same genes. For instance, TransBac clone library can be used for verification for the results of the Gateway® entry clone library . This ORF library is as yet unpublished. It is based on low-copy vector derived by F plasmid. The main advantage of TransBac library is maintenance of authentic structure of coding region: encoding protein does not have additional spacers with amino acids at the ends.

It is also important to consider that the whole method is based on natural physiological compensation. This may be the reason explaining why the genes whose compensatory possibilities have already described in the literature were not found. For example, the compensatory abilities of Era were established by overexpression of it in the *ybeY* null-mutants, which does not prove the preference of this compensatory pathway in natural selection (Ghosal *et al*., 2018).

Given our sequencing results, the quality of gene detection should also be improved. Colony PCR is a better method for preparing material for further sequencing: samples are cleaner and provide more accurate information. However, this method is more suitable for fully pooled

library. In the case of the partially pooled library, where plasmid diversity is quite high even in the last cycles, colony PCR can take more time and resources. Total plasmids PCR is a quick method, which works properly with partially pooled library, but there are difficulties with the quality of the samples cut from the gel. This problem can probably be solved by additional PCR of these samples, or by increasing in the number of selection cycles, what will reduce the diversity of plasmids in culture. As mentioned earlier, changing conditions can increase the speed of selection, then it may not even have to increase the number of cycles to reduce the variety of plasmids.

It is very important to understand in general the goal of this kind selections. This kind of experiment merely provides preliminary results, which need to be verified by further work. It is not enough to merely repeat the selection several times to become sure of its conclusions. Selection is the key that opens the door to new potential experiments. The results obtained can affect both the direct interactions of the protein of interest and alternative molecular pathways independent of it. When researching multifunctional protein, one can not be sure, how large is the circle of other proteins involved into its direct and indirect interactions. One can sort through all theoretically possible genes one by one, but it would take a lot of time and resources. Furthermore, important genes that theoretically have not yet been linked with the studied gene may be missed. Pooled library growth selection can narrow this circle of genes or shed light on previously unknown alternative pathways. This method has great potential, and flexibility, which means that the details of the experiment can be changed depending on the characteristics of the studied gene.

CONCLUSION

Present thesis gives a theoretical overview on a prokaryotic ribosome, especially focusing on rRNA processing. The main attention is paid on YbeY functions and its compensation possibilities. The literature review part also includes an overview on genomic libraries and gives the description of The Gateway® cloning system.

The goals set earlier were achieved in the practical part. The protocol for selection growth was optimized: the electroporation method was used for further experiments. Two selection options were developed and applied for different types of pooled libraries. The data obtained shows phased compensation during growth selection in both cases. The main differences were in a speed of selection and in the level of compensation: as expected, fully pooled library showed faster and almost complete compensation compared to partially pooled library. Colony screening was performed for fully pooled library, and total plasmids PCR – for partially pooled library. According to the sequencing results, seven genes were detected with different accuracy rate.

The objectives of the work were fulfilled. The main conclusion from this thesis is a statement that pooled ORF library can be used as a method to research the compensatory opportunities of deleted genes.

Geenieksрressiooni raamatukogu kasutamise võimalused *Escherichia coli` ybeY* geeni kompensatsiooni uuringutes

Amata Žukova

Resümee

Ribosoom on oluline raku organell, mille põhifunktsioon on valkude süntees ehk translatsioon. Ribosoomi korrektne küpsemine mängib olulist rolli bakteri raku normaalses funktsioneerimises. See on keeruline protsess, milles osaleb palju erinevaid valke. YbeY on multifunktsionaalne valk, mille peamine funktsioon on 16S rRNA 3′ otsa töötlemine, seega tal on ka suur mõju ribosoomi küpsemisele ja translatsiooni protsessile (Sulthana and Deutscher, 2013).

Geen *ybeY* on väga konserveerunud nii gramm-negatiivsetel, kui gramm-positiivsetel bakteritel. Ta kuulub minimaalse bakteri genoomikomplekti (Gil *et al*., 2004). YbeY valgu funktsioonid on mitmekesised ja kriitiliselt olulised bakteri funktsioneerimisele. *ybeY* geeni deletsioon *Escherichia coli* genoomist viib selleni, et bakteritel ilmnevad fenotüübilised defektid ja ribosoomi funktsiooni vead, sealhulgas on vähenenud ribosoomi aktiivsus ja translatsiooniline täpsus (Davies *et al.*, 2010). Kuigi YbeY valgu molekulaarseid mehhanisme pole veel täielikult välja selgitatud, võib väita, et YbeY valgul on laialdane interaktsioonide võrgustik teiste valkudega. Geenieksрressiooni raamatukogu on hea vahend selliste interaktsioonide uurimiseks.

Töö esimene osa annab teoreetilise ülevaate ribosoomi küpsemisest, YbeY valgu funktsioonidest, selle kompensatsiooni võimalustest ja geneetilise kloonikogude kasutamisest.

Praktiline osa koosneb kolmest osast: protokolli optimiseerimine, kahe erineva kasvul põhineva selektsiooni rakendamine ja geenide tuvastamine. Esimesel etapil sai välja selgitatud, et elektroporatsioon on parim valik edasiste katsete jaoks. Täielikult ja osaliselt ühendatud Gateway® kloonikogud olid elektroporeeritud *ybeY* nullmutantidesse. Edaspidi rakendati kahte selektsiooni strateegiat. Katsete tulemus oli, et kasvukompensatsioon toimub mõlemal juhul, kuigi kompensatsiooni tase on erinev. Mõned enamlevinud geenid selekteeritud bakterikultuurides sekveneeriti, saades nii materjali edasiseks uurimiseks.

REFERENCES

Ades, S.E., Grigorova, I.L., Gross, C.A., 2003. **Regulation of the Alternative Sigma Factor σE during Initiation, Adaptation, and Shutoff of the Extracytoplasmic Heat Shock Response in Escherichia coli.** J. Bacteriol. 185, 2512–2519. h

Aiba, H., 2007. **Mechanism of RNA silencing by Hfq-binding small RNAs.** Curr. Opin. Microbiol., Cell regulation (RNA special issue) 10, 134–139.

Ansaldi, M., Théraulaz, L., Méjean, V., 2004. **TorI, a response regulator inhibitor of phage origin in Escherichia coli.** Proc. Natl. Acad. Sci. U. S. A. 101, 9423–9428.

Aramini, J.M., Swapna, G.V.T., Huang, Y.J., Rajan, P.K., Xiao, R., Shastry, R., Acton, T.B., Cort, J.R., Kennedy, M.A., Montelione, G.T., 2005. **1H, 13C, and 15N Resonance Assignments for Escherichia coli ytfP, a Member of the Broadly Conserved UPF0131 Protein Domain Family.** J. Biomol. NMR 33, 197–197.

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

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A., 2000. **The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution**. Science 289, 905–920.

Baumgardt, K., Gilet, L., Figaro, S., Condon, C., 2018. **The essential nature of YqfG, a YbeY homologue required for 3′ maturation of Bacillus subtilis 16S ribosomal RNA is suppressed by deletion of RNase R**. Nucleic Acids Res. 46, 8605–8615.

Bianchi, A.A., Baneyx, F., 1999. **Hyperosmotic shock induces the σ32 and σE stress regulons of Escherichia coli.** Mol. Microbiol. 34, 1029–1038.

Boettcher, M., Covarrubias, S., Biton, A., Blau, J., Wang, H., Zaitlen, N., McManus, M.T., 2019. **Tracing cellular heterogeneity in pooled genetic screens via multi-level barcoding.** BMC Genomics 20.

Brennan, R.G., Link, T.M., 2007. **Hfq structure, function and ligand binding.** Curr. Opin. Microbiol., Cell regulation (RNA special issue) 10, 125–133.

Brimacombe, R., Stiege, W., 1985. **Structure and function of ribosomal RNA**. Biochem. J. 229, 1–17.

Coussens, M.J., Corman, C., Fischer, A.L., Sago, J., Swarthout, J., 2011. MISSION LentiPlex **Pooled shRNA Library Screening in Mammalian Cells**. J. Vis. Exp. JoVE.

Davies, B.W., Köhrer, C., Jacob, A.I., Simmons, L.A., Zhu, J., Aleman, L.M., RajBhandary, U.L., Walker, G.C., 2010. **Role of Escherichia coli YbeY, a highly conserved protein, in rRNA processing.** Mol. Microbiol. 78, 506–518.

Davis, J.H., Williamson, J.R., 2017. **Structure and dynamics of bacterial ribosome biogenesis.** Philos. Trans. R. Soc. B Biol. Sci. 372.

Deutscher, M.P., 2009. Chapter 9 **Maturation and Degradation of Ribosomal RNA in Bacteria, in: Progress in Molecular Biology and Translational Science.** Elsevier, pp. 369–391.

Diaconu, M., Kothe, U., Schlünzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., Stark, H., Rodnina, M.V., Wahl, M.C., 2005. **Structural Basis for the Function of the Ribosomal L7/12 Stalk in Factor Binding and GTPase Activation**. Cell 121, 991–1004.

Ghora, B.K., Apirion, D., 1978. **Structural analysis and in vitro processing to p5 rRNA of a 9S RNA molecule isolated from an rne mutant of E. coli**. Cell 15, 1055–1066.

Ghosal, A., Babu, V.M.P., Walker, G.C., 2018. **Elevated Levels of Era GTPase Improve Growth, 16S rRNA Processing, and 70S Ribosome Assembly of Escherichia coli Lacking Highly Conserved Multifunctional YbeY Endoribonuclease.** J. Bacteriol. 200.

Gibson, M.M., Bagga, D.A., Miller, C.G., Maguire, M.E., 1991. **Magnesium transport in Salmonella typhimurium: the influence of new mutations conferring Co2+ resistance on the CorA Mg2+ transport system.** Mol. Microbiol. 5, 2753–2762.

Gil, R., Silva, F.J., Peretó, J., Moya, A., 2004. **Determination of the Core of a Minimal Bacterial Gene Set**. Microbiol. Mol. Biol. Rev. 68, 518–537.

Green, M.R., Sambrook, J., Sambrook, J., 2012. **Molecular cloning: a laboratory manual, 4th ed.** ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Grinwald, M., Ron, E.Z., 2013. **The Escherichia coli Translation-Associated Heat Shock Protein YbeY Is Involved in rRNA Transcription Antitermination.** PLoS ONE 8. h

Gupta, S.D., Gan, K., Schmid, M.B., Wu, H.C., 1993. **Characterization of a temperature-sensitive mutant of Salmonella typhimurium defective in apolipoprotein N-acyltransferase**. J. Biol. Chem. 268, 16551–16556.

Gupta, S.D., Wu, H.C., 1991. **Identification and subcellular localization of apolipoprotein N-acyltransferase in Escherichia coli**. FEMS Microbiol. Lett. 62, 37–41.

Henderson, P.J., Giddens, R.A., Jones-Mortimer, M.C., 1977. **Transport of galactose, glucose and their molecular analogues by Escherichia coli K12**. Biochem. J. 162, 309–320.

Hsu, P.-L., Ross, W., Landy, A., 1980. **The λ phage att site: functional limits and interaction with Int protein.** Nature 285, 85–91.

Inoue, H., Nojima, H., Okayama, H., 1990. **High efficiency transformation of Escherichia coli with plasmids**. Gene 96, 23–28.

Jacob, A.I., Köhrer, C., Davies, B.W., RajBhandary, U.L., Walker, G.C., 2013. **Conserved Bacterial RNase YbeY Plays Key Roles in 70S Ribosome Quality Control and 16S rRNA Maturation.** Mol. Cell 49, 427–438.

James D. Watson, Tania A. Baker, Michael Levine and Richard Losick, Stephen P. Bell, Alexander Gann, 2004. **Molecular Biology of the Gene**. Published by Pearson.

Kaczanowska, M., Rydén-Aulin, M., 2007. **Ribosome Biogenesis and the Translation Process in Escherichia coli.** Microbiol. Mol. Biol. Rev. MMBR 71, 477–494.

King, T.C., Schlessinger, D., 1983. **S1 nuclease mapping analysis of ribosomal RNA processing in wild type and processing deficient Escherichia coli**. J. Biol. Chem. 258, 12034–12042.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., Mori, H., 2005. **Complete set of ORF clones of Escherichia coli ASKA library (A Complete S et of E. coli K -12 ORF Archive)**: Unique Resources for Biological Research. DNA Res. 12, 291–299.

Kooiker, M., Xue, G.-P., 2014. **cDNA Library Preparation, in: Henry, R.J., Furtado, A. (Eds.), Cereal Genomics: Methods and Protocols**, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 29–40.

Kweon, J., Kim, D., Jang, A.-H., Kim, Y., 2018. **CRISPR/Cas-based customization of pooled CRISPR libraries.** PLOS ONE 13, e0199473.

Li, G.-W., Oh, E., Weissman, J.S., 2012. **The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria**. Nature 484, 538–541.

Li, Z., Pandit, S., Deutscher, M.P., 1999a. **RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA**. EMBO J. 18, 2878–2885.

Li, Z., Pandit, S., Deutscher, M.P., 1999b. **Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T**. RNA 5, 139–146.

McAteer, S.P., Sy, B.M., Wong, J.L., Tollervey, D., Gally, D.L., Tree, J.J., 2018. **Ribosome maturation by the endoribonuclease YbeY stabilizes a type 3 secretion system transcript required for virulence of enterohemorrhagic Escherichia coli**. J. Biol. Chem. 293, 9006–9016.

Mecsas, J., Rouviere, P.E., n.d. **The activity of an Escherichia coli heat-inducible r-factor, is modulated by expression of outer membrane proteins 12.**

Morgan, E.A., 1986. **Antitermination mechanisms in rRNA operons of Escherichia coli.** J. Bacteriol. 168, 1–5.

Nierhaus, K.H., Wilson, D.N. (Eds.), 2004. **Protein synthesis and ribosome structure: translating the genome.** Wiley-VCH, Weinheim.

Nierman, W.C., Feldblyum, T.V., 2001. **Genomic Library, in: Encyclopedia of Genetics**. Elsevier, pp. 865–872.

Orelle, C., Carlson, E.D., Szal, T., Florin, T., Jewett, M.C., Mankin, A.S., 2015. **Protein synthesis by ribosomes with tethered subunits**. Nature 524, 119–124.

Pandey, S.P., Winkler, J.A., Li, H., Camacho, D.M., Collins, J.J., Walker, G.C., 2014. **Central role for RNase YbeY in Hfq-dependent and Hfq-independent small-RNA regulation in bacteria**. BMC Genomics 15, 121.

Petrov, A.S., Bernier, C.R., Hershkovits, E., Xue, Y., Waterbury, C.C., Hsiao, C., Stepanov, V.G., Gaucher, E.A., Grover, M.A., Harvey, S.C., Hud, N.V., Wartell, R.M., Fox, G.E., Williams, L.D., 2013. **Secondary structure and domain architecture of the 23S and 5S rRNAs.** Nucleic Acids Res. 41, 7522–7535.

Pontes, M.H., Yeom, J., Groisman, E.A., 2016. **Reducing Ribosome Biosynthesis Promotes Translation during Low Mg2+ Stress**. Mol. Cell 64, 480–492.

Quail, M.A., Matthews, L., Sims, S., Lloyd, C., Beasley, H., Baxter, S.W., 2012a. **Genomic Libraries: I. Construction and Screening of Fosmid Genomic Libraries**, in: Orgogozo, V., Rockman, M.V. (Eds.), Molecular Methods for Evolutionary Genetics. Humana Press, Totowa, NJ, pp. 37–58.

Quail, M.A., Matthews, L., Sims, S., Lloyd, C., Beasley, H., Baxter, S.W., 2012b. **Genomic Libraries: II. Subcloning, Sequencing, and Assembling Large-Insert Genomic DNA Clones**, in: Orgogozo, V., Rockman, M.V. (Eds.), Molecular Methods for Evolutionary Genetics. Humana Press, Totowa, NJ, pp. 59–81.

Rajagopala, S.V., Yamamoto, N., Zweifel, A.E., Nakamichi, T., Huang, H.-K., Mendez-Rios, J.D., Franca-Koh, J., Boorgula, M.P., Fujita, K., Suzuki, K., Hu, J.C., Wanner, B.L., Mori, H., Uetz, P., 2010. **The Escherichia coli K-12 ORFeome: a resource for comparative molecular microbiology**. BMC Genomics 11, 470.

Rasouly, A., Davidovich, C., Ron, E.Z., 2010. **The Heat Shock Protein YbeY Is Required for Optimal Activity of the 30S Ribosomal Subunit**. J. Bacteriol. 192, 4592–4596.

Rasouly, A., Schonbrun, M., Shenhar, Y., Ron, E.Z., 2009. **YbeY, a Heat Shock Protein Involved in Translation in Escherichia coli.** J. Bacteriol. 191, 2649–2655.

Reece-Hoyes, J.S., Walhout, A.J.M., 2018. **Gateway Recombinational Cloning**. Cold Spring Harb. Protoc. 2018, pdb.top094912.

Rheinberger, H.J., Sternbach, H., Nierhaus, K.H., 1981. **Three tRNA binding sites on Escherichia coli ribosomes.** Proc. Natl. Acad. Sci. 78, 5310–5314.

Rodnina, M.V., Beringer, M., Wintermeyer, W., 2007. **How ribosomes make peptide bonds**. Trends Biochem. Sci. 32, 20–26.

Shine, J., Dalgarno, L., 1974. **The 3′-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites**. Proc. Natl. Acad. Sci. U. S. A. 71, 1342–1346.

Smit, S., Widmann, J., Knight, R., 2007. **Evolutionary rates vary among rRNA structural elements.** Nucleic Acids Res. 35, 3339–3354.

Storz, G., Vogel, J., Wassarman, K.M., 2011. **Regulation by Small RNAs in Bacteria: Expanding Frontiers**. Mol. Cell 43, 880–891.

Sugai, R., Shimizu, H., Nishiyama, K., Tokuda, H., 2001. **Overexpression of yccL (gnsA) and ydfY (gnsB) increases levels of unsaturated fatty acids and suppresses both the**

temperature-sensitive fabA6 mutation and cold-sensitive secG null mutation of Escherichia coli. J. Bacteriol. 183, 5523–5528.

Sulthana, S., Deutscher, M.P., 2013. **Multiple Exoribonucleases Catalyze Maturation of the 3′ Terminus of 16S Ribosomal RNA (rRNA)**. J. Biol. Chem. 288, 12574–12579.

Taglicht, D., Padan, E., Oppenheim, A.B., Schuldiner, S., 1987. **An alkaline shift induces the heat shock response in Escherichia coli**. J. Bacteriol. 169, 885–887.

Thurlow, B., Davis, J.H., Leong, V., Moraes, T.F., Williamson, J.R., Ortega, J., 2016. **Binding properties of YjeQ (RsgA), RbfA, RimM and Era to assembly intermediates of the 30S subunit**. Nucleic Acids Res. 44, 9918–9932.

Tu, C., Zhou, X., Tropea, J.E., Austin, B.P., Waugh, D.S., Court, D.L., Ji, X., 2009. **Structure of ERA in complex with the 3′ end of 16S rRNA: Implications for ribosome biogenesis**. Proc. Natl. Acad. Sci. U. S. A. 106, 14843–14848.

VanBogelen, R.A., Kelley, P.M., Neidhardt, F.C., 1987. **Differential induction of heat shock, SOS, and oxidation stress regulons and accumulation of nucleotides in Escherichia coli.** J. Bacteriol. 169, 26–32.

Vercruysse, M., Köhrer, C., Davies, B.W., Arnold, M.F.F., Mekalanos, J.J., RajBhandary, U.L., Walker, G.C., 2014. **The Highly Conserved Bacterial RNase YbeY Is Essential in Vibrio cholerae, Playing a Critical Role in Virulence, Stress Regulation, and RNA Processing**. PLOS Pathog. 10, e1004175.

Vercruysse, M., Köhrer, C., Shen, Y., Proulx, S., Ghosal, A., Davies, B.W., RajBhandary, U.L., Walker, G.C., 2016. **Identification of YbeY-Protein Interactions Involved in 16S rRNA Maturation and Stress Regulation in Escherichia coli.** mBio 7.

Warner, J.R., 2013. **YbeY: The Jealous Tailor**. Mol. Cell 49, 422–423.

Waters, L.S., Storz, G., 2009. **Regulatory RNAs in Bacteria**. Cell 136, 615–628.

Williamson, J.R., 2003. **After the ribosome structures: How are the subunits assembled?** RNA 9, 165–167.

Wilson, D.N., Blaha, G., Connell, S.R., Ivanov, P.V., Jenke, H., Stelzl, U., Nierhaus, Y.T. and K.H., 2002. P**rotein Synthesis at Atomic Resolution: Mechanistics of Translation in the Light of Highly Resolved Structures for the Ribosome** Curr. Protein Pept. Sci.

Wissenbach, U., Six, S., Bongaerts, J., Ternes, D., Steinwachs, S., Unden, G., 1995. **A third periplasmic transport system for L-arginine in Escherichia coli: molecular characterization of the artPIQMJ genes, arginine binding and transport.** Mol. Microbiol. 17, 675–686.

Zhan, C., Fedorov, E.V., Shi, W., Ramagopal, U.A., Thirumuruhan, R., Manjasetty, Babu.A., Almo, S.C., Fiser, A., Chance, M.R., Fedorov, A.A., 2005. **The ybeY protein from Escherichia coli is a metalloprotein.** Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 61, 959–963.

Zhang, W., Fu, J., Liu, J., Wang, H., Schiwon, M., Janz, S., Schaffarczyk, L., von der Goltz, L., Ehrke-Schulz, E., Dörner, J., Solanki, M., Boehme, P., Bergmann, T., Lieber, A., Lauber, C., Dahl, A., Petzold, A., Zhang, Y., Stewart, A.F., Ehrhardt, A., 2017. **An Engineered Virus Library as a Resource for the Spectrum-wide Exploration of Virus and Vector Diversity.** Cell Rep. 19, 1698–1709.

INTERNET REFERENCES

- Summary of Escherichia coli K-12 substr. MG1655, version 19.5

http://vm-trypanocyc.toulouse.inra.fr/ECOLI/organism-summary?object=ECOLI&detaillevel=2 (15.01.2020)

- Database of rRNA secondary structures

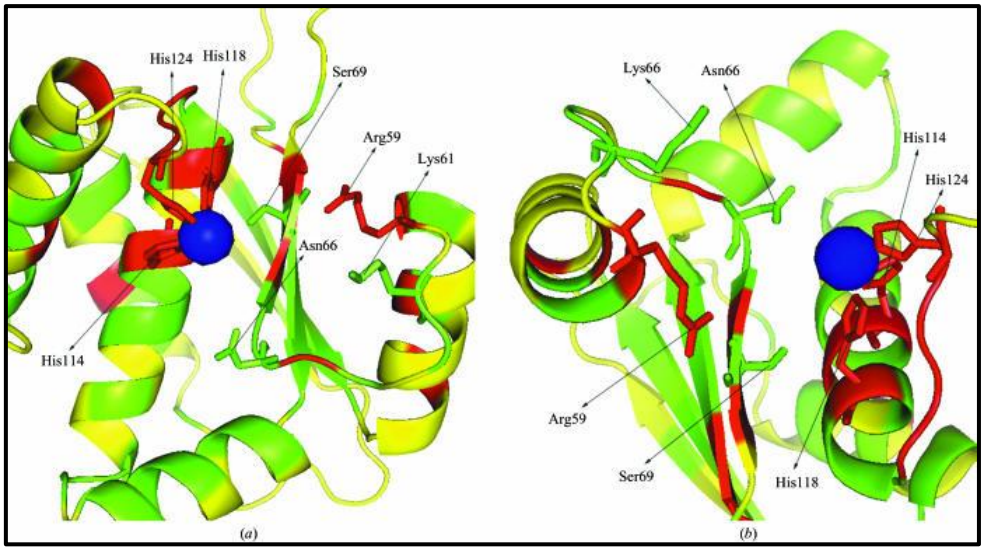
http://apollo.chemistry.gatech.edu/RibosomeGallery/index.html (13.12.2019)

Program resources

- T-Coffee: M-Coffee alignment. Version_11.00.d625267 (20.04.2020)
- UniProt: Blast and Align (20.04.2020)

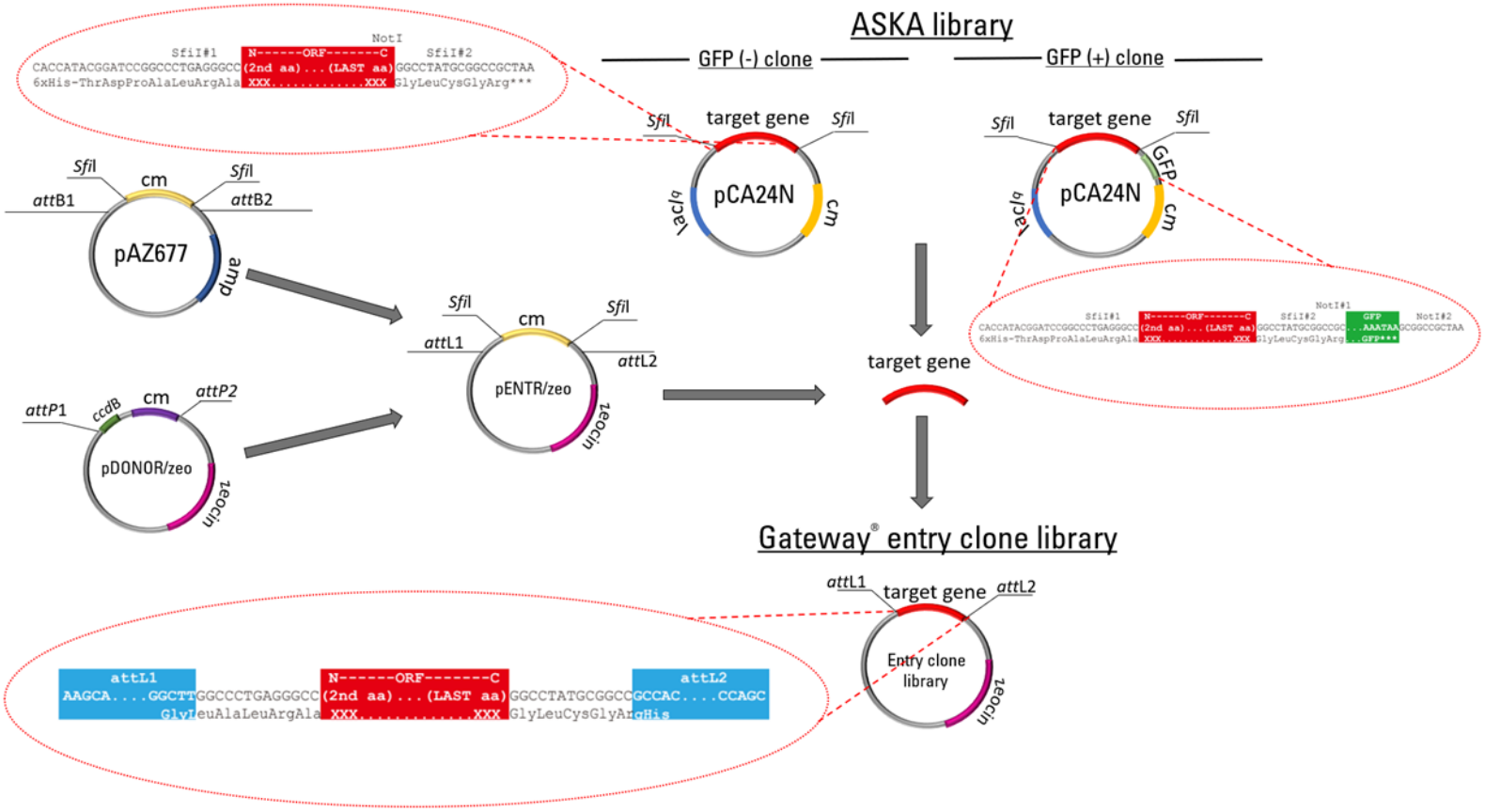
SUPPLEMENTAL MATERIALS

Annex 1



Annex 1 | The crystal structure of YbeY. The protein is shown here in secondary structure. The blue dot is the Ni^{2+} ion surrounded by 3 histidine residues (His114, His124, His118) and arginine (Arg59) (Zhan *et al.*, 2005).

Annex 2



Annex 2 | Schematic representation of creating the Gateway® entry clone library. ASKA library includes two types of clones depending on the availability of GFP. pCA24N is a multi-copy vector with chloramphenicol resistance, can be induced by IPTG, *Sfi*I restriction sites are located on both sides of target ORF. The target gene was digested with *Sfi*I to release it from pCA24N. pENTR/zeo was created using BP recombination (part of The Gateway® cloning system) between pAZ677 (chloramphenicol and ampicillin resistant) and pDONOR/zeo (zeocin and chloramphenicol resistant, contains *ccd*B gene, which is necessary for Gateway® method cloning). As a result, pENTR/zeo has *Sfi*I and *att*L restriction sites, zeocin and chloramphenicol resistance. The target ORF was ligated into *Sfi*I-digested pENTR/Zeo vector, creating entry clone library with zeocin resistance and *att*L restriction sites on both sides of ORF.

NON-EXCLUSIVE LICENSE

Non-exclusive licence to reproduce thesis and make thesis public

I, Amata Žukova,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

The possibilities of using plasmid ORF clone library for *ybeY* gene compensation studies in *Escherichia coli*

supervised by Ismail Sarigül, and co-supervised by Ülo Maiväli

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Amata Žukova

08/06/2020