5-2017

Using Rb-TnSeq to Analyze Natural Variation in Saccharomyces cerevisiae

Syed Raza Mahmood
University of Arkansas, Fayetteville

Using Rb-TnSeq to Analyze Natural Variation in Saccharomyces cerevisiae

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Cell and Molecular Biology

by

Syed Raza Mahmood
Institute of Business Administration Karachi
Bachelor of Business Administration, 2008
University of Karachi
Bachelor of Science in Biology, 2013

May 2017
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Dr. Jeffrey A. Lewis
Thesis Director

_____
Dr. Young Min Kwon
Committee Member

_____
Dr. David S. McNabb
Committee Member

ABSTRACT

One of the main challenges in biology today is the characterization of millions of genes of unknown function being continuously identified in sequencing studies. Transposon mutagenesis is a technique that has been widely used for annotating gene function and has now been combined with next-generation sequencing (Tn-Seq) to assess mutant fitness on a genome wide level. However, Tn-Seq approaches are often constrained by laborious library preparation protocols which limit the number of organisms or conditions that can be assessed. Random bar code transposon-site sequencing (RB-TnSeq), is a transposon sequencing technique that streamlines library preparation and increases the throughput of mutant fitness profiling by incorporating random DNA barcodes into Tn5 Transposons. Rb-TnSeq has been successfully used for high throughput mutant fitness assays in diverse bacterial species. However, this technique is yet to be applied to a eukaryotic model organism. The goal of this study is to develop tools that allow the construction of barcoded mutant libraries in *saccharomyces cerevisiae* and describes methods for producing barcoded mutant libraries using a plasmid based or transposome based approach. These library construction protocols can prove to be powerful tools for studying gene function in *S. cerevisiae* on a genome wide basis.

ACKNOWLEDGMENTS

I would like to thank Dr. Jeffrey Lewis for accepting me in the lab and giving me an opportunity to pursue a career in biology research despite not having any prior experience in the sciences. I also want to thank Dr. Tara Stuecker and graduate students Rebecca Sides and Amanda Scholes for helping me learn the basics of biology research especially during my initial time in the lab and for always making me feel welcome while transitioning to a new field and a new culture. I am grateful to the Fulbright scholarship program for sponsoring my education in the United States and the United States Educational Foundation in Pakistan for supporting me through the admission process despite numerous delays and unexpected problems. Finally, I am thankful to my family for always supporting me in my decision to pursue the career I had always wanted to.

TABLE OF CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

TABLE OF FIGURES

1. INTRODUCTION

The budding yeast *saccharomyces cerevisiae* is one of the most common model organisms in molecular biology (Giaever and Nislow, 2014). Being a unicellular eukaryotic organism that can exist in both haploid and diploid states and carry out mitosis and meiosis, yeast is ideally suited to genetic manipulation in a laboratory setting. Furthermore, its ability to adapt its physiology to extreme environmental changes, survive exposure to toxic substances and ability to consume a variety of carbon and nitrogen sources make yeast an ideal model for investigating eukaryotic biology at the cellular level. In addition to being an important model organism, *S. cerevisiae* is is also extensively used in industry with commercial applications in areas as diverse as brewing and baking to pharmaceuticals. Moreover, recent studies have shown a high degree of functional similarity between yeast and human genes with nearly half of essential yeast genes being replaceable with their human orthologs (Kachroo et al. 2017).

It is therefore not surprising that the first eukaryotic genome to be sequenced belonged to *S. cerevisiae* (Giaever and Nislow, 2014). However, even before the yeast genome was completed, it became obvious that the next challenge would be to assign function to the large number of newly discovered and sequenced genes. As the creation of mutants is a well-established method of investigating gene function, this problem gave rise to the yeast deletion collection project (Giaever and Nislow, 2014). The yeast deletion collection comprises more than 21,000 mutant strains that carry precise start-to-stop deletions of every one of the ~6000 open reading frames present in the yeast genome. The project used a PCR based strategy to delete each open reading frame and replace it with a KANMX cassette tagged to a unique 20mer sequence (Giaever and Nislow, 2014).

Since its creation the yeast deletion collection has been used for a number of genome wide phenotypic studies aimed towards understanding gene function, environmental stress response, and mechanism of drug action in yeast (Giaever and Nislow, 2014). However, despite being an invaluable tool for genome wide screens, a number of factors limit its utility for future studies. Firstly, since the strain used for the yeast genome project was S288c, the yeast deletion collection was constructed in the same genetic background. While S288c is one of the most commonly used yeast strains in a laboratory setting and has long been used as a model for yeast physiology and basic biology, a number of studies have suggested that it is aberrant compared to wild strains of yeast and is phenotypically distinct in terms of stress sensitivity, gene expression, mitochondrial content, and gene-dosage control (Kvitek, Will and Gasch 2017; Young and Court 2008). Making general conclusions about yeast biology based on studies conducted using the lab strain can therefore prove to be problematic and studies pertaining to a specific strain should ideally be conducted in the relevant genetic background. The development of new deletion collections in different genetic backgrounds, however, requires significant investment of time and resources. Since the construction of the original yeast deletion collection took over 3 years with funding of over $2 million (Giaever and Nislow, 2014), the development of new collections using the same techniques is not practical.

This study therefore aims to develop a technique that allows the construction of a deletion mutant library in *S. cereviisae* with minimum use of time and resources. Such a technique can not only allow rapid development of deletion libraries in multiple strains of yeast making possible comparative studies of gene function on a genome wide basis but can also potentially be extended to other organisms in future. One such technique based on a combination of transposon mutagenesis and massively parallel sequencing that has been successfully used to make deletion

mutants for genetic screening in bacterial cells is Random Barcode Transposon Site Sequencing (Rb-TnSeq). This goal of this study is to modify and validate Rb-TnSeq library construction techniques for the creation of deletion libraries in *S. cerevisiae.*

Since Rb-TnSeq is based on transposon mutangenesis the first section of this work will introduce transposable elements, their types, mechanism of action and regulation. The properties, mechanism of action and regulation of the hyperactive Tn5 transposon system used in this study will then be described. An overview of the transposon sequencing techniques that combine transposon mutagenesis with massively parallel sequencing will be provided followed by the experimental strategies used for constructing Rb-TnSeq libraries in yeast, the results obtained and possible future research avenues.

1.1 Transposable Elements

Transposable elements are DNA sequences that can move from one location to another in the genome. First discovered by Barbara McClintock in the 1940's in maize, they have since been found in identified in eukaryotes and prokaryotes and can often form a large proportion of an organism's genome. Transposable elements comprise 3% of the yeast genome, 37% of the mouse genome, 45% of the human genome, over 80% of some plant genomes such as maize and have played an important role in the evolution of these genomes (Castanera et al., 2016; Muñoz-López&García-Pérez, 2010). The movement of transposable elements can impact genomes through various mechanisms. Insertion of a transposable element in an exonic region, for instance, may inactivate a gene while insertion in a regulatory region may alter its expression. The transposition process can produce genetic alterations ranging from insertions, excisions, duplications or translocations (Castanera et al., 2016; Muñoz-López&García-Pérez, 2010). Transposable elements have been shown to generate new proteins by exon shuffling and are an important source of regulatory sequences that can modify existing networks. Similarly, due to their repetitive nature transposable elements can promote homologous recombination while their ability to duplicate has been a key driver of variation in genome size. Thus while initially considered 'junk DNA' the role of transposable elements in driving genome evolution is now widely recognized (Castanera et al., 2016).

1.2 Types of Transposable Elements

Transposable elements can be divided into two main classes based on their mechanism of transposition. Class I transposable elements or retrotransposons move through a replicative 'copy and paste' mechanism involving reverse transcription of the transposon RNA and integration of

the resulting cDNA into another locus (Muñoz-López&García-Pérez, 2010). Class II transposable elements or DNA transposons on the other hand move through a 'cut and paste' mechanism where the transposon is excised from one location in the genome and inserted at another (Muñoz-López&García-Pérez, 2010).

Retrotransposons are further subdivided into LTR and Non LTR retrotransposons. LTR retrotransposons are one of the main constituents of the eukaryotic genome and have a structure similar to retroviruses (Havecker, Gao, &Voytas, 2004). They are named after the long terminal repeats flanking the internal coding region of the retrotransposon which contains the *pol* and *gag* genes essential to the transposition process. Non LTR retrotranspsons on the other hand lack the terminal repeats characteristic of LTR retrotransposons and behave like an integrated mRNA containing reverse transcriptase and endonuclease domains (Han, 2010).

DNA transposons consist of a transposase gene flanked by two Terminal Inverted Repeats (TIRs). These TIRs are recognized by the transposase which binds to them, excises the transposon DNA and inserts it at another genomic location. Since the transposes make a staggered cut at the target insertion site, the resulting single stranded DNA is repaired by the host DNA polymerases and leads to a target site duplication flanking the transposable element. These target site duplications are characteristic of specific transposases and can act as indicators or footprints of transposition (Muñoz-López&García-Pérez, 2010).

1.3 Transposition Mechanism

Since this study is based on the use of DNA transposons as genomic tools, this section will focus on the transposition mechanism of Class II DNA transposons such as the bacterial Tn5

transposon and only briefly describe the transposition mechanism of Class I LTR and Non LTR retrotransposons.

LTR retrotransposons are first transcribed by the host's RNA polymerase and translated in the cytoplasm (Havecker, Gao, &Voytas, 2004). This includes the *gag* gene which encodes for several structural proteins that form a virus like particle (VLP) and the *pol* gene which encodes the reverse transcriptase and integrase required for transposition. The transposon RNA is then packaged into the VLP, reverse transcribed into double stranded cDNA by the reverse transcriptase and integrated into the host genome with the help of integrase (Havecker, Gao, &Voytas, 2004). This process results in the addition of a new copy of the retrotransposon to the host genome. Non LTR retrotransposons on the other hand are transcribed, translated and assembled into a ribonuleoprotein particle which is then imported back into the nucleus and integrated into the genome via a process called target-primed reverse transcription (Han, 2010).

Class II DNA transposons such as the bacterial Tn5 move through a cut and paste mechanism where they are excised from their original location and inserted randomly at another location in the genome. The molecular mechanism of transposition in DNA transposons as exemplified by the bacterial Tn5 transposon can be described as follows (Reznikoff, 2008):

1. The transposase(Tnp) identifies and binds to the terminal inverted repeat (TIR) or end sequences (ES). Since the transposase has to identify these sequences in a DNA polymer consisting of millions of bases, the precise targeting mechanism is still not fully understood and is an active area of investigation.
2. The Tnp forms a binary synaptic complex composed of two ES-bound Tnps constituting a scaffold for the subsequent catalytic steps.

3. The Tnp catalyzes cleavage of the ES sequences at each end of the transposon. This involves water mediated nicking of one strand to generate a 3' OH group which then attacks the opposite strand to generate a hairpin structure followed by Tnp catalyzed cleavage of the hairpin using water as a nucleophile.
4. The dimeric Tnp-ES complex cleaved from its original location then moves to the target site to form a target capture complex.
5. Nucleophillic attack of both 3' OH groups of the Tnp-ES complex on either side of the target sequence results in insertion of the transposon into the new location.

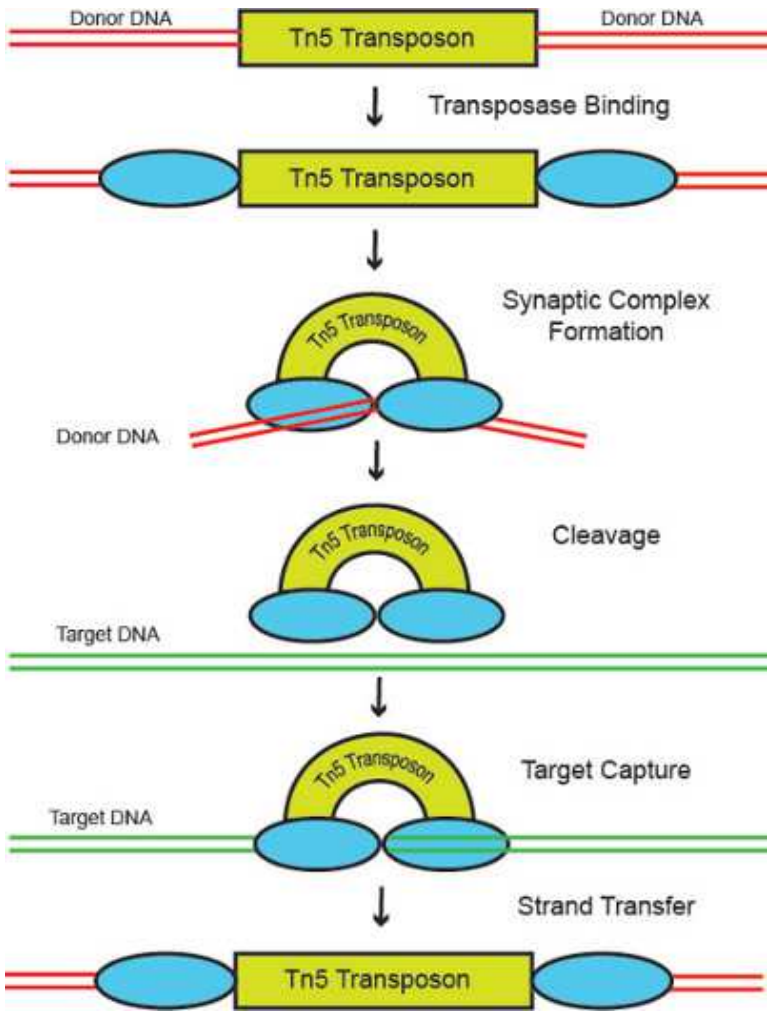


Figure 1. Tn5 Transposition Mechanism (Adapted from Reznikoff, 2003)

1.4 Regulation of Transposable Elements

While transposable elements have played a key role in genome evolution, excessive transposition can be potentially deleterious to the host genome if transposable elements colonize functionally important regions of the genome. Host genomes and especially the genomes of higher eukaryotes which harbor a large percentage of transposable elements have therefore evolved various strategies to suppress and regulate the proliferation of transposable elements. Similarly, as transposable elements depend on the survival of the host for their continued proliferation, transposable elements themselves have also evolved mechanisms to limit the impact of transposition on host fitness. Some of these strategies include (Muñoz-López&García-Pérez, 2010):

Overproduction Inhibition (OPI)

The increase of transposase concentration above a threshold level can lead to an inhibition of transposase production. While the exact mechanism of this inhibition is still being investigated, it is proposed that transposase monomers can combine to form oligomers with reduced or no activity. This mechanism serves to limit the deleterious proliferation of transposable elements as an increase in the copy number of these elements reduces the activity of transposase.

Vertical Inactivation

Vertical inactivation is a mechanism down regulating transposition activity through the accumulation of inactivating mutations in transposable elements. A large number of transposable elements in eukaryotes produce termination, deletion, frameshift and missense mutations that disrupt the open reading frame or result in the production of inactive transposase. While some of

these mutations may accumulate due to chance alone, it has been suggested that others may have been positively selected as they reduce the fitness cost of transposition. Similarly, as two transposase molecules are required for transposition, mutations producing inactive transposases can not only act as inhibitors of the process but also downregulate transposition through OPI.

Epigenetic Mechanisms

Host genomes have evolved various epigenetic mechanisms to suppress the proliferation of transposable elements. These include epigenetic silencing signals such as DNA methylation and histone modifications that specifically target transposable elements. Similarly, posttranscriptional silencing mechanisms such as RNA interference have also been proposed to regulate transposition.

1.5 Tn5 Transposon

The transposon mutagenesis system used in this study is based on the bacterial transposable element Tn5. The Tn5 transposon mutagenesis system has been chosen for this study based on a number of reasons. Firstly, Tn5 based systems have been successfully used in transposon mutagenesis studies in a range of organisms including *E. coli, Phaeobacterinhibens, Pseudomonas stutzeri, Shewanellaamazonensis,* and *Shewanellaoneidensis*(Wetmore et al. 2015) and Tn5 based studies have also obtained transposition events in *S. cerevisiae* (Reznikoff et al. 2000). Secondly, Tn5 based systems are also easily available commercially and use a hyperactive form of the Tn5 transposase containing several mutations to increase the efficiency of the transposition process. Similarly, unlike other transposon mutagenesis systems such as Mu, Tn7 and Ty1 which have reduced transposition activity *in vivo*, Tn5 has been demonstrated to be equally efficient for both *in vivo* and *in vitro* transposition (Goryshin 1998).

The Tn5 transposable element is an example of a composite transposable element in which three antibiotic resistance genes (kanamycin, bleomycin and streptomycin resistance) are flanked by two insertion elements IS50R and IS50L (Reznikoff,1993,2003, 2008). IS50L encodes an inactive truncated version of the transposase and is not relevant for the purposes of this discussion. IS50R encodes functional transposase (Tnp) as well as an inhibitor of transposition (Inh) (Reznikoff,1993,2003, 2008). The transposase encoded by IS50R interacts with 19bp inverted repeat sequences (end sequences or ESs) flanking the IS50R to catalyze transposition (Reznikoff,1993,2003, 2008). The transposase has limited target sequence specificity resulting in relatively random insertions and does not require any host proteins or energy in the form of ATP as the energy required for transposition is derived from cleavage of the target DNA (Reznikoff,1993,2003, 2008). These properties make Tn5 suitable for both *in vivo* and *in vitro* mutagenesis studies as the only requirements for transposition are the right pH and salt conditions, a donor and target DNA, an active transposase protein, and a divalent metal ion cofactor such as Mg^{2+} or Mn^{2+}.

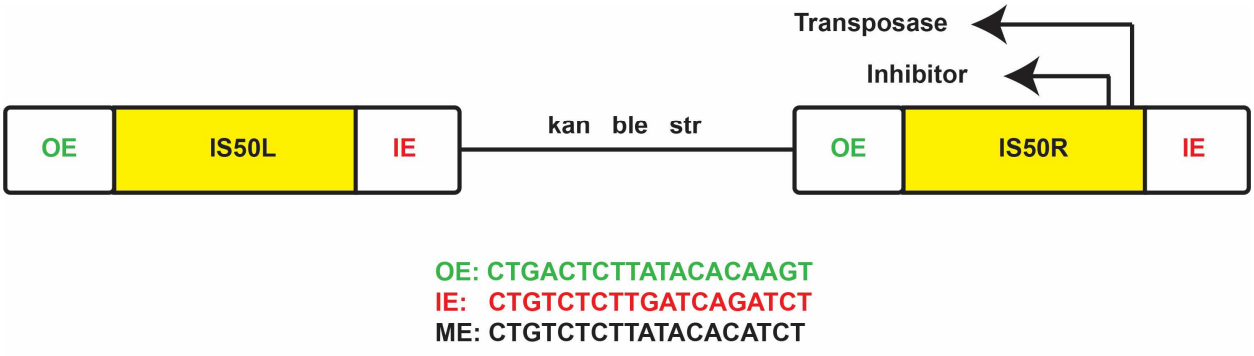


Figure 2. Structure of Tn5 Transposposonshowing end sequences andthehyperactive mosaic end sequence. (Adapted from Reznikoff 2008)

1.6 Regulation of Tn5 transposition

The wild type Tn5 transposase is a very inactive protein with the purified enzyme showing almost no activity *in vivo* or *in vitro* (Reznikoff 2003). As previously discussed transposable elements are highly regulated by the host genome to prevent deleterious proliferation and multiple mechanisms downregulate Tn5 transposition efficiency to less than 10^{-5} cells per generation making the system inherently inactive(Reznikoff, 2003).

The first of these mechanisms is the strong *cis*-bias displayed by the transposase. It has been shown that ES bound transposase is more than 50 fold more active at catalyzing transposition of the sequence that encoded the enzyme than ES bound sequences located elsewhere in the genome (Reznikoff, 2008). This mechanism ensures sufficient transposition to allow maintenance of the transposon while protecting the host genome from potentially deleterious mutations caused by excessive proliferation of the transposon.

Another mechanism regulating expression of the Tn5 transposase stops expression of the transposase from read through transcripts (sequences that have been transcribed beyond their normal termination sequence) (Reznikoff, 2008). Read through transcription in genomic regions where the transposon is located can lead to over expression of the transposase. However, the presence of a symmetrical DNA sequence near the termini of Tn5 leads to the formation of a read through mRNA secondary structure that prevents translation of the read through transcript (Reznikoff, 2008). Read through transcripts also lead to the synthesis of a truncated inactive transposase called an inhibitor (Inh) that can impair transposition by dimerizing with the active enzyme (Reznikoff, 2008). Finally, production of the transposase is further controlled by the Dam methylation of the transposase promoter.

1.7 Hyperactive Tn5 Mutants

While the low transposition efficiency of the Tn5 system is essential for host fitness, it makes its use as a genomics tool in mutagenesis studies impractical. As a result a number of hyperactive Tn5 mutants have been isolated that increase the transposition efficiency of the system by up to four folds. Using site directed mutation studies a number of activity enhancing point mutations have been identified that can increase the activity of the transposase even more when combined together.

The first class of these mutations includes L372P introduces a proline near the C-terminal end of the protein (Reznikoff, 2003). The N-terminal DNA binding domain and the C-terminal dimerization domain of the transposase are located close to each other and are thought to inhibit each other's activity. It is proposed that introduction of the proline near the C-terminal end helps separate the two domains and enhances the activity of the enzyme (Reznikoff, 2003).

A second class of hyperactive mutation is exemplified by E54K (Reznikoff, 2003). The E54K mutation increases the affinity of the enzyme to the 19bp terminal inverted repeat sequence and it has been shown that other mutations near residue 54 can further enhance the DNA binding affinity of the enzyme (Reznikoff, 2003).

In addition to L372P and E54K, many other mutations such as E110K, E345K, M56A, Y64I, K200R, and S303G have also been shown to increase transposition activity (Reznikoff, 1993, 2003, 2008). Furthermore, the activity of the transposase can be further increased by replacing the DNA end sequences with a 19-bp site termed mosaic end (ME) which differs from the original sequence at three positions and results in a 10–50-fold increase in transposition rate when used in conjunction with the hyperactive mutations (Reznikoff, 2003, 2008). These

modifications have made the Tn5 system highly efficient both in vivo and in vitro and an ideal tool for genomic studies.

1.8 Transposon Sequencing

Transposon sequencing (Tn-seq) is a technique that combines transposon insertional mutagenesis with massively parallel sequencing of insertion sites to identify genes contributing to functions of interest on a genome wide basis (van Opijnen and Camilli 2013). Tremendous advancements in sequencing technologies have led to the generation of enormous amounts of sequencing data. This availability of sequenced genomes for a large variety of organisms coupled with the rapidly declining cost of sequencing have made possible large scale studies linking observed phenotypes to specific genotypes and made transposon mutagenesis and ideal tool for gene identification.

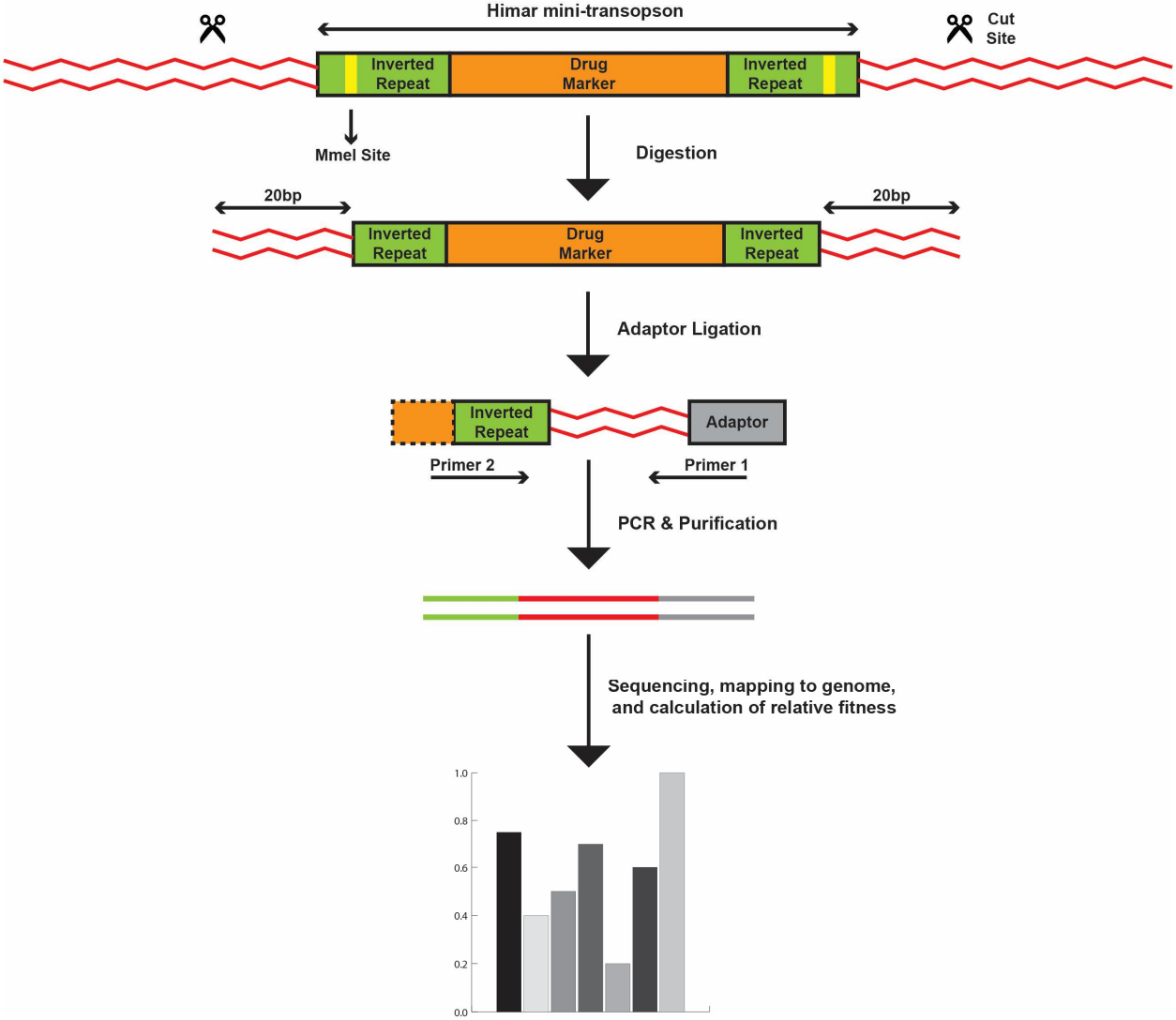
A transposon sequencing experiment is based on the construction of a mutant library where most or all non-essential genes have been knocked out using transposon insertions. This is followed by the growth of the library in specific experimental conditions under investigation. The relative frequency of each mutant in the library before and after the experiment is then calculated by sequencing the insertion sites and serves as a proxy for the fitness contribution of a gene under the condition being investigated.

1.9 Types of Tn-Seq

While the basic outline of a transposon sequencing experiment remains the same, a number of variations exist in how the transposon insertion library is constructed. These include library construction techniques such as high-throughput insertion tracking by deep sequencing

(HITS), transposon-directed insertion site sequencing (TraDIS), insertion sequencing (INSeq) and transposon sequencing (Tn-seq).

Figure3. Overview of TnSeq and INSeq. Adapted from van Opijnen and Camilli 2013.



Insertion sequencing (INSeq) and transposon sequencing (Tn-seq) are similar techniques that utilize the Himar I *Mariner* transposon for insertional mutagenesis. Both these techniques take advantage of the fact that a single base change in the inverted repeat sequence of the *Mariner* transposon can generate *MmeI* restriction sites (van Opijnen and Camilli 2013).

Since the *MmeI* restriction endonuclease cuts 20bp downstream of its recognition sites, digestion of DNA from a transposon mutant library with *MmeI* produces fragments containing the insertion site and 16bp of flanking DNA (van Opijnen and Camilli 2013). As fragments containing the insertion site are identical in size, they can be amplified by PCR and gel purified based on size. Sequencing of the 16bp flanking region then allows mapping of the insertion sites to the genome. Since this technique results in the generation of DNA fragments of identical length, it simplifies sample preparation but unlike other techniques is limited to *Mariner* transposons.

Unlike INSeq and Tn-seq, HITS and TraDIS use transposons that lack a restriction site that may be used for isolating the transposon insertion sites. As a result these techniques rely on random shearing of DNA followed by adapter ligation to produce DNA fragments of variable length (van Opijnen and Camilli 2013). Primers specific to the transposon insertion and adaptor are then used to amplify the DNA, which is then sequenced and mapped to the genome (van Opijnen and Camilli 2013). While this method can result in shorter fragments being preferentially amplified over longer ones, it has the advantage of being compatible with a range of transposable elements and is not limited to *Mariner* transposons.

1.10 Rb-TnSeq

While transposon sequencing protocols described previously are an excellent experimental tool for studies annotating gene function, they have not been used in a wide range of experimental conditions and organisms due to the highly labor intensive nature of the library preparation process. TRADIS and HITS for instance require DNA shearing, end repair, adaptor ligation, and PCR including several purification steps. Similarly, INSeq and Tn-Seq also rely on

multiple purification steps. More importantly, these techniques require the library preparation steps to be repeated for every experimental condition tested and hence limit the number of experimental conditions that can be realistically investigated. Random bar code transposon-site sequencing (RB-TnSeq) is a recent technique that retains the advantages of the transposon sequencing approaches while greatly reducing the time and resources required for library preparation (Wetmore et al. 2015).

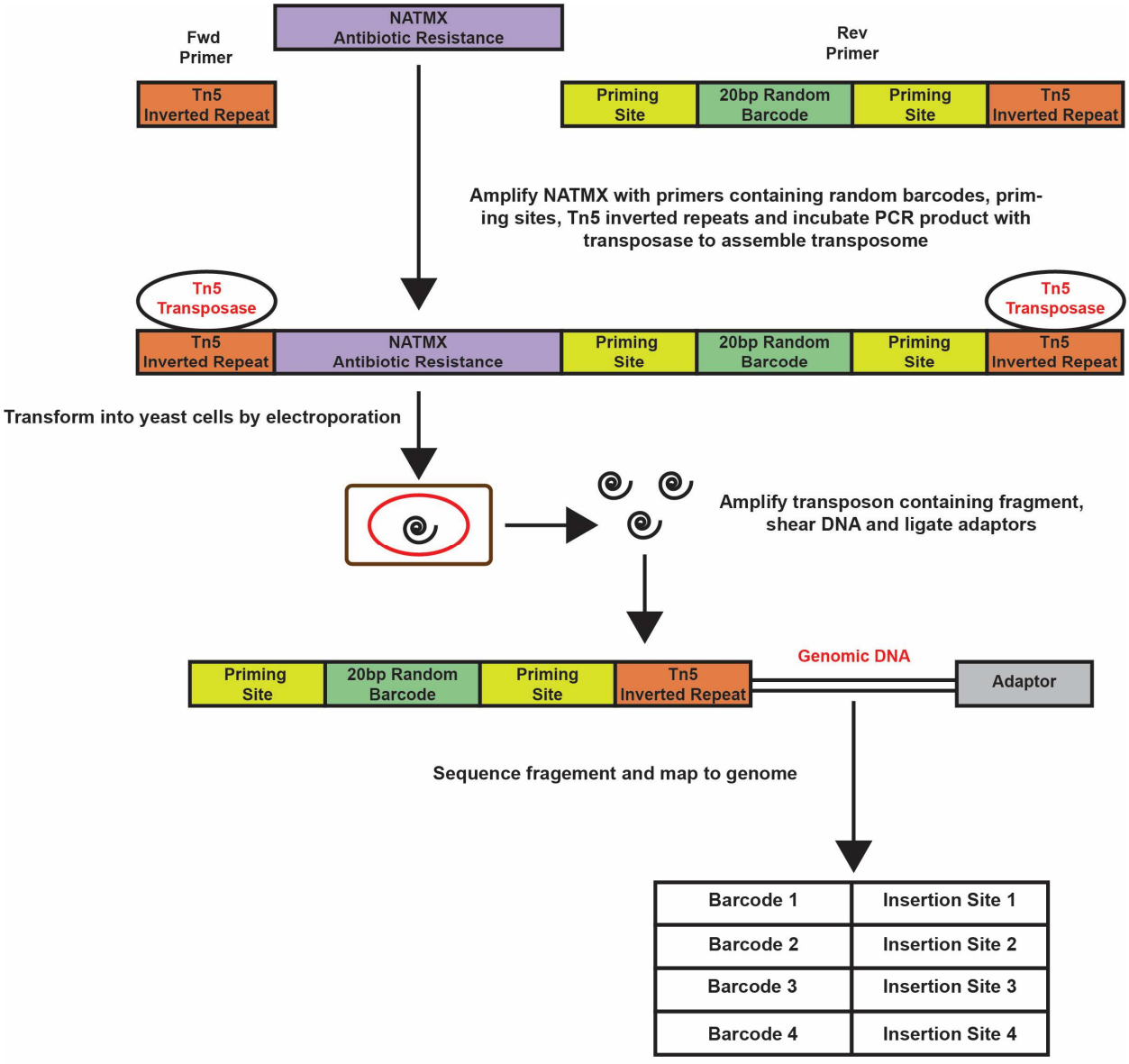


Figure 4. Overview of Rb-TnSeq. Adapted from Wetmore et al. 2015

RB-TnSeq streamlines the transposon sequencing protocol by decoupling the construction of transposon mutant libraries from the evaluation of mutant fitness under different experimental conditions (Wetmore et al. 2015). It achieves this by incorporating millions of random DNA barcodes and common PCR priming sites coupled to a drug resistance marker between the inverted repeat sequences of the transposon. A transposon mutant library is then constructed using these barcoded transposons with the goal that each mutant will carry a single transposon insertion linked to a unique barcode (Wetmore et al. 2015). The transposon mutants are isolated by selection on the relevant drug, DNA from the library is pooled and insertions mapped to the genome using any of the protocols described previously. This approach ensures that each specific mutation is associated with a unique DNA barcode and hence characterization of the mutant library is required once per organism rather than once per experimental condition tested (Wetmore et al. 2015). Once constructed, a mutant library can be used repeatedly for a wide range of experiments as subsequent mutant fitness assays rely on sequencing of the barcodes associated with each mutation and only require a simple PCR to amplify the barcodes instead of the multi step library preparation protocol. Furthermore, RB-TnSeq can be adapted for use with any of the transposon sequencing techniques and can allow for the multiplexing of multiple experiments by incorporating an index in the transposon.

While the RB-TnSeq protocol has been successfully applied to bacterial systems, this study focuses on adapting the protocol for the construction of mutant libraries in *S. cerevisiae.*

2. TRANSPOSON SEQUENCING

2.1 EXPERIMENTAL DESIGN

While most transposon sequencing techniques are similar in how the mutant fitness assays are conducted, the library construction process can differ significantly between different studies. The method of transposon delivery, transposition system used and location of transposition (whether *in vivo* or *in vitro*) will therefore depend on the protocol being followed. As Rb-TnSeq uncouples the library construction step from fitness assays and barcode sequencing, it can be adapted for use with different transposition systems and transposon delivery techniques. However, the transposition strategy used must fulfill certain requirements to ensure efficient mutagenesis (Reznikoff and Winterberg 2008). Firstly, the technique must ensure that the efficiency of transposition is sufficiently high to target every gene in the genome but not high enough to knock out multiple genes. Secondly, the transposition system should not display significant sequence bias and be random enough to ensure that every gene target has an equal chance of being targeted. Finally, the transposase specific to the transposition system should not be available in the target cell after the transposition event has taken place to ensure that the transposition is genetically stable. Based on these considerations this study focuses on testing two different transposon delivery techniques for constructing Rb-TnSeq libraries.

2.1.1 Plasmid based transposition

In vivo transposition strategies rely on plasmid transformation or conjugation for the introduction of the transposon into the target organism. However, the plasmid used for *in vivo* transposition must contain a number of adaptations to make it a suitable transposon vector. Firstly, the plasmid should act as a suicide vector and must not be able to replicate in the target

cells under the relevant experimental conditions. Secondly, the gene encoding the transposase should be located on the plasmid but outside the transposon itself to ensure that no transposase is present after loss of the suicide vector and the transposition is genetically stable. Finally, the plasmid should contain selection markers that allow selection of transposon mutants and a second marker that allows counter selection against cells containing the plasmid if a suicide vector is not used.

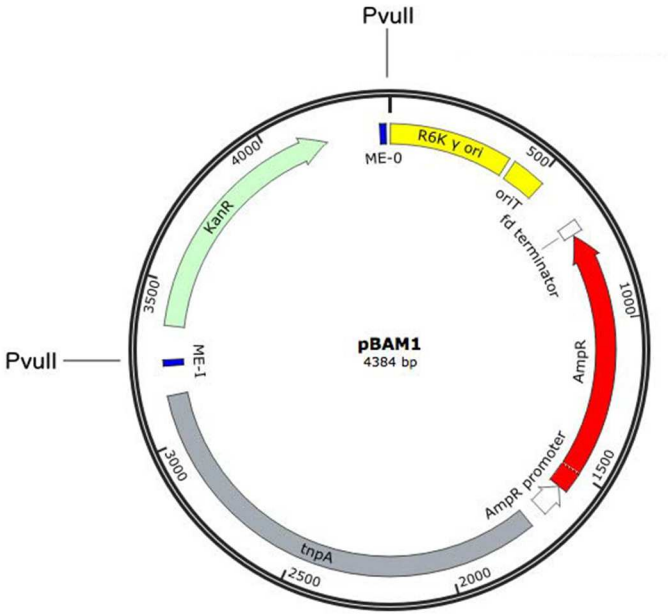


Figure 5.pBAM1 plasmid map showing antibiotic markers (Ap, ampicillin, Km, kanamycin), transposase (tnpA), origin of replication (R6K), the origin of transfer region (oriT), mosaic element O (ME-O), and mosaic element I (ME-I). Adapted from Zhang et al. 2008

For the purpose of this study the plasmid used is a modified version of the transposon vector pBAM1. pBAM1 is a mini transposon vector that can be used to make stable insertions of foreign DNA in gram negative bacteria (Zhang et al. 2008). It contains the R6K origin of replication that can only replicate in the presence of the π protein encoded by pir gene of R6K (Zhang et al. 2008). The plasmid can therefore only replicate in bacterial strains containing the pir gene and acts as a suicide vector in strains lacking it. It also contains the Tn5 tnpA gene encoding a hyper active version of the transposase due to the presence of the previously described E54K, M56A and L372P mutations (Zhang et al. 2008). Finally, to ensure genetically

stable transposition, the transposase gene is located outside the transposon itself and the 19bp inverted terminal repeats recognized by the transposase instead flank a kanamycin resistance cassette.

While pBAM1 has been successfully used to make insertion libraries in bacterial species such as *Pseudomonas putida* (Zhang et al. 2008), it requires a number of modifications for efficient transposition in yeast. This aim of this study is therefore to construct and test a number of different plasmids based on the pBAM1 backbone to maximize transposition efficiency in yeast cells. Firstly, the kanamycin resistance marker has been replaced with an antibiotic resistance marker functional in yeast. The plasmids used in this study therefore contain a NATMX cassette, which allows growth on the antibiotic nourseothricin (NAT). Secondly, the Tn5 tnpA gene in pBAM1 has been replaced with a tnpA gene codon optimized for expression in yeast and placed under an active yeast promoter. The glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter has been be used for the expression of the transposase gene since it has been successfully used for constitutive expression of heterologous proteins in yeast (Zhang et al. 2008). Furthermore, as this plasmid only contains a bacterial R6K replication origin it cannot replicate in yeast and acts as a suicide vector ensuring that no transposase is present after transposition. A second plasmid has also been constructed containing a yeast origin of replication. As this plasmid is able to replicate in yeast, it allows a comparison of transposition efficiency between the suicide vector and a vector capable of replication and continued transposase expression. Finally, a third plasmid has been constructed without an antibiotic resistance marker which can be co-transformed into yeast cells along with the transposon DNA. This plasmid can be used to evaluate the impact of co-transforming varying amounts of transposon DNA (containing the barcoded NATMX) on transposition efficiency and can be used

with different antibiotic cassettes as the antibiotic resistance gene is not present on the plasmid itself. A URA3 cassette has been included in all plasmid constructs to allow for the removal of plasmids through counter selection on 5-Fluoroorotic Acid that is converted to the toxic 5-fluorouracil in the presence of a functional URA3 gene.

The transposition efficiency of the different plasmid constructs can be tested using the following protocols:

Plasmid 1:

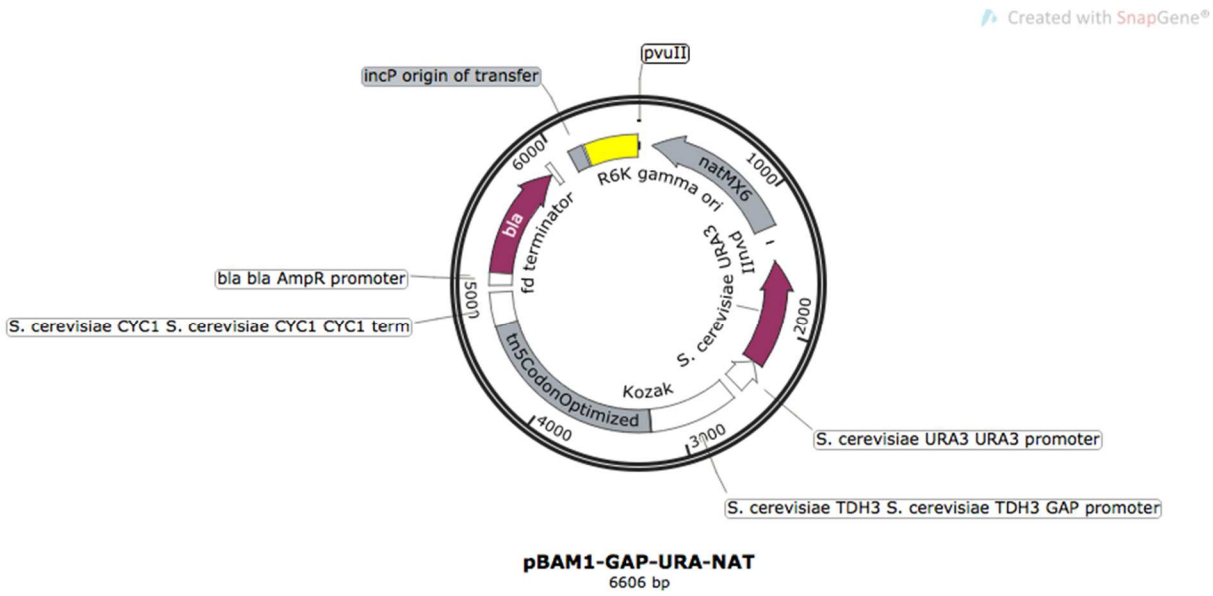


Figure 6. pBAM1 suicide vector with NAT resistance marker and Tn5 under GAP promoter

- The NATMX cassette was amplified using primers containing a 20bp random barcode, priming sites for amplification of insertion site, 19bp inverted repeats recognized by the transposase and 15bp sequence homologous with pBAM1 digested with pvuII.
- The barcoded NATMX cassette was cloned into the yeast optimized pBAM1 vector using pvuII restriction sites to replace the kanamycin cassette. Random primers allow

generation of millions of unique barcodes and multiple infusion cloning reactions lead to the creation of plasmid libraries containing different barcodes

- The plasmid library is transformed into yeast using standard transformation protocols.
- Transposon mutants are selected by replica plating on nourseothricin and 5-Fluoroorotic Acid

Plasmid 2:

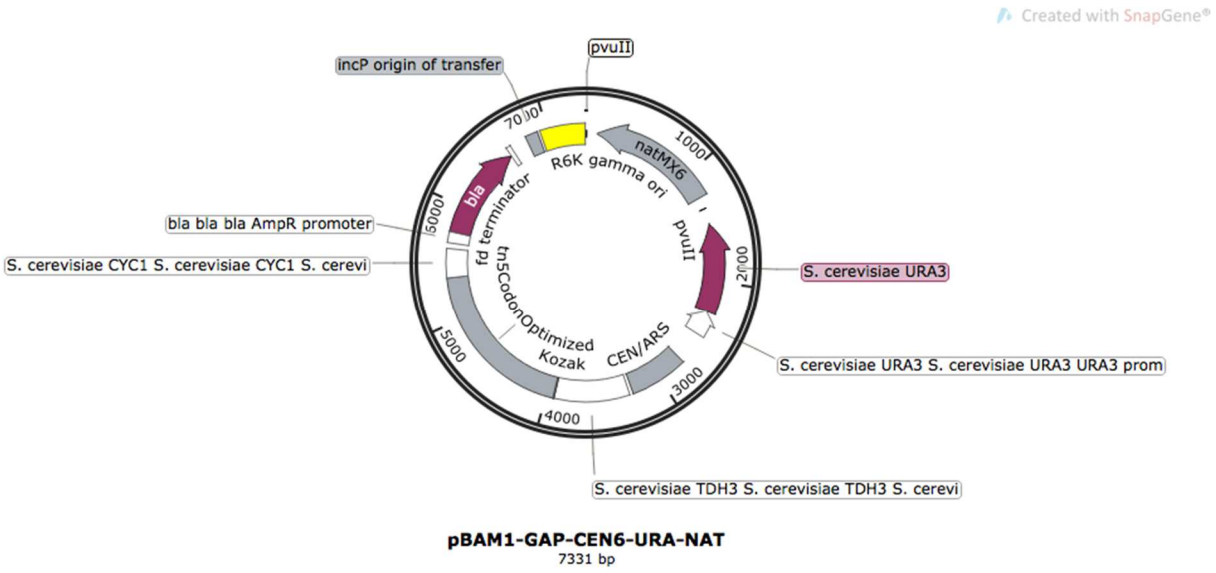


Figure 7.pBAM1 with NAT resistance marker, Tn5 under GAP promoter and Cen6 yeast replication origin

- The barcoded NATMX cassette was amplified as previously described and cloned into pBAM1 containing yeast origin of replication.
- The plasmids are transformed into yeast cells using standard transformation protocols.
- Transformants are grown in YPD overnight without selection to allow gradual loss of plasmid and plated on 5-Fluoroorotic Acid to ensure complete removal.
- Transposon mutants are selected by replica plating on nourseothricin

Plasmid 3:

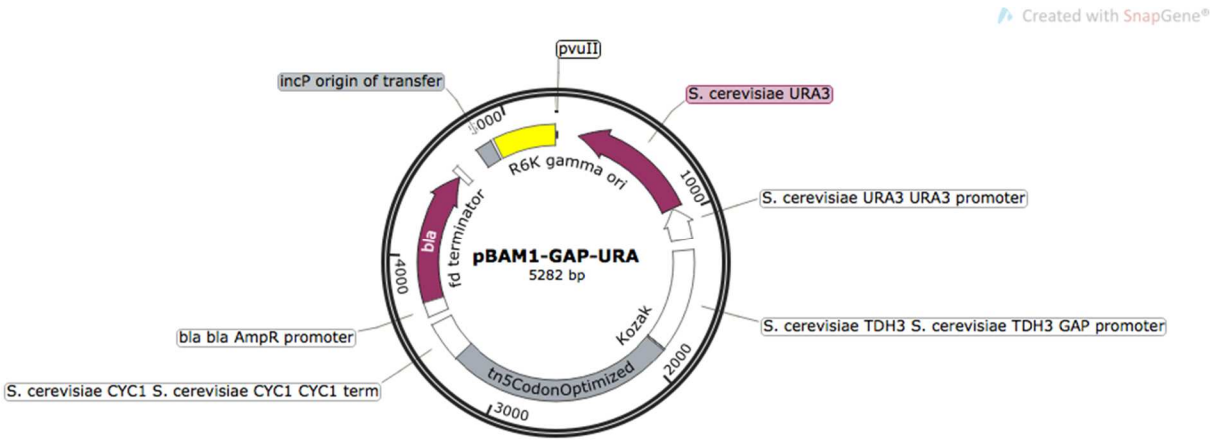


Figure 8. pBAM1 suicide vector with Tn5 under GAP promoter but without NAT resistance marker

- Plasmid 2 was digested with pvuII and religated to generate a plasmid lacking an antibiotic resistance cassette.
- The plasmid was co-transformed into yeast cells with varying amounts of barcoded NATMX cassette using standard transformation protocols.
- Transposon mutants were selected by replica plating on nourseothricin and 5-Fluoroorotic Acid

As previously described the main advantage of Rb-TnSeq is the uncoupling of library preparation from transposon sequencing which allows libraries to be constructed once per organism instead of once per condition. This requires the generation of mutants containing unique barcodes that can be associated with the specific gene knocked out in each mutant. Since the first two plasmids rely on the incorporation of a randomly barcoded NATMX cassette into the plasmid vector by infusion cloning, it is essential to ensure the successful cloning of the barcoded insert into every plasmid before transformation of the plasmid library into yeast cells. This study therefore utilizes the blue white screening technique to estimate the efficiency of

incorporation of the barcoded insert into the plasmid. To test the cloning efficiency the original antibiotic resistance marker in pBAM1 was replaced with the lacZ gene flanked by pvuII restriction sites. The plasmid containing the lacZ gene was then digested with pvuII followed by cloning of the NATMX cassette into the same restriction sites. Clones were then screened on plates containing X-Gal, which is converted into an insoluble blue pigment (5,5'-dibromo-4, 4'-dichloro-indigo) in the presence of functional β-galactosidase. The ratio of blue colonies containing the lacZ gene to white colonies containing the barcoded insert was then used to ensure that majority of the plasmids contained the barcoded insert..

2.1.2 Transposome based transposition

A transposome-based strategy combines both *in vivo* and *in vitro* steps to generate transposon mutants. A transposome is a complex of the Tn5 transposase enzyme bound to the terminal inverted repeats contained in transposon DNA. This technique relies on the *in vitro* assembly of transposome complexes followed by the electroporation of these complexes into the target cells. The *in vivo* step involves the activation of transposomes by the intracellular Mg^{2+}causing incorporation of the transposon DNA into the cell's genome. An advantage of this technique is that it does not require the construction of a transposon vector and since the transposase is not produced on a plasmid inside the cell no subsequent steps are required for removing the plasmid to ensure stable transposition. While this technique has been shown to work in *S. cerevisiae* (Reznikoff et al. 2000) it has some disadvantages that can limit transposition efficiency in yeast. As the technique requires the transformation of purified Tn5 transposase into yeast cells, it cannot be used with standard high efficiency yeast transformation protocols which include incubation at temperatures that may denature the enzyme. The technique must therefore rely on the electroporation of the transposome into yeast cells. However, since the

electroporation efficiency of yeast cells is lower than bacterial cells, the technique might require a large amount of purified enzyme and multiple electroporation reactions to achieve the desired transposition efficiency.

For the purpose of this study the following basic protocol was utilized for testing this approach:

- Tn5 transposase was overexpressed in *E. coli* and purified using chitin affinity chromatography (Picelli et al. 2014)
- NATMX cassette was amplified using primers containing a 20bp random barcode, priming sites for the amplification of insertion sites and 19bp terminal inverted repeats recognized by the transposase. The use of random primers allowed the generation of barcoded NAT cassette containing millions of unique barcodes
- The barcoded transposon DNA (100ng/ul) was mixed with purified Tn5 transposase (A280>3) and 100% glycerol in the ratio of 1:2:1 and incubated at room temperature for 1hour to allow transposome assembly
- 1ul of transposomes were electroporated into yeast using a modified version of the protocol described by (Benatuil et al. 2010) and cells were plated on selective media

2.2 RESULTS

2.2.1 Plasmid based transposition

The plasmids were constructed using infusion cloning as previously described. While plasmid number one and two have been constructed, they are still in the process of being tested. This section will therefore focus on plasmid 3 (pBAM1-GAP-Tn5-URA). The plasmid was tested by transformation in BY (S288C) using standard high efficiency yeast transformation protocols.

Since this plasmid lacks an antibiotic resistance marker, it was co-transformed into yeast cells along with NATMX cassette amplified using primers containing random barcodes and Tn5 inverted repeats. 1ug of the barcoded NATMX transposon DNA and 400ng of the plasmid were transformed into BY using standard yeast transformation protocol. 1ug of NATMX transposon DNA without the plasmid was transformed into BY as a negative control. ~50 mutant colonies were obtained on YPD plates containing 100ug/ml NAT while ~10 colonies were obtained on the negative control plate. While further work is required to optimize to maximize transposition efficiency sufficient for library construction, this result proves that it is possible to obtain *in vivo* transposition in yeast using a plasmid based approach. Furthermore, since the antibiotic marker is not encoded by the vector, this plasmid can be used with different antibiotic cassettes and allows the efficiency of transposition to be optimized by changing the amounts of transformed plasmid and transposon DNA. The two plasmids encoding the antibiotic resistance marker and containing the yeast origin of replication are still being tested and will allow a comparison of the three plasmid based approaches to maxixmize transposition efficiency.

2.2.2 Transposome based transposition

While transposome based transposition has been previously reported in *S. cerevisiae* (Reznikoff et al. 2000), this study was not able to successfully generate transposon mutants in yeast cells using the transposome-based techniques described previously.

The first challenge in optimizing transposome-based mutagenesis in yeast was maximizing electroporation efficiency in yeast cells to levels sufficient for transposition. According to the conditions recommended by Epicentre for its EZ-Tn5^{TM}Custom Transposome Construction Kits, a transformation efficiency of at least 10^6cfu/ug is essential for efficient mutagenesis. A range of different electroporation conditions was therefore tested to maximize transformation efficiency. The initial transformations were performed according to the protocol described by Benatuil et al. 2010 using plasmids containing different antibiotic resistance cassettes and resulted in transformation efficiencies in the range of 10^4cfu/ug. To improve transformation efficiency to levels sufficient for transposition, a number of modifications to the electroporation protocol were tested. Firstly, the impact of plating transformants either directly on selective plates versus replica printing was analyzed. Secondly, carrier DNA was added to the electroporation reaction. Thirdly, the impact of starting OD_{600} on transformation efficiency was analyzed and electroporation efficiency of auxotrophic selection was compared with antibiotic selection. Finally, various concentrations of different antibiotics were tested to ensure minimum background and maximum efficiency in three different strains of yeast (NAT was the only antibiotic with no background in all strains). Together these modifications yielded plasmid electroporation efficiencies ranging from 10^5 to 10^6.

100ng PAG36 into BY OD_{600}~3.00	Colonies	Efficiency
Directly on NAT	~ 500	7.0E+04
Replica Print on NAT	~ 800	1.1E+05
20ug Cr.DNA Directly on NAT	~ 1200	1.7E+05
20ug Cr.DNA Replica Print on NAT	~ 1800	2.5E+05

Figure9. Electroporation efficiency of transforming 100ng of PAG36 plasmid into yeast (S288C) with and without carrier DNA and replica plating (on 100ug/ml NAT).

Dilution	OD_{600} 10		OD_{600} 6.5		OD_{600} 3.4	
	URA-	NAT	URA-	NAT	URA-	NAT
Projected	8500	870	7000	1300	6200	3500
1 in 10	850	87	700	130	620	350
1 in 100	147	4	81	8	140	42
1 in 1000	17	2	14	3	20	1
Efficiency	1.19E+06	1.22E+05	9.80E+05	1.82E+05	8.68E+05	4.90E+05

Figure 10. Impact of staring OD600 on electroporation efficiency with antibiotic or auxotrophic selection markers using a range of dilutions. (Using 100ng of PAG36, replica printing, 20ug Carrier DNA. Electroporation efficiency calculated using 1 in 10 dilution on 100ug/ml NAT).

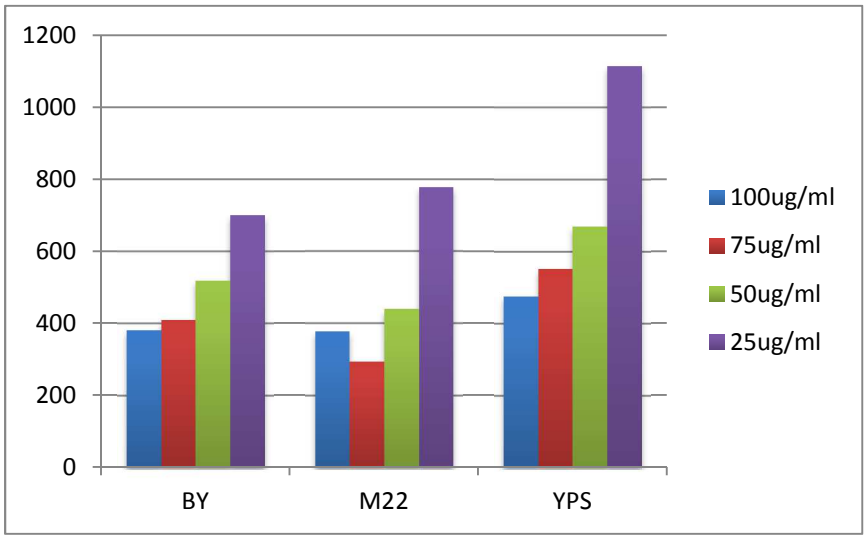


Figure 11. Impact of different concentrations of NAT on number of colonies obtained in BY (S288C), M22 (wild vineyard strain), and YPS (wild oak strain). (Using 100ng of PAG36, replica printing, 20ug Carrier DNA. Electroporation efficiency calculated using 1 in 10 dilution)

The second obstacle in optimizing transposome-based mutagenesis in yeast was the production of active transposase enzyme. Tn5 transposase was purified using the protocol described by Picelli et al. 2014 with NEB C3013 competent cells and Addgene pTXB1-Tn5 plasmid. However, significant overexpression was not observed with the pTXB-1 plasmid under conditions described by Picelli et al. 2014 (growth at 37°C to OD_{600} = 0.9,addition of 0.25 mM IPTG and induction for 4 h at 23°C) and the purified protein showed no activity in yeast or *E.coli*. A range of different conditions was then tested for maximizing overexpression. Growth at 37°C to OD_{600} = 0.9, followed by overnight induction with 1mM IPTG at 37°C produced the best results but protein purified under these conditions did not show any activity in yeast or *E.coli*.

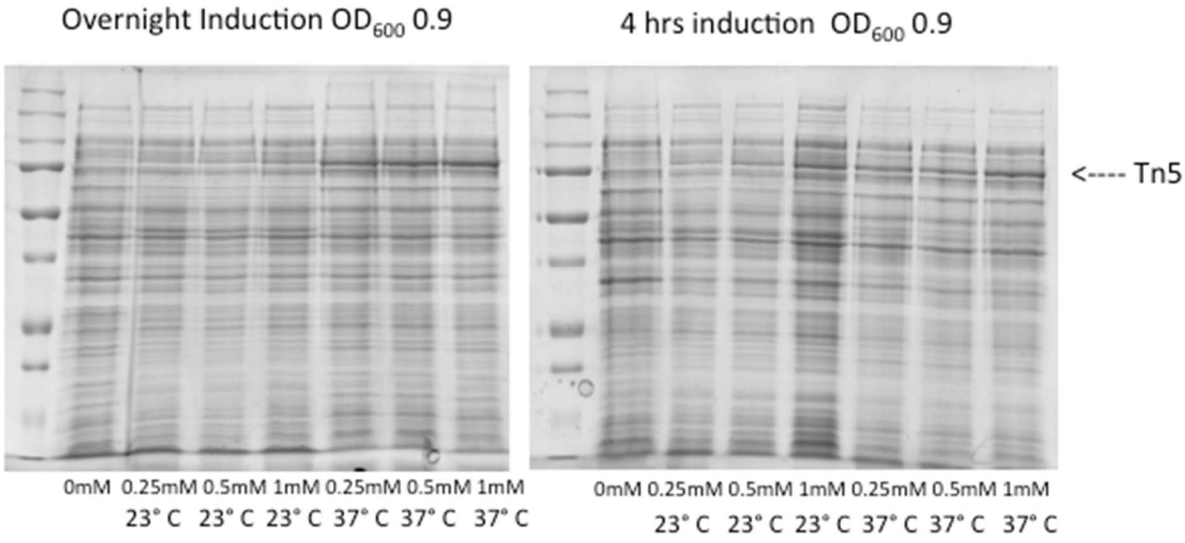


Figure 12. Tn5 Overexpression using different IPTG concentrations, temperatures, and induction times. Growth at 37°C to OD_{600} = 0.9, 0.25 mM IPTG and induction for 4 h at 23°C worked best.

To ensure that no mutations were present in the original plasmid, the Tn5 gene was sequenced and re-cloned into pTXB1 vector and the protein purified again. While considerably

more overexpression was observed with this plasmid using the original conditions described by Picelli et al. 2014, the purified protein remained inactive in yeast and *E.coli*. To ensure efficient transposome assembly while testing the purified transposase, two different methods of transposome assembly were tested. Transposomes were assembled either as described by Picelli et al. 2014 (7ul of 55% glycerol stock of Tn5 A280>3 was mixed with 1ul transposon DNA at 100ng/ul and incubated at room temperature for one hour) or using the protocol recommended by Epicentre for its EZ-Tn5^{TM} Custom Transposome Construction Kits (1ul transposon DNA (100ng/ul) was mixed with 2ul purified Tn5 transposase (A280>3) and 1ul 100% glycerol and incubated at room temperature for 30 minutes). However, no activity was observed in yeast or *E.coli* using transposomes assembled by either approach.

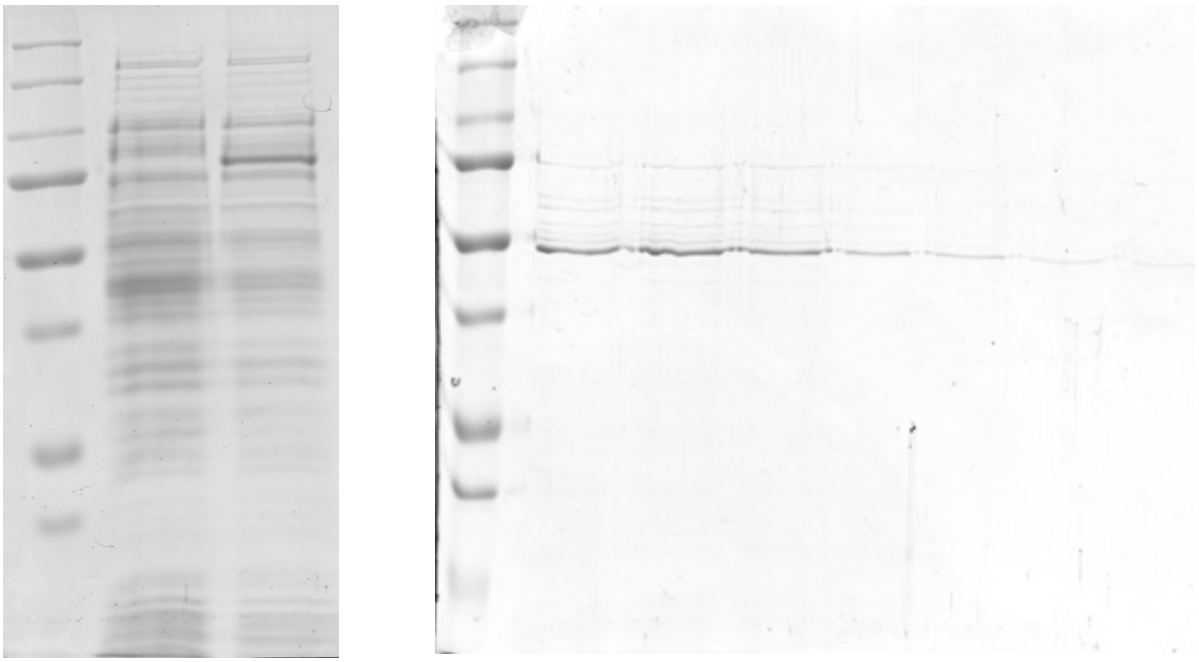


Figure 13. Tn5 overexpressed and purified with recloned plasmid using conditions described by Picelli et al.

2.3 MATERIALS AND METHODS

2.3.1 Tn5 Purification

Tn5 transposase was purified using protocol described by Picelli et al2014 as follows:

- pTXB1-Tn5 plasmid was transformed into C3013 competent cells for overexpression (protocol for making chemically competent C3013 competent cells is described separately).
- One litre culture was grown to A600=0.9 at 37°C in LB media contacting 100ug/ml ampicillin.
- Culture was chilled to 10°C in an ice bath and IPTG was added to a concentration of 0.25mM.Culture was grown for another 4 hours at 23°C to an A600 of ~3.0.
- 3-4gr of cells were collected by centrifugation and frozen at -80°C. Samples were run on SDS-PAGE gel to confirm overexpression of protein.
- Cells were thawed on ice and resuspended in 15ml of column buffer containing protease inhibitors and DNaseI (20mM HepesNaOH pH 7.2, 0.8M NaCl, 1mM EDTA, 10% Glycerol, 0.2% Triton X-100) for every 5g of cells.
- Sonication was carried out using five cycles (6 bursts of 10sec with 20sec intervals) at 80% probe intensity in a beaker packed in ice. Protein concentration was analyzed using Bradford assay to ensure complete lysis.
- Lysate was pelleted using Beckman JA17 rotor at 15,000 rpm for 30 min at 4°C and 2.1ml of 10% neutralized PEI (pH~7) was added drop wise on a magnetic stirrer at 4°C.Precipitate was removed using centrifugation at 12,000rpm for 10min at 4°C.

- Supernatant was filtered through 0.45 micron filer and loaded on a chitin column (NEB) with 5ml bed volume at ~0.4ml/min. Once all lysate had passed, the column was washed with 20 volumes of column buffer.
- Column buffer containing 100mM DTT was added to the column and buffer was drained until the flow through smelled (~20-30ml).
- Column was left closed for 36-48 hours at 4°C to effect cleavage of Tn5 from intein and eluted with column buffer in 1ml aliquots which were analyzed using SDS-PAGE gel and pooled together.
- Protein samples were dialysed overnight at 4°C in a litre of dialysis buffer (100mM HEPES-KOH pH 7.2, 0.2M NaCl, 0.2mM EDTA, 2mM DTT, 0.2% Triton x-100, 20% Glycerol). Buffer was changed twice at 2-hour intervals next day.
- A280 after dialysis was measured using nanodrop (Tn5 M.W = 53.3kD Eu=86,525) and protein was concentrated using Amicon 30 filters to A280 greater than 3.00 if required.
- Protein was frozen at -20°C as 55% glycerol stock (1.1 vol 100% glycerol, 0.33 vol dialysis buffer, and 1 vol protein).

2.3.2 Yeast Electroporation

Yeast electroporation was performed using a modified version of electroporation protocol described by Benatuil et al 2010 as follows:

- Cell were grown overnight in 2ml YPD at 30°C.Overnight culture was used to start a fresh culture in 100ml YPD and cells were grown overnight to OD_{600} of ~ 3.2 ($1.6x10^{\wedge 7}$ cells/ml)

- Cells were collected by centrifugation in 50ml falcon tubes and washed twice with ice cold sterile water and once by ice cold electroporation buffer (1M Sorbitol, 1mM $CaCl_2$)
- Cells were conditioned by resuspending the pellet in 20ml 0.1M LiAc, 10mM DTT and incubated at 30°C with shaking at 270rpm
- Cells were collected by centrifugation, washed once with ice cold electroporation buffer, and resuspended in 200ul of electroporation buffer
- Transposomes were assembled by mixing 2ul of DNA at 100ng/ul, 4ul of purified Tn5, 2ul of 100% glycerol and incubating at room temperature for 1hour
- 100ng of plasmid or 1ul of transposome was added to 400ul of cells in pre-chilled epitubes, incubated on ice for 5 minutes and transferred to pre-chilled 0.2cm electroporation cuvettes
- Electroporation was carried out at 2.5 kV, 25 μF and 200ohms.Typical time constant ranged from 3.0 to 4.5 milliseconds
- 1mL of 1:1 mix of 1 M sorbitol and YPD media (pre warmed to 30C) was immediately added to the cuvette. Cells were incubate with shaking at 270 rpm and 30°C for 1 hour
- 100ul of cells were plated directly on antibiotic plates and incubated at 30°C until colonies appeared (~1-2days)

2.3.3 Yeast Transformation

Yeast transformations were performed using protocol described by Gietz et al 2007

- Transformation mix was prepared as follows:

  240μL50%PEG 3350

  18 μL 2M Lithium Acetate

  10μL Carrier DNA (10 mg/mL) (Denatured for 10min at 95°C)

58 µL water

---

326 µL total volume

- Cells were grown overnight in 2ml YPD at 30°C and used to start a fresh culture in 5ml 2X YPD at an OD_{600} of 0.5 (~5 x 10^6 cells/mL)
- Cells were grown at 30°Cand 270rpm for at least 2 doublings (~4hrs)
- Cells were collected by centrifugation at 1500rcf for 3 minutes, washed twice with 2.5ml of sterile water and resuspended in 100ul of sterile water
- Cells were transferred to 1.5ml microcentrifuge tube, and collected by centrifugation at 13,000rcf for 30sec.
- Cells were washed with 100ul of sterile water and resuspended in 326ul of transformation mix.
- Up to 34ul of plasmid or PCR fragment was added to the transformation mix and sterile water was added to final volume of 360ul
- Tubes were vortexed and placed in a water bath at 42°C for 60 minutes
- Cells were centrifuged at 1500rcf for 30sec to remove supernatant, resuspended in 100ul of sterile YPD (or 100ul of 2% glucose and plated directly on selective media for auxotrophic selection) and plated on YPD
- Plate were incubated overnight at 30°C, replica printed on antibiotic plates and incubated at 30°C until colonies appeared (2-3 days)

2.3.4 Competent cells

Chemically competent C3013 cells for Tn5 production and chemically competent Dh5α-λ pir cells for pBAM1 were prepared as follows:

- Fresh plate of relevant *E.Coli* strain was used to start a culture in 0.5ml SOB media at 37°C
- Overnight culture was used to start a culture in 100ml SOB which was grown at 37°C with shaking at 270rpm to $OD_{600}$of 0.425 (cells were grown in a 1litre flask to provide abundant aeration)
- Culture was chilled on ice for 10 minutes and cells were collected by centrifugation at 3,000xg for 10 minutes at 4°C
- Cells were resuspended in 15ml CM (50mM $CaCl_2$, 50mM $MgCl_2$, 10% Glycerol) and chilled on ice for 15min
- Cells were collected by centrifugation at 3,000xg for 10 minutes at 4°C, resuspended in 3.6ml CM and chilled on ice for 5 min
- 125ul of DMSO was added twice with 5min on ice in between
- 50ul aliquots were frozen at -80°C for future use
- Before transformation cells were thawed on ice and incubated with relevant plasmid for 15-30min on ice
- Cells were heat shocked for 45sec in 42°C heat block, placed on ice for 2 minutes, transferred to 0.5ml LB media and incubated for 1hr at 37°C with shaking at 270rpm
- Cells were resuspended in 100ul LB and plated on selective media

2.3.5 Cloning

All cloning reactions were performed using Clonetech In-Fusion® HD Cloning Kit using standard protocols. Vectors digestion was performed using Thermo Scientific™ FastDigest™ restriction enzymes. PCR reactions for cloning were performed using EMD MILLIPORE KOD Hot Start DNA Polymerase. DNA purification was performed using ThermofisherPureLink®PCR Purification Kit. Plasmids were isolated using ZymoPURE™ Plasmid Midiprep Kit. Custom DNA oligos and primers for cloning were ordered from Integrated DNA Technologies.

2.3.6 Primers

MX pBAM1 F	ATGGGGCGCGCCCAGCTGTCTCTTATACACATCTCGGTGTCGGTCTCGTAG-NNNNNNNNNNNNNNNNNNATCGATGAATTCGAGCTCG	For infusion cloning barcoded NATMX from pAG36 into pbam1 pvuII sites
MX pBAM1 R	TGTAGTTTAAACCAGCTGTCTCTTATACACATCTCGTACGCTGCAGGTCGAC	
LacZ pBAM1 F	ATGGGGCGCGCCCAGCTGTCTCTTATACACATCTGGAGAAAATACCGCATCAGG	For infusion cloning LacZ from prs426 into pbam1 pvuII sites
LacZ pBAM1 R	TGTAGTTTAAACCAGCTGTCTCTTATACACATCTCAGGTTTCCCGACTGGAAAG	
Fwd_Inf_CenUrapBAM1	AGAGACAGCTGGTTTAAACCCTGATGCGGTATTTTCTC	For infusion cloning Cen6 and Ura from PAG36 into pbam1 pmeI site
Rev_Inf_CenUrapBAM1	CTGCAGTGCACGTTTAAACATCACGTGCTATAAAAATAATT	
Fwd_Inf_TDH3pBAM1	AGAGACAGCTGGTTTAAACGTGCACTGCAGCTC	For infusion cloning codon optimized Tn5 into pbam1 pmeI and swaI sites
Rev_Inf_TDH3pBAM1	GCGGGCTACTAATTTAAATGCAAATTAAAGCCTTCG	
Tn5_bar_KanMX_F	CTGTCTCTTATACACATCTNNNNNNNNNNNNNNNNNNCGTACGCTGCAGGTCGAC	For amplifying barcoded NATMX with Tn5 inverted repeats
Tn5_bar_KanMX_R4	CTGTCTCTTATACACATCTCGGTGTCGGTCTCGTAGNNNNNNNNNNNNNNNNNNATCGATGAATTCGAGCTCG	
Tn5 pTXB1 Nde F IF	GAAGGAGATATACATATGATTACCAGTGCACTGCATC	For infusion cloning Tn5 gene into ptxb1 spaI and ndeI sites
Tn5 pTXB1 Sap1 R IF	ATCTCCCGTGATGCAGATTTTAATGCCCTGCGCCATC	
Rev_Inf_UrapBAM1	CTGCAGTGCACGTTTAAACGCATCAGAGCAGATTGTAC	Use with Fwd_Inf_CenUrapBAM1 to clone URA only
MX pBAM1 R2	GCACGTTTAAACCAGCTGTCTCTTATACACATCTCGTACGCTGCAGGTCGAC	Use with MX pBAM1 F for pbam1 containing GAP and optimized Tn5
MX pBAM1 R3	CAGGGTTTAAACCAGCTGTCTCTTATACACATCTCGTACGCTGCAGGTCGAC	Use with MX pBAM1 F for pbam1 containing GAP,Cen6,Ura& Tn5

3. NATURAL VARIATION IN STRESS RESPONSE

Since one of the main advantages of adapting Rb-TnSeq for use in *S. cerevisiaei* is the ability to construct and compare mutant libraries in different strains of yeast, it is important to identify specific phenotypic differences among these strains that can potentially be used for mutant fitness assays in future. A number of preliminary experiments were therefore performed analyzing natural variation in the stress response of lab and wild strains of yeast. Strains used included DBY (S288C - lab strain), YPS (oak strain), M22 (vineyard strain) and YJM339 (clinical isolate).

The aim of these experiments was to identify unique phenotypic differences between the commonly used lab strain and wild yeast isolates in response to heat, salt or oxidative stress. While the primary goal was to analyze natural variation in stress tolerance of yeast isolates and identify broad phenotypic differences that could potentially be analyzed using transposon sequencing, these experiments were also used for analyzing the role of posttranslational modifications (acetylation) in the acquisition of stress resistance.

To analyze potential differences in stress response pathways involving posttranslational modifications, all strains were treated with a protein synthesis inhibitors before exposure to a mild heat, salt or oxidative stress followed by a more severe secondary stress. Acquisition of resistance in the absence of protein synthesis was used as an indication of the possible role of post-translational regulation in the acquisition of stress resistance.

3.1 EXPERIMENTAL DESIGN

All strains were grown to mid log phase (OD_{600} of 0.3-0.6) in shaking incubators at 30ºC (strains being exposed to 37ºC primary heat stress were grown at 25ºC instead of 30ºC to ensure a temperature differential sufficient to induce a stress response). To eliminate any previous epigenetic memory of stress resistance all strains were grown for at least 8 doublings. Cycloheximide (CHX), which inhibits protein synthesis in eukaryotic cells by interfering with eukaryotic translation, was used as a protein synthesis inhibitor for all stress assays. Once in the required OD range, one 5ml sample of each strain was incubated with 10ug/ml of CHX for 20 minutes at the original growth temperature. Another 5ml sample will be incubated without CHX as a control. All samples were then exposed to the primary stress for one hour. 0.7M NaCl, and 4mM H_2O_2 were used as salt and oxidative stress respectively while growth at 37ºC was used as heat stress.

Upon completion of primary stress, all samples were centrifuged and re-suspended in YPD to an OD of 0.6 and exposed to secondary stress for 1hr. Temperatures ranging from 42C to 48C, NaCl concentrations ranging from zero to 3.2 M, and H_2O_2 concentrations ranging from zero to 5 mM were used as the secondary heat salt and oxidative stress respectively. On completion of secondary stress 1:50 dilutions of all cultures were plated on YPD and incubated at 30ºC until colonies appeared. A semi quantitative score for each sample was calculated. For experiments analyzing variation in acetylation patterns, cells were collected and frozen for future analysis via western blotting.

3.2 RESULTS

The main stress experiment that identified a reproducible phenotypic difference between the wild and lab strains of yeast compared the acquisition of thermotolerance among wild and lab strains in the absence of protein synthesis. Results indicated that while all strains displayed some acquisition of thermotolerance after exposure to the initial primary stress, the commonly used lab strain DBY (S288C) was unique in being the only strain to exhibit acquisition of thermotolerance in the presence of protein synthesis inhibitor CHX. It is therefore possible that while protein synthesis plays an important role in the acquisition of thermotolerance in wild yeast strains, the lab strain may have acquired a novel stress response pathway relying on post-translational protein modifications. Western blot analysis of the wild & lab strains grown at 25ºC and exposed to heat shock at 37ºC has also revealed acetylation differences between the strains that correlated with this phenotype. Furthermore, considerable variation in acetylation patterns was observed both among the strains and on exposure to different stresses.

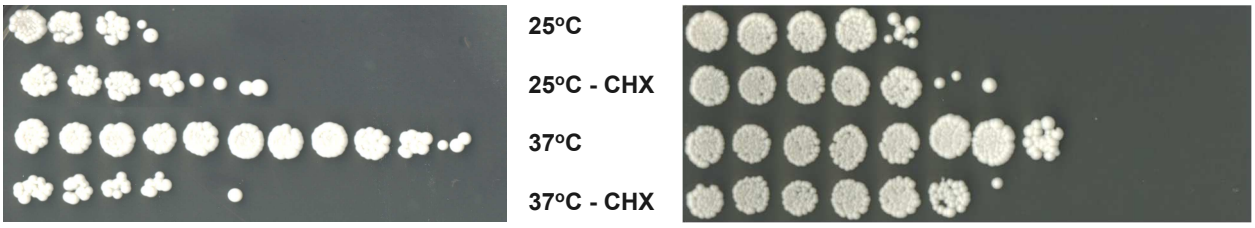


Figure 14. Acquisition of thermotolerance in BY in the presence of CHX

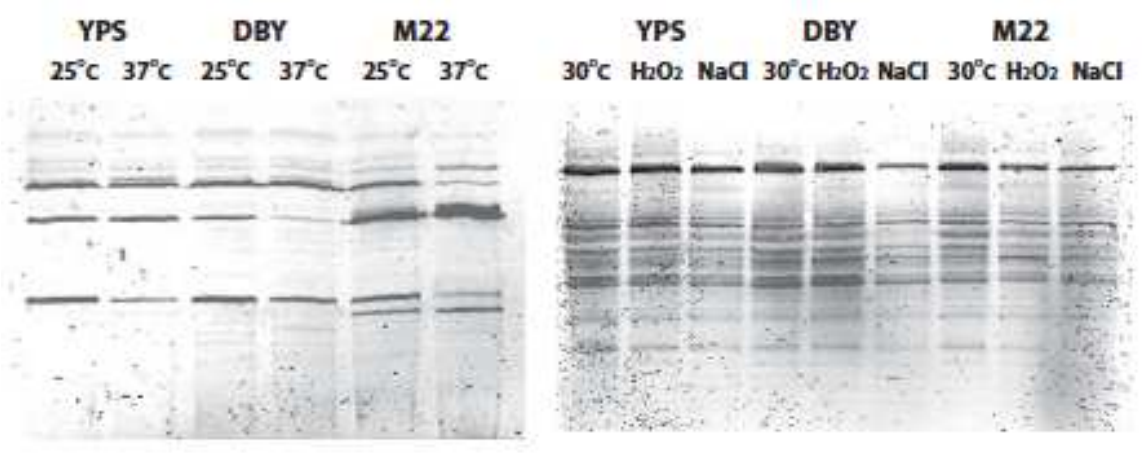


Figure 15. Western blots showing acetylation patterns in lab and wild strains after heat, salt or oxidative stress

While the lab strain was unique in being able to acquire thermotolerance in the absence of protein synthesis, additional stress assays were also performed to confirm if this acquisition was specific to heat stress or also applicable to salt and oxidative stresses. Furthermore, different combinations of heat, salt and oxidative primary and secondary stresses were used to check if exposure to one type of stress can cross protect against a different stress in the presence of protein synthesis inhibition and how these responses vary among yeast isolates from different environments. Results of these assays are shown in Figures 16 to 18. The results of these experiments demonstrated that the acquisition of resistance in the presence of protein synthesis inhibition was both unique to the lab strain and only occurred using heat as both the mild and severe stress. While all strains displayed some acquisition of resistance to heat, salt, and oxidative stress after pretreatment with mild heat stress at 37ºC, this acquisition was not observed in the presence of cycloheximide, indicating that protein synthesis independent acquisition of resistance was specific to both the strain and the type of stress used. Acquisition of

resistance, however, was observed in all strains in the absence of cycloheximide when using heat pretreatment at 37ºC in combination with salt or hydrogen peroxide as the secondary stress indicating that it was possible for mild heat stress to cross protect against salt and oxidative stress. An intriguing phenotype was also observed when testing the acquisition of resistance to salt stress after prior heat treatment in the presence of cycloheximide with all strains surviving much higher concentrations of salt in the presence of protein synthesis inhibition.

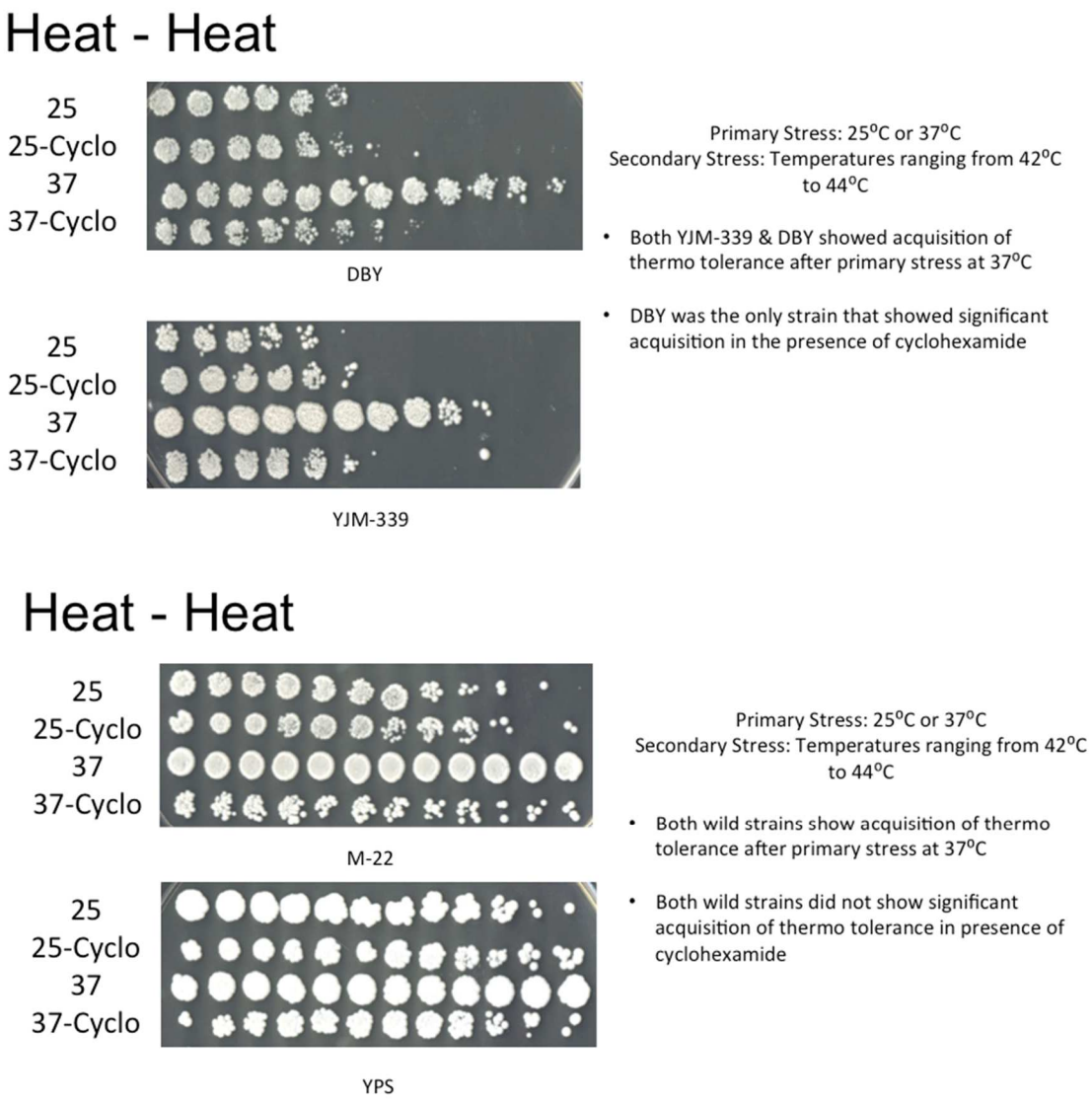


Figure 16. Acquisition of resistance in lab and wild strains using heat as primary (mild) and secondary (severe) stress

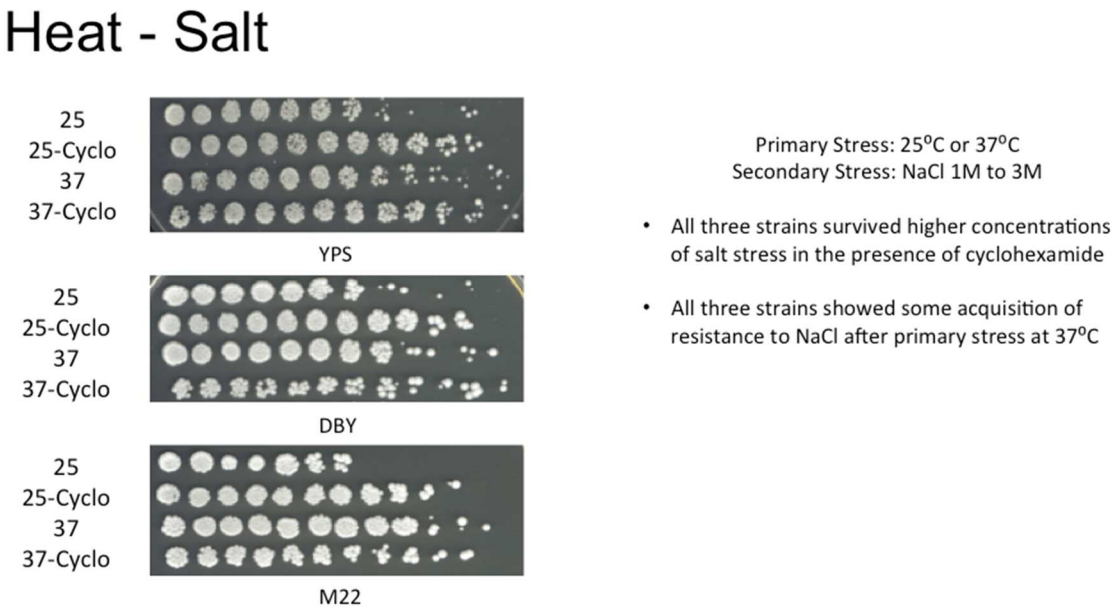


Figure 17. Acquisition of resistance in lab and wild strains using mild heat shock as primary and salt as secondary stress

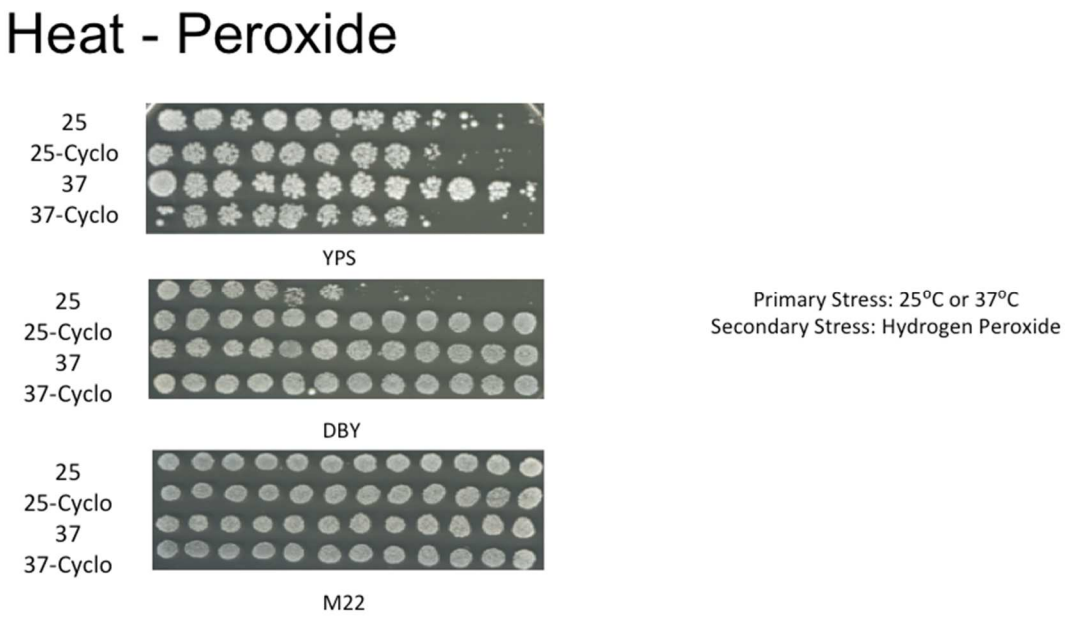
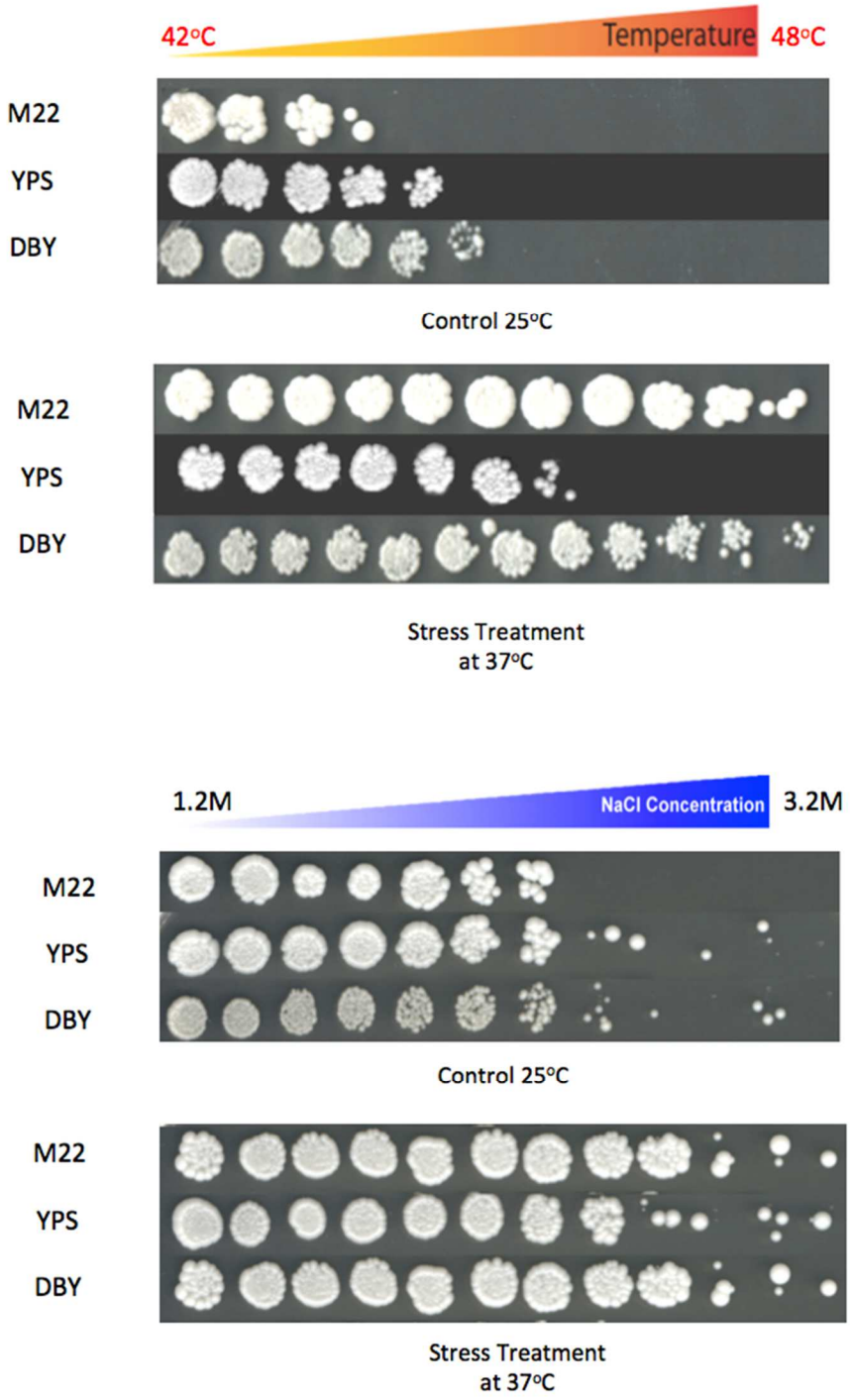


Figure 18. Acquisition of resistance in lab and wild strains using mild heat shock as primary and hydrogen peroxide as secondary stress

Figure 19. Acquisition of resistance to heat and salt stress in lab and wild strains after exposure to mild heat shock at 37ºC



Finally, the natural variation in the basal resistances and degree of acquisition was compared between lab and wild strains of yeast as shown in Figure 19. In the absence of any prior stress treatment all strains survived temperatures ranging from 42ºC to 45ºC with no growth on higher temperatures. However, strains displayed natural variation in basal resistances with the lab strain surviving slightly higher temperatures than the wild strains. All strains displayed significant acquisition of resistance after being pretreated to mild heat stress at 37ºC and survived temperatures of up to 48ºC. Natural variation was also observed in the degree of acquisition among the strains with the vineyard strain M22 and lab strain DBY displaying greater acquisition than the oak strain YPS.

While not as prominent as the natural variation in the basal resistance of the different strains to temperature, some variation was also observed in the basal resistance of strains to salt stress. Acquisition of resistance to salt stress after mild heat shock was also observed in all strains with strains surviving salt concentrations of up to 2.8M after pretreatment at 37ºC compared to 2.2M in the absence of prior heat treatment.

2.3 MATERIALS & METHODS

2.3.1 Strains and standard growth conditions

- Yeast strains used included BY (S288C), M22 (wild vineyard strain), and YPS (wild oak strain)
- Yeast strains were grown at 30°C with 270rpm shaking in YPD media unless stated otherwise (2X YPD media was used for yeast transformations)
- YPD media was made as follows: 10g Yeast Extract and 20g Peptone were dissolved in water and volume made up to 900ml. Solution was autoclaved and 100ml of sterile 20% glucose was added after autoclaving

- *E.Coli* strains (Dh5α-λ pir) were grown at 37°C with 270rpm shaking in Luria-Bertani broth (LB) unless stated otherwise
- LB media was made as follows: 5g Yeast Extract, 10g Tryptone and 5g NaCl was dissolved in water to a volume of 1 litre and autoclaved
- For experiments requiring antibiotic selection following concentrations of antibiotics were used unless stated otherwise:
  - Ampicillin: 100ug/ml
  - Kanamycin: 50ug/ml
  - Nourseothricin: 100ug/ml
- For blue white screening 40ul of X-Gal stock solution (20mg/ml in DMSO) and 40ul of 0.1M IPTG was spread plated on LB-Amp plates

2.3.2 Western blotting

- Cells were exposed to relevant stress, flash frozen in liquid nitrogen & stored at -80° C
- Cells were thawed on ice and previously recorded OD_{600} was used to determine the volume of buffer needed to suspend cells. 100µl of sample buffer (0.06M Tris-HCl pH6.8, 10% glycerol, 2%SDS, 5% 2-mercaptoethanol, 0.025% bromophenol blue) containing the protease inhibitor cocktail was used for every 1.05 OD units
- 150ul of cells were transferred to 1.7ml tubes containing 100µl of glass beads and shaken at 1500rpm at 95° for 10 min with vortexing at intervals to ensure efficient lysis.
- Samples were spun down and 15ul of supernatant was loaded on gel for SDS PAGE
- Gel was equilibrated in cold transfer buffer (3.03g Tris, 14.4g Glycine, 200ml methanol, in1L water) for 10 min

- Transfer was conducted at 100V for 1hour at 4°C
- Nitrocellulose membrane was incubated with 10ml blocking solution for 60min at 4°C with constant shaking (5% milk in TBST)
- Primary antibody solution was prepared as follows:
    - 10ml of 5% milk in TBST
    - 10ul of mouse anti acetyl lysine antibody (1:1000 dilution; Cell Signalling)
    - 2ul of rabbit anti actin antibody (1:5000 dilution, GeneTex)
- Membrane was incubated with antibody solution with gentle shaking at 4°C overnight
- Membrane was washed with 10ml TBST for 5 minutes with constant shaking at room temperature for a total of 4 washes
- Secondary antibody solution was prepared as follows:
    - 10ml of 5% milk in TBST
    - 2ul of donkey anti-mouse 800 antibody (1:5000 dilution; LiCor)
    - 2ul of donkey anti-rabbit 680 antibody (1:5000 dilution, LiCor)
- Membrane was incubated in secondary antibody for 1hr at room temperature with shaking
- Membrane was washed with 10ml TBST for 5 minutes with constant shaking at room temperature for a total of 4 washes and scanned

DISCUSSION & FUTURE RESEARCH

One of the main challenges in biology today is the characterization of the millions of genes of unknown function identified in sequencing studies. The rapidly declining cost of sequencing coupled with the increasing use of sequencing based techniques has resulted in the generation of tremendous amounts of genomic data. The development of increasingly sophisticated techniques for the functional analysis of this data is therefore essential to keep pace with the advancements in sequencing technology.

This study has presented two strategies for adapting Rb-TnSeq in yeast using either a transposome based or plasmid based approach. While this study was not able to obtain transposition in yeast using a transposome based approach, further work on this technique can potentially yield better results. The main challenge faced in optimizing this technique in yeast was the purification of active Tn5 transposase enzyme. Since the purified enzyme did not display significant activity it was not possible to optimize the amounts of enzyme, transposon DNA and transformation protocols necessary for library construction in yeast. Similarly, the cost of commercial Tn5 transposase coupled with the low electroporation efficiency of the tested strains made the validation of these protocols or the construction of multiple libraries using the commercial enzyme impractical. However, this study was able to demonstrate that it is possible to increase the electroporation efficiency of yeast strains above the miniumum threshold required for transposition by optimizing the electroporation protocols and using appropriate antibiotic or auxotrophic markers. Future work on optimizing the overexpression and purification techniques for producing active transposase can therefore make this a viable technique for library construction in yeast.

This study was also able to demonstrate that it is possible to obtain transposition on yeast using a plasmid-based approach. Further work, however, is required to increase the efficiency of this approach to levels sufficient for library construction. Future research comparing the transposition efficiency of the different plasmids constructed in this study and optimization of the transformation protocols, and the amount of transposon and transposase can potentially maximize transposition in yeast to levels sufficient for library construction.

The adaptation of Rb-TnSeq in yeast using either of the approaches described above offers a number of advantages over traditional transposon sequencing techniques allowing the testing of mutant fitness on a much larger scale and across a range of experimental conditions. While transposon sequencing techniques have been utilized extensively in bacterial studies, the application of these techniques to a eukaryotic model organism like budding yeast can be beneficial in many areas of future research. One example of the potential use of this technique is in the study of intra species natural variation. As previously described a major obstacle in the study of natural variation between different strains of yeast on a genome wide level is the lack of deletion libraries for wild strains. The ease of library preparation using the techniques described in this study can therefore allow the construction of mutant libraries for a variety of strains and simultaneous analysis of the behavior of these strains under a range of experimental conditions. Such studies can provide novel insight into the genetic differences among these strains and the mechanisms by which environmental conditions produce natural variation within a species. This will also shed light on the pitfalls of using a single model strain as representative of a species and allow more extensive studies of less commonly used strains. Study of eukaryotic cellular stress is another area of research that can benefit from this approach. While the stress response of eukaryotic model organisms such as yeast has been extensively studied and distinct stress

induced phenotypes related to specific stressors can be easily identified, isolating the genetic underpinnings of these phenotypes can prove to be much more challenging. The transposon sequencing approach provides a quick method for validating experimental findings by correlating these phenotypes with specific genetic regions. Furthermore, the use of the same library across a range of stressors can allow the analysis of genes involved in response to each stress and also shed light on the genetic interactions that permit one stress to cross protect against another.

It has also been suggested that transposon sequencing approaches can be modified to produce multiples insertions within a mutant as a means of studying genetic interactions. Similarly, it has been proposed that transposon mutagenesis can be a useful tool for the study of regulatory regions. Techniques that simplify the construction of barcoded mutant libraries can therefore prove to be extremely useful for the development of such tools. Finally, the optimization of these library construction techniques in yeast can pave the way for the development of similar techniques in higher eukaryotic model organisms and allow mutational analysis of clinically important phenotypes.

LITERATURE CITED

1. Benatuil, L., et al. "An Improved Yeast Transformation Method for the Generation of Very Large Human Antibody Libraries." Protein Eng Des Sel 23.4 (2010): 155-9. Print.

2. Castanera, R., et al. "Transposable Elements Versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles." PLoS Genet 12.6 (2016): e1006108. Print.

3. Giaever, G., and C. Nislow. "The Yeast Deletion Collection: A Decade of Functional Genomics." Genetics 197.2 (2014): 451-65. Print.

4. Goryshin, Igor Y., et al. "Insertional Transposon Mutagenesis by Electroporation of Released Tn5transposition Complexes." Nature Biotechnology 18.1 (2000): 97-100. Print.

5. Goryshin, Igor Yu, and William S. Reznikoff. "Tn5 in Vitro Transposition." (1998). Print.

6. Han, Jeffrey S. "Non-Long Terminal Repeat (Non-Ltr) Retrotransposons: Mechanisms, Recent Developments, and Unanswered Questions." Mobile DNA 1.1 (2010): 15. Print.

7. Havecker, Ericka R, Xiang Gao, and Daniel F Voytas. "The Diversity of Ltr Retrotransposons." Genome Biology 5.6 (2004): 225. Print.

8. Kachroo, A. H., et al. "Evolution. Systematic Humanization of Yeast Genes Reveals Conserved Functions and Genetic Modularity." Science 348.6237 (2015): 921-5. Print.

9. Kvitek, D. J., J. L. Will, and A. P. Gasch. "Variations in Stress Sensitivity and Genomic Expression in Diverse S. Cerevisiae Isolates." PLoS Genet 4.10 (2008): e1000223. Print.

10. MartÌnez-GarcÌa, Esteban, et al. "Pbam1: An All-Synthetic Genetic Tool for Analysis and Construction of Complex Bacterial Phenotypes." BMC Microbiology 11.1 (2011): 38. Print.

11. Munoz-Lopez, M., and J. L. Garcia-Perez. "Dna Transposons: Nature and Applications in Genomics." Curr Genomics 11.2 (2010): 115-28. Print.

12. Opijnen, Tim van, and Andrew Camilli. "Transposon Insertion Sequencing: A New Tool for Systems-Level Analysis of Microorganisms." Nature Reviews Microbiology 11 (2013): 435-42. Print.
13. Picelli, Simone, et al. "Tn5 Transposase and Tagmentation Procedures for Massively Scaled Sequencing Projects." (2014). Print.

14. Reznikoff, W. S. "The Tn5 Transposon." Annu Rev Microbiol 47 (1993): 945-63. Print.

15. Reznikoff, W. S., and K. M. Winterberg. "Transposon-Based Strategies for the Identification of Essential Bacterial Genes." Methods Mol Biol 416 (2008): 13-26. Print.

16. Reznikoff, William S. "Transposon Tn5." http://dx.doi.org/10.1146/annurev.genet.42.110807.091656 (2008). Print.

17. Reznikoff, William S., and University of Wisconsin Department of Biochemistry, Madison, WI 53706, USA. "Tn5 as a Model for Understanding Dna Transposition." Molecular Microbiology 47.5 (2017): 1199-206. Print.

18. Wetmore, Kelly M., et al. "Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons." (2015). Print.

19. Young, M. J., and D. A. Court. "Effects of the S288c Genetic Background and Common Auxotrophic Markers on Mitochondrial Dna Function in Saccharomyces Cerevisiae." Yeast 25.12 (2008): 903-12. Print.

20. Zhang, A. L., et al. "Recent Advances on the Gap Promoter Derived Expression System of Pichia Pastoris." Mol Biol Rep 36.6 (2009): 1611-9. Print.