Jasmin Elgin Fischer, BSc

Development of new Methods for reliable cDNA Library Generation and Expression

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieurin

Master's degree programme: Biotechnology

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder

Institute of Molecular Biotechnology

Faculty of Technical Chemistry, Chemical and Process Engineering, Biotechnology

Graz, August 2016

TABLE OF CONTENTS

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als Solche gekennzeichnet habe.

Graz, August 2016 …….

Unterschrift

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, August 2016 …….

signature

DANKSAGUNG

I want to thank Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder for the supervision and the possibility to make this thesis in his working group.

I also want to thank Lukas Sturmberger for the awesome supervision, much support and for always being there.

I would also like to thank acib and Clariant for giving me the possibility to write this thesis on a very interesting project.

Also a big “thank you” to the whole Gliedergroup for the nice working atmosphere, helpfulness and funny lunch breaks and Dönerstage. You are awesome and I am glad to have you as colleagues!

This work was supported by Clariant and the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (BMVIT), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT – Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

In addition, the work for vector design and construction has been supported by the IMI cofunded CHEM21 project (Chemical Manufacturing Methods for the 21st Century Pharmaceutical Industries). CHEM21 has received funding from the Innovative Medicines Initiative Joint Undertaking under grant agreement n°115360, resources of which are composed of financial contribution from the European Union’s framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution.

Last but not least I want to thank my parents, Hans-Peter and Lilia Fischer, my godparents Wolfgang and Eleonore Steffel and also my partner Daniel for their support during my studies.

ZUSAMMENFASSUNG

cDNA (complementary DNA) Bibliotheken sind eine viel verwendete Methode, um Informationen über transkribierte Gene von Organismen oder ganzer Umweltproben zu einem bestimmten Zeitpunkt oder unter bestimmten Bedingungen zu erlangen. In dieser Arbeit wurde eine Methode entwickelt, um cDNA unkompliziert generieren und in heterologen Organismen exprimieren zu können. Dieser Ansatz beruht auf der Isolierung der bereits transkribierten messenger-RNA in einer Probe, die anschließend von einem Enzym, der reversen Transkriptase, in cDNA übersetzt wird.

Die Technik für die reverse Transkription, die hier optimiert wurde, ist die des „template switching“. Mit dieser Methode können cDNA Moleküle generiert werden, die der vollen Länge der transkribierten Gene entsprechen. Nach der Etablierung dieser Methode wurde zusätzlich ein Plasmid-System entwickelt, um diese cDNA in den zwei Hefe-Spezies *Pichia pastoris* und *Schizosaccharomyces pombe* exprimieren zu können. Die Entwicklung dieses Systems bestand aus der Wahl der Promotoren, Selektionsmarkern und Art der Erhaltung der Expressionskassette in den Zellen. Es wurden verschiedene Promotoren für die Genexpression in beiden Hefen getestet, von welchen sich zwei aus *P. pastoris*, P_{CAT1} und P_{ENO2} als besonders geeignet für beide Spezies herausstellten. Auch wurde die autonom replizierende Sequenz aus *S. pombe* in beiden Spezies getestet, um eine Integration der Kassette in das Genom des Expressionsorganismus zu umgehen. Diese autonom replizierende Sequenz lieferte vor Allem in *P. pastoris* mit einem Fluoreszenzprotein und einer Lipase sehr uniforme Expressionslevels und wurde daher als geeignet für die Expression von cDNA Bibliotheken in beiden Hefearten befunden.

ABSTRACT

cDNA (complementary DNA) libraries are a widely used tool to obtain information about transcriptomes of an organism or whole environmental samples at certain conditions. During this thesis a method for reliable and reproducible cDNA library generation and subsequent expression in heterologous organisms was developed. This technique relies on the isolation of already transcribed messenger-RNA and subsequent transcription into cDNA by an enzyme called reverse transcriptase.

As eukaryotic mRNA was the source for the cDNA, an appropriate system for heterologous library expression in two different yeast species, budding yeast *Pichia pastoris* and fission yeast *Schizosaccharomyces pombe*, was established. The development of the expression vector included promoter and selection marker analysis in both yeast species as well as the type of maintenance of the expression cassette in the cells. During the analyses two promoters from *P. pastoris,* P_{CAT1} and P_{ENO2} turned out to be suitable for both species as well as the selection with zeocin. Furthermore, the autonomously replicating sequence from *S. pombe* was tested in both species to circumvent integration of the expression cassette into the hosts genome. Especially in *P. pastoris* uniform expression levels of a reporter protein and a lipase could be obtained by inclusion of this ARS element in the expression vector and was therefore found as a suitable method for cDNA library expression in both yeast species.

1 INTRODUCTION

An integral part of industrial biotechnology nowadays is the discovery of novel enzymes and proteins for catalyzing reactions which are otherwise challenging to produce with classical protein engineering. These biocatalysts can either be used in the medical or pharmaceutical sector or more broadly speaking for industrially relevant processes. The probably most difficult part in enzyme discovery is where to start searching.

In general, there are two different approaches to identify novel enzymes based on either sequence similarity to known enzymes or by screening whole libraries for desired functions when sequence information is not available. The first approach mainly relies bioinformatics tools with which certain properties and catalyzed reactions of an enzyme are predicted based on sequence or structural similarities to known enzymes. On the other hand, there is the functional approach. Here, enzymes or whole cells are directly screened for their functionality towards a certain reaction.

Functional screenings use genomic or cDNA libraries, which are produced from gDNA or mRNA, respectively, and subsequently expressed to screen for enzyme activities or protein interactions [1,2]. Genomic libraries are a useful tool to gather sequence information of an organisms genome but are only to a limited extent useful for functional screenings because they yield unspliced genes and hence mostly used for bacterial gDNA [3,4]. If the expression organism employs different splicing signals compared the original organism, this can result in incorrectly spliced mRNA and subsequently disfunctional proteins.

Complementary DNA libraries, called cDNA libraries, are a useful tool to explore the transcriptome of an organism or whole environmental samples (metatranscriptome) [5]. In comparison to a genomic DNA library they lack information about introns and untranslated regions, but provide exact information about the transcribed genes at certain conditions.

cDNA Generation

cDNA is generated from mRNA. This starts with the isolation of the whole RNA from a sample, followed by a primed reverse transcription by an enzyme called reverse transcriptase (RT) [6,7].

Key-enzyme reverse transcriptase

A reverse transcriptase is a RNA-dependent DNA polymerase that synthesizes the complementary DNA strand to a single stranded RNA template in 3' to 5' direction, contrary to other enzymes like DNA or RNA polymerases which synthesize 5' to 3'. Reverse transcriptases have their origin in retroviruses but also exist in other viruses like hepatitis B [7,8]. The retroviruses use this enzyme to convert their single stranded RNA genome to double stranded cDNA, followed by integration into the hosts genome [9]. This ability of the enzyme is nowadays used for *in vitro* reverse transcription of RNA. Different reverse transcriptases originating from different viruses are commercially available for cDNA generation. The most established ones are Avian Myeloblastosis Virus (AMV) reverse transcriptase and Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase [10]. A variation of the M-MLV RT is the RNAse H- point mutant. M-MLV has an internal RNase H activity to degrade the RNA template after finishing the so called first strand synthesis to enable the second strand synthesis. For *in vitro* cDNA library generation, where full-length transcripts are required, this activity is undesirable because occasional internal digestion of RNA template is resulting in shorter, prematurely terminated transcripts. By a point mutation in this RNAse H domain, this activity is inhibited [11–13].

Methods of cDNA generation – state of the art

For synthesizing cDNA from a mRNA template, different methods exist. In the following I want to give a short overview of the commonly used techniques.

cDNA generation using paramagnetic beads and therefore binding cDNA to a solid surface makes use of immobilizing oligo(dT) primers to a solid surface being able to bind polyadenylated RNA. Subsequent a reverse transcription is done, followed by a PCR reaction for amplification of the cDNAs second strand [14, 15]. This method is useful especially if the amount of sample RNA as starting material is limited [16]. The mechanism is schematically shown in Figure 1.1.

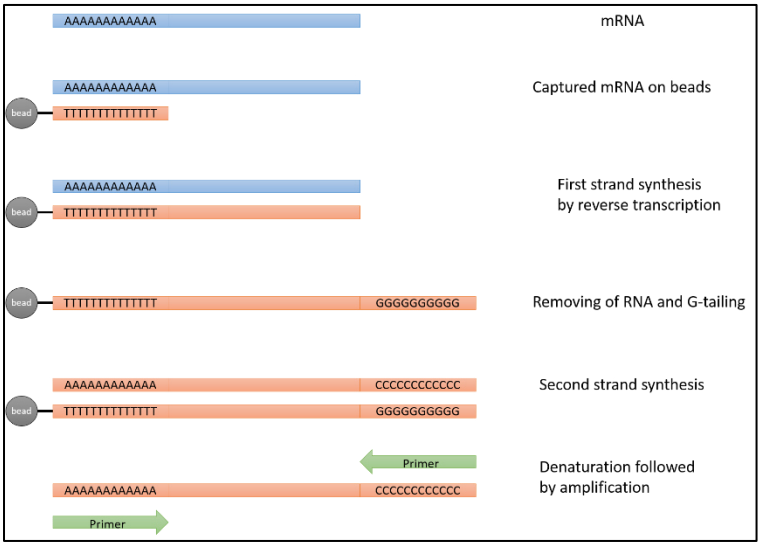


Figure 1.1 Schematic representation of the cDNA generation with paramagnetic beads.

The mRNA is captured by a paramagnetic bead on the one side bound to a magnetic matrix and on the other side carrying a poly-d(T) oligonucleotide which binds to the poly-A tail of the mRNA. By reverse transcription the first strand of the cDNA is synthesized, followed by RNase digest and G-tailing. The poly-T and poly-G overhang are then used as anchor points for second strand synthesis after being removed from the beads.

Another commonly used method represents adapter ligation to the 5'-end of the cDNAs first strand. The mechanism is shown in Figure 1.2.

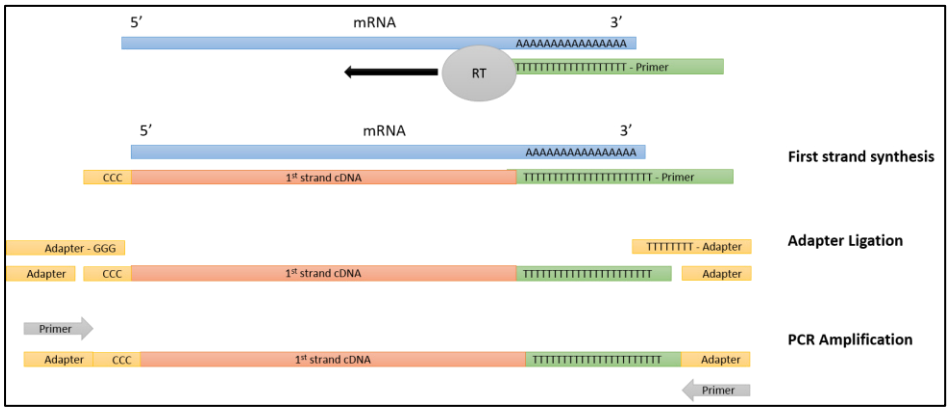


Figure 1.2 Schematic representation of cDNA generation employing adapter ligation to the ends of the first strand cDNA.

After capturing mRNA by an oligo-d(T) primer and reverse transcription including addition of the non-templated cytosine residues to the 3' end, mRNA is digested and adapters containing for example restriction sites are ligated to both ends of the first strand. These adapters are then used as anchor points for PCR amplification.

First strand cDNA synthesis is carried out by oligo(dT) primed reverse transcription. M-MLV reverse transcriptase has the property to add three non-templated ribonucleotide cytosins at the 5'-end of the first strand cDNA by its terminal transferase activity [17–19]. This property is used then for ligation of a double stranded adapter, which has a GGG-overhang. This adapter then serves as primer binding site in addition to the poly-dT stretch for second strand cDNA synthesis.

Another commonly used method employs the template-switching mechanism [20].

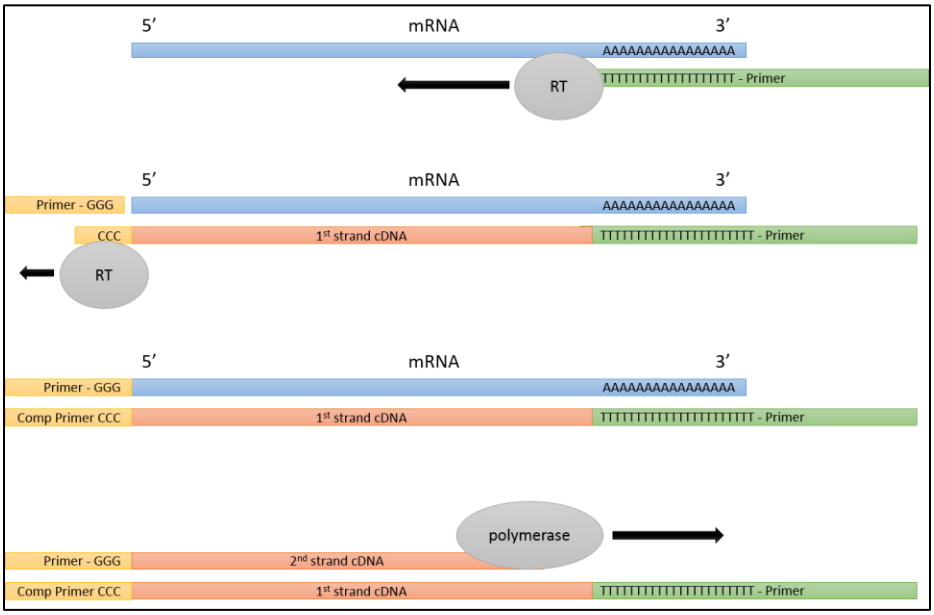


Figure 1.3 Schematic representation of the template switching mechanism.
This is employing a reverse transcriptase for first strand synthesis primed by an poly-d(T) primer as well as a template switching oligonucleotide. Second strand synthesis can then be carried out by any high fidelity polymerase.

As described above, the CCC-addition on the 5' end of the first strand cDNA serves here as a template for a so-called "template switching oligonucleotide", a primer which has three ribonucleotide guanines at its 3'-end. These guanines serve as an enhanced template for the reverse transcriptase that subsequently adds the complement to the rest of the primer attached to the first strand cDNA. This enables the possibility to add vector overhangs or restriction sites on both ends of the newly synthesized cDNA, which allows more cloning efficiency by avoiding application of blunt-end restriction enzymes. This method is described as very effective, fast and

capable to yield full length cDNAs [20]. For that reason, this method was the technique of choice for this thesis.

After successful synthesis of cDNA and cloning into vectors, there is the need to evaluate the quality generated cDNA library. This can be done by sequencing the inserts, or directly screening the library for expression of enzymes or proteins. The use of a suitable vector for library expression is an essential part for successful cDNA library screening. Hence, the second part of this thesis was dedicated to develop these expression vectors for screening in eukaryotic hosts. For this purpose, I chose the yeast species *Pichia pastoris* and *Schizosaccharomyces pombe.*

Expression system development for fission and budding yeasts

For screening of cDNA libraries, different hosts can be used. Prokaryotic hosts like *E. coli,* which is rather easy to cultivate, fast growing and well established. However, depending on the source of the cDNA prokaryotic expression hosts might lack machinery for posttranslational modifications or employ different codon usage if the cDNA is derived from a eukaryotic organism. On the other hand, eukaryotic hosts such as *P. pastoris* and *S. pombe* offer the existence of posttranslational modification machineries, as well as the possibility to secrete recombinant proteins. *P. pastoris* is a well-known, methylotrophic buddying yeast offering simple cultivation and the usage of strong and well established promoter systems for heterologous protein expression [21,22]. Thousands of heterologous proteins were already produced in this buddying yeast. *P. pastoris* which was originally used for single-cell protein production for animal feedstocks by the Philipps Petroleum Company who started to cultivate the yeast on methanol yielding high cell densities up to 130 g/L CDW [23]. At the beginning of the 1980s, the *AOX1* gene and its promoter were isolated from *P. pastoris,* SIBIA and Philipps Petroleum company developed an expression system, transformation methods and strains to have a heterologous expression system that is easy to manipulate and cultivate [24–26].

Nevertheless, also the application of *P. pastoris* has its limits concerning posttranslational modifications of heterologous proteins and splicing. As the goal of this thesis was to screen cDNA libraries obtained from fungi or plant RNA, the splicing machinery should be as similar to the hosts organism as possible to obtain correctly spliced mRNA and functional proteins. Buddying yeasts

like *Saccharomyces cerevisiae* or *Pichia pastoris* often have problems with correct splicing of foreign proteins [27,28]. *Schizosaccharomyces pombe* has a sophisticated splicing machinery that is more similar to that of higher eukaryotes [29, 30].

S. pombe was originally isolated from an east African millet beer in the 1890s and first described in 1893. It has a rod-shaped morphology with 7-15 µm length and 4 µm in diameter [30]. Its genome harbors 12.57 Mbp in 5125 genes on three chromosomes [31].

It's introns are, like that of higher plants, shorter and the whole splicing machinery shows high similarity to higher plants in terms of sequence and structure [32]. Also, *S. pombe* has the advantages of monocellular eukaryotic organisms, such as short cell division times and the availability of tools for straightforward genetic manipulation.

Hence *S. pombe* was the second chosen organism for library screening in this thesis. In general, genomic integration of expression cassettes in *S. pombe* and *P. pastoris* as well as episomal heterologous protein expression is possible in these yeast species. The integration of expression cassettes into the host organisms genome offers more stable maintenance of genetic material compared to episomal propagation. It therefore enables heterologous protein expression without selection pressure in minimal as well as complex media. Genomic integration can either be random or locus-targeted by employing homologous sequences to the host's genome which are incorporated on the plasmid DNA in addition to the expression cassette, as well as a selection marker for identification of successfully transformed clones. In *S. pombe* the *LEU2* or *URA4* locus are widely used for targeted integration, in *P. pastoris AOX1* or *HIS4* are used [33,34]. Figure 1.4 shows an example of such an integrated cassette.

Figure 1.4 Example of an expression cassette for genomic integration in P. pastoris.
The cassette contains the coding sequence of the gene of interest flanked by promoter and terminator, a selection marker expression cassette as well as a homologous region to the hosts genome for targeted integration, in this example the 3'AOX1 untranslated region.

Also ectopic integration of expression cassettes is used where the cassette integrates randomly into the genome. This approach is often used for heterologous protein expression to identify the best expression clone of a protein or enzyme due to positioning effects, highly transcribed regions or activation or inactivation of genes. For library screenings the scope is not to find the best expressing clone, but to gather uniform expression that guarantees the identification of variants of proteins or enzymes with higher activity and therefore allows the screening of variant libraries as employed in protein engineering. Such uniform expression is usually achieved by site-specific integration by homologous recombination.

The episomal approach, where the vector is located extrachromosomal as an autonomously replicating plasmid, is a good alternative to the conventional genomic integration. The sequence elements responsible for extrachromosomal maintenance of plasmids, autonomously replicating sequence elements, were identified in the buddying yeast *Saccharomyces cerevisiae* in 1980 [35]. Furthermore, the ARS element from *S. pombe* was characterized 1995 by Clyne and colleagues [36]. Protein expression from episomal plasmids shows a similar uniformity as site-specific integration, as it was shown in this thesis, without the need to linearize the vector and eventually destroy cDNA fragments when screening a library because of unknown restriction sites within the fragment.

Figure 1.5 shows an example of an episomal expression vector as it was developed and used during this thesis.

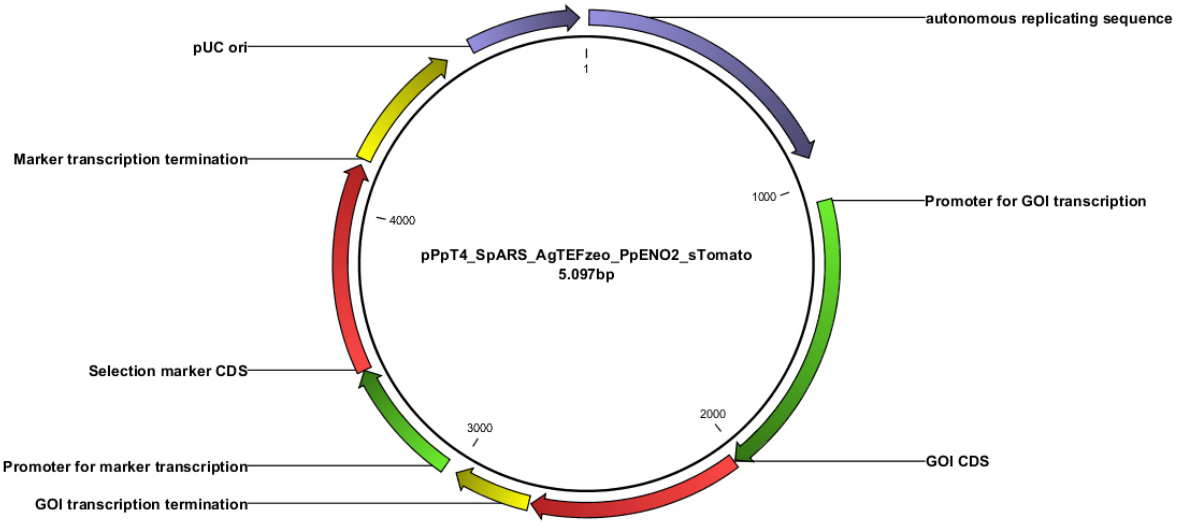


Figure 1.5 Plasmid map of an episomal expression vector for yeast.
It contains an expression cassette for a gene of interest as well as one for selection marker expression. In addition, it contains origins of replication for E. coli as well as for yeast for extrachromosomal propagation.

Shuttle vectors: Using the physically same vector for different species

In the 1980 shuttle vectors were developed to be able to transfer DNA between yeast (*S. cerevisiae)* and bacteria (*E. coli)*, to enable cloning and subsequent amplification of the plasmid in *E. coli* followed by transfer to a yeast expression system for heterologous protein expression [37,38]. Bacteria-yeast shuttle vectors are now state-of-the-art and commonly applied.

Shuttling vectors not only between bacteria and yeast but also between different yeast species are a useful tool for determining the best host system for expression of proteins in terms of posttranslational modifications, codon usage or splicing events. To have a functional vector in various yeast species, all parts of the vector have to be adapted. Such new multi-host vectors as they were also developed by Steinborn et al. 2006 [39] should allow to evaluate heterologous protein or biosynthetic pathway expression in different yeasts by single genetic construct.

Choosing the right vector parts

As it is described in the section above, different parts have to be integrated into an expression vector to enable heterologous protein expression. The promoter, which enables the transcription

of plasmid encoded genes plays an important role for the efficiency of the expression. Promoters are DNA sequences that are located upstream of the genes to recruit transcription factors and polymerases to transcribe the DNA into RNA. Promoter sequences show 5-20 bp long transcription factor binding sites and are usually 100-1000 bp in size [40].

The two promoters for protein expression used in this thesis were P_{CAT1}, the promoter of *P. pastoris* catalase 1 gene (also known as *CTA1*[41]), and P_{ENO2}, promoter of enolase 2 gene. The *CAT1* promoter was characterized by Vogl et al., 2015. The catalase 1 protein is part of the ROS defense mechanism in *P. pastoris* and located in the peroxisome. The *CAT1* promoter is repressed by glucose and glycerol, and the downstream located gene is actively transcribed if the carbon source is depleted and the cells get stressed [42]. P_{ENO2} is active constitutively in *P. pastoris*, especially when grown on glucose [43]. These two promoters of different strengths were then used to characterize the library expression vector for both yeast species *P. pastoris* and *S. pombe.*

Also the selection marker had to be functional in both species, here the marker of choice was phleomycin resistance encoded by the Sh *ble* gene from *Streptoalloteichus hindustanus*. The antibiotic zeocin can be used in *P. pastoris* as well as in *S. pombe* [44–47].

Aim of this thesis

The aim of this thesis was to develop a simple, fast and reliable method for cDNA library generation, cloning and an appropriate expression system for functional library screenings. The developed expression vector should be functional in two different yeast species, *S. pombe* and *P. pastoris*, to enable screening in both species without modifications of the library.

2 MATERIAL AND METHODS

2.1 DEVICES & EQUIPMENT

Table 2.1 All Devices used during this thesis including the manufacturer.

Device	Manufacturer
96 well PS Microplater sterile	Greiner Bio-One GmbH, Frickenhausen, D
96 well PS Microplater unsterile	Greiner Bio-One GmbH, Frickenhausen, D
Applied Biosystems 2720 Thermal Cycler	Applied Biosystems Inc., Foster City, USA
Bel-Art 96-Well Deep Well Plates	Bel-Art Products, Wayne, NJ, USA
Biohit Proline 50-1200 µL	Biohit Oyj, Helsinki, FI
Biohit Proline 5-100µL	Biohit Oyj, Helsinki, FI
Bio-Rad BioRad Gene Pulser 1652076	Bio-Rad Laboratories Inc., Hercules, CA, USA
BioRad PowerPac Basic Power Supply	Bio-Rad Laboratories Inc., Hercules, CA, USA
Certoclav LVEL 12L	CertoClav GmbH, Traun, AT
Eppendorf Centrifuge 5415D	Eppendorf AG, Hamburg, D
Eppendorf Centrifuge 5415R	Eppendorf AG, Hamburg, D
Eppendorf Centrifuge 5810R	Eppendorf AG, Hamburg, D
Eppendorf Research plus 0.1-2.5 µL	Eppendorf AG, Hamburg, D
Eppendorf Thermomixer Comfort	Eppendorf AG, Hamburg, D
Gilson PIPETMAN Classic P1000	Gilson Inc., Middletown, WI, USA
Gilson PIPETMAN Classic P20	Gilson Inc., Middletown, WI, USA
Gilson PIPETMAN Classic P200	Gilson Inc., Middletown, WI, USA
Heidolph MR 2002 Mangentic Stirrer	Heidolph Instruments, Schwabach, D
Heidolph MR 3000 Magnetic Stirrer	Heidolph Instruments, Schwabach, D
Heidolph Titramax 1000 plate shaker	Heidolph Instruments, Schwabach, D
HT Infors Multitron Shaker	Infors AG, Bottmingen, CH
HT Infors Orbitron shaker	Infors AG, Bottmingen, CH
HT Infors RS306 shaker	Infors AG, Bottmingen, CH
Ikamag RCT Magnetic Stirrer	IKA-Werke GmbH & Co. KG, Staufen, D
inoLab pH 720 pH-Meter	WTW GmbH, Weilheim, D
MicroPulser™ Electroporator	Bio-Rad Laboratories Inc., Hercules, CA, USA
MT PG12001-S DeltaRange Balance	Mettler Toledo Inc., Greifensee, CH
NanoDrop 2000 UV-Vis Spectrophotometer	Thermo Fisher Scientific Inc., Waltham, MA, USA
Nunc™ MicroWell™ 96-Well Optical-Bottom Plates with Polymer Base	Thermo Fisher Scientific Inc., Rochester, NY, USA

RVC 2-18 Rotational Vacuum Concentrator	Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, D
Sartorius B2 120S Analytical Lab Scale	Sartorius AG, Göttingen, D
SynergyMx Plate Reader	Biotek Inc., Winooski, USA
Vortex Genie 2	Scientific Industries Inc., Bohemia, NY, USA

2.2 CHEMICALS

Table 2.2 Chemicals used during this thesis.

Chemical	**Supplier**
4-Nitrophenyl butyrate (pNPB)	Sigma-Aldrich Chemie GmbH, Vienna, AT
4-Hydroxybenzoic acid hydrazide	Sigma-Aldrich Chemie GmbH, Vienna, AT
Adenosine triphosphate	New England Biolabs, Ipswich, MA, USA
Agar-Agar	Carl Roth GmbH, Karlsruhe, D
Ampicillin	Sigma-Aldrich Chemie GmbH, Vienna, AT
Aqua bidest. "Fresenius"	Fresenius Kabi Austria GmbH, Graz, AT
Avicel® PH-101	Sigma-Aldrich Chemie GmbH, Vienna, AT
BactoTM peptone	Becton, Dickinson and Company, Sparks, USA
BactoTM yeast extract	Becton, Dickinson and Company, Sparks, USA
Bicine	Fluka Chemia AG, Basel, CH
Biozym LE Agarose	Biozym Biotech Trading GmbH, Vienna, AT
Citric acid monohydrate	Carl Roth GmbH, Karlsruhe, D
D-biotin	Fluka Chemia AG, Basel, CH
D-glucose monohydrate	Carl Roth GmbH, Karlsruhe, D
D-sorbitol	Carl Roth GmbH, Karlsruhe, D
Deoxyadenosine triphosphate	Thermo Scientific – Austria GmbH, Vienna, AT
Deoxycytidine triphosphate	Thermo Scientific – Austria GmbH, Vienna, AT
Deoxyguanosine triphosphate	Thermo Scientific – Austria GmbH, Vienna, AT
Deoxythymidine triphosphate	Thermo Scientific – Austria GmbH, Vienna, AT
Di-potassium hydrogen phosphate	Carl Roth GmbH, Karlsruhe, D
DifcoTM yeast nitrogen base w/o amino acids	Becton, Dickinson and Company, Sparks, USA
Dimethyl sulphoxide	Carl Roth GmbH, Karlsruhe, D
Dithiothreitol	Carl Roth GmbH, Karlsruhe, D
Ethanol absolute	Merck KGaA, Darmstadt, D
Ethidium bromide (1%)	Carl Roth GmbH, Karlsruhe, D
Ethlenediaminetetraacetic acid	Carl Roth GmbH, Karlsruhe, D
Ethylene glycol	Sigma-Aldrich Chemie GmbH, Vienna, AT

Glycerol	Carl Roth GmbH, Karlsruhe, D
Hydrochloric acid (37%)	Merck KGaA, Darmstadt, D
Kanamycin	Carl Roth GmbH, Karlsruhe, D
LB-media	Carl Roth GmbH, Karlsruhe, D
Liquid nitrogen	Air Liquide Austria GmbH, Graz, AT
Magnesium chloride	Promega Madison, WI, USA
Methanol	Carl Roth GmbH, Karlsruhe, D
Potassium dihydrogen phosphate	Carl Roth GmbH, Karlsruhe, D
Potassium hydroxide	Carl Roth GmbH, Karlsruhe, D
Roti® phenol/chloroform/isoamylalcohol	Carl Roth GmbH, Karlsruhe, D
SOB-Medium	Carl Roth GmbH, Karlsruhe, D
Sodium chloride	Carl Roth GmbH, Karlsruhe, D
Sodium dodecyl sulphate	Carl Roth GmbH, Karlsruhe, D
Sodium hydroxide	Carl Roth GmbH, Karlsruhe, D
Tris	Carl Roth GmbH, Karlsruhe, D
Zeocin™	InvivoGen-Eubio, Vienna, AT

2.3 STRAINS

Table 2.3 Strains used for transformations in this thesis.

Organism	**Name**	**Properties**
E. coli	K12 Top10 F'	*E. coli* expression strain
P. pastoris	CBS 7435	Wildtype
P. pastoris	BG10ΔGUT1	*GUT1* knockout strain
S. pombe	972h-	Wildtype

2.4 PRIMERS & GBLOCKS

Table 2.4 Primers used in this thesis

#	Name	Sequence
P168	T4_SpARS_fwd	AAATACACAGTTATTATTCATTTAAATGACAGAATGGGATACAAGGGCATCG
P169	T4_SpARS_rev	CTCCTTAACCACTCGGACATAGTGA
P170	T4_pCAT1_fwd	ACTATGTCCGAGTGGTTAAGGAGAGATCTGAACTCCGAATGCGGTTCTCC
P171	T4_pCAT1_kozak_rev	ACCCTTAGAAACCATCGTTTCGAGAATTCCTAGTTTTAATTGTAAGTCTTGACTAGAGC
P172	T4_sTom_kozak_fwd	ATTAAAACTAGGAATTCTCGAAACGATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGT
P173	T4_sTom_rev	TGGCATTCTGACATCCTCTTGAGCGGCCGCTTACTTATAAAGCTCGTCCATACCGTACAAG
P194	T4_pCAT1_CalB_fwd	AGACTTACAATTAAAACTAGGAATTCTCGAAACGATGCTTCCATCTGGTTCTGATCCTG

P195	T4_pCAT1_alphaCalB_fwd	ACTTACAATTAAAACTAGGAATTCTCGAAACGATGAGATTCCCATCTATTTTCACCGCT
P196	T4_CalB_rev	ATTCTGACATCCTCTTGAGCGGCCGCTTATGGGGTCACGATACCGGA
P197	pCAT1_library rev_1	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTAG
P198	pCAT1_library rev_2	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTAA
P199	pCAT1_library rev_3	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTAC
P200	pCAT1_library rev_4	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTAT
P201	pCAT1_library rev_5	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTGA
P202	pCAT1_library rev_6	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTGG
P203	pCAT1_library rev_7	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTGC
P204	pCAT1_library rev_8	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTGT
P205	pCAT1_library rev_9	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTCA
P206	pCAT1_library rev_10	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTCG
P207	pCAT1_library rev_11	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTCT
P208	pCAT1_library rev_12	TTCTGACATCCTCTTGAGGCCGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTTTTCC
P218	pT4_Ptef1_fwd	AAGTGAGACCTTCGTTTGTGCGGATCCGATCTAACGACATGGAGGCCCA
P219	pT4_Ptef1_rev	GCAGAGGTGAGTTTAGCCATGGTTGTTTATGTTCGGATGTGATGTGAGA
P220	pT4_tef1_zeo_fwd	ACATCCGAACATAAACAACCATGGCTAAACTCACCTCTGCTGTTC
P221	pT4_tef1TT_rev	TCTGACGCTCAGTGGTACCTGCAGTGGATGGCGGCGTTAGTATCG
P222	pT4_TEF1_AOXTT_rev	TCTGGGCCTCCATGTCGTTAGATCGGATCCGCACAAACGAAGGTC
P223	pT4_SpeI_sTom	TTACAATTAAAACTAGACTAGTTCGAAACGATGGTTTCTAAGGGTGAGG
P226	pT4_agTEF1_zeo_fwd	AGGAAAAGACTCACGTTAACCTCGAGATGGCTAAACTCACCTCTGCTGTT
P229	pT4_CAT1_SpeI_rev	CGAACTAGTCTAGTTTTAATTGTAAGTCTTGACTAGAGCAAGTGTTATGG
P259	Ribo_pCAT1_fwd	CTCTAGTCAAGACTTACAATTAAAACTAGACTAGTTCGAAACG(GGG)
P272	seq_library_fwd	GCTCTAGTCAAGACTTACAATTAAAACTAGACTAGTTCG
P273	seq_library_rev	ATGGCATTCTGACATCCTCTTGAGC
P280	pT4_ARS_PpENO2-1000_fwd	TGTCCGAGTGGTTAAGGAGAGATCTATGAAAGAGTGAGAGGAAAGTACCTGGG
P281	pT4_ARS_PpENO2-1000_rev	AGAAACCATCGTTTCGAACTAGTTTTTAGATGTAGATTGTTATAATTGTGTGTTTCAAC
P300	GUT1_fwd	ATTAAAACTAGGAATTCTCGAAACGATGGGAAAAGACTATACACCACTAGTTGC
P301	GUT1_rev	TTCTGACATCCTCTTGAGCGGCCGCTTAAGCAGTGTCCTTAAGCCAGC
P315	T4_ampR_fwd	AAAAATACACAGTTATTATTCATTTTAAAAAAAATCCTTAGCTTTCGCTAAGGATTTAC
P316	T4_ampR_rev	ATGCCCTTGTATCCCATTCTGTCATTTCTAGAGTGTTGTTACTTTATACTTCCGGC

Table 2.5 Sequencing primers used in this thesis

#	Name	Sequence
P166	seq_pUC_fwd	CCT TTT GCT GGC CTT TTG CTC A
P174	seq_pCAT1_rev	CGAAAAGTCCAGGCCAATCGA

P175	seq_pCAT1_fwd	AAG CTT CTC GCA CGA GAC C
P272	seq_library_fwd	GCTCTAGTCAAGACTTACAATTAAAACTAGACTAGTTCG
P273	seq_library_rev	ATGGCATTCTGACATCCTCTTGAGC
P317	pUC_seq_rev	GCAGAGCGAGGTATGTAGGCG
C43	AOXTT_seq_rev	TCCCAAACCCCTACCACAAG
C173	pAOX_fw	AGATCTAACATCCAAAGACGAAAGGTTGAATGAAAC
C372	5AOX_seq	GACTGGTTCCAATTGACAAGC
C387	sTomato_seq_rev	ACCATCTGGTGGGAAGTTGGTAC

Table 2.6 gBlocks gene fragments used during this thesis.

Name	**Sequence**
AgTEF1_promoter	GACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCaCGCGGGTCACCCGGCCAGCGACATGGAGGCCCAGAT ACCCTCCTTGACAGTCTTGACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACAT CCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCGACGGCGCAAGCAAAAATTACGGCTCCTCGCTG CAGACCTGCGAGCAGGGAAACGCTCCCCTCAGCAACGCGTTGAATTCTCCCCACGGCGCGCCCCTGTAGAGAAA TATAAAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCA CTCACATCCGAACATAAACAAAAATGGGTAAGGAAAAGACTCACGTTAACCTCGAGATGGCTAAACTCACC
AgTEF1_terminator	AAGAGCAGGACTAAGTCGACAAGGGGTAAGAAATAGAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACT TGTCATTTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTGCCTCGAC ATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGTCAATCGTATGTGAATGCTGGTCGCTAT ACTGCTGTCGATTCGATACTAACGCCGCCCCAGTGTCTACCTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGAT AGAATATGCTCGAACTGTTGAAGTCCATCAACAAGGATAACCCATATGCTCTATCGGCGGAGAAAACGTTCAGA GCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCTAGATGAGAAGTACGGGGTTTAGTGTTTCCAGGCCTCGTA AATGCCGCAATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTCAAGGACCGC CTTTTCATATATGGCCACCATCAATATAGAGCCAGCGAACGATGATGCATTCTGGTATATTGAACGCGAGGATCC CTGCAGGTACCACTGAGCGTCA

2.5 ENZYMES

Table 2.7 Enzymes used in this thesis

Enzyme	**Type**	**Supplier**
FD *Bam*HI	RE	Thermo Scientific – Austria GmbH, Vienna, AT
FD *Bcu*I/*Spe*I	RE	Thermo Scientific – Austria GmbH, Vienna, AT
FD *Bgl*II	RE	Thermo Scientific – Austria GmbH, Vienna, AT
FD *Eco*RI	RE	Thermo Scientific – Austria GmbH, Vienna, AT
FD *Not*I	RE	Thermo Scientific – Austria GmbH, Vienna, AT
FD *Smi*I	RE	Thermo Scientific – Austria GmbH, Vienna, AT
FD *Swa*I	RE	Thermo Scientific – Austria GmbH, Vienna, AT
GoTaq DNA Polymerase	Polymerase	Promega GmbH, Mannheim, D

M-MLV H- reverse transcriptase	RT	Promega GmbH, Mannheim, D
M-MLV reverse transcriptase	RT	New England Biolabs, Ipswich, MA, USA
Phusion High Fidelity DNA Polymerase	Polymerase	Thermo Scientific – Austria GmbH, Vienna, AT
Protoscript II	RT	New England Biolabs, Ipswich, MA, USA
RNase H	RNase	New England Biolabs, Ipswich, MA, USA
*Sfi*I	RE	New England Biolabs, Ipswich, MA, USA
T5 Exonuclease	Nuclease	New England Biolabs, Ipswich, MA, USA
Taq DNA Ligase	Ligase	New England Biolabs, Ipswich, MA, USA

2.6 KITS

Table 2.8 Commercial available kits used in this thesis

Kit	**Supplier**
GeneJET Gel Extraction Kit	Thermo Scientific – Austria GmbH, Vienna, AT
GeneJET Plasmid Miniprep Kit	Thermo Scientific – Austria GmbH, Vienna, AT
SuperScript® III First-Strand Synthesis System	Thermo Scientific – Austria GmbH, Vienna, AT
Wizard® SV Gel and PCR Clean-Up System	Promega GmbH, Mannheim, D
CHROMA SPIN™+TE-1000 Columns	Takara Bio Europe, St-Germain-en-Laye, F
Plant/Fungi RNA Purification Kit	Norgen Biotek Corporation, Thorold, CA

2.7 DNA LADDERS & DYES

Table 2.9 Dyes and Ladders used in this thesis

Ladder/Dye	**Supplier**
DNA Gel Loading Dye (6X)	Thermo Scientific – Austria GmbH, Vienna, AT
DNA Loading Dye & SDS Solution (6X)	Thermo Scientific – Austria GmbH, Vienna, AT
GeneRuler 1 kb DNA Ladder	Thermo Scientific – Austria GmbH, Vienna, AT
GeneRuler DNA Ladder Mix	Thermo Scientific – Austria GmbH, Vienna, AT
Quick-Load® Purple 2-Log DNA Ladder (0.1 - 10.0 kb)	New England Biolabs, Ipswich, MA, USA

2.8 MEDIA AND SOLUTIONS

2.8.1 *E. coli* media

SOC (Super Optimal Broth + Glucose)

SOB	30.7 g/L	Autoclaved
Glucose	3.6 g/L	Autoclaved

2.8.2 Yeast media

YPD

Bacto yeast extract	10 g/L	Autoclaved
Bacto peptone	20 g/L	
glucose	1% (w/v)	Autoclaved

Bromphenole-blue YPD

Bacto yeast extract	10 g/L	Autoclaved
Bacto peptone	20 g/L	
glucose	1% (w/v)	Autoclaved
Bromphenole blue	0.2 g/L	Sterile filtered

ABTS/$CuSO_4$ YPD

Bacto yeast extract	10 g/L	Autoclaved
Bacto peptone	20 g/L	
glucose	1% (w/v)	Autoclaved
ABTS	0.1 g/L	Sterile filtered
$CuSO_4$	25 mg/L	Sterile filtered

Rhodamine B YPD

Bacto yeast extract	10 g/L	Autoclaved
Bacto peptone	20 g/L	
glucose	1% (w/v)	Autoclaved
Rhodamine B	10 mg/L	Sterile filtered
Olive oil	1.5% (v/v)	Added to hot medium

YES

Bacto yeast extract	5 g/L	Autoclaved
SP Supplements	1 g/L	
glucose	3% (w/v)	Autoclaved

BMG

Yeast nitrogen base w/o aa	13.4 g/L	Autoclaved
1 M potassium phosphate buffer pH 6	200 mL/L	Autoclaved
Biotin	0.4 mg/L	Sterile filtered
Glycerol	1% (w/v)	Autoclaved

PMG

Potassium phthallate	3 g/L	Autoclaved
Na_2HPO_4	2.2 g/L	
L-glutamic acid	3.26 g/L	
50x salts	20 mL/L	Autoclaved
1000x vitamins	1 mL/L	Sterile filtered
10000x minerals	0.1 mL/L	Sterile filtered
Glucose	2% (w/v)	Autoclaved

SLD

Yeast nitrogen base	0.67% (w/v)	Autoclaved
Glucose	0.5% (w/v)	Autoclaved

2.8.3 Solutions

1 M potassium phosphate buffer (pH 6)

K_2HPO_4	30 g/L
KH_2PO_4	118 g/L

pH was adjusted with KOH and solution autoclaved.

50x salts

$MgCl_2*6H_2O$	260 mM
$CaCl_2*2H_2O$	5 mM
KCl	670 mM
Na_2SO_4	14 mM

Solution was autoclaved.

1000x vitamins

Panthotenic acid	4.2 mM
Nicotinic acid	81.2 mM
Inositol	55.5 mM
Biotin	40.8 µM

Solution was sterile filtered.

10000x minerals

Boric acid	80.9 mM
$MnSO_4$	23.7 mM
$ZnSO_4$	13.9 mM
$FeCl_2$	7.4 mM
Molybdic acid	2.47 mM
KI	6.02 mM
$CuSO_4$	1.6 mM
Citric acid	47.6 mM

Solution was sterile filtered.

5X RT buffer (pH 8.3)

Tris	250 mM
$MgCl_2$	30 mM
KCl	375 mM
DTT	10 mM

pH was adjusted with 37% HCl and solution was sterile filtered.

10X TE buffer

Tris	100 mM
EDTA	10 mM

Solution was autoclaved.

2.9 PROTOCOLS

2.9.1 Agarose gel electrophoresis

For separating DNA fragments by size an agarose gel electrophoresis was performed with 1% w/w agarose in 1xTAE buffer (40 mM Tris; 20 mM acetate; 1 mM EDTA; pH 8) at 100-120 V with 0.5 µg/mL ethidium bromide for visualization of DNA fragments.

2.9.2 pNPB assay

For testing if the cells had expressed and secreted a functional *Candida antarctica* lipase B, the following assay was performed. 0.84% (v/v) *p*-nitrophenyl-butyrate in 300 mM Tris/HCl (pH 7) was used as substrate for the *CalB.* 180 µL of the reaction solution were mixed with 20 µL of culture supernatant and the absorption at 405 nm was immediately measured over 3 minutes with 20 seconds intervall. The volumetric activity was calculated using the molar extinction coefficient of pNPB for these conditions (9.594 mM^{-1} *cm^{-1}) and the following formula:

$$vol.activity\ \left[\frac{U}{mL}\right] = \frac{\Delta\ Abs\ [\frac{mAU}{min}] * \ 0.2\ mL}{1\ min * 9.594\ mM^{-1} * 0.54\ cm^{-1} * 0.02\ mL}$$

2.9.3 Competent *E. coli* cells

To produce competent *E. coli* K12 Top10 F' cells 50 mL LB medium was inoculated with a single colony and incubated overnight at 37°C while shaking. The next morning 400 mL LB medium were inoculated with 2 mL of the ONC and incubated at 37°C and 200 rpm shaking to OD_{600} 0.7. As the culture reached the OD_{600} the cells were transferred into pre-chilled 500 mL centrifuge bottles and incubated on ice for 30 minutes. Afterwards the cells were pelleted by centrifugation at 2000*g for 15 minutes at 4°C. The cells were washed with approximately 200 mL ice-cold, sterile ddH_2O three times. After the last centrifugation step the cells were resuspended in 10% glycerol. After another centrifugation step at 4000*g for 15 minutes the cells were resuspended to a cell density of 10^9 cells/mL and aliquoted to 80 µL. The cells were shock-frozen in liquid nitrogen and stored at -80°C until use.

2.9.4 *E. coli* K12 Top10 F' electroporation

Electroporation in competent *E. coli* K12 Top10 F' cells was done by applying 2.5 kV followed by adding 900 µL SOC medium (0.5% w/v yeast extract; 2% w/v tryptone; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM glucose; pH 7) and regenerating at 37°C, 650 rpm for 1 hour. After regeneration the positive transformants were selected on LB agar plates (0.5% w/v yeast extract; 1.5% w/v agar; 1% w/v peptone; 1% w/v NaCl) containing respective antibiotic. Positive transformants were streaked out again and after plasmid isolation (GeneJET Plasmid Miniprep Kit, Thermo Scientific) the isolated plasmids were sent for sequencing (Microsynth AG, Vienna, Austria).

2.9.5 *P. pastoris* electroporation

To produce competent *P. pastoris* cells, 5 mL YPD medium was inoculated with cells. From the overnight culture a main culture was inoculated to OD_{600} 0.1 and grown for about 5 hours at 28°C, 90 rpm and 50% humidity. At OD_{600} ~1 the cells were spun down for 5 minutes at 3220*g, 4°C. The pellet was resuspended in 9 mL ice-cold BEDS (10 mM bicine-NaOH; 3% v/v ethylene glycol; 5% v/v dimethyl sulfoxide; 1 M sorbitol) and 1 mL DTT and mild shaking until the pellet was completely solved. The cells were spun down again and resuspended in 1mL BEDS and kept on ice. For transformation, 80 µL of competent cells were mixed with different amounts of DNA (100-200 ng circular, 1 µg linearized plasmid DNA) and electroporated with 2 kV. For regeneration, 500 µL 1 M Sorbitol and 500 µL YPD medium were added immediately and the cells were incubated at 28°C, 500 rpm for 2-3 hours. For selecting positive transformants, 100 µL of cell suspension were plated on YPD agar plates containing 50 µg/mL zeocin or BMG1 agar plates (100 mM potassium phosphate buffer pH 6; 1.34% yeast nitrogen base w/o amino acids; 0.0004% biotin; 1% glycerol) in case of complementation marker *GUT1* without regeneration time [48].

2.9.6 Assembly Cloning

For ligation of vector and inserts, Assembly Cloning was used. Therefore, 100 ng of enzyme digested vector DNA and insert in molar vector:insert ratio 1:3 were added to 15 µL of assembly cloning master mix containing reaction buffer (5% w/v PEG-8000; 100 mM Tris/Cl pH 7.5; 10 mM $MgCl_2$; 10 mM DTT; 0.2 mM dA/C/T/GTP; 1 mM NAD), 0.075 U T5 exonuclease, 0.45 U Phusion

DNA polymerase and 80 U *Taq* DNA ligase and incubated at 50 °C for one hour. The assembly was performed according to the protocol of Gibson and colleagues [49].

2.9.7 PCR

Polymerase chain reactions were used to amplify DNA fragments from either genomic or plasmid DNA as template. Primers were designed to have a T^m of 57.5-58.5 °C based on the algorithm provided by the IDT Oligo Analyzer Software tool and ordered from IDT (Integrated DNA Technologies, CA, USA). The PCRs were performed using the following agents:

Table 2.10 Agents for PCR reactions

Vol [µL]	**agent**
10	5x Phusion HF Buffer
2	Primer fw (10 µM)
2	Primer rv (10 µM)
5	dNTPs (2 mM)
0.4	Phusion DNA Polymerase (2 U/µL)

The template DNA (10 ng) was added and the reaction mixture was filled up to 50 µL with ddH_2O. The PCR was performed with the following cycle parameters:

Table 2.11 Temperature profile for PCR reactions

reaction	**# of cycles**	**Time [s]**	**Temperature [°C]**
Initial denaturation	1	60	98
Denaturation	30	8	98
Annealing		30	58
Extension		15/kbp	72
Final extension	1	450	72
Holding	1	∞	4

For size evaluation of the PCR products, they were mixed with 5 µL 6X Loading Dye + SDS Solution and applied on a 1% agarose gel next to an appropriate DNA Ladder. Electrophoresis was done for 40 minutes at 120 V. If the PCR products showed the right size, they were cut out of the gel, purified via GeneJET Gel Extraction Kit or Wizard® SV Gel and PCR Clean-Up System and eluted in

35 µL ddH_2O. The concentration (in ng/µL) was determined by a spectrophotometric measurement using NanoDrop.

2.9.8 Colony-PCR of *E. coli*

As a control for correct vector ligation, colony-PCRs were performed with GoTaq® G2 DNA Polymerase. Single colonies of the transformation plate were picked with a pipette tip, transferred to a new selection plate and then transferred into 25 µL of PCR Master Mix by pipetting up and down. The Master Mix consisted of:

Table 2.12 Agents used for colony-PCR

Vol [µL]	**agent**
5	5x Green GoTaq® Reaction Buffer
2.5	dNTPs (2 mM)
1	Primer 1 (10 µM)
1	Primer 2 (10 µM)
0.15 (0.75 U)	GoTaq® DNA Polymerase (5 U/µL)
15.35	ddH2O

The PCR was performed with the following cycle parameters:

Table 2.13 Temperature profile for colony-PCR

reaction	**# of cycles**	**Time [s]**	**Temperature [°C]**
Initial denaturation	1	300	95
Denaturation	20	15	95
Annealing		30	58
Extension		30/kbp	72
Final extension	1	450	72
Holding	1	∞	4

For size evaluation of the PCR products the DNA was applied on a 1% agarose gel next to an appropriate DNA Ladder. Electrophoresis was done for 40 minutes at 120 V. If the PCR products showed the corresponding size of the insert, the correct transformants were streaked on LB plates containing respective antibiotics. Minipreparations were made and sent for sequencing.

2.9.9 *S. pombe* transformation of cryopreserved cells [50]

To insert plasmid DNA in *S. pombe* 972h- cells, an overnight culture was inoculated with a single colony in 50 mL YES medium. The main culture was inoculated from the ONC to OD_{600} 0.3 and grown at 30°C and 120 rpm to OD_{600} 1. The cells were harvested by centrifugation for 5 minutes at 1600*g, then washed three times with ice-cold, sterile ddH_2O and finally diluted with 0.1 M LiAc in 30% glycerol to a cell density of 10^9 cells/mL. 50 µL were aliquoted into microcentrifuge tubes, placed on ice for 30 minutes and then directly frozen at -80°C.

For transformation, cells were thawed in a thermomixer for 2 minutes at 40°C, mixed with 50 µg carrier DNA, 5 µg plasmid DNA and 145 µL 50% PEG-4000 and heat shocked for 15 minutes at 43°C. Finally the cells were spun down at 3000*g for 2 minutes, the pellet resuspended in 500 µL YES and regenerated for 3-4 hours at 30°C. After regeneration the cells were plated on appropriate 100 µg/mL containing YES plates [50].

2.9.10 *S. pombe* transformation [51]

Additionally we performed a transformation method for *S. pombe* using lithium acetate without freezing after Paul Nurse and colleagues [51]. An ONC in 50 mL YES medium was cultivated at 30°C and shaking. The main culture (50 mL YES medium) was inoculated to OD_{600} 0.2 and grown to 0.8-1 at 30°C and 120 rpm. The cells were harvested by centrifugation at 500*g for 5 minutes, resuspended in 50 mL sterile ddH_2O and harvested by a 5 minutes centrifugation at 500*g. The cells were resuspended in 5 mL LiAc-TE (0.1 M LiAc in TE buffer) and again collected by centrifuging at 500*g for 5 minutes. Following, the cells were resuspended in 800 µL LiAc-TE and 100 µL of the suspension was then mixed gently with 5 µL carrier DNA at 10 mg/mL and 1-5 µg plasmid DNA and incubated at room temperature for 10 minutes. Afterwards the cells were mixed with 260 µL PEG/LiAc-TE (40% PEG in 0.1 M LiAc-TE) and incubated at 30°C for 30 minutes. 43 µL of DMSO were mixed to the cells and they were heat-shocked at 42°C for 5 minutes. Afterwards the cells were pelleted by short centrifugation, resuspended in 1 mL YES and regenerated for 2 h at 30°C. Finally, 100 µL of cells were plated on YES agar plates with 100 µg/mL zeocin for selection.

2.9.11 Fluorescence measurements in MTPs

For determining sTomato and eGFP fluorescence, the cultures were diluted 1:20 in 200 µL ddH_2O in microtiter plates and absorbance was measured in a plate reader at 581 nm with the excitation wavelength of 554 nm (sTomato) and 507 nm with the excitation wavelength of 488 nm (eGFP). The resulting values were normalized by the measured OD_{600}.

2.9.12 *P. pastoris* cultivation in 96-well plates

400 µL YPD (1% yeast extract, 2% peptone, 0.5-1% glucose) containing 50 mg/L zeocin were inoculated with single colonies from the transformation plates and incubated for 60 h at 28°C and 320 rpm in a HT Infors Orbitron shaker.

2.9.13 *S. pombe* cultivation in 96-well plates

400 µL YES medium containing 100 µg/mL zeocin were inoculated with single colonies from the transformation plates, sealed with gas-permeable plate sealers and incubated at 28°C and 320 rpm in a HT Infors Orbitron shaker 60 h.

2.9.14 ABTS plate assay

For functional screening of cDNA libraries in *P. pastoris* an ABTS plate assay for detecting clones expressing active laccases was performed. Therefore selective plates (YPD with 50 µg/mL zeocin) containing in addition 100 mg/L ABTS (2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)) and 25 mg/L $CuSO_4$ were added to the liquid agar before pouring the plates. For the screening itself, 100 ng of library plasmid were transformed in *P. pastoris* CBS7435 electrocompetent cells and 100 µL of the regeneration culture were plated onto the ABTS plates directly after regeneration. The plates were incubated at 28°C for three days and kept at 4°C for further 8-10 days. In case of a positive clone, a green halo around the colony should be visible [52].

2.9.15 cDNA library generation (final protocol; optimized from Kooiker and Xue, 2013 [53])

2.9.15.1 RNA Isolation

As starting material for cDNA library generation total RNA of different organisms was used. The RNA was isolated with the Plant/Fungi RNA Purification Kit. Therefore, about 1-2 g of freeze-dried

fern or plant tissue were doused with liquid nitrogen in a mortar and pestled until the tissue was completely pulverized. Then, from 150 mg of this tissue total RNA was purified according to the kits manual. Concentration of the RNA after purification as well as 260/280 and 230/260 values were determined by spectrophotometric measurements with nanodrop.

2.9.15.2 First strand synthesis

In general, only RNase-free materials were used to avoid any degradation of the RNA before cDNA was synthesized.

For first strand synthesis of the cDNA the following ingredients were mixed together in a RNase-free PCR tube on ice:

- 1 µg total RNA (obtained as described in 2.9.15.1)
- 1 µL Poly-d(T) primer mix (10 µM)
- Up to 4 µL RNase-free ddH_2O

The mixture was incubated at 72°C for 2 minutes, then at 42°C for 3 minutes in a thermocycler and placed on ice afterwards. The following mixture was added immediately:

Table 2.14 cDNA first strand synthesis step 2

Vol [µL]	**agent**
5	5x RT Buffer
1	dNTPs (10 mM)
1	M-MLV reverse transcriptase (2 U/µL)
0.25	RNase out
1	Ribo-Primer (12 µM)
0.75	ddH2O

Mixture occurred by pipetting and the tube was again placed in a thermocycler at 42°C for 90 minutes, followed by heating to 70°C for 10 minutes and then stored at 4°C.

1 µL RNase H (2U) were added and mixed by pipetting and incubated for 20 minutes at 37°C followed by heating to 85°C for 5 minutes to inactivate the RNase.

2.9.15.3 Second strand synthesis and amplification

For second strand synthesis the following mixture was prepared on ice:

Table 2.15 Agents for cDNA second strand synthesis

Vol [µL]	agent
20	5x HF Buffer
10	dNTP's (2 mM)
4	Anchor primer fwd (10 µM)
4	Anchor primer rev (10 µM)
0.6	Phusion DNA polymerase (2 U/µL)
10	ss cDNA (from 2.9.15.2)
51.4	ddH_2O

The reaction was placed in a thermocycler and run with the following cycle parameters:

Table 2.16 Temperature profile for cDNA second strand synthesis and amplification

Reaction	# of cycles	Time [s]	Temperature [°C]
Initial denaturation	1	120	98
Denaturation	15	30	98
Annealing		30	58
Extension		300	72
Final extension	1	450	72
Holding	1	∞	4

2.9.15.4 Size selection of ds cDNA and Cloning

To ensure only full length or at least transcripts longer than 1000 basepairs are used for cloning, a size selection had to be performed. Two different methods were used:

- Chromaspin TE columns

  20 µL 6X loading dye + SDS were added to the ds cDNA and loaded onto a Chromaspin TE 1000 column. The purification occurred according to the manufacturers manual.

- Agarose gel

  For size selection via a 1% agarose gel, also 20 µL 6x loading dye + SDS were added to the ds cDNA and the mixture was applied onto the gel. The separation occurred at 120 V for 40 minutes in 1x TAE buffer. After the 40 minutes the area between 1000 and 10000 basepairs was excised of the gel and purified via a Gel Extraction Kit (Thermo Scientific).

The concentration of the purified DNA was determined by photometric measurement via NanoDrop.

To clone the cDNA into the expression vector, the vector was digested with *Bcu*I and *Not*I and also purified via an agarose gel.

About 30 ng of digested vector (1 µL) and 4 µL cDNA were assembled by assembly cloning. The whole assembly was then desalted for 20 minutes on Millipore DNA filter paper (0.025 µm pore size) and transformed into *E. coli* K12 Top10 F' according to 2.9.3. After regeneration 100 µL of cell suspension were plated on a selection plate, with the residual 900 µL cell suspension 10 mL of liquid selection medium was inoculated and incubated at 37°C overnight while shaking.

The next day a colony PCR with 20 clones (section 2.9.8) was performed from the selection plate to evaluate the size of the inserted cDNA. Three Minipreparations of plasmid DNA were performed from the liquid ONC and their concentration was measured photometrically via NanoDrop.

2.9.16 Restriction digest

All restriction digestions in this thesis were performed for 1 h at 37°C. 1 µL of fast digest restriction enzyme per 1 µg of DNA was used, since all enzymes were fast digest enzymes from Thermo Scientific Fisher. The volume of the digest was adjusted in a way that the glycerol content did not exceed 10% to avoid activity loss of the enzyme. For a 20 µL digestion the following ingredients were used.

Table 2.17 Restriction digestion composition

amount	**agent**
2 µL	10x FD buffer
1 µL	Restriction enzyme
1 µg	DNA
Up to 20 µL	ddH_2O

For double digestions 1 µL of each enzyme per 1 µg of DNA were used and the digest was done in at least 40 µL to avoid high glycerol concentrations.

2.9.17 *P. pastoris* Bust'n'Grab gDNA isolation [54]

In order to gain gDNA of *P. pastoris* for amplification of promoters or genes from the genome, Bust'n'Grab protocol was used. A single *P. pastoris* colony was inoculated in 5 mL YPD and incubated overnight at 28°C and 90 rpm. The next day 2 mL of ONC were pelleted by centrifugation and gDNA was isolated according to the protocol of Harju and colleagues. Subsequently DNA was precipitated in 400 µL 100% EtOH, which was done overnight at -20°C.

2.9.18 Sequencing and plate sequencing

For sequencing verification of amplified DNA fragments, samples were sent to Microsynth AG, Switzerland. 1.5 µg DNA and 4 µL sequencing primer were filled up to 15 µL with ddH_2O.

For whole plate sequencing, to evaluate the cDNA fragments generated (2.9.15), 48 wells of a 96-well plate were filled with 400 µL LB medium supplemented with 25 mg/L zeocin and inoculated with single *E. coli* colonies from the transformation plates. After 24 h of growth 120 µL of each well were transferred two times in a prepaid sequencing plate (e. g. row 1 and 6 of the sequencing plate were inoculated from row 1 of the deep-well plate) and sent for sequencing to Microsynth AG, Switzerland. Row 1-6 were sequenced using primer #272, row 7-12 were sequenced using primer #273.

3 RESULTS AND DISCUSSION

3.1 cDNA LIBRARY GENERATION

The first goal of this master thesis was to develop a method for fast and reliable cDNA library generation. The cDNA (complementary DNA) is the DNA complementary to the mRNA transcribed. There exist various methods to gain this cDNA from either mRNA or total RNA isolated from an organism. As described in the introduction, different methods for generation of such libraries are available. The method used in this thesis was the "template-switching" mechanism, also called SMART^{TM} approach.

cDNA libraries are widely used to screen for new enzymes or gather sequence information about various transcripts from one organism as well as from environmental samples. From the methods existing for cDNA library generation the template-switching method is the fastest. Additionally the technique is suitable to generate a high number of full-length transcripts because only transcripts at which the reverse transcriptase reaches the 5' end are captured with the ribonucleotide primer to add a vector overhang and therefore the second strand can be cloned into the vector subsequently [55].

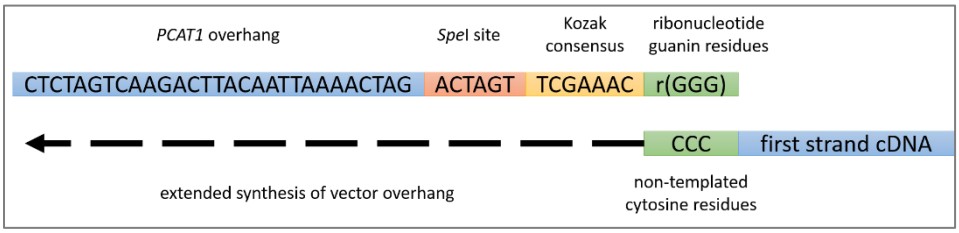


Figure 3.1 Extended first strand synthesis by M-MLV RT.
The upper line shows the primer sequence of the ribo-primer with vector overhang, restriction site *Spe*I for subsequent cloning, *P. pastoris* consensus Kozak sequence and the three ribonucleotide guanine residues binding to the non-templated cytosine residues at the 3' end of the newly synthesized cDNA. By extension of first strand synthesis and incorporation of the *Spe*I restriction site *the* cloning of the full length cDNA fragments into the expression vector is enabled.

For functional screening actives enzymes and so full length transcripts. If only fragments of genes are gained during cDNA synthesis, proteins may be disfunctional because of missing or broken conserved domains essential for their activity. Reviewing all different cDNA generation methods available the template switching mechanism method after Zhu and colleagues was investigated more thoroughly [20].

3.1.1 Method of choice: template-switching [20]

3.1.1.1 The mechanism

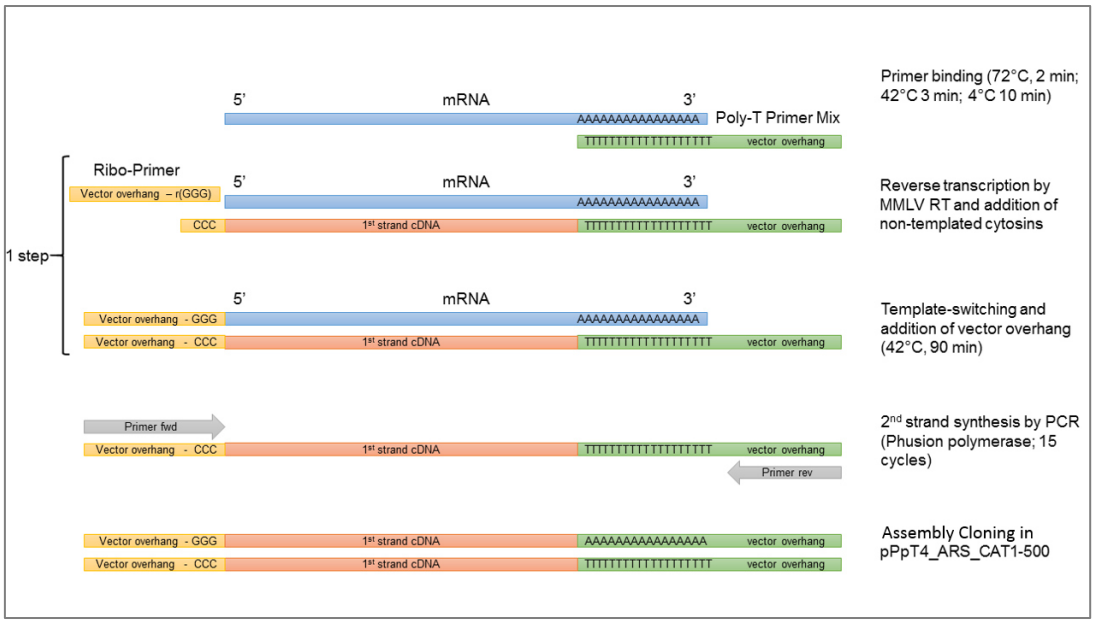


Figure 3.2 Template-switching mechanism for cDNA generation.
The mRNA is captured using poly-d(T) primers. The template-switching mechanism for cDNA generation makes use of the property of MMLV to add three cytosine bases, when it reaches the end of the template, to the new synthesized cDNA first strand which work in turn as a prolonged RNA template and lead the reverse transcriptase again to synthesize the complementary to the nucleotides located on the primer. The second strand is then synthesized by a 15-cycle PCR using primers complementary to the vector overhangs at the 3' and 5' end of the ss cDNA.

The template switching mechanism relies on the natural 3' terminal cytosine addition of Moloney Murine Leukemia Virus reverse transcriptase. The enzyme adds three cytosine bases to the ss cDNA strand synthesized after it reaches the 3' end of the template RNA with 79% probability [18]. This CCC can function as an anchor point for adapter primers for directional cloning. For cDNA generation with this method, a mixed DNA-RNA primer termed "Ribo-primer" had to be designed to have an overhang to the 5' promoter region of the vector as described in 3.1.1.2. The 5' end consisted of three guanine-ribonucleic acid residues. These r(GGG) are recognized by the reverse transcriptase resulting in extended synthesis of the complementary DNA to the vector overhang on the Ribo-primer.

3.1.1.2 Primer design

For the cDNA generation 4 different primers had to be designed. The primer containing the poly-d(T) stretch to capture polyadenylated mRNA, the DNA-RNA primer for template switching and a forward and reverse primer for second strand synthesis and amplification of the generated cDNA.

The poly-d(T) primer was designed to have a 27 bp long poly-d(T) stretch its 3' end to bind to the poly-A tail of the polyadenylated mRNA. If there is no anchor point provided at the 5' end of the poly-d(T) stretch the annealing characteristics of such sequences tend to show a mixture of internal and terminal poly-A priming. To decrease the probability of undesirable internal priming and therefore ensure full-length cDNA transcripts we designed anchored poly-d(T) primers [56]. We used anchored poly-d(T) primers where the 3' end consists of 27 thymidines and each possible combination of A, C and G at the second position as well as A, C, G or T on the first position.

This resulted in a total of 12 primers combinations that were synthesized and mixed together to generate the poly-d(T) primer mix. This variety in nucleotides at the 3' end of the primer decreases the internal poly-A priming and therefore the occurrence of truncated cDNAs dramatically [56]. To allow directional cloning the 12 different primers were designed to harbor an overhang to the transcription terminator on the 5' end of the expression vector. Also the concentration of the primer was kept low at 10 µM to avoid over-saturation in priming the poly-A tail. If two high amounts of poly-d(T) primers are present, few thymidines of a primer can pair with few adenosins on the poly-A stretch resulting in several primers next to each other on one single mRNA molecule. This can hinder the reverse transcriptase to bind to the mRNA.

As described in 3.1.1.1 the Ribo-primer had to contain 3 ribonucleotide guanosines at its 3' end to enable the template switching of the M-MLV. Zajac et al. investigated in 2013 the probabilities for the M-MLV to add the different bases as non-templated nucleotides to the 5' end of the new synthesized cDNA first strand as it is visualized in Figure 3.3.

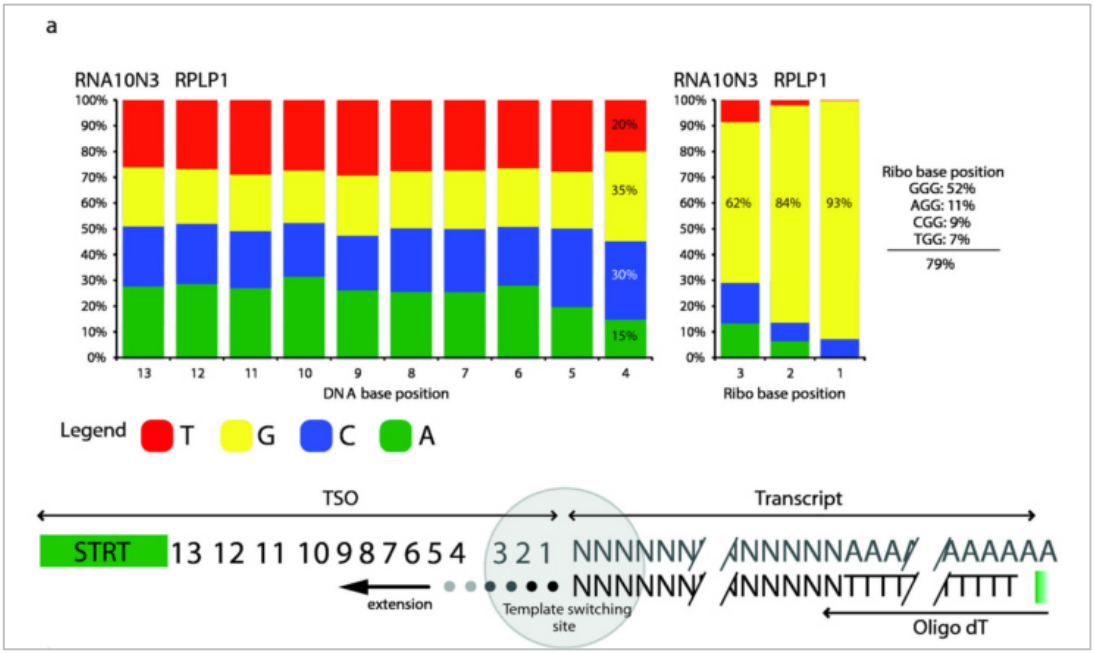


Figure 3.3 Preferred bases for M-MLV to add as non-templated nucleotides to the 5' end of the new synthsized first strand of the cDNA. From Zajac et al., 2013.

As Figure 3.3 shows Zajac et al. found that on position 1 a C is added with 93% probability and for position 2 the probability is still above 80%. And even with the 64% of position 3 the overall probability for the CCC is 79%. For that reason we designed the Ribo-primer with a 5’ overhang to the *P. pastoris* Kozak sequence (TCGAAACG) and 3’ end of the promoter region to support ribosome binding if the natural Kozak sequence is not present on the mRNA or not recognized by *P. pastoris* translation machinery. The Kozak sequence is followed by a *Spe*I restriction site and three guanosines (r(GGG)).

For the second strand synthesis two further primers had to be designed, termed anchor fwd and anchor rev with their binding sites located at the 3’ and 5’ end of the synthesized vector overhang of the cDNA molecule. These primers were used for 15-cycle PCR procedure (see section 2.9.15.3) as it is shown in Figure 3.2. The anchor fwd primer is binding to the 5’ end of the promoter region including the *P. pastoris* Kozak sequence, which corresponds to the Ribo-primer sequence lacking the r(GGG). The anchor rev primer showed the same sequence as each of the poly-d(T) primers but lacking the poly-d(T)-tail.

The following sections describe in detail the different elements of the cDNA synthesis procedure and how they were optimized to generate cDNA libraries in a reliable, reproducible manner. For all following experiments total RNA from *Pteridium aquilinum*, an eagle fern endemic to north-western Europe as well as north africa[57], was used as starting material.

3.1.2 First strand synthesis

3.1.2.1 Reverse transcriptases

During this thesis reverse transcriptases from three different suppliers were tested for cDNA generation: M-MLV, M-MLV RNase H- and Protoscript II. M-MLV is the reverse transcriptase, a RNA-dependent DNA polymerase, coded on the *pol* gene of Moloney Murine Leukemia Virus [13,58]. M-MLV RNase H- is a point mutant of the previously described enzyme lacking the natural RNase H activity. This activity is undesired especially when long transcripts, such as full-length genes, are synthesized [11–13,59]. The described enzymes are both provided by Promega GmbH. Protoscript II is the same point mutant as RNase H- but from another supplier (New England Biolabs). Most experiments were done with the M-MLV because this enzyme yielded the most reliable and reproducible results. The first cDNA generation with this enzyme yielded cDNA, but the average insert length was too short to code for full length genes, as Figure 3.4 shows.

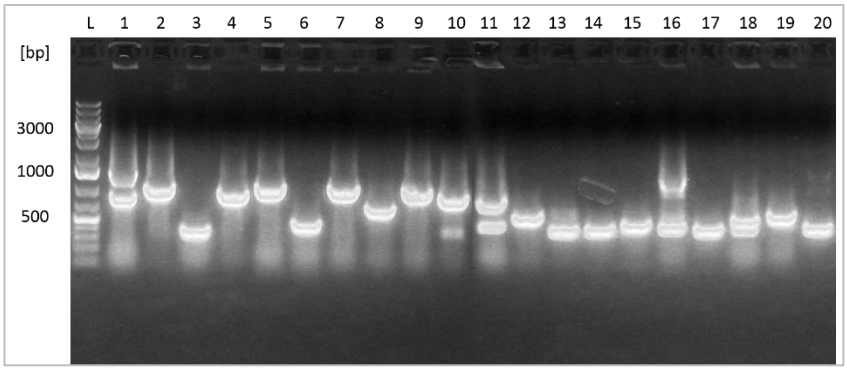


Figure 3.4 Agarose gel electrophoresis of cPCR of 20 cDNA library *E. coli* clones.
For the cPCR 20 *E. coli* transformants were used and the cDNA insert, derived from *P. aquilinum* RNA, was amplified via PCR with primers binding directly upstream and downstream of the insert. The figure shows that most of the inserts have sizes around 500 bp. As ladder the 2-log DNA ladder was applied.

The average insert length of these clones was around 500 bp which is distinctively too short for full length genes [60]. A possible explanation for this result is the presence of RNase activity observed in wildtype M-MLV reverse transcriptase. Due to this activity a certain number of RNA strands might be enzymatically digested before they can serve as a template for reverse transcription. In addition the shorter RNA fragments generated during the digest can easily serve again as a template for reverse transcription, and so increase the short fragments dramatically [61]. As there are already different mutants of M-MLV lacking RNase H activity available improving reverse transcript length either by preventing internal RNA digestion or ensuring increased thermal stability [62], two enzymes from two different suppliers, as described above, were tested in this thesis.

- M-MLV H- (Promega)

  This reverse transcriptase is especially designed for long transcripts >5 kbp and completely lacks the native RNase H activity [12]. This transcriptase did not yield any cDNA in all of our experiments, neither in the supplied buffer nor in our self-made RT buffer. This was also tested with two different RNA samles with different concentrations, which again did not make any difference.

- Protoscript II (NEB)

  This reverse transcriptase has reduced RNase H activity as well as increased thermostability, remaining active up to 48°C, as the manufacturer describes. With Protoscript II cDNA generation was successful but the cDNA molecules generated are too short to code for eukaryotic full-length genes as shown in Figure 3.5 [60].

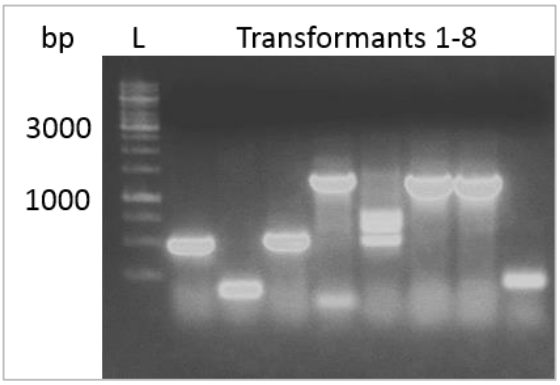


Figure 3.5 Colony PCR of cDNA clones where cDNA was generated with Protoscript II from *P. aquilinum* RNA.

The cDNA was generated using Protoscript II reverse transcriptase and cloned into the expression vector. Colony PCR was performed using primers 272 and 273. PCR products were visualized on a 1% agarose gel.

Figure 3.5 shows the result of the colony PCR of the cDNA clones from Protoscript II. Protoscript II is reported to generate larger cDNA fragments than the unmodified M-MLV reverse transcriptase and is also more thermostable due to lacking the RNase H activity. The longest cDNA fragments we gained using this reverse transcriptase were about 1000 bp which is no improvement compared to the wildtype M-MLV.

Furthermore, a systematic test was performed with all three reverse transcriptases mentioned above. In total 8 reactions were performed as listed in Table 3.1.

Table 3.1 Scheme for systematic reverse transcriptases test. In addition to the three different reverse transcriptases, also two ways of priming the second strand synthesis were tested. Adaptor means the primer pair binding to the synthesized vector overhangs, Poly-d(t) primer is the primer mix as described in section 3.1.1.2.

# of reaction	RT	Type of 2nd strand synthesis	# of cycles for 2nd strand synthesis	Label in Figure 3.6
1	M-MLV RNase H-	Adaptor	15 cycles	H-1
2		Poly-d(T)	15 cycles	H-2
3	Protoscript II	Adaptor	15 cycles	PS1
4		Poly-d(T)	15 cycles	PS2
5	M-MLV wildtype	Adaptor	8 cycles	WT1
6			15 cycles	WT2
7		Poly-d(T)	8 cycles	WT3
8			15 cycles	WT4

The agarose gel, showing the cDNA generated after second strand synthesis, is shown in Figure 3.6.

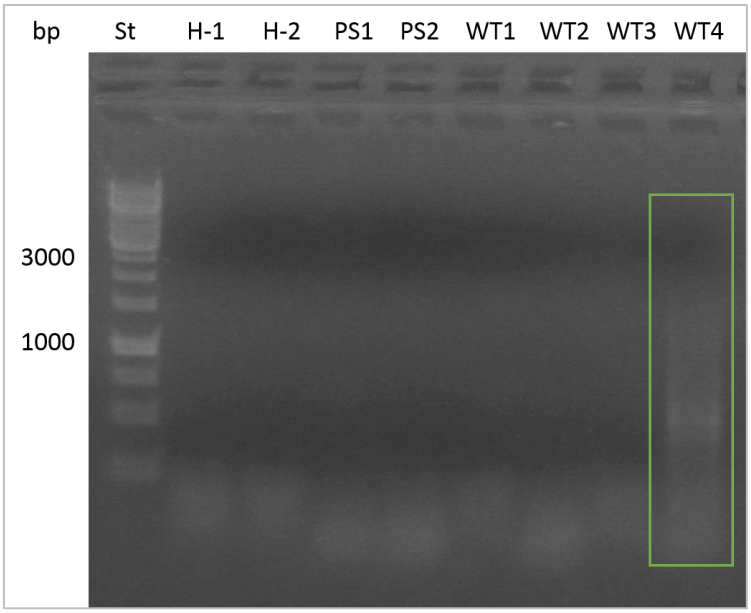


Figure 3.6 Result of the reverse transcriptases systematic test for cDNA generation from *P. aquilinum* RNA. cDNA generated as described in Table 3.1 after 2nd strand synthesis was visualized on a 1% agarose gel. The only reaction where reverse transcription yielded cDNA was #8 where the wildtype M-MLV was used together with poly-T 2nd strand synthesis in 15 cycles (green square).

As can be seen in Figure 3.6 the reverse transcriptions with M-MLV RNase H- as well as with Protoscript II did not yield any amounts of cDNA visible using agarose gel electrophoresis. Only reverse transcription with M-MLV, poly-d(T) primer and 15 thermal cycles for second strand synthesis yielded sufficient amounts of cDNA for subsequent purification and cloning as can be seen in Figure 3.6 (WT4). These results led us to the conclusion that RNase H activity lacking reverse transcriptases don't offer an improvement in the lengths of transcripts for our experimental procedure. We decided to perform further experiments with M-MLV and addition further select the transcripts by size to gain longer cDNA fragments that could code for full length genes as described in 3.1.3.1.

3.1.2.2 Buffers

A closer analysis of potential factors influencing cDNA generation efficiency revealed that the buffer employed during the reverse transcription procedure proved to be a critical one. The buffer recommended for M-MLV contains 50 mM Tris/HCl, 3 mM $MgCl_2$, 75 mM KCl and 10 mM DTT at pH 8.3. In comparison the buffer of the SMART kit (Clontech Laboratories Inc.), which is the commercially available kit for generating cDNA via the template-switching mechanism contains 50 mM Tris/HCl, 6 mM $MgCl_2$, 75 mM KCl and 2 mM DTT. We hypothesized if the lower DTT concentration might support the reverse transcriptase activity while not having to lower the RNase activity since the reverse transcriptase contained in the kit is lacking this activity anyway. Subsequently we set up an 5x RT buffer containing 250 mM Tris/HCl, 6 mM $MgCl_2$, 75 mM KCl and 10 mM DTT. To lower or eliminate RNase activity 0.25 µL (10 U) RNase out were added to the reaction instead of a higher DTT concentration. By using this buffer the first successful cDNA generation with the M-MLV (Promega) as well as with Protoscript II (NEB) was performed. Successful cDNA generation was seen as a smear on an agarose gel between 1000 and 5000 bp as can be seen in Figure 3.8. This buffer was then used for further experiments.

3.1.3 Second strand synthesis and amplification of ss cDNA

3.1.3.1 Size selection

As described in section 3.1.2.1 the reverse transcription procedure resulted in transcript lengths too short for incorporation of a high quality full length cDNA library. To avoid cloning of smaller fragments than 1000 bp into the expression vector, two different methods for size selection were tested:

- CHROMA SPIN™+TE-1000 Columns

  These columns enable nucleic acid purification of fragments larger than 1000 bp and are based on a gel filtration principle. They contain resins holding back molecules smaller than the desired size while larger molecules are excluded and can be collected by centrifugation. They are recommended for size selection of large cDNA fragments (Protocol PT1300-1, Clontech Laboratories Inc.).

For size selection of the synthesized cDNA, 20 µL of 6x loading dye + SDS solution were mixed with the double stranded cDNA and applied to the column. As the total volume after purification was 120 µL with an average concentration of 2-3 ng/µL, nucleic acids were concentrated in a rotational vacuum concentrator to 10-15 µL with an average concentration of 20-30 ng/µL. This was then used to clone into the expression vector by assembly cloning, followed by transformation into electrocompetent *E. coli* K12 Top10 F' cells. An example of a colony-PCR after 24 h of growth on selection plates is shown in Figure 3.7.

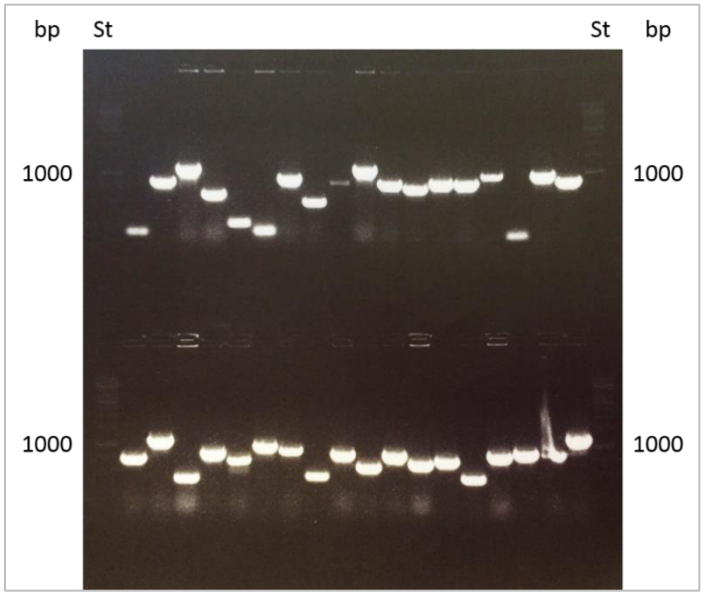


Figure 3.7 Colony-PCR of 36 cDNA clones, from *P. aquilinum* RNA, after size selection via CHROMASPIN^{TM}+TE-1000 Columns.
St= GeneRuler 1 kb DNA ladder. Although the used columns should eliminate fragments smaller than 1000 bp in size, 50 % of the clones have an insert smaller than 1000 bp.

As it is shown in Figure 3.7, despite the size selection with CHROMA SPIN™+TE-1000 Columns half of all tested clones had an insert size smaller than 1000 bp. Independent repletion of column binding experiments resulted in the same average insert size below 1000bp, which would not support the generation of a high quality full length cDNA library.

- Size selection via agarose gel [63]

  Another possibility for size selection is a simple agarose gel electrophoresis. The double stranded cDNA was applied on a 1% agarose gel and electrophoresis was done for 30 minutes at 100 V. Afterwards the area of the desired cDNA fragment size was excised. A positive and negative example for cDNA generation and the following excision of the cDNA fragments is shown in Figure 3.8.

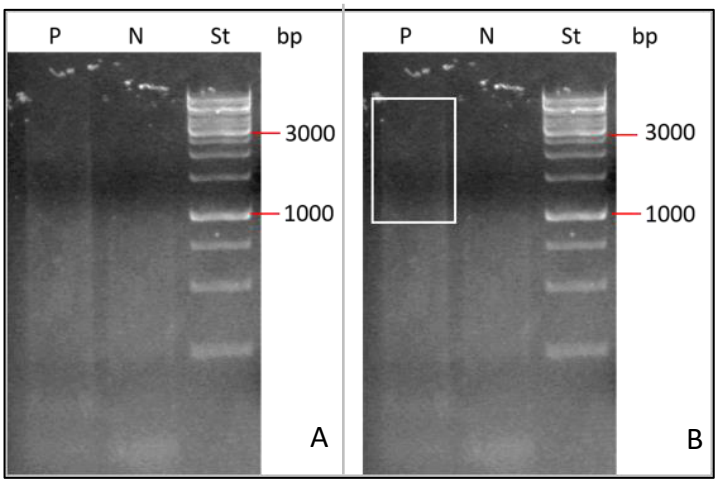


Figure 3.8 Size selection via agarose gel of double stranded cDNA derived from *P. aquilinum* RNA. P=positive, N=negative, St=GeneRuler 1 kb DNA ladder. In slot P a smear is visible from 0.05 to 6 kpb, which represents a successful cDNA synthesis. In slot N the smear is only visible up to 1 kpb, cDNA synthesis was not succesful here. The light grey square on the right side shows the area from 1 to 6 kpb cut out of the gel for purification.

Figure 3.8 shows a positive (P) and a negative (N) result of cDNA generation. In slot P a smear is visible from 50 to 10 000 bp, which represents the successfully generated cDNA. The grey square in panel B shows the area where cDNA fragments from 1 000 – 10 000 bp are located. This area was cut out of the gel and purified via GeneJET Gel Extraction Kit. The resulting eluate typically showed concentrations from 8-15 ng/µL, after concentration in a rotational vacuum concentrator the concentration could be increased to 20-30 ng/µL. 4 µL of this solution were then used for assembly cloning to clone the cDNA fragments into the expression vector. The following figure shows a colony-PCR from cDNA clones obtained with the agarose gel size selection method.

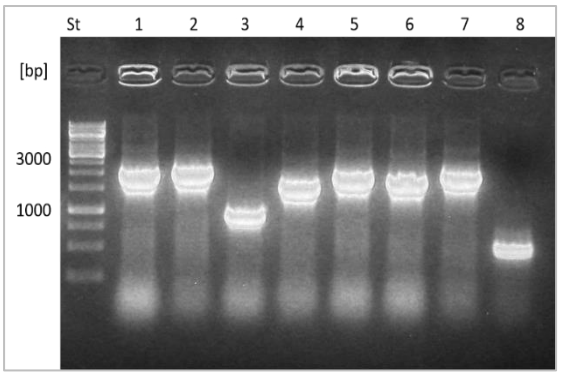


Figure 3.9 Colony-PCR of *E. coli* cDNA clones from *P. aquilinum* RNA where the cDNA fragments cloned into the vector were obtained with agarose gel size selection method.
St= GeneRuler 1 kb DNA ladder. Slots 1-8 show the result of the colony-PCR of 8 clones after successful size selection via an agarose gel electrophoresis.

As it is shown in Figure 3.9 the size selection of the generated cDNA was clearly more successful than obtained with the CHROMA SPIN™+TE-1000 Columns. Six out of eight clones contain an insert larger than 1000 bp. 48 clones of this library were sequenced to obtain more information about average insert sizes, ribosomal contaminations and recurrence of any sequence of the cDNA as it is described in 0.

The size selection via agarose gel electrophoresis performed much better than the CHROMA SPIN™+TE-1000 Columns. This method was then included as general size-selection method in the protocol.

3.1.4 Cloning and amplification in *E. coli*

After successful cDNA generation, the next step was to clone the gained cDNA fragments into the expression vector, as described in 3.2.1.. 15 µL of assembly cloning reaction mix were mixed with 30 ng *Bcu*I and *Not*I digested vector and 4 µL (size selected) cDNA and incubated for one hour at 50 °C. Afterwards, the whole reaction was desalted for 20 minutes and then transformed into electrocompetent *E. coli* K12 Top10 F' cells. After regeneration 100 µL of culture were plated on a zeocin containing LB agar plate for a colony PCR the next day. With the residual 900 µL cell culture 10 mL liquid LB medium supplemented with 25 mg/L zeocin were inoculated and incubated at 37°C overnight. The next day three Mini-Preps were done from the liquid culture and a colony PCR was performed from the agar plate to evaluate the size of the cDNA fragments.

3.1.5 Library Sequencing

To obtain more specific sequence information and evaluate the occurrence of ribosomal RNA contaminations and repetitive sequences in the generated library, two different libraries, one without size selection and one with size-selected cDNA, were sent for sequencing at Microsynth AG. The first 96 well plate was inoculated with clones of the library shown in Figure 3.9, where wildtype M-MLV reverse transcriptase was used for cDNA generation followed by agarose gel size selection. The clones were inoculated in duplicates to gather forward and reverse sequencing read. After receipt of the sequencing results, blastn search was done for every clone and also insert lengths were examined. The blastn results of the size-selected cDNA library are listed in Table 3.2.

Table 3.2 blastn results of sequenced cDNA library obtained from *P. aquilinum* RNA (48 clones, forward and reverse read each).
The table comprises the insert length, the highest score hit of the blastn search as well as the identity.

Clone	Insert length	blastn result	E-value	Identity
A1	880	E.coli-K.pastoris shuttle vector pPpHIS4, complete sequence	0.0	100%
A2	139	Polypodium vulgare 28S ribosomal RNA gene, partial sequence	5,00E-62	99%
A3	1557	Theobroma cacao Fructose-bisphosphate aldolase 1 (TCM_006785) mRNA, complete cds	2,00E-122	74%
A4	167	Polypodium vulgare 28S ribosomal RNA gene, partial sequence	1,00E-74	99%
A5	1423	PREDICTED: Pyrus x bretschneideri 26S protease regulatory subunit 6A homolog (LOC103930611), mRNA	0.0	77%
A6	850	PREDICTED: Camelina sativa uncharacterized protein At5g49945-like (LOC104760982), transcript variant X2, mRNA	7,00E-04	100%
B1	905	E.coli-K.pastoris shuttle vector pPpHIS4, complete sequence	0,00E+00	100%
B2		no single colony		
B3	869	no result		
B4	391	Nipponolejeunea subalpina 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and 26S ribosomal RNA gene, partial sequence	3,00E-153	94%
B5	1504	PREDICTED: Brachypodium distachyon UDP-arabinopyranose mutase 1 (LOC100838265), mRNA	6,00E-176	78%
B6	0	empty		
C1	1627	no result		
C2	1365	Hordeum vulgare subsp. vulgare cultivar Bonus magnesium chelatase 40-kDa subunit (xantha-h) gene, complete cds	1,00E-166	73%
C3	380	Polypodium vulgare 28S ribosomal RNA gene, partial sequence	2,00E-161	99%
C4	1418	no result		
C5	1654	no result		
C6	796	PREDICTED: Malus x domestica protein THYLAKOID FORMATION1, chloroplastic (LOC103404755), mRNA	1,00E-10	74%
D1	1704	no result		
D2	1804	Selaginella moellendorffii hypothetical protein, mRNA	1,00E-153	79%
D3	1687	Picea glauca clone GQ04008_O10 mRNA sequence	1,00E-104	73%
D4	835	no result		
D5	1624	no result		
D6	2500	PREDICTED: Vitis vinifera probable methyltransferase PMT24 (LOC100265973), transcript variant X2, mRNA	3,00E-65	82%
E1	old insert			
E2	1560	no result		
E3	1690	Picea glauca clone GQ03105_L15 mRNA sequence	1,00E-13	78%
E4	1566	Pteridium aquilinum subsp. aquilinum chloroplast, complete genome	0.0	94%
E5	old insert			
E6	847	PREDICTED: Nicotiana tomentosiformis glutamate--glyoxylate aminotransferase 2 (LOC104114985), mRNA	1,00E-65	80%
F1	1322	PREDICTED: Nicotiana sylvestris Golgi to ER traffic protein 4 homolog (LOC104236108), transcript variant X4, misc_RNA	0.004	75%
F2	0	empty		
F3	1651	Cyprinus carpio genome assembly common carp genome, scaffold 000000952	7,00E-07	92%
F4	1687	PREDICTED: Lepisosteus oculatus S phase cyclin A-associated protein in the endoplasmic reticulum-like (LOC102689104), mRNA	0.25	90%
F5	1040	Physcomitrella patens subsp. patens predicted protein (PHYPADRAFT_148997) mRNA, complete cds	2,00E-80	75%
F6	864	Pteridium aquilinum subsp. aquilinum chloroplast, complete genome	0.0	99%
G1	361	Nipponolejeunea subalpina 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and 26S ribosomal RNA gene, partial sequence	1,00E-151	94%
G2	853	Selaginella moellendorffii hypothetical protein, mRNA	7,00E-79	76%
G3	1104	PREDICTED: Gossypium raimondii bifunctional epoxide hydrolase 2-like (LOC105783703), mRNA	6,00E-06	78%
G4	1645	PREDICTED: Fragaria vesca subsp. vesca elongation factor 1-alpha (LOC101305005), mRNA	0.0	85%
G5	934	PREDICTED: Nelumbo nucifera probable 2-carboxy-D-arabinitol-1-phosphatase (LOC104587139), mRNA	6,00E-15	70%
G6	565	PREDICTED: Elaeis guineensis protein ASPARTIC PROTEASE IN GUARD CELL 1-like (LOC105047320), mRNA	6,00E-08	86%
H1	149	Polypodium vulgare 28S ribosomal RNA gene, partial sequence	3,00E-69	99%
H2	864	E.coli-K.pastoris shuttle vector pPpHIS4, complete sequence	0.0	100%
H3	145	Polypodium vulgare 28S ribosomal RNA gene, partial sequence	3,00E-65	99%
H4	1372	O.sativa hsp70 gene for heat shock protein 70	0.0	83%
H5	1590	Arabidopsis thaliana similar to dihydroflavonol reductase (T23G18.6) mRNA, complete cds	4,00E-153	77%
H6	839	PREDICTED: Vitis vinifera probable RNA-binding protein 46 (LOC100853981), transcript variant X4, mRNA	7,00E-04	76%

Two clones (E1, F2) contained the stuffer fragment arising due to incompletely digested vector. Two clones showed no insert and one clone resulted in an overlay of two sequencing spectra signals most likely due to a mixed colony. Five clones yielded no result at blastn search although containing an insert larger than 1000 bp. Two further clones were empty and blastn resulted in *E. coli-K.pastoris* shuttle vector where probably the transcription terminator was sequenced due to the lack of an insert. Also chloroplast DNA from *P. aquilinum*, the organism from which the RNA was isolated for cDNA generation, was found within the library. 11.6% were ribosomal RNA contaminations, 96% of all clones had an insert. The average insert size disregarding the ribosomal RNA contaminations was 1302 bp. These numbers correspond to the quality criteria offered by LGC Genomics for library generation and sequencing. The sequencing results of the non size-selected library did not correspond to the standard quality criteria and are therefore not included.

3.2 VECTOR DEVELOPMENT AND CHARACTERIZATION

The successful screening of large cDNA libraries requires the efficient expression of genes of interest on a suitable expression vector. In order to be able to functionally screen the cDNA libraries generated in 3.1 in at least the two different yeast species *Pichia pastoris* and *Schizosaccharomyces pombe* an appropriate expression vector was developed. This vector contains an antibiotic resistance cassette for selection of positive transformants in yeast as well as in *E. coli*, a promoter and terminator for expression of the cDNA in yeast and an *E. coli* origin of replication for plasmid propagation.

As starting point for expression vector development the in-house T4-based plasmid was used [64]. This parental plasmid contained the inducible P_{AOX1} from *P. pastoris* and also the *AOX1* transcription terminator, as well as the pUC origin of replication and a zeocin resistance cassette for selection in *E. coli* and yeast.

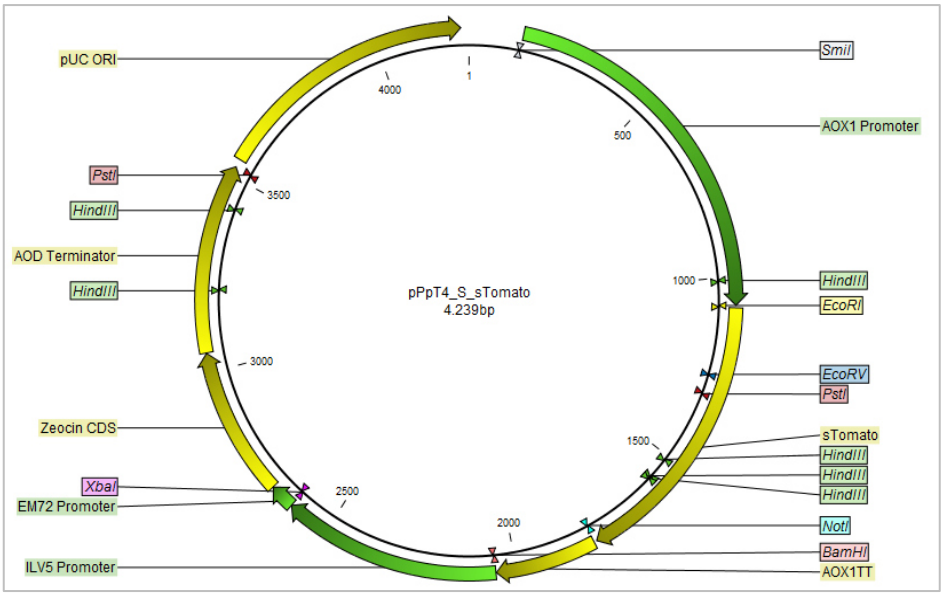


Figure 3.10 In-house T4-based *E. coli – P. pastoris* shuttle vector used as starting point for library expression vector development.
pUC= *E. coli* origin of replication for multicopy retainment and amplification.
AOX1 promoter= *P. pastoris* promoter from *AOX1* gene for gene of interest expression.
AOX1TT= transcription terminator from *AOX1* gene for GOI transcription termination
sTom= sTomato CDS, red fluorescence protein as reporter gene
ILV5 Promoter= *P. pastoris* promoter of the ILV5 gene for resistance protein expression in yeast
EM72= syntethic promoter for resistance protein expression in *E. coli*
AODTT= terminator for resistance protein expression

3.2.1 Cloning of the expression vector

For heterologous gene expression in *P. pastoris* linearized DNA fragments are usually integrated into the host's genome targeted to a specific locus or randomly integrated throughout the chromosomes. This requires linearization of the previously in *E. coli* amplified vector. For linearization of cDNA libraries this is a critical point due to lack of knowledge about insert sequence and therefore the undesired digestion within the gene of interest. For linearization of the plasmid *Smi*I is often used which has the recognition site 5'-ATTT↓AAAT-3' and is an eight basepair cutter which cuts statistically every 65536 bp.

The plasmid system designed in this study should predominantly be used to screen for interesting enzymes in plants and fungi. The genomes of these organisms have sizes of several Mbp, which results in a quite high probability to find a restriction site within the cDNA fragment cloned into the vector. For the same reason episomal plasmids were used as expression vector as described in 3.2.2.

The same consideration was applied for the cloning itself. Using conventional restriction site cloning with commonly employed enzymes such as *Eco*RI or *Not*I would potentially fragment the cDNA library insert since these enzymes have a cutting probability that would severely limit the usage of these vectors in a screening.

One alternative to circumvent the fragmentation probability issue is to design unique linearization site exploiting the existence of type 2S restriction enzymes such as *Sfi*I. This enzyme has two recognition sites *Sfi*IA and *Sfi*IB (5'- GGCCNNNNGGCC-3') where NNNN can be chosen randomly. However, the enzyme needs a stuffer of at least 170 bp to from a loop structure and cut the DNA properly [65]. As a stuffer fragment we decided to use the 750 bp long CDS of sTomato, a red fluorescence protein, as it gives us also the possibility to characterize the plasmid concerning promoter/terminator activity as well as expression levels as a reporter gene.

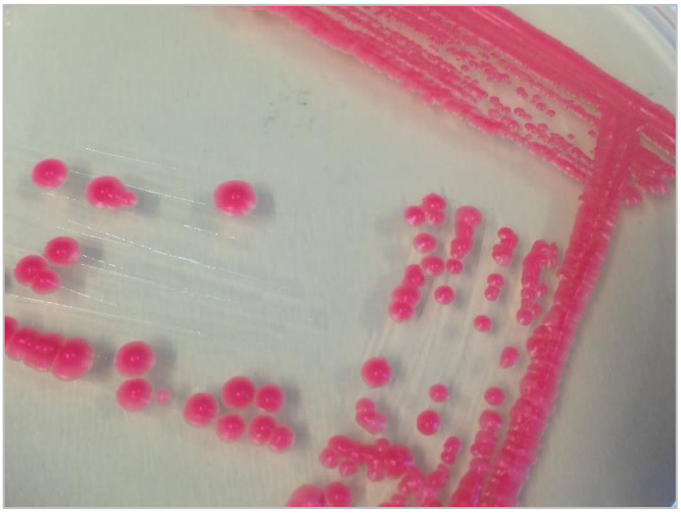


Figure 3.11 *S. pombe* cells expressing the reporter gene sTomato.
Cells were cultivated on YES agar plates supplemented with 100 mg/L zeocin for three days.

We designed primers containing *Sfi*IA (fwd primer) and *Sfi*IB (rev primer) and amplified sTomato, as well as P_{AOX1} and *AOX1TT* to construct a vector backbone containing the appropriate restriction sites. The sTomato PCR product was digested with *Sfi*i and ligated into the expression vector. This also acted as a test to evaluate the *Sfi*I cloning efficiency. As the cloning yielded positive transformants in *E. coli*, *P. pastoris* CBS7435 cells were transformed with 1 µg of *Smi*I linearized plasmid DNA, positive transformants were selected on zeocin containing agar pates and then inoculated in YPD for fluorescence screening.

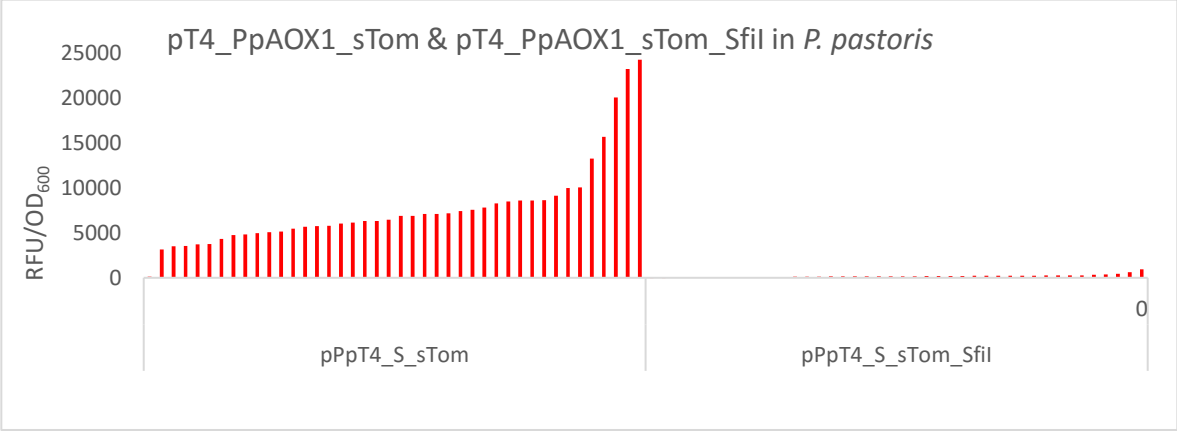


Figure 3.12 Comparison of sTomato expression values with and without *Sfi*I restriction site before and after the CDS in *P. pastoris.*
Cells were cultivated for 60 h on glucose followed by 24 h of methanol induction. In presence of *Sfi*I restriction site directly in upstream of the sTomato CDS hardly any fluorescence protein is measurable.

Figure 3.12 shows that hardly any reporter gene expression is observable if the *Sfi*I restriction site is present directly upstream of the reporter gene CDS. Compared to the construct without the restriction site, where the average fluorescence level was 8000 RFU/OD_{600}, the *Sfi*I construct showed 32 times lower expression. This indicates that the sequence of the restriction site *Sfi*IA, which has a high GC content (62.5%: GGCCATTTTGGCC) results in a decreased expression efficiency either through lowering transcript levels of alternatively affecting translation processes. As it was already known that GC rich sequences at the 3' end of the promoter region may decrease promoter efficiency, the constructed *Sfi*IA site was already as AT-rich as possible.

The testing of *Sfi*I restriction site influence on expression levels was also performed with the pGAZ2 plasmid (see dissertation of Alexandra Andryushkova) and *S. pombe.* The expression of the reporter gene was driven by P_{PGK1} and we observed a similarly strong influence on expression levels in this background.

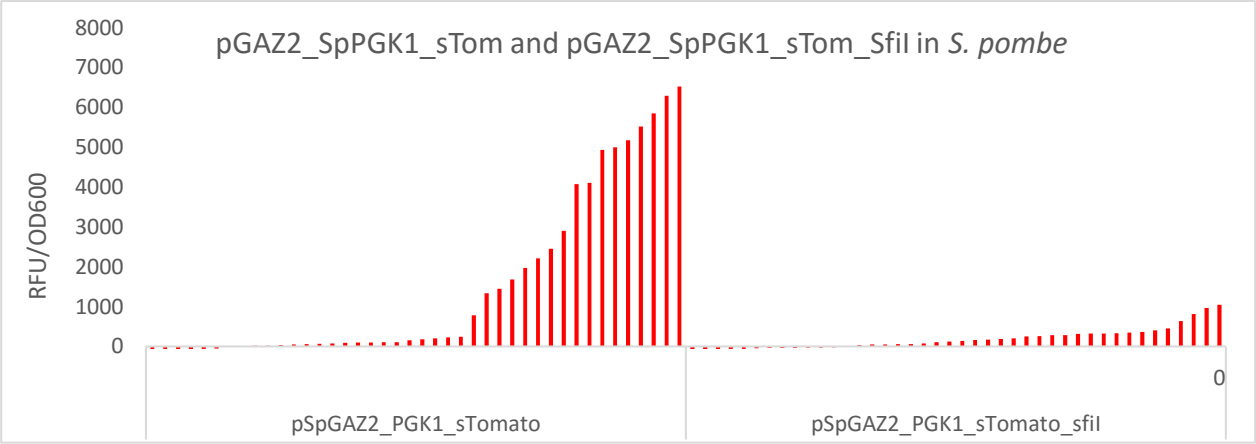


Figure 3.13 Comparison of sTomato expression values with and without *Sfi*I restriction site upstream and downstream the reporter gene CDS in *S. pombe.*

Cells were cultivated for 60 h in 400 µL YES. In presence of the *Sfi*I restriction site directly upstream of the sTomato CDS hardly any fluorescence protein was detectable.

Also for *S. pombe* the comparison of the same constructs with and without the *Sfi*I restriction site at the 3' end of the promoter region (upstream of the start ATG) the same effect was visible as Figure 3.13 shows. Fluorescence levels of the P_{PGK1} driven expression with the *Sfi*I restriction site directly upstream of the start ATG showed 7.5 times lower expression level as without *Sfi*I. These

results indicated the restriction site to be the reason for the significantly lower expression for both yeast species.

If the high GC content of the *Sfi*I restriction site was the reason for the decreased expression, we wanted to test if an additional TATA box downstream the restriction site and upstream of the start ATG can increase the expression levels. The following constructs were designed.

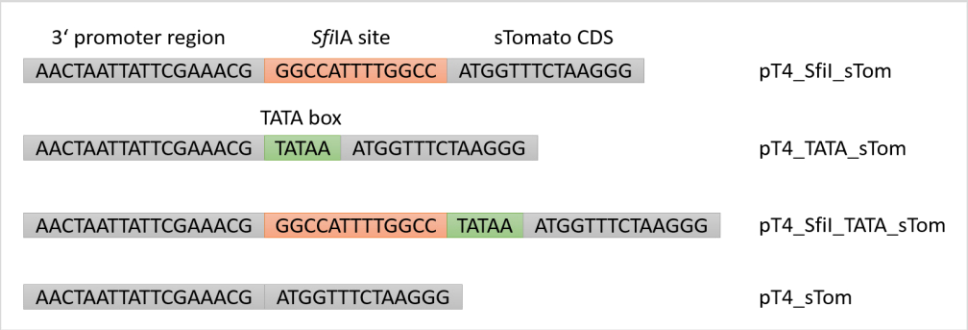


Figure 3.14 Constructs for testing TATA box influence on expression levels.

pT4_SfiI_sTom was the previously screened construct, whereas in pT4_TATA_sTom the *Sfi*I restriction site was replaced by TATAA to observe also the influence of only a TATA box on the reporter gene expression level. pT4_SfiI_TATA_sTom contained the *Sfi*I restriction site followed by the TATA box to see if expression can be improved. pT4_sTom contained neither the *Sfi*I site nor the TATA box.

The constructs were cloned in the T4 plasmid (shown in Figure 3.10), *P. pastoris* cells were transformed with linearized plasmid DNA and screened for reporter gene expression. The fluorescence levels of the different constructs are shown in Figure 3.15.

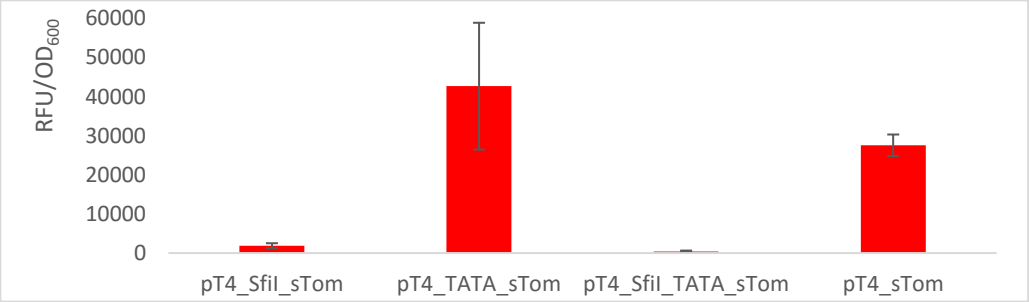


Figure 3.15 Reporter gene expression in *P. pastoris* with different comibinations of *Sfi*I restriction site and additional TATA boxes.

The figure shows the reporter gene expression normalized by OD_{600} after 60 h of methanol induction for P_{AOX1}-*Sfi*I-sTomato, P_{AOX1}_TATA_sTomato, P_{AOX1}_*Sfi*I_TATA-sTomato and the original P_{AOX1}_sTomato. The highest expression is possible without *Sfi*I site and additional TATA box, whereas the lowest expression was observed by the construct containing *Sfi*I and TATA box.

The results obtained show again hardly any expression for the constructs containing the *Sfi*I restriction site even with additional TATA box added as can be seen in Figure 3.15. GC-rich restriction sites downstream P_{AOX1} have in general a significant influence on promoter activity in terms of decrease in expression (pers. communication with Dr. Claudia Ruth). Although the GC content of *Sfi*IA was kept as low as possible – the 4 bases in between the recognition site were chosen GGCCATTTGGCC – it seems the GC content is still high enough to avoid proper expression. The addition of a TATA box directly upstream of *Sfi*IA did not result in any increase in reporter gene expression. This observation poses serious obstacles to efficient cDNA library expression and was therefore ruled out as potential method for library cloning.

To circumvent restriction sites influencing promoter activity we chose cutting sites already tested in prior studies, e.g. *Eco*RI and *Spe*I. The reporter gene flanked by an *Eco*RI and *Not*I restriction site was then cloned into the existing vector and tested again for fluorescence levels. No influence on promoter activity could be observed. However, due to the higher probability of digestion compared to *Sfi*I cloning and therefore a potential loss of cDNA inserts a restriction free cloning technique in the form of assembly cloning was employed [49].

To allow the screening of identical cDNA libraries without the need of subcloning into different expression vectors suitable parts comprising of promoters, selection markers and autonomous replicating sequences functionally in both species had to be designed.

3.2.2 Autonomously replicating sequence

Autonomously replicating sequence elements (ARS) are sequence elements that promote extrachromosomal maintenance of plasmid DNA. In eukaryotes they were first identified in *Saccharomyces cerevisiae* and function as an origin of replication [35,66]. Autonomous replicating plasmids have a big advantage for our purpose: The lack of necessary linearization for genomic integration in yeast prevents any loss of sequence due to restriction enzyme mediated fragmentation and therefore allows a more efficient cDNA library screening.

The family of pGAZ2 plasmids contain the *ars1* from *S. pombe* [36] in a shortened version (868 bp). We first tested if this autonomous replicating sequence is also functional in *P. pastoris*. 100 ng of plasmid DNA, which contained the zeocin resistance cassette under control of *A. gossypii* P_{TEF1}, were transformed in *P. pastoris* CBS7435 cells and cultivated on zeocin selection plates. This resulted in a transformation efficiency of 10^5/µg plasmid DNA and the cells were able to grow on selective plates. After this positive result, the *ars1* was cloned into our library expression vector.

3.2.3 Promoters for GOI expression

The P_{AOX1} promoter is driving the expression of the GOI in the basic plasmid, which is the most widely used promoter for heterologous gene expression in *P. pastoris* [67], the first idea was to use this promoter also for the library expression. Though the inducibility of P_{AOX1} by methanol is a big advantage for protein production at high levels, methanol is an expensive and inflammable compound. In addition, the strength of P_{AOX1} is not suitable for expressing unknown enzymes in terms of their complexity in folding or possible toxicity in too high amounts.

In first instance the promoter of choice was P_{CAT1}, the promoter of catalase 1 from *P. pastoris* [42]. The 500 bp version of the *CAT1* promoter was cloned into the library expression vector replacing P_{AOX1} and expression of the reporter gene sTomato was analyzed in *P. pastoris* in media containing different zeocin concentrations to examine which antibiotic concentration should be used for further experiments.

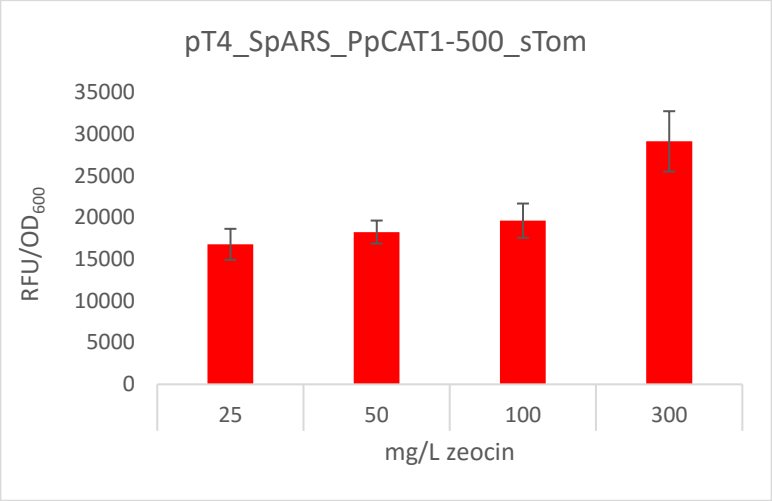


Figure 3.16 Average fluorescence values of sTomato normalized by OD_{600} of *P. pastoris* cells in YPD containing 25-300 mg/L zeocin.

For each antibiotic concentration 40 clones were cultivated in 96-well plates in 400 µL YPD. 25 mg/L weren't an option because the wildtype still grew. The difference in expression level with 50 mg/L zeocin was only 7% lower than with 100 mg/L.

Figure 3.16 shows the reporter gene expression levels normalized by OD_{600} of *P. pastoris* clones harboring pT4_SpARS_PpCAT1-500_sTom cultivated in YPD with different antibiotic concentrations. As we were developing a plasmid for library screening where thousands of clones have to be screened, our focus was also to find an antibiotic concentration that is feasible.

25 mg/L zeocin were not able to avoid wildtype growth completely, while on the other hand the high fluorescence values with 300 mg/L do not warrant the added push in fluorescence rptein expression. This result led us to the decision to use 50 mg/L zeocin for all further experiments in *P. pastoris*.

In *S. pombe* the promoter was also characterized. 1 µg of plasmid DNA was used for transformation of *S. pombe* 972h- cells and transformants were cultivated for 60 h before fluorescence and OD_{600} were measured.

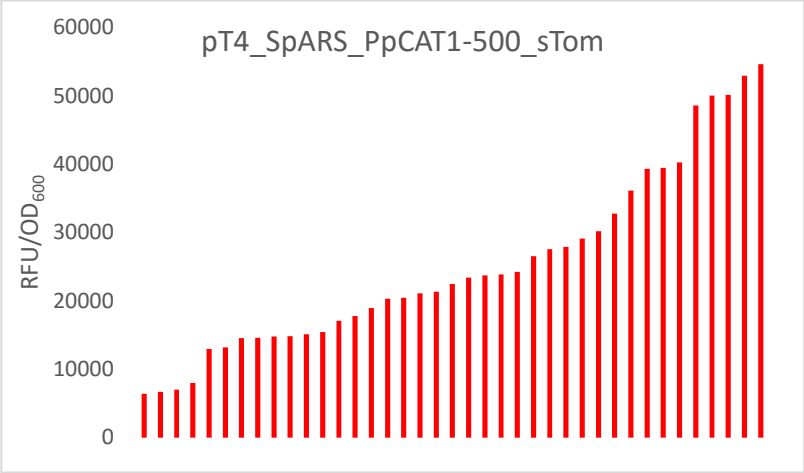


Figure 3.17 Fluorescence values normalized by OD_{600} of 40 *S. pombe* clones harboring pT4_SpARS_PpCAT1-500_sTom.
Cells were cultivated in 96-well plates in 400 µL YES supplemented with 100 mg/L zeocin. Fluorescence and OD_{600} were measured after 60 h of growth.

As Figure 3.17 shows the landscape of 40 observed *S. pombe* clones that yield a standard deviation of 50% in comparison to *P. pastoris* with a standard deviation of 7% with the same plasmid. This leads to the conclusion that there might be differences in the copy number [36]. Also the difference in colony size on the transformation plates gives indications into this direction. Whether this effect originates from differences in transformation efficiency or develops during plasmid replication still needs to be determined in future experiments.

Also the methanol inducibility of P_{CAT1} on the episomal plasmid was examined in order to have an option for higher expression levels.

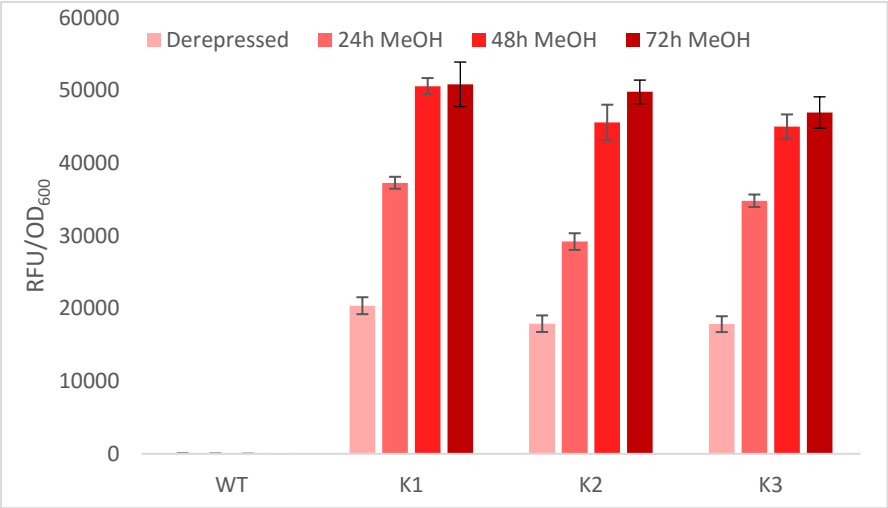


Figure 3.18 Average fluorescence values of sTomato normalized by OD_{600} in *P. pastoris* cultivated for 60 h in YPD followed by 72 h MeOH induction.
The inducibility of P_{CAT1} is also given for episomal expression. Three clones harboring pT4_SpARS_PpCAT1-500_sTom were cultivated in 50 mL YPD supplemented with 50 mg/L zeocin, as control the wildtype CBS7435 without plasmid was cultivated in YPD.

Figure 3.18 shows the expression levels of sTomato normalized by OD_{600} of *P. pastoris* cultivated in YPD to observe derepressed activity and also induced activity. The observed expression values after 72 h of methanol induction exceed the values of genomic integrated constructs shown by Thomas Vogl and colleagues 2015 by two times [42].

Although P_{CAT1} is tightly repressed on glucose, which allows the cells to reach the logarithmic growth phase before expression starts and makes it therefore suitable for potentially harmful enzymes, it is a rather strong promoter. In general, medium strong or rather weak promoters are more desirable for our screening for system. A strong promoter may lead to death of the cells by stress or toxicity of the expressed protein. We wanted to find an additional weaker promoter functional in both yeasts. Literature research led us to Wang and colleagues, where they reported in 2014 the promoter of *ENO101* from *S. pombe* to be as active as P_{NMT1}. Homology search in *P. pastoris* yielded over 70% identity in the *ENO101* coding sequence to *P. pastoris ENO2* as Figure 3.19 shows.

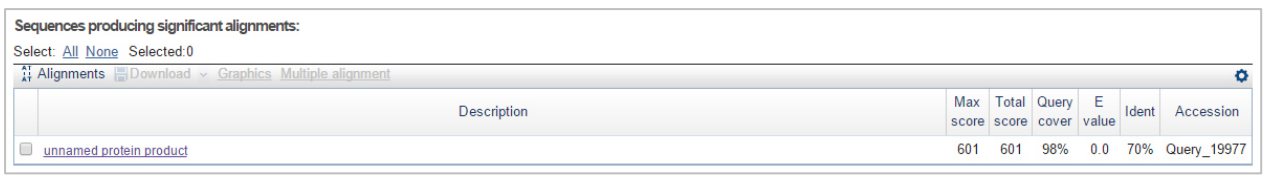


Figure 3.19 blastp alignment of *S. pombe ENO101* and *P. pastoris ENO2.*
The two sequences showed an identity of 70% as well as 98% query cover.

The first idea to amplify P_{ENO101} from *S. pombe* genome was discarded because none of the *S. pombe* promoters tested showed sufficient activity in *P. pastoris* as Figure 3.20 shows:

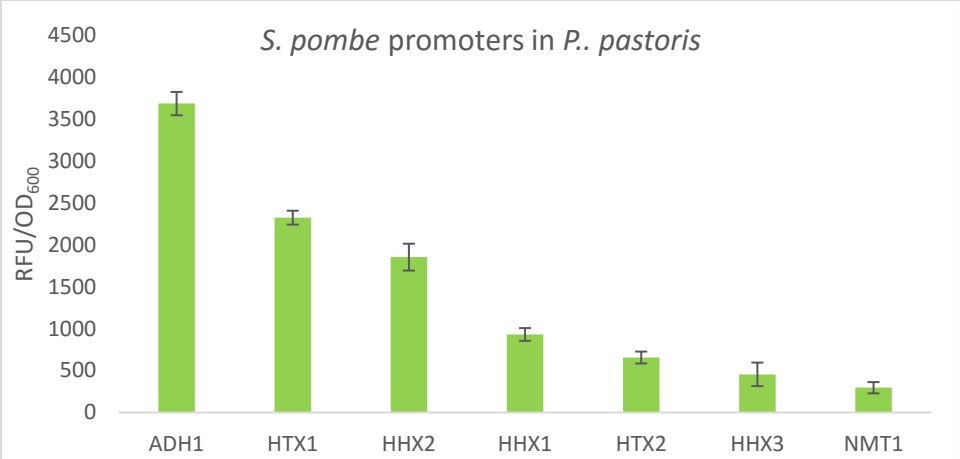


Figure 3.20 *S. pombe* promoters activity in *P. pastoris.*
As reporter gene eGFP was used. The cells were cultivated for 60 h in 400 µL YPD supplemented with 50 mg/L zeocin.

After this result the last 1000 bp upstream from *ENO2* from the *P. pastoris* genome were amplified and cloned into our expression vector, used for transformation of *P. pastoris* and *S. pombe* and screened for reporter gene expression.

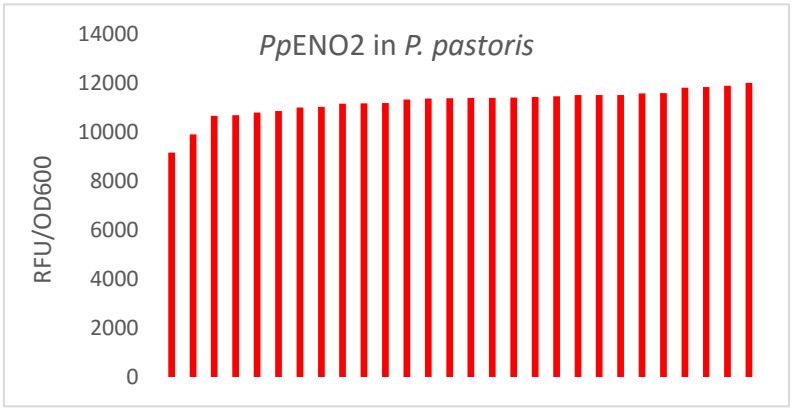


Figure 3.21 Fluorescence values normalized by OD_{600}, expression was driven by P_{ENO2} from *P. pastoris.*
The cells were cultivated for 60 h in YPD supplemented with 50 mg/L zeocin.

As can be seen in Figure 3.21 P_{ENO2} of *P. pastoris* shows an average fluorescence of 11200 normalized by OD_{600}. This can be confirmed as a medium strong promoter, compared to P_{CAT1} that yielded an average fluorescence of 18300. We hypothesized weak or medium strong promoters are more suitable for library screenings to avoid problems concerning stress of the cells because

of product toxicity or protein aggregation due to incorrect folding. For further characterization of the plasmid we kept using P_{CAT1}.

As the promoter should be functional also in *S. pombe,* competent cells were also transformed 2.5 µg of plasmid DNA via LiAc method (section 2.9.10), positive clones were inoculated in selection medium and fluorescence protein expression was measured after 60 h of growth.

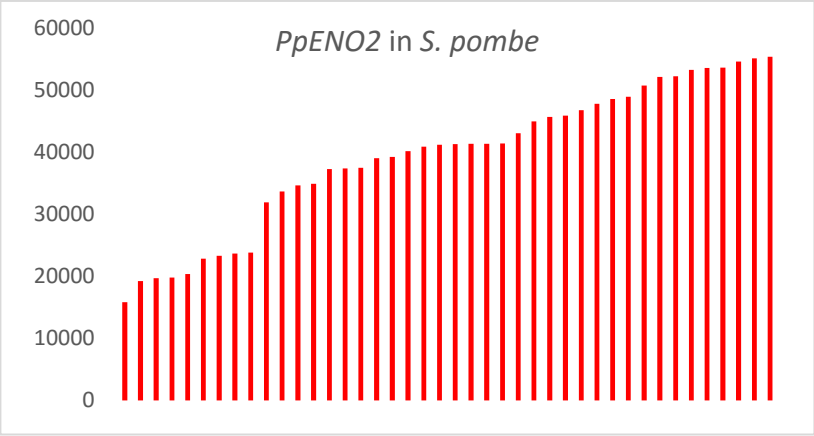


Figure 3.22 Fluorescence values normalized by OD_{600}, expression was driven by P_{ENO2} from *P. pastoris* in *S. pombe.*
The cells were cultivated for 60 h in 400 µL YES supplemented with 100 mg/L zeocin.

The Figure 3.22 shows the fluorescence measurement of *S. pombe* expressing sTomato under control of P_{ENO2}. The expression level is not as uniform as for *P. pastoris* (Figure 3.21) but has only 29% standard deviation. With this result a second promoter was found suitable for expression in both yeast species.

3.2.4 Selection marker

As described in section 3.2.2 autonomously replicating plasmids have great advantages for library screening. The biggest disadvantage, at least for the ARS used during this thesis, is the necessity to keep the selection pressure upright during the whole cultivation to retain the plasmid. As the purpose for the developed plasmid system was including functionality in *S. pombe* and *P. pastoris,* the selection marker of choice was antibiotic selection with zeocin. The *S. h. ble* gene is reported to deliver zeocin resistance in *P. pastoris* [44] as well as in *S. pombe* [47].

3.2.4.1 Promoter for Sh *ble* expression

On the basic vector described in 3.2.1 the promoter driving the expression of the zeocin resistance protein was $P_{ILV\text{-}5}$ for selection in yeast, fused with P_{EM72} for selection in *E. coli.* As Figure 3.16 shows, the fusion promoter was functional in *P. pastoris.* In *S. pombe* antibiotic selection with this promoter was not possible, since no transformants were observed after transformation. On pGAZ2 plasmids, the promoter driving zeocin resistance protein expression is P_{TEF1} of *A. gossypii* as well as the corresponding terminator. The *TEF1* gene and its promoter region were first described by Steiner and Philippsen in 1994 [69] and claimed the promoter and terminator region of transcription elongation factor 1α [70]. As described in section 3.2.2, where the functionality of *S. pombe ars1* was examined by transformation of *P. pastoris* with the pGAZ2 plasmid, antibiotic selection was possible in *P. pastoris* as well. This indicated that P_{TEF1} also enables functional zeocin resistance protein expression in *P. pastoris.* For enabling the usage of the vector for both yeast species, the original resistance cassette was replaced by P_{TEF1} resistance cassette in the parental vector and transformed in *P. pastoris* and *S. pombe.*

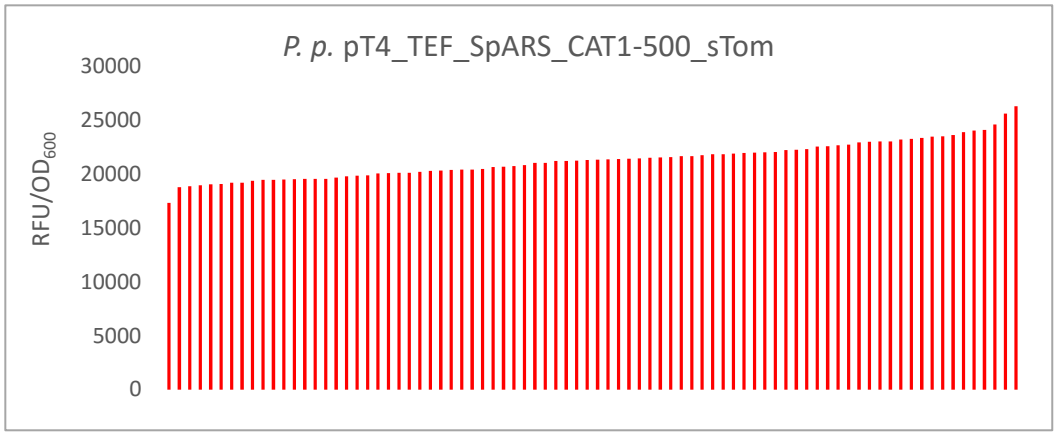


Figure 3.23 Reporter gene expression in *P. pastoris* with P_{TEF1} resistance cassette.
sTomato fluorescence normalized by OD_{600} after 60 h of growth in YPD supplemented with 50 mg/L zeocin. The figure shows that the P_{TEF1} resistance cassette was also functional in *P. pastoris*.

The fluorescence levels shown in Figure 3.23 were similar to that with P_{ILV-5} resistance cassette (21000 for P_{TEF1} and 18300 for P_{ILV-5}) as well as the standard deviation (7% for P_{TEF1} and 10% for P_{ILV-5}), the results for P_{ILV-5} are shown in Figure 3.16. In addition, P_{TEF1} was also functional in *E. coli,* where the vector was amplified after cloning, because no additional *E. coli* promoter was present on the vector. Interestingly no consensus Shine-Dalgarno sequence can be found in P_{TEF1}. The BPROM promoter prediction tool recognized though a promoter region for *E. coli* as can be seen in Figure 3.24.

```
>PTEF1
 Length of sequence-       446
 Threshold for promoters -  0.20
 Number of predicted promoters -      1
 Promoter Pos:    169 LDF-  7.21
 -10 box at pos.    154 ATGTATAAT Score    82
 -35 box at pos.    137 TTTACC    Score    25

 Oligonucleotides from known TF binding sites:

 For promoter at    169:
       narL:  TCCCCATG at position     149 Score -   6
     rpoD16:  TGTATAAT at position     155 Score -  13
       argR:  ATAATCAT at position     158 Score -   9
        fur:  ATTTGCAT at position     164 Score -   7
       argR:  CATCCATA at position     169 Score -   6
       lexA:  TCCATACA at position     171 Score -   9
```

Figure 3.24 BPROM promoter prediction result of P_{TEF1} for *E. coli*.
Promoter prediction by BPROM [76] predicted a promoter region for *E. coli* within *A. gossypii* P_{TEF1} at position 169.

Although no Shine-Dalgarno sequence was present, P_{TEF1} was a functional promoter in *E. coli* as growth on antibiotic containing plates was possible after transformation with the developed vector. Scharff et al. showed in 2011 that a typical Shine-Dalgarno consensus sequence is not obligatory for ribosome binding in *E. coli*. This seems to be the case for P_{TEF1}.

3.2.4.2 *GUT1* complementation

Disregarding the possibility to be able to use one vector in two yeast species, another purpose was to test the plasmid system developed in this thesis in terms of feasibility for industrial applications. For industrial purposes a plasmid system would be in general feasible because of easy transformation and uniform expressing clones. The only hindering fact with our system is the necessity to use antibiotic during the whole cultivation to force the cells to retain the plasmid.

For *P. pastoris* the *GUT1* complementation marker offers the possibility to circumvent the use of classical antibiotic selection markers. *GUT1* codes for a glycerol kinase, an essential enzyme for carbon metabolism. The strain *P. pastoris* BG10ΔGUT1 lacks this glyercol kinase in its genome and is therefore only able to grow on glycerol as sole carbon source when harboring the plasmid complementing this knockout.

For that purpose, we replaced the Sh *ble* CDS by *GUT1* and *P. pastoris* BG10ΔGUT1 cells were transformed with 100 ng of plasmid DNA. Positive transformants were selected on BMG agar plates for three days at 28°C, because only cells expressing the glycerol kinase coded on the vector were able to grow on glycerol as sole carbon source. For reporter gene fluorescence screening 96 well plates with 400 µL liquid BMG medium were inoculated with positive transformants and grown for 60 hours. The result of the fluorescence measurement is shown in Figure 3.25.

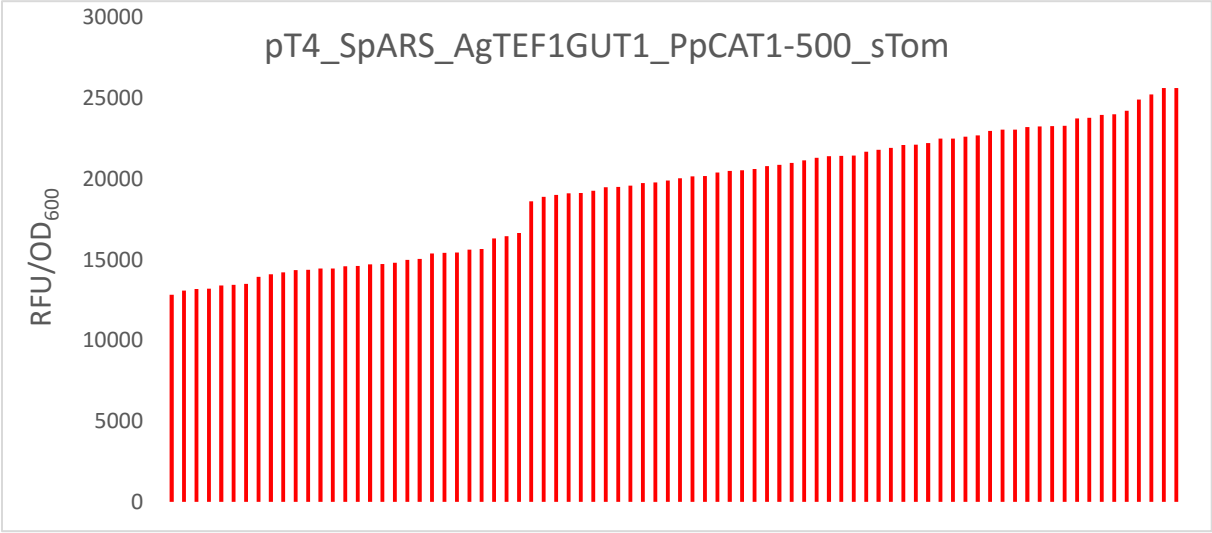


Figure 3.25 Fluorescence measurement of *P. pastoris* BG10ΔGUT1 cells expressing sTomato.
The cells were cultivated for three days in BMG which contained 1% glycerol as sole carbon source.

The average fluorescence of the reporter gene normalized by OD_{600} is 19200. This is even higher than in complex media like YPD supplemented with 50 µg/mL zeocin, where the average fluorescence was 18300. This result demonstrates that complementation marker are a good alternative to antibiotic resistance markers, also for episomal expression. Especially for libraries where ten thousands of clones are usually screened, application of antibiotic markers is expensive. For application of the *GUT1* marker only minimal medium with glycerol as sole carbon source is required that is also cheaper than a complex media.

3.2.5 Functional enzyme expression and secretion of *C. antarctica* lipase B

As the vector is intended to be used for functional screening for enzymes, it was also examined the functional enzyme expression with the ARS plasmid system presented here. Therefore, *Candida antarctica* lipase B was cloned into the vector under the control of P_{CAT1} as well as upstream of the *CalB* the alpha factor from *S. cerevisiae* was included for proper secretion of the enzyme. *P. pastoris* CBS 7435 cells were transformed with 100 ng of plasmid DNA and an enzymatic assay with pNPB as substrate was performed to screen for functional lipase expression. Figure 3.26 shows the results of the enzymatic assay.

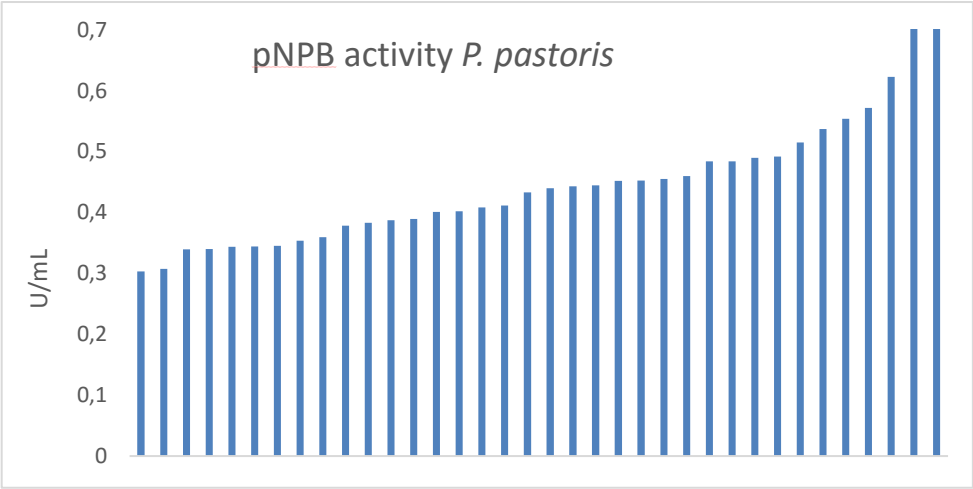


Figure 3.26 Activity of functionally expressed *CalB* on p-nitrophenole butyrate secreted by *P. pastoris*. Clones that functionally expressed from the autonomous replicating plasmid pT4_SpARS_AgTEF1zeo_pCAT1-500_alphaCalB and secreted *CalB* were cultivated in 96-well plates under selective pressure. The assay was done with culture supernatant, absorbance was measured over 3 minutes at 405 nm wavelength.

As Figure 3.26 shows functional enzyme expression and secretion is possible with the ARS plasmid system for *P. pastoris*. The average *CalB* activity is 0.5 U/mL with 22% standard deviation when cultivation is done in YPD supplemented with 50 mg/L zeocin. Former results for genomically integrated expression cassettes that were used for *CalB* expression in *P. pastoris* in the minimal medium BMD yielded 0.1-0.2 U/mL activity on pNPB, which is 3-4 times lower [72]. This may be due to the effect of lower cell densities in minimal media, because the activity was not normalized to OD_{600}. Nevertheless, the expression with the ARS plasmid could be still increased by applying

higher antibiotic concentrations like it was shown for the reporter gene expression in Figure 3.16 or by methanol induction like shown in Figure 3.18.

Secretion of *CalB* was not possible in *S. pombe* due to the malfunctioning alpha factor from *S. cerevisiae* in that species. For enabling the secretion also in *S. pombe* other endogenous signal peptides of this yeast like *Dpp* or *cyp* could also be tested in *P. pastoris* to again enable the possibility to use the plasmid system for secretion of proteins in both yeast species [73].

For the purpose of this thesis an additional secretion signal on the vector is anyways not necessary. It was designed for naturally secreted enzymes like lipases, laccases or cellulases that should contain a native secretion signal.

3.2.6 Violacein pathway expression

The next step in plasmid characterization finally was to test if whole pathway expression is also feasible with the developed vector. For this experiment the violacein pathway of *Chromobacterium violaceum* was used. The 5 genes necessary for pathway expression were expressed in a polycicstronic way by use of a single promoter and viral 2A linker sequences as Figure 3.27 shows [74].

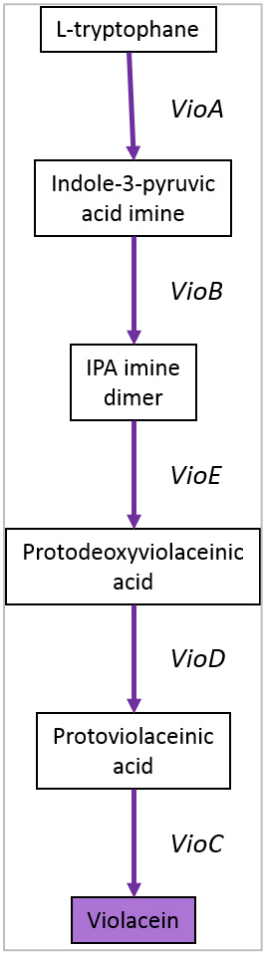


Figure 3.27 Violacein pathway with interspersing 2A sequences for polycistronic expression.
The different genes for VioA-VioE were arranged polycistronic with viral 2A sequences inbetween [74].

Therefore, I cloned the Violacein pathway with interspersing 2A sequences for polycistronic expression into the plasmid under control of P_{CAT1}, transformed *P. pastoris* CBS7435 cells with it and cultivated the transformants in YPD (0.5% glucose) with three different zeocin concentrations

for 60 h in deepwell plates. Successfully produced violacein accumulates in the cells expressing the pathway. Such successful whole pathway expression results in purple colored cells.

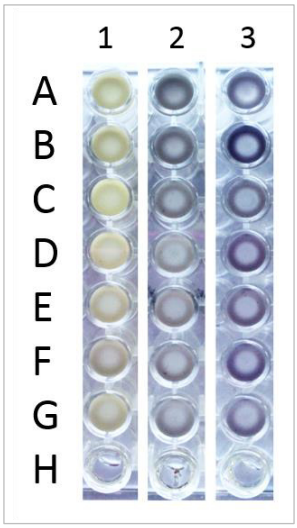


Figure 3.28 Episomal violacein pathway expression in *P. pastoris*.
Shown are cell pellets of 150 µL culture volume after centrifugation in a MTP. Slot 1 represents YPD1% with 50 mg/L, slot 2 200 mg/L and slot 3 300 mg/L zeocin. Row H was inoculated with *P. pastoris* CBS7435 wildtype.

Figure 3.28 shows the pellet of *P. pastoris* cell cultures after 60 h of growth in YPD supplemented with 50, 200 and 300 mg/L zeocin. As it is visible in the figure, cultures cultivated with 50 mg/L zeocin show hardly any visible violacein expression (slot 1). 200 mg/L (slot 2) result in a distinct violet color of the cells, as well as 300 mg/L (slot 3). These results show that also whole pathway expression was succesful in *P. pastoris* with the developed episomal multi host plasmid and fine-tuning of the expression can be implemented by adjusting the antibiotic concentration in the medium.

3.2.7 Plasmid stability

As described in 3.2.2 autonomously replicating plasmids have big advantages such as higher transformation rates and more uniform expression, however, the severe possible plasmid instability might challenge optimal protein expression. For this purpose the retention of the developed plasmid under non-selective conditions was examined over time and the percentage of cells harboring the plasmid was calculated as it is shown in Figure 3.29. Therefore, a liquid culture without selective pressure was inoculated from a preculture supplemented with 50 mg/L zeocin to OD_{600} 0.01. Every 3 hours samples were taken, for which the OD_{600} was measured. Based on these OD_{600} measurements the dilution to obtain approximately 50 cells/100 µL was calculated and the volume was plated on selective and non-selective plates. From these plates the percentage of cells retaining the plasmid was calculated.

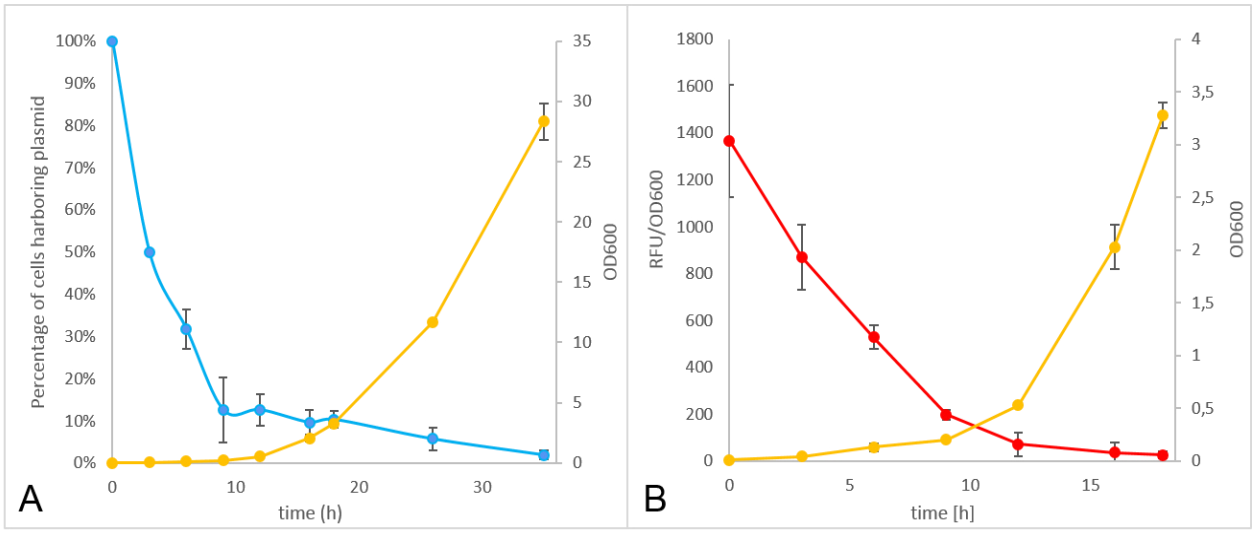


Figure 3.29 Plasmid stability in *P. pastoris* without selection pressure.
Panel A shows the percentage of cells harboring the plasmid (blue line) and so growing on selection plates compared to the total CFU on agar plates without antibiotic. The yellow line shows the OD_{600}. Panel B shows the relative fluorescence units of the reporter gene sTomato (red line) per OD_{600} compared to the OD_{600} (yellow line).

Figure 3.29 shows the plasmid stability of the developed expression vector without selection pressure. As it is visible in panel A after 3 hours of growth only 50% of the cells were retaining the plasmid, after 9 hours only 11%. These values correlated with the fluorescence measurements of

the reporter gene expression in panel B. Here about 64% of the fluorescence was measurable after 3 hours of growth and 15% after 9 hours of growth without selection pressure. The slightly higher percentage in the fluorescence compared to the percentage in panel A can be explained in the postponed expression compared to the growth, as the reporter gene expression is controlled by P_{CAT1} which is repressed by glucose. As well, the reporter protein is not degraded completely immediately after loss of the plasmid and also mRNA might remain for some time.

3.2.8 Library expression and test screening techniques

The vector characterized in sections 3.2.1 to 3.2.6 was then used to functionally screen for enzyme activity. Therefore, I tested two different methods for transforming cDNA libraries in *P. pastoris.* First, I tried to avoid the amplification of cDNA libraries in *E. coli* to avoid amplification within the library. For this purpose, the assembly cloning, done as described in 2.9.6, was used for direct transformation of electrocompetent *P. pastoris* CBS 7435 cells, and different volumes of the regeneration culture (100, 200 and 700 µL) were plated on antibiotic selection plates. After 3 days of incubation no colonies at all were visible on these plates. This result indicated that cDNA library amplification in *E. coli* is necessary.

Next, we wanted to prove if functional enzyme screening is possible with the cDNA libraries obtained as described in 3.1, which was generated from fungal RNA (organisms name is confidential). Since this organism is predicted to express and secrete lipases, an assay with *p*NPB assay was performed. In addition, I also screened for laccases in the *Pteridium aquilinium* cDNA library.

Four different assays were performed to screen for enzyme expressing cDNA clones.

- pNPB microtiter plate assay for lipase screening

  For this assay the cells were cultivated in 96 well plates in YPD supplemented with 50 mg/L zeocin. After 60 h of growth the cells were pelleted and the supernatant was transferred into a microtiter plate, where the substrate was added. The conversion was measured over three minutes in a spectrophotometric plate reader. This was done for 252 clones, with none of them expressing a lipase. The highest activity measured was 0.008 U/mL.

- ABTS plate assay for laccase screening [52]

  To screen for laccases, the cells were plated directly after transformation on ABTS and $CuSO_4$ containing plates. The plates were incubated for three days at 30°C and afterwards kept at 4°C for further seven days. A positively expressed laccase should result in a green-blue halo around the expressing colony. With ten agar plates, each containing 80-100 colonies, there was no laccase activity detected.

- Bromphenole-blue plate assay for laccase screening [75]

For laccase screening a bromphenole-blue plate assay was performed. As described above for the ABTS assay, the cells were plated directly after transformation on bromphenole-blue containing plates. Laccase expressing clones should degrade the dye which results in a decolorization. Also with this assay no laccases activity could be detected.

These were initial experiments to test if functional screenings are possible on microtiter plate and agar plate level. For a reliable library screening about ten thousand clones should be screened.

3.3 CONCLUSION AND OUTLOOK

During this thesis a multi host vector system had been developed to screen cDNA libraries in at least two different yeast species which can also be applied in other fields. Multi host vectors are a useful tool for the transformation of different hosts with the physically same vector in order to compare protein expression directly. Steinborn and colleagues developed such a multi host vector system in 2006 for *Arxula adeninivorans, Hansenula polymorpha* and other yeasts [39]. The vector system developed in this thesis contained a zeocin selection marker where resistance protein expression was driven by *A. gossypii TEF1* promoter and terminated by the *TEF1* terminator, as promoter for GOI expression P_{CAT1} of *P. pastoris* and the autonomously replicating sequence *ars1* from *S. pombe*. All of these components were shown to be functional in *P. pastoris* as well as in *S. pombe.* With this system also whole pathway expression and functional enzyme expression was possible in *P. pastoris,* similar to the vector system developed by Steinborn and colleagues[39], as well as expression level fine-tuning by adjusting selective pressure in the culture media.

An interesting future experiment would be to express whole pathways as well as secreted enzymes with the developed vector system in *S. pombe,* employing a functional secretion signal for this organism, and maybe also in other yeasts like *S. cerevisiae.* In addition, more different regulated promoters, as well as selection markers, terminators and secretion signals that are functional in several yeast species would extend the possible applications of this multi host vector.

The advantages of an episomal expression system for cDNA fragments are not only the simple transformation procedure but also preventing the library from undesired digestion and sequence loss during linearization of the plasmid DNA due to lack of knowledge about restriction sites on the cDNA fragment. Due to this uncomplicated transformation possibility also functional screenings via plate assays can be performed directly on the selection plates after transformation, such as a pNPB assay for lipases or ABTS and bromphenole-blue assays for laccases.

Concerning the cDNA library generation method developed here it can be summed up that a reliable and simple method for cDNA generation had been established to be independent from sometimes non-reliable or available kits from commercial suppliers. In addition, by custom designed primers one is able to clone the generated cDNA in every in-house vector and

subsequent expression in preferred in-house hosts. During this thesis different reverse transcriptases, primers, thermal cycling protocols and buffers were tested for cDNA generation. The best results were obtained by employing the wildtype M-MLV reverse transcriptase, a mixture of 12 poly-d(T) primers with every combination of two different anchor nucleotides, the self-made buffer where the DTT concentration was lowered to 2 mM and 15 cycles for second strand synthesis.

From the RNA isolation to cDNA generation to library expression in the desired host ten days of working time are necessary. Another positive aspect of this method is that RNA from any organism can be used for the cDNA generation, as the starting material is always total RNA. The suitable developed vector described above completes the flexibility of this system.

To sum up results it can be stated that using this method including the adaptable primer design it is possible to generate expression libraries from any source organism and clone it into any vector.

The cDNA libraries generated using this method were consistent with the quality criteria for cDNA libraries from LGC genomics although the length of the generated cDNA fragments could still be optimized. The average insert length was 1302 bp, 11% were ribosomal RNA contaminations and 96% of the examined clones contained a cDNA fragment as insert as sequencing showed. Extending the cDNA fragment lengths should increase the probability to find full-length genes in order to find functional expressed enzymes. This could be done by further adjusting of the reaction parameters during cDNA synthesis.

4 APPENDIX

4.1 GLYCEROL STOCKS

Table 4.1 List of glycerol stocks of strains generated during this thesis including selection marker.

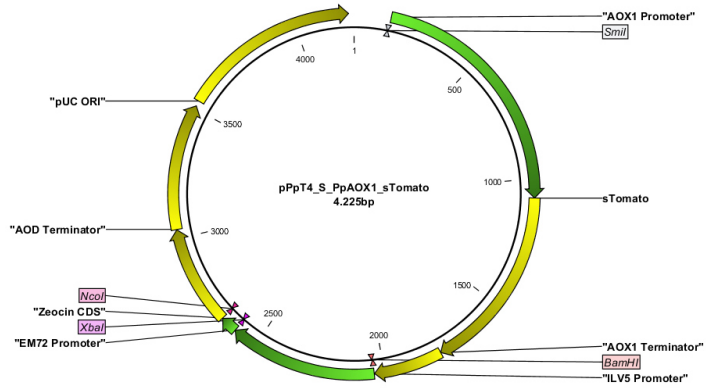
Strain	Plasmid/Cassette	Marker	IMBT #
S. pombe 972h-	-	-	
S. pombe 972h-	pT4_SpARS_AgTEF1zeo_PpCAT1-500_sTom	Zeo	
S. pombe 972h-	pT4_SpARS_AgTEF1zeo_PpENO2_sTom	Zeo	
P. pastoris CBS7435	pT4_SpARS_AgTEF1zeo_PpCAT1-500_sTom	Zeo	
P. pastoris CBS7435	pT4_SpARS_AgTEF1zeo_PpENO2_sTom	Zeo	
P. pastoris CBS7435	pT4_SpARS_AgTEF1zeo_PpCAT1-500_Vio2A	Zeo	
P. pastoris BG10ΔGUT1	pT4_SpARS_AgTEF1GUT1_PpCAT1-500_sTom	*GUT1*	
E. coli K12 Top10 F'	pT4_SpARS_AgTEF1zeo_PpCAT1-500_sTom	Zeo	7925
E. coli K12 Top10 F'	pT4_SpARS_AgTEF1zeo_PpENO2_sTom	Zeo	7926
E. coli K12 Top10 F'	pT4_SpARS_AgTEF1GUT1_PpCAT1-500_sTom	Amp	7927

4.2 VECTOR MAPS & SEQUENCES

Table 4.2 Vectors developed during this thesis.

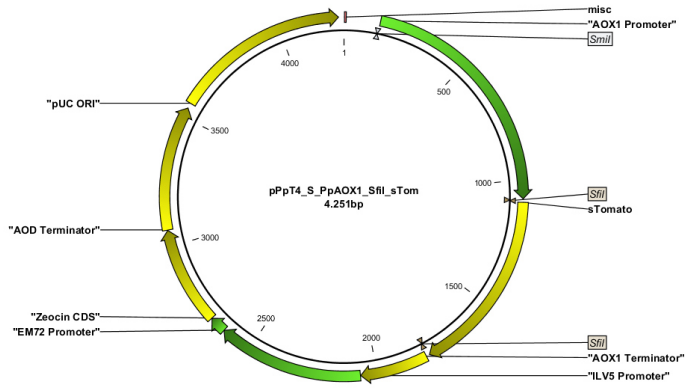
Name	Marker
pPpT4_S_PpAOX1_sTom	Zeo
pPpT4_S_PpAOX1_SfiI_sTom	Zeo
pPpT4_S_ARS_agTEF_PpCAT1_alphaCalB	Zeo
pPpT4_S_ARS_agTEF_PpCAT1_sTomato	Zeo
pPpT4_S_ARS_agTEF_PpCAT1_Vio2A	Zeo
pPpT4_S_ARS_agTEF_PpENO2_sTom	Zeo
pPpT4_S_ARS_agTEFGUT1_PpCAT1_sTom	Amp

pPpT4_S_PpAOX1_sTom



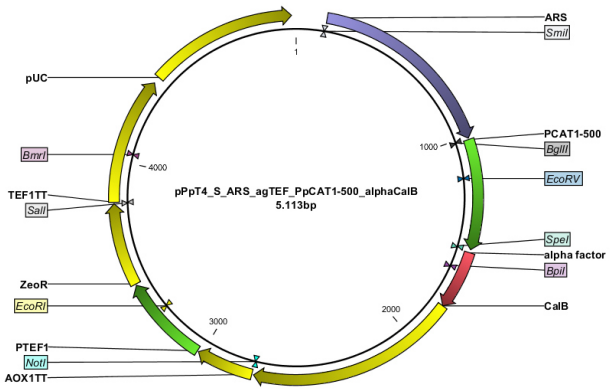
AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTAAATAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATCCACAGGTCCATTCTCACACATAA
GTGCCAAACGCAACAGGAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGC
CCAGTTATTGGGCTTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCA
TGTTTGTTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGCCAGTT
GGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGCGCAGTCTCTCTATC
GCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGG
GAATACTGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTT
TTTATCATCATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTA
ATTATTGAAAGAATTCATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGTTCATGCGTTTCAAGGTCAGAATGGAAGGTTCTATGAACGGTCATGAGTTCGAGA
TTGAAGGAGAAGGTGAAGGAAGACCATATGAGGGTACTCAAACCGCAAAGTTGAAGGTTACTAAAGGAGGTCCTTTACCATTCGCTTGGGATATCCTGTCTCCA
CAATTCATGTATGGTTCTAAGGCATACGTTAAGCATCCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGAGGGTTTCAAGTGGGAAAGAGTCATGAAC
TTCGAAGACGGTGGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGATGCGTGGTACCAACTTCCCACCAGATGGT
CCTGTTATGCAGAAAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATCCAAGAGATGGTGTTTTGAAGGGTGAGATTCACCAAGCTTTGAAGCTTAA
AGATGGAGGTCACTACTTGGTTGAGTTCAAGACCATTTACATGGCTAAGAAACCAGTTCAACTTCCTGGATACTATTACGTTGACACTAAGCTGGACATTACCTCT
CACAACGAAGACTACACCATCGTTGAGCAATACGAGAGATCCGAAGGTAGACACCACTTGTTCTTGTACGGTATGGACGAGCTTTATAAGTAAGCGGCCGCTCA
AGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCT
CGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACT
CCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCTTCAGTAATGTCTTGTTTCTTTTGTTGCAGTGGTGAGCCATTTTGACTTCGTGAAAGT
TTCTTTAGAATAGTTGTTTCCAGAGGCCAAACATTCCACCCGTAGTAAAGTGCAAGCGTAGGAAGACCAAGACTGGCATAAATCAGGTATAAGTGTCGAGCACT
GGCAGGTGATCTTCTGAAAGTTTCTACTAGCAGATAAGATCCAGTAGTCATGCATATGGCAACAATGTACCGTGTGGATCTAAGAACGCGTCCTACTAACCTTCG
CATTCGTTGGTCCAGTTTGTTGTTATCGATCAACGTGACAAGGTTGTCGATTCCGCGTAAGCATGCATACCCAAGGACGCCTGTTGCAATTCCAAGTGAGCCAGTT
CCAACAATCTTTGTAATATTAGAGCACTTCATTGTGTTGCGCTTGAAAGTAAAATGCGAACAAATTAAGAGATAATCTCGAAACCGCGACTTCAAACGCCAATATG
ATGTGCGGCACACAATAAGCGTTCATATCCGCTGGGTGACTTTCTCGCTTTAAAAAATTATCCGAAAAAATTTTCTAGAGTGTTGTTACTTTATACTTCCGGCTCGT
ATAATACGACAAGGTGTAAGGAGGACTAAACCATGGCTAAACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTGGACTG
ATAGACTCGGTTTCTCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCAGGACCAGGTTGTGCCAG
ACAACACTCTGGCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACTTCCGTGATGCATCTGGTCCAGCTATGA
CCGAGATCGGTGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTGCGTGCATTTCGTCGCAGAAGAGCAGGACTAACAATTGACACCT
TACGATTATTTAGAGAGTATTTATTAGTTTTATTGTATGTATACGGATGTTTTATTATCTATTTATGCCCTTATATTCTGTAACTATCCAAAAGTCCTATCTTATCAAG
CCAGCAATCTATGTCCGCGAACGTCAACTAAAAATAAGCTTTTTATGCTCTTCTCTCTTTTTTTCCCTTCGGTATAATTATACCTTGCATCCACAGATTCTCCTGCCAA
ATTTTGCATAATCCTTTACAACATGGCTATATGGGAGCACTTAGCGCCCTCCAAAAACCCATATTGCCTACGCATGTATAGGTGTTTTTTCCACAATATTTTCTCTGTG
CTCTCTTTTTATTAAAGAGAAGCTCTATATCGGAGAAGCTTCTGTGGCCGTTATATTCGGCCTTATCGTGGGACCACATTGCCTGAATTGGTTTGCCCCGGAAGAT
TGGGGAAACTTGGATCTGATTACCTTAGCTGCAGGTACCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT
CTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCA
GATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCC
CAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTA
AGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTC
GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTT
TCCTGCGGTACCC

pPpT4_S_PpAOX1_SfiI_sTom



AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTAAATAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATCCACAGGTCCATTCTCACACATAA
GTGCCAAACGCAACAGGAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGC
CCAGTTATTGGGCTTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCA
TGTTTGTTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGCCAGTT
GGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGCGCAGTCTCTCTATC
GCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGG
GAATACTGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTT
TTTATCATCATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTA
ATTATTGAAAGAATTCACTAGTGGCCGAAACGGGCCATTTTGGCCATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGTTCATGCGTTTCAAGGTCAGAATGGA
AGGTTCTATGAACGGTCATGAGTTCGAGATTGAAGGAGAAGGTGAAGGAAGACCATATGAGGGTACTCAAACCGCAAAGTTGAAGGTTACTAAAGGAGGTCCT
TTACCATTCGCTTGGGATATCCTGTCTCCACAATTCATGTATGGTTCTAAGGCATACGTTAAGCATCCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGA
GGGTTTCAAGTGGGAAAGAGTCATGAACTTCGAAGACGGTGGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGAT
GCGTGGTACCAACTTCCCACCAGATGGTCCTGTTATGCAGAAAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATCCAAGAGATGGTGTTTTGAAGG
GTGAGATTCACCAAGCTTTGAAGCTTAAAGATGGAGGTCACTACTTGGTTGAGTTCAAGACCATTTACATGGCTAAGAAACCAGTTCAACTTCCTGGATACTATTA
CGTTGACACTAAGCTGGACATTACCTCTCACAACGAAGACTACACCATCGTTGAGCAATACGAGAGATCCGAAGGTAGACACCACTTGTTCTTGTACGGTATGGA
CGAGCTTTATAAGTAAGGCCGCCTCGGCCGCGCCGGCGCGCCGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGAT
ACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTG
GTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCTTC
AGTAATGTCTTGTTTCTTTTGTTGCAGTGGTGAGCCATTTTGACTTCGTGAAAGTTTCTTTAGAATAGTTGTTTCCAGAGGCCAAACATTCCACCCGTAGTAAAGTG
CAAGCGTAGGAAGACCAAGACTGGCATAAATCAGGTATAAGTGTCGAGCACTGGCAGGTGATCTTCTGAAAGTTTCTACTAGCAGATAAGATCCAGTAGTCATG
CATATGGCAACAATGTACCGTGTGGATCTAAGAACGCGTCCTACTAACCTTCGCATTCGTTGGTCCAGTTTGTTGTTATCGATCAACGTGACAAGGTTGTCGATTC
CGCGTAAGCATGCATACCCAAGGACGCCTGTTGCAATTCCAAGTGAGCCAGTTCCAACAATCTTTGTAATATTAGAGCACTTCATTGTGTTGCGCTTGAAAGTAAA
ATGCGAACAAATTAAGAGATAATCTCGAAACCGCGACTTCAAACGCCAATATGATGTGCGGCACACAATAAGCGTTCATATCCGCTGGGTGACTTTCTCGCTTTA
AAAAATTATCCGAAAAAATTTTCTAGAGTGTTGTTACTTTATACTTCCGGCTCGTATAATACGACAAGGTGTAAGGAGGACTAAACCATGGCTAAACTCACCTCTG
CTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTGGACTGATAGACTCGGTTTCTCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGT
ACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCAGGACCAGGTTGTGCCAGACAACACTCTGGCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGA
GTGGTCTGAGGTCGTGTCTACCAACTTCCGTGATGCATCTGGTCCAGCTATGACCGAGATCGGTGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGC
TGGTAACTGCGTGCATTTCGTCGCAGAAGAGCAGGACTAACAATTGACACCTTACGATTATTTAGAGAGTATTTATTAGTTTTATTGTATGTATACGGATGTTTTA
TTATCTATTTATGCCCTTATATTCTGTAACTATCCAAAAGTCCTATCTTATCAAGCCAGCAATCTATGTCCGCGAACGTCAACTAAAAATAAGCTTTTTATGCTCTTC
TCTCTTTTTTTCCCTTCGGTATAATTATACCTTGCATCCACAGATTCTCCTGCCAAATTTTGCATAATCCTTTACAACATGGCTATATGGGAGCACTTAGCGCCCTCC
AAAACCCATATTGCCTACGCATGTATAGGTGTTTTTTCCACAATATTTTCTCTGTGCTCTCTTTTTATTAAAGAGAAGCTCTATATCGGAGAAGCTTCTGTGGCCGTT
ATATTCGGCCTTATCGTGGGACCACATTGCCTGAATTGGTTTGCCCCGGAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGGTACCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATA
GTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAG
CTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGG
AAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAG
CAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGGTACCC

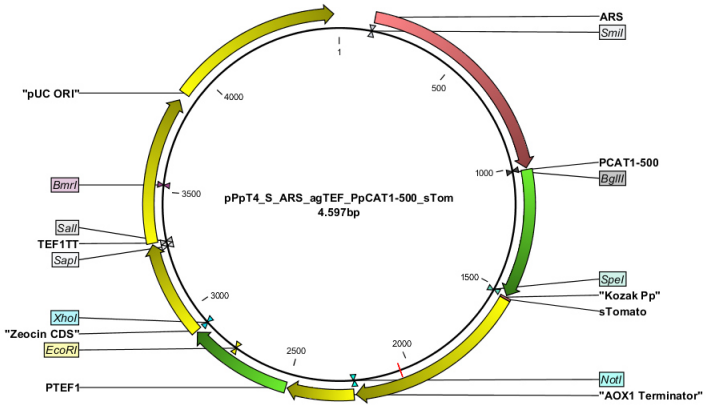
pPpT4_S_ARS_agTEF_PpCAT1_alphaCalB



AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTAAATGACAGAATGGGATACAAGGGCATCGTCTATAATTATAGCTAAAAATTGTATTTTAATTTGTATTTTTTGTAAT
TTTATTTCTCATTGTTTTACTTAAAATGAATGCGAATTAGAGAAAACTTCAACGAAATGTCAAAATAAGCTCAGCAAAATATACAATTTTAGGGAAAGCGATCAGC
AACTTATTCTCCGTTGTAGATTCTTTTTCTTCATAAATTCCAACAACGTAACTGCCAATTGATTTTTCTCTAGATGCTAATAAATTGAAACTAATGAGTATAATAACA
CAGGTTTTGATAAATTGTAAACTCTTACTCAAAAAAAGAATGTTAAAACTAATTTAACCTTTTTATAAAGGCTCATGGCGTTAAAGAGGATAGTAAAAATTCTACA
ATTTTTTAATTTAATGTCGCAATTGTCTTGCGAATATAATTCTATTAATTTTCGCTAACTGCTGTTTTAGTACTATTACTTTCAATATGCCTTATATGCAACTTTAATTC
ATGAAGAATGCAATATTTTTGCAATCTTTCTTGTACTGAGCGCTTTTTATCATTTAGTACTCATTATTTAATTTTTCAGTAAAGGGCAATAAATAAATTTGTAGAAGA
TGCAATGTAATCTCCTCTATCCTTTTGCTCATATATGTTTATGAGTATACCTAGTCTAGAAAGGCTTGTATTTAAAATATTATTCAATAAAAATTCACAATTTTTACG
ACATGTGCTACCATTCACCTAACTTCCTGATTATAAAATTGGTTCGTTTATACTAATTACTTAAGTACCTTTAACTAAACAAAATGCCTATATATATATTAATTTACA
ATGAGTGTCAGATAAGTCACTATGTCCGAGTGGTTAAGGAGAGATCTGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTAAATAAAC
TCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTGAAGTCTTTCTATTGTTCCGGTCGGCATTACCTAATAAATTCGAATCGAGATTGCTAGT
ACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTACTAATGGAATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATACTAT
GCACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTGGCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCC
AGCGGGGTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGCACTTCAATTACTTGAAATTCACCATAACACTTGCTCTAGTCAAGA
CTTACAATTAAAACTAGACTAGTTCGAAACGATGAGATTCCCATCTATTTTCACCGCTGTCTTGTTCGCTGCCTCCTCTGCATTGGCTGCCCCTGTTAACACTACCAC
TGAAGACGAGACTGCTCAAATTCCAGCTGAAGCAGTTATCGGTTACTCTGACCTTGAGGGTGATTTCGACGTCGCTGTTTTGCCTTTCTCTAACTCCACTAACAAC
GGTTTGTTGTTCATTAACACCACTATCGCTTCCATTGCTGCTAAGGAAGAGGGTGTCTCTCTCGAGAAGAGAGAGGCCGAAGCTCTTCCATCTGGTTCTGATCCTG
CTTTCTCCCAGCCAAAGTCCGTCCTTGACGCAGGTTTGACCTGTCAAGGTGCTTCTCCATCTTCTGTCTCTAAGCCAATTTTGTTGGTTCCAGGTACTGGTACCACT
GGACCACAATCTTTCGACTCCAACTGGATTCCTCTGTCTACTCAGTTGGGTTACACTCCATGCTGGATTTCCCCACCTCCATTCATGTTGAACGACACCCAAGTTAA
CACCGAATACATGGTCAACGCAATCACCGCTCTTTACGCTGGTTCTGGTAACAACAAGTTGCCAGTCCTTACTTGGTCTCAAGGTGGATTGGTTGCTCAATGGGGT
TTGACCTTCTTCCCATCTATCCGTTCCAAGGTTGACAGATTGATGGCCTTCGCACCTGACTACAAGGGAACTGTTCTTGCTGGACCATTGGACGCATTGGCTGTTTC
CGCTCCTTCTGTGTGGCAACAAACCACCGGTTCTGCTTTGACTACCGCCTTGAGAAACGCAGGAGGTTTGACCCAAATTGTCCCTACCACTAACCTGTACTCTGCC
ACTGACGAGATCGTTCAGCCTCAAGTGTCTAACTCCCCATTGGACTCCTCTTACTTGTTCAACGGAAAGAACGTCCAAGCACAAGCTGTTTGTGGTCCTTTGTTCG
TTATTGACCACGCCGGTTCTCTTACCTCCCAATTCTCTTACGTCGTTGGTAGATCCGCACTTAGATCCACCACTGGTCAAGCTCGTTCTGCTGATTACGGTATCACT
GACTGTAACCCTTTACCAGCAAACGACCTTACTCCAGAGCAAAAGGTCGCTGCCGCTGCATTGTTGGCTCCTGCTGCCGCAGCTATTGTTGCTGGTCCTAAGCAAA
ACTGTGAACCAGATTTGATGCCATATGCTAGACCTTTCGCTGTTGGTAAGAGAACTTGTTCCGGTATCGTGACCCCATAAGCGGCCGCTCAAGAGGATGTCAGAA
TGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCTC
CTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTACA
GAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCaCGCGGGTCACCCGGCCAGCGACATGGAGGCCCAGATACCCTCCTTGACAGTCTT
GACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCGACGGCG
CAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCAGCAACGCGTTGAATTCTCCCCACGGCGCGCCCCTGTAGAGAAATATA
AAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACTCACATCCGAACATAAACAAAAATGGGTAAGG
AAAAGACTCACGTTAACCTCGAGATGGCTAAACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTGGACTGATAGACTCGG
TTTCTCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCAGGACCAGGTTGTGCCAGACAACACTCTG
GCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACTTCCGTGATGCATCTGGTCCAGCTATGACCGAGATCGG
TGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTGCGTGCATTTCGTCGCAGAAGAGCAGGACTAAGTCGACAAGGGGTAAGAAATA
GAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTT
GCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAA
CGCCGCCCCAGTGTCTACCTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGATAGAATATGCTCGAACTGTTGAAGTCCATCAACAAGGATAACCCATATGCTCTA
TCGGCGGAGAAAACGTTCAGAGCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCTAGATGAGAAGTACGGGGTTTAGTGTTTCCAGGCCTCGTAAATGCCGCA
ATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTCAAGGACCGCCTTTTCATATATGGCCACCATCAATATAGAGCCAGCGAAC
GATGATGCATTCTGGTATATTGAACGCGAGGATCCCTGCAGGTACCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTG
CGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGC

AGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC
CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAG
GTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGA
CTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGTTCTTTCCTGCGGTACCC

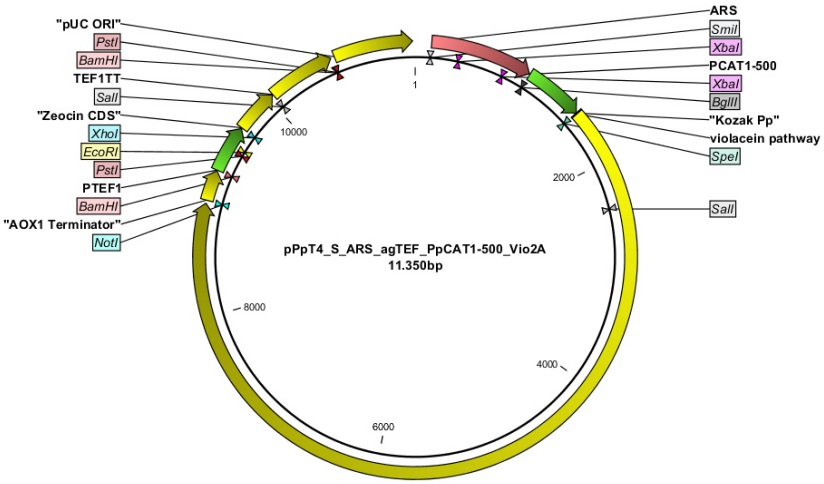
pPpT4_S_ARS_agTEF_PpCAT1_sTomato



AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTAAATGACAGAATGGGATACAAGGGCATCGTCTATAATTATAGCTAAAAATTGTATTTTAATTTGTATTTTTTGTAAT
TTTATTTCTCATTGTTTTACTTAAAATGAATGCGAATTAGAGAAAACTTCAACGAAATGTCAAAATAAGCTCAGCAAAATATACAATTTTAGGGAAAGCGATCAGC
AACTTATTCTCCGTTGTAGATTCTTTTTCTTCATAAATTCCAACAACGTAACTGCCAATTGATTTTTCTCTAGATGCTAATAAATTGAAACTAATGAGTATAATAACA
CAGGTTTTGATAAATTGTAAACTCTTACTCAAAAAAAGAATGTTAAAACTAATTTAACCTTTTTATAAAGGCTCATGGCGTTAAAGAGGATAGTAAAAATTCTACA
ATTTTTTAATTTAATGTCGCAATTGTCTTGCGAATATAATTCTATTAATTTTCGCTAACTGCTGTTTTAGTACTATTACTTTCAATATGCCTTATATGCAACTTTAATTC
ATGAAGAATGCAATATTTTTGCAATCTTTCTTGTACTGAGCGCTTTTTATCATTTAGTACTCATTATTTAATTTTTCAGTAAAGGGCAATAAATAAATTTGTAGAAGA
TGCAATGTAATCTCCTCTATCCTTTTGCTCATATATGTTTATGAGTATACCTAGTCTAGAAAGGCTTGTATTTAAAATATTATTCAATAAAAATTCACAATTTTTACG
ACATGTGCTACCATTCACCTAACTTCCTGATTATAAAATTGGTTCGTTTATACTAATTACTTAAGTACCTTTAACTAAACAAAATGCCTATATATATATTAATTTACA
ATGAGTGTCAGATAAGTCACTATGTCCGAGTGGTTAAGGAGAGATCTGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTAAATAAAC
TCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTGAAGTCTTTCTATTGTTCCGGTCGGCATTACCTAATAAATTCGAATCGAGATTGCTAGT
ACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTACTAATGGAATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATACTAT
GCACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTGGCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCC
AGCGGGGTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGCACTTCAATTACTTGAAATTCACCATAACACTTGCTCTAGTCAAGA
CTTACAATTAAAACTAGACTAGTTCGAAACGATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGTTCATGCGTTTCAAGGTCAGAATGGAAGGTTCTATGAACG
GTCATGAGTTCGAGATTGAAGGAGAAGGTGAAGGAAGACCATATGAGGGTACTCAAACCGCAAAGTTGAAGGTTACTAAAGGAGGTCCTTTACCATTCGCTTG
GGATATCCTGTCTCCACAATTCATGTATGGTTCTAAGGCATACGTTAAGCATCCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGAGGGTTTCAAGTGG
GAAAGAGTCATGAACTTCGAAGACGGTGGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGATGCGTGGTACCAAC
TTCCCACCAGATGGTCCTGTTATGCAGAAAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATCCAAGAGATGGTGTTTTGAAGGGTGAGATTCACCA
AGCTTTGAAGCTTAAAGATGGAGGTCACTACTTGGTTGAGTTCAAGACCATTTACATGGCTAAGAAACCAGTTCAACTTCCTGGATACTATTACGTTGACACTAAG
CTGGACATTACCTCTCACAACGAAGACTACACCATCGTTGAGCAATACGAGAGATCCGAAGGTAGACACCACTTGTTCTTGTACGGTATGGACGAGCTTTATAAG
TAAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTGT
CATTTTGTTTCTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCT
TGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCaCGCGGGTCACCCGGCCAGCGACATG
GAGGCCCAGATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACATCCCCATGTATAATCATTTGC
ATCCATACATTTTGATGGCCGCGACGGCGCAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCAGCAACGCGTTGAATTCTCC
CCACGGCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACTCACA
TCCGAACATAAACAAAAATGGGTAAGGAAAAGACTCACGTTAACCTCGAGATGGCTAAACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTG
CTGTTGAGTTCTGGACTGATAGACTCGGTTTCTCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCA
GGACCAGGTTGTGCCAGACAACACTCTGGCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACTTCCGTGATG
CATCTGGTCCAGCTATGACCGAGATCGGTGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTGCGTGCATTTCGTCGCAGAAGAGCAG
GACTAAGTCGACAAGGGGTAAGAAATAGAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTATAGTTTTTTTATATTGTAGTTGTTCTATTTTA
ATCAAATGTTAGCGTGATTTATATTTTTTGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGTCAATCGTATGTGAATGCTGGT
CGCTATACTGCTGTCGATTCGATACTAACGCCGCCCCAGTGTCTACCTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGATAGAATATGCTCGAACTGTTGAAGTC

CATCAACAAGGATAACCCATATGCTCTATCGGCGGAGAAAACGTTCAGAGCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCTAGATGAGAAGTACGGGGTTT
AGTGTTTCCAGGCCTCGTAAATGCCGCAATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTCAAGGACCGCCTTTTCATATAT
GGCCACCATCAATATAGAGCCAGCGAACGATGATGCATTCTGGTATATTGAACGCGAGGATCCCTGCAGGTACCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAG
CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATA
GTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGGTACCC

pPpT4_S_ARS_agTEF_PpCAT1_Vio2A

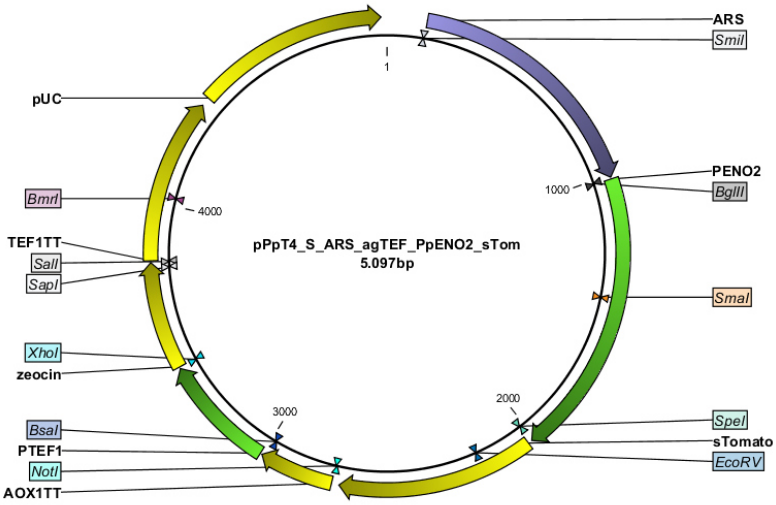


AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTAAATGACAGAATGGGATACAAGGGCATCGTCTATAATTATAGCTAAAAATTGTATTTTAATTTGTATTTTTTGTAAT
TTTATTTCTCATTGTTTTACTTAAAATGAATGCGAATTAGAGAAAACTTCAACGAAATGTCAAAATAAGCTCAGCAAAATATACAATTTTAGGGAAAGCGATCAGC
AACTTATTCTCCGTTGTAGATTCTTTTTCTTCATAAATTCCAACAACGTAACTGCCAATTGATTTTTCTCTAGATGCTAATAAATTGAAACTAATGAGTATAATAACA
CAGGTTTTGATAAATTGTAAACTCTTACTCAAAAAAAGAATGTTAAAACTAATTTAACCTTTTTATAAAGGCTCATGGCGTTAAAGAGGATAGTAAAAATTCTACA
ATTTTTTAATTTAATGTCGCAATTGTCTTGCGAATATAATTCTATTAATTTTCGCTAACTGCTGTTTTAGTACTATTACTTTCAATATGCCTTATATGCAACTTTAATTC
ATGAAGAATGCAATATTTTTGCAATCTTTCTTGTACTGAGCGCTTTTTATCATTTAGTACTCATTATTTAATTTTTCAGTAAAGGGCAATAAATAAATTTGTAGAAGA
TGCAATGTAATCTCCTCTATCCTTTTGCTCATATATGTTTATGAGTATACCTAGTCTAGAAAGGCTTGTATTTAAAATATTATTCAATAAAAATTCACAATTTTTACG
ACATGTGCTACCATTCACCTAACTTCCTGATTATAAAATTGGTTCGTTTATACTAATTACTTAAGTACCTTTAACTAAACAAAATGCCTATATATATATTAATTTACA
ATGAGTGTCAGATAAGTCACTATGTCCGAGTGGTTAAGGAGAGATCTGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTAAATAAAC
TCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTGAAGTCTTTCTATTGTTCCGGTCGGCATTACCTAATAAATTCGAATCGAGATTGCTAGT
ACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTACTAATGGAATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATACTAT
GCACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTGGCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCC
AGCGGGGTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGCACTTCAATTACTTGAAATTCACCATAACACTTGCTCTAGTCAAGA
CTTACAATTAAAACTAGACTAGTTCGAAACGATGAAACACTCTTCCGACATTTGTATTGTCGGAGCTGGTATCTCAGGTTTGACTTGCGCCTCTCACTTGCTGGATT
CTCCAGCCTGTAGAGGTTTGTCCCTTAGAATCTTTGACATGCAACAGGAGGCTGGAGGTAGAATTAGATCAAAGATGTTGGATGGTAAGGCATCTATTGAATTG
GGTGCTGGTAGATACTCTCCACAATTGCACCCACACTTTCAGTCTGCTATGCAACATTACTCCCAAAAGTCTGAGGTTTACCCTTTCACTCAGCTTAAGTTCAAATC
TCACGTTCAACAGAAGCTTAAGCGTGCTATGAACGAATTGTCCCCACGTTTGAAGGAGCACGGTAAAGAGTCTTTCTTACAGTTCGTGTCCAGATACCAAGGACA
TGACTCTGCTGTTGGTATGATTAGATCAATGGGTTACGATGCATTGTTTCTTCCAGACATCTCTGCTGAGATGGCATACGACATTGTTGGTAAGCATCCTGAGATT
CAATCTGTTACCGATAACGACGCTAACCAGTGGTTCGCAGCTGAGACTGGTTTTGCTGGATTGATTCAGGGAATCAAAGCTAAGGTCAAGGCAGCTGGTGCCAG
ATTCTCCCTTGGTTACAGATTGCTGTCTGTGAGAACTGACGGTGATGGATATCTGTTGCAGCTTGCCGGTGACGACGGATGGAAGCTTGAGCACCGTACTAGACA
CCTGATCCTTGCTATTCCACCTTCTGCAATGGCAGGTTTGAATGTTGACTTCCCAGAAGCTTGGTCTGGAGCCAGATACGGTTCCTTGCCTTTGTTTAAGGGATTCT
TAACCTACGGTGAACCTTGGTGGTTGGACTACAAGCTTGATGACCAAGTCCTGATCGTCGACAATCCATTGCGTAAAATCTACTTCAAGGGAGACAAGTACCTTT
TCTTCTACACTGACTCTGAGATGGCCAACTACTGGAGAGGTTGCGTCGCCGAAGGTGAGGATGGTTACTTAGAGCAAATCCGTACCCACTTGGCTTCTGCTCTTG

GAATCGTCAGAGAAAGAATCCCTCAACCTTTGGCCCACGTTCATAAGTACTGGGCACACGGTGTTGAGTTCTGTAGAGACTCAGATATCGACCACCCATCCGCAT
TGTCACATAGAGACTCTGGTATCATTGCTTGTTCTGATGCCTATACCGAACACTGCGGATGGATGGAGGGAGGACTGTTGTCCGCCAGAGAAGCCTCCCGTTTGC
TTTTGCAAAGAATTGCTGCCCGTGCAGAGGGCCGTGGTTCCTTACTTACCTGCGGTGATGTGGAAGAAAATCCAGGACCCTCTATTTTGGACTTCCCAAGAATCC
ACTTTCGTGGTTGGGCAAGAGTTAACGCTCCAACCGCCAACAGAGATCCACACGGTCACATCGATATGGCCTCCAACACCGTCGCTATGGCCGGTGAACCATTCG
ACCTGGCCAGACATCCTACCGAGTTCCACAGACACTTGAGATCCTTGGGTCCAAGATTTGGATTAGATGGAAGAGCTGACCCTGAAGGACCTTTCTCACTGGCCG
AGGGTTACAACGCCGCTGGAAACAATCACTTCTCTTGGGAGTCAGCTACTGTTTCTCACGTGCAATGGGACGGTGGAGAAGCTGATAGAGGTGACGGATTGGTT
GGTGCCAGACTGGCTCTTTGGGGTCATTACAATGACTACTTGAGAACCACTTTCAATAGAGCTAGATGGGTCGATTCCGACCCTACCAGAAGAGATGCTGCACAA
ATCTACGCCGGACAGTTCACCATTTCTCCAGCAGGTGCAGGTCCAGGAACCCCTTGGTTGTTTACCGCCGACATTGACGATTCTCACGGTGCTCGTTGGACCCGTG
GAGGTCACATTGCTGAAAGAGGAGGTCACTTCTTAGATGAAGAGTTCGGTCTTGCTAGACTGTTTCAGTTCTCAGTTCCAAAAGACCACCCTCACTTCCTTTTCCA
CCCTGGTCCATTTGATTCTGAGGCTTGGAGAAGACTGCAACTTGCTTTGGAGGACGATGACGTTTTGGGACTTACTGTTCAATACGCACTGTTCAACATGTCTACT
CCTCCTCAACCAAACTCTCCAGTGTTCCATGACATGGTCGGTGTTGTCGGTCTTTGGAGAAGAGGTGAGCTTGCATCTTACCCAGCTGGTAGACTGTTGCGTCCAA
GACAACCAGGTCTGGGTGACCTTACTTTGAGAGTCAACGGTGGTAGAGTCGCTTTGAACTTAGCCTGTGCCATTCCATTCTCAACTAGAGCTGCTCAGCCTTCAGC
ACCTGATAGACTGACTCCAGACCTGGGAGCAAAGTTGCCATTGGGTGATTTGTTGTTACGTGACGAAGACGGAGCCCTGTTGGCTAGAGTCCCACAAGCACTTT
ACCAAGATTACTGGACTAACCACGGTATCGTTGACTTGCCTCTGTTGCGTGAGCCAAGAGGTTCCCTTACCTTGTCTTCCGAGTTGGCAGAGTGGAGAGAGCAGG
ACTGGGTGACTCAATCTGATGCATCCAACCTTTACTTAGAGGCTCCAGACAGACGTCACGGTCGTTTCTTTCCAGAGTCTATTGCTTTGCGTTCTTACTTCAGAGGT
GAAGCAAGAGCCAGACCTGACATTCCACACCGTATTGAGGGAATGGGTCTTGTGGGAGTTGAGTCAAGACAGGATGGTGACGCTGCCGAATGGCGTCTGACCG
GTTTGCGTCCAGGTCCTGCTAGAATCGTTCTGGACGATGGTGCAGAAGCTATTCCTTTGAGAGTCCTTCCAGACGACTGGGCCTTGGATGACGCTACTGTTGAAG
AGGTTGACTACGCTTTCTTGTACCGTCACGTTATGGCCTACTATGAGTTGGTGTACCCTTTCATGTCTGATAAAGTCTTTTCTCTGGCTGACCGTTGTAAGTGTGAA
ACTTACGCCCGTTTGATGTGGCAAATGTGCGACCCACAAAACAGAAACAAGTCCTACTACATGCCTTCCACTAGAGAGTTGTCCGCTCCAAAGGCTAGATTGTTTC
TTAAGTACTTGGCTCACGTTGAAGGTCAAGCTAGATTGCAAGCTCCACCTCCAGCTGGTCCTGCACGTATTGAGTCTAAAGCTCAATTGGCCGCAGAATTGCGTA
AGGCTGTTGACCTGGAGCTTTCCGTTATGCTTCAATACCTTTACGCTGCCTACTCTATCCCTAACTATGCACAGGGTCAGCAAAGAGTTAGAGACGGTGCCTGGAC
TGCTGAGCAGTTACAATTGGCATGTGGTTCTGGAGATAGACGTAGAGACGGAGGTATCAGAGCTGCATTGCTTGAGATTGCTCACGAGGAAATGATTCACTATC
TGGTTGTCAACAACTTGTTGATGGCTCTTGGTGAACCTTTCTACGCCGGTGTTCCATTGATGGGTGAAGCCGCTAGACAAGCATTCGGTTTAGACACCGAGTTTG
CCTTAGAACCTTTCTCTGAGTCCACCCTTGCCCGTTTCGTTCGTCTGGAATGGCCTCACTTCATCCCAGCACCTGGTAAGTCCATTGCTGACTGTTATGCTGCCATC
AGACAAGCTTTCTTGGACCTGCCAGATTTGTTTGGAGGTGAAGCAGGTAAGAGAGGAGGTGAGCACCATCTTTTCTTGAACGAGTTGACCAATAGAGCCCACCC
TGGATACCAATTGGAAGTCTTTGACAGAGACTCTGCTCTTTTCGGTATTGCTTTCGTTACTGACCAAGGTGAGGGTGGTGCTTTGGACTCTCCTCATTACGAACAC
TCCCACTTTCAAAGATTGCGTGAGATGTCCGCTCGTATCATGGCTCAATCCGCTCCTTTCGAACCAGCTTTGCCTGCTCTTCGTAACCCAGTCTTGGACGAATCCCC
AGGATGCCAAAGAGTCGCAGACGGTAGAGCCCGTGCATTGATGGCCTTGTACCAGGGTGTTTACGAATTGATGTTCGCTATGATGGCTCAACATTTCGCTGTCAA
GCCACTGGGTTCTTTGAGACGTTCAAGACTTATGAACGCAGCTATTGACCTTATGACTGGTTTGCTTAGACCTTTGTCTTGCGCTTTGATGAATCTTCCATCTGGTA
TCGCTGGTAGAACTGCTGGTCCACCTCTGCCAGGTCCTGTTGACACTAGATCCTACGACGACTATGCTTTAGGTTGTAGAATGTTGGCTAGAAGATGTGAGAGAT
TACTTGAGCAAGCCTCCATGTTGGAACCTGGTTGGCTTCCTGATGCTCAGATGGAGTTGTTGGACTTCTACCGTAGACAGATGTTGGACTTGGCTTGTGGAAAGT
TATCTCGTGAAGCTCGTGCCGAGGGTAGGGGATCACTTCTTACATGTGGAGACGTCGAGGAGAACCCTGGTCCAGAAAACCGTGAGCCACCTTTGCTTCCAGCT
AGATGGTCTTCCGCATACGTTTCATATTGGTCTCCAATGTTGCCTGACGATCAGCTGACTTCCGGTTACTGTTGGTTTGACTACGAGAGAGACATTTGCCGTATTG
ATGGACTTTTCAATCCTTGGTCAGAGAGAGACACCGGTTACAGATTGTGGATGTCTGAAGTTGGAAACGCCGCTTCCGGTAGAACTTGGAAACAGAAGGTTGCT
TACGGTAGAGAACGTACCGCTTTAGGTGAGCAACTGTGTGAAAGACCACTTGATGACGAGACTGGTCCATTCGCCGAGTTGTTCCTTCCTCGTGACGTTTTGAGA
AGACTGGGTGCTAGACACATCGGTCGTAGAGTCGTTTTGGGTAGAGAAGCCGATGGATGGAGATATCAAAGACCTGGAAAGGGTCCATCTACCCTTTACTTGGA
CGCAGCTTCTGGTACTCCTCTTAGAATGGTGACTGGTGATGAGGCATCTCGTGCTTCCTTGAGAGACTTTCCAAACGTGTCTGAGGCTGAAATCCCAGACGCTGT
CTTCGCCGCTAAGAGAAGAGCTGAAGGAAGGGGTTCCCTGTTAACGTGTGGCGATGTTGAAGAGAACCCCGGTCCTAAGATCCTTGTGATTGGTGCAGGACCA
GCCGGTTTGGTGTTTGCTTCTCAACTTAAGCAGGCTAGACCATTGTGGGCTATTGACATCGTTGAGAAGAACGACGAACAAGAGGTGTTGGGATGGGGAGTCGT
TCTTCCTGGTAGACCAGGACAACACCCTGCCAACCCACTGTCTTATCTTGATGCTCCTGAAAGATTGAATCCACAGTTCTTAGAGGACTTCAAGCTTGTCCACCAC
AACGAACCTTCCTTGATGTCAACTGGTGTGCTGTTGTGCGGAGTTGAGCGTAGAGGTTTAGTCCATGCCTTGAGAGACAAATGTCGTTCTCAAGGTATCGCCATT
AGATTTGAGTCTCCTTTGCTTGAGCACGGTGAACTTCCATTGGCTGACTACGATCTTGTTGTGTTGGCTAACGGTGTCAACCATAAGACCGCCCACTTCACTGAAG
CTCTGGTTCCACAGGTTGACTATGGTAGAAACAAGTACATTTGGTACGGAACCTCTCAGTTGTTTGACCAAATGAACTTGGTTTTCAGAACCCACGGTAAGGACA
TCTTCATTGCCCACGCCTACAAGTACTCCGATACTATGTCAACCTTCATTGTCGAATGCTCTGAGGAAACTTATGCAAGAGCTAGACTGGGTGAAATGTCCGAGG
AGGCCTCTGCTGAGTACGTTGCTAAGGTTTTCCAAGCCGAATTGGGTGGACACGGTCTGGTTTCTCAGCCTGGTCTTGGTTGGCGTAACTTCATGACCTTGTCCCA
CGATCGTTGTCATGATGGAAAGTTGGTTCTGTTAGGTGATGCACTTCAATCTGGTCACTTCTCAATCGGTCACGGAACCACTATGGCCGTGGTTGTCGCTCAGTTG
CTGGTCAAAGCACTTTGTACTGAAGACGGTGTTCCTGCTGCTTTGAAACGTTTTGAGGAAAGAGCATTGCCACTGGTTCAACTGTTCAGAGGTCATGCTGACAAT
TCCCGTGTCTGGTTTGAAACTGTTGAAGAGAGAATGCACCTGTCCTCTGCTGAGTTCGTCCAATCCTTCGACGCTAGACGTAAGTCTCTTCCACCTATGCCAGAAG
CTTTGGCACAAAACCTTAGATACGCCTTGCAAAGAAGGGCAGAAGGCAGAGGATCTCTGTTGACTTGTGGTGATGTAGAGGAGAATCCCGGCCCAAAGAGAGC
TATCATTGTTGGTGGAGGTTTAGCTGGTGGTTTGACCGCTATCTATCTGGCAAAGAGAGGTTACGAAGTGCACGTTGTCGAGAAGAGAGGTGACCCACTTAGAG
ATTTGTCTTCCTACGTTGACGTCGTTTCCTCACGTGCCATTGGTGTGTCTATGACTGTTAGAGGTATTAAATCAGTCCTGGCTGCCGGTATTCCACGTGCTGAATTG
GACGCTTGTGGTGAGCCTATCGTTGCTATGGCCTTTTCTGTGGGTGGACAATACAGAATGAGAGAATTGAAACCTTTGGAAGACTTCCGTCCTCTTTCATTGAACA
GAGCTGCATTTCAAAAGTTACTGAACAAGTATGCTAACTTGGCTGGTGTTAGATATTACTTCGAGCACAAGTGCCTTGATGTTGATCTGGACGGTAAGTCTGTCTT
GATTCAGGGTAAGGACGGTCAACCACAAAGATTACAGGGTGATATGATCATTGGTGCCGATGGTGCACACTCAGCTGTGCGTCAAGCTATGCAATCTGGTTTGA
GACGTTTCGAGTTTCAACAGACCTTCTTCAGACACGGATACAAAACTTTGGTTCTTCCAGACGCTCAAGCCTTGGGTTACAGAAAAGATACCTTGTACTTCTTTGG
TATGGACTCCGGTGGATTGTTCGCAGGTAGAGCCGCTACTATTCCAGACGGTTCTGTCTCTATCGCTGTTTGTCTTCCATACTCCGGATCTCCATCTTTGACTACTA
CCGACGAACCAACTATGCGTGCTTTCTTTGACAGATACTTCGGAGGTCTGCCAAGAGACGCTCGTGACGAGATGTTAAGACAGTTCCTGGCTAAGCCTTCCAACG
ACTTGATCAATGTTCGTTCTTCCACTTTTCACTACAAGGGTAACGTCTTGCTGTTAGGAGACGCCGCTCATGCTACTGCTCCTTTCCTTGGTCAAGGTATGAACATG
GCATTGGAGGACGCAAGAACCTTCGTCGAATTGCTTGACAGACATCAAGGAGATCAGGATAAGGCATTCCCTGAGTTCACTGAATTGAGAAAGGTTCAAGCCGA
TGCTATGCAGGATATGGCCAGAGCTAACTATGACGTCCTGTCCTGTTCTAATCCAATCTTCTTCATGAGAGCCAGATACACTCGTTACATGCATTCCAAGTTTCCTG
GTCTTTACCCACCTGATATGGCAGAGAAACTGTACTTCACCTCCGAACCATACGATCGTCTTCAACAGATTCAAAGAAAGCAGAACGTCTGGTACAAGATTGGTA
GAGTTAACTAAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGAT
TTTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGA
TGTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCACGCGGGTCACCCGGCCA
GCGACATGGAGGCCCAGATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACATCCCCATGTATA
ATCATTTGCATCCATACATTTTGATGGCCGCGACGGCGCAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCAGCAACGCGTT
GAATTCTCCCCACGGCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTC

TCACTCACATCCGAACATAAACAAAAATGGGTAAGGAAAAGACTCACGTTAACCTCGAGATGGCTAAACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATG
TTGCTGGTGCTGTTGAGTTCTGGACTGATAGACTCGGTTTCTCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTC
CGCAGTTCAGGACCAGGTTGTGCCAGACAACACTCTGGCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACT
TCCGTGATGCATCTGGTCCAGCTATGACCGAGATCGGTGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTGCGTGCATTTCGTCGCAG
AAGAGCAGGACTAAGTCGACAAGGGGTAAGAAATAGAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTATAGTTTTTTTATATTGTAGTTGT
TCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGTCAATCGTATGTGA
ATGCTGGTCGCTATACTGCTGTCGATTCGATACTAACGCCGCCCCAGTGTCTACCTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGATAGAATATGCTCGAACTG
TTGAAGTCCATCAACAAGGATAACCCATATGCTCTATCGGCGGAGAAAACGTTCAGAGCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCTAGATGAGAAGTA
CGGGGTTTAGTGTTTCCAGGCCTCGTAAATGCCGCAATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTCAAGGACCGCCTTT
TCATATATGGCCACCATCAATATAGAGCCAGCGAACGATGATGCATTCTGGTATATTGAACGCGAGGATCCCTGCAGGTACCACTGAGCGTCAGACCCCGTAGA
AAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG
CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAG
CACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA
AGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGG
TATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCC
TTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGGTACCC

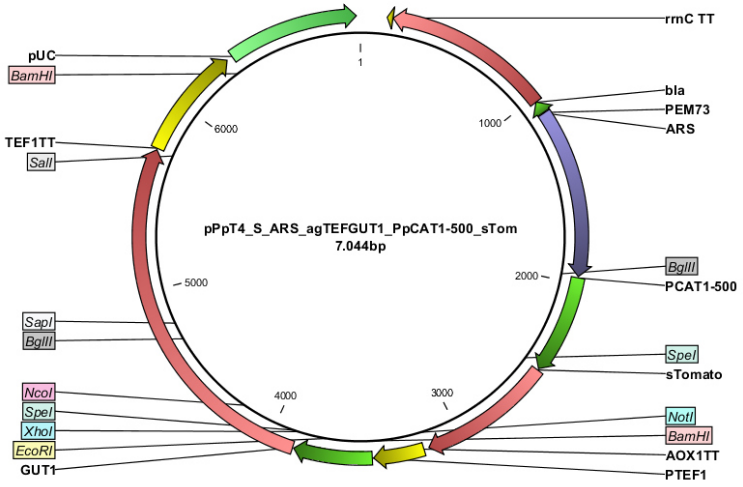
pPpT4_S_ARS_agTEF_PpENO2_sTom



AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTAAATGACAGAATGGGATACAAGGGCATCGTCTATAATTATAGCTAAAAATTGTATTTTAATTTGTATTTTTTGTAAT
TTTATTTCTCATTGTTTTACTTAAAATGAATGCGAATTAGAGAAAACTTCAACGAAATGTCAAAATAAGCTCAGCAAAATATACAATTTTAGGGAAAGCGATCAGC
AACTTATTCTCCGTTGTAGATTCTTTTTCTTCATAAATTCCAACAACGTAACTGCCAATTGATTTTTCTCTAGATGCTAATAAATTGAAACTAATGAGTATAATAACA
CAGGTTTTGATAAATTGTAAACTCTTACTCAAAAAAAGAATGTTAAAACTAATTTAACCTTTTTATAAAGGCTCATGGCGTTAAAGAGGATAGTAAAAATTCTACA
ATTTTTTAATTTAATGTCGCAATTGTCTTGCGAATATAATTCTATTAATTTTCGCTAACTGCTGTTTTAGTACTATTACTTTCAATATGCCTTATATGCAACTTTAATTC
ATGAAGAATGCAATATTTTTGCAATCTTTCTTGTACTGAGCGCTTTTTATCATTTAGTACTCATTATTTAATTTTTCAGTAAAGGGCAATAAATAAATTTGTAGAAGA
TGCAATGTAATCTCCTCTATCCTTTTGCTCATATATGTTTATGAGTATACCTAGTCTAGAAAGGCTTGTATTTAAAATATTATTCAATAAAAATTCACAATTTTTACG
ACATGTGCTACCATTCACCTAACTTCCTGATTATAAAATTGGTTCGTTTATACTAATTACTTAAGTACCTTTAACTAAACAAAATGCCTATATATATATTAATTTACA
ATGAGTGTCAGATAAGTCACTATGTCCGAGTGGTTAAGGAGAGATCTATGAAAGAGTGAGAGGAAAGTACCTGGGCAAAATCACACAATTCCAAACCATGCTAA
ATGAGATTTAAAGAACAAACGATGGCAAAAGGCAACCGTTATAAATGTGATCTTTCTTGGCAGTTATCTGTCAATTTTTCTAAGGAACAGTGAATTCATCATAGG
AGAGATGTTATACGTTACATAATCATACATACTGCATGTATCTCACCTACTTTACCTCATCAACTCTAAAACAGTTCTAGTCCCAACCCCAGATTCCTAGTCATGACA
CAAGTCCGCACCGGACAGGACTCACAACCAGCAAGAGAAGCTAACAAATTTACGCCCCGGTAAAACATTCTTTAGGGGCCGTTCAATGGTAATTTTCCTCTCACC
CGTTTAAACTTACCTCCGGGCGGTATCTTCAATAACCTCTGTTGTCCCCGGGTATCATTGGAAACAGTGAGGGACGTTGAACAGAAGAGAGGATCACCGTAAATT
TGCCTTGCAATTGGCCCTAACCACGGATGGTTAACTTCAAGCCATCACGACAGCAATTGAGTCGGCGCATAGCTACCCTCCTCTTCTTGACCCCATGCATAGGACC
AACCTTAACCGATGGAACAGGTTCCTCCGCTCCGTCCCCTGGTAGTGTCTCTGCGCAAGAAATAGTTAAGGTATGAAGACTGATCTCTCGCACCCCCCTCACAGTA

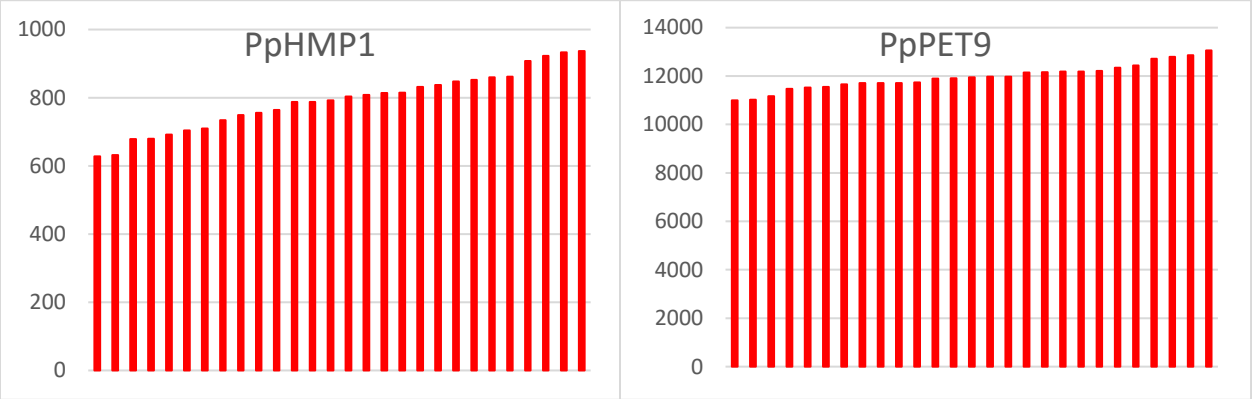
CTGTTATGGTGAATTGACAAAGCCATTGGCTAGATTGAAACATGTAATTCATATGTAATCTTGTTCAATTAACGAGCTTCGTACAGTCTCAATCTAGACGTCTGAT
AATGGCGTTTGTGCTCCTAATCGATGAGCCATCTCATGTGACGTCTATACGCTTCGATGGCTTCCGTCGCGAATATAGAACCACTTGAAATATGCTGCAAACCACG
ATCCACCCTGGTCCTGAAAAGATATAAATACAGCACATCTAGCAGGCTTTTGTCTTCTTGGTTGAAACACACAATTATAACAATCTACATCTAAAAACTAGTTCGA
AACGATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGTTCATGCGTTTCAAGGTCAGAATGGAAGGTTCTATGAACGGTCATGAGTTCGAGATTGAAGGAGAA
GGTGAAGGAAGACCATATGAGGGTACTCAAACCGCAAAGTTGAAGGTTACTAAAGGAGGTCCTTTACCATTCGCTTGGGATATCCTGTCTCCACAATTCATGTAT
GGTTCTAAGGCATACGTTAAGCATCCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGAGGGTTTCAAGTGGGAAAGAGTCATGAACTTCGAAGACGGT
GGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGATGCGTGGTACCAACTTCCCACCAGATGGTCCTGTTATGCAGA
AAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATCCAAGAGATGGTGTTTTGAAGGGTGAGATTCACCAAGCTTTGAAGCTTAAAGATGGAGGTCAC
TACTTGGTTGAGTTCAAGACCATTTACATGGCTAAGAAACCAGTTCAACTTCCTGGATACTATTACGTTGACACTAAGCTGGACATTACCTCTCACAACGAAGACT
ACACCATCGTTGAGCAATACGAGAGATCCGAAGGTAGACACCACTTGTTCTTGTACGGTATGGACGAGCTTTATAAGTAAGCGGCCGCTCAAGAGGATGTCAGA
ATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCT
CCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTACA
GAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCaCGCGGGTCACCCGGCCAGCGACATGGAGGCCCAGATACCCTCCTTGACAGTCTT
GACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCGACGGCG
CAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCAGCAACGCGTTGAATTCTCCCCACGGCGCGCCCCTGTAGAGAAATATA
AAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACTCACATCCGAACATAAACAAAAATGGGTAAGG
AAAAGACTCACGTTAACCTCGAGATGGCTAAACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTGGACTGATAGACTCGG
TTTCTCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCAGGACCAGGTTGTGCCAGACAACACTCTG
GCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACTTCCGTGATGCATCTGGTCCAGCTATGACCGAGATCGG
TGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTGCGTGCATTTCGTCGCAGAAGAGCAGGACTAAGTCGACAAGGGGTAAGAAATA
GAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTT
GCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAA
CGCCGCCCCAGTGTCTACCTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGATAGAATATGCTCGAACTGTTGAAGTCCATCAACAAGGATAACCCATATGCTCTA
TCGGCGGAGAAAACGTTCAGAGCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCTAGATGAGAAGTACGGGGTTTAGTGTTTCCAGGCCTCGTAAATGCCGCA
ATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTCAAGGACCGCCTTTTCATATATGGCCACCATCAATATAGAGCCAGCGAAC
GATGATGCATTCTGGTATATTGAACGCGAGGATCCCTGCAGGTACCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTG
CGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGC
AGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC
CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAG
GTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGA
CTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGTTCTTTCCTGCGGTACCC

pPpT4_S_ARS_agTEFGUT1_PpCAT1_sTom



AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATAGACGCAGATCGGGAA
CACTGAAAAATACACAGTTATTATTCATTTTAAAAAAAATCCTTAGCTTTCGCTAAGGATTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCTAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCG
GGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTT
CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCT
GGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATGTTTAGTCCTCCTTACACCTTGTCGTATTA
TACGAGCCGGAAGTATAAAGTAACAACACTCTAGAAATGACAGAATGGGATACAAGGGCATCGTCTATAATTATAGCTAAAAATTGTATTTTAATTTGTATTTTTT
GTAATTTTATTTCTCATTGTTTTACTTAAAATGAATGCGAATTAGAGAAAACTTCAACGAAATGTCAAAATAAGCTCAGCAAAATATACAATTTTAGGGAAAGCGA
TCAGCAACTTATTCTCCGTTGTAGATTCTTTTTCTTCATAAATTCCAACAACGTAACTGCCAATTGATTTTTCTCTAGATGCTAATAAATTGAAACTAATGAGTATAA
TAACACAGGTTTTGATAAATTGTAAACTCTTACTCAAAAAAAGAATGTTAAAACTAATTTAACCTTTTTATAAAGGCTCATGGCGTTAAAGAGGATAGTAAAAATT
CTACAATTTTTTAATTTAATGTCGCAATTGTCTTGCGAATATAATTCTATTAATTTTCGCTAACTGCTGTTTTAGTACTATTACTTTCAATATGCCTTATATGCAACTTT
AATTCATGAAGAATGCAATATTTTTGCAATCTTTCTTGTACTGAGCGCTTTTTATCATTTAGTACTCATTATTTAATTTTTCAGTAAAGGGCAATAAATAAATTTGTA
GAAGATGCAATGTAATCTCCTCTATCCTTTTGCTCATATATGTTTATGAGTATACCTAGTCTAGAAAGGCTTGTATTTAAAATATTATTCAATAAAAATTCACAATTT
TTACGACATGTGCTACCATTCACCTAACTTCCTGATTATAAAATTGGTTCGTTTATACTAATTACTTAAGTACCTTTAACTAAACAAAATGCCTATATATATATTAAT
TTACAATGAGTGTCAGATAAGTCACTATGTCCGAGTGGTTAAGGAGAGATCTGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTAAAT
AAACTCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTGAAGTCTTTCTATTGTTCCGGTCGGCATTACCTAATAAATTCGAATCGAGATTGCT
AGTACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTACTAATGGAATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATAC
TATGCACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTGGCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTAT
CCCAGCGGGGTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGCACTTCAATTACTTGAAATTCACCATAACACTTGCTCTAGTCA
AGACTTACAATTAAAACTAGACTAGTTCGAAACGATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGTTCATGCGTTTCAAGGTCAGAATGGAAGGTTCTATGA
ACGGTCATGAGTTCGAGATTGAAGGAGAAGGTGAAGGAAGACCATATGAGGGTACTCAAACCGCAAAGTTGAAGGTTACTAAAGGAGGTCCTTTACCATTCGC
TTGGGATATCCTGTCTCCACAATTCATGTATGGTTCTAAGGCATACGTTAAGCATCCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGAGGGTTTCAAGT
GGGAAAGAGTCATGAACTTCGAAGACGGTGGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGATGCGTGGTACC
AACTTCCCACCAGATGGTCCTGTTATGCAGAAAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATCCAAGAGATGGTGTTTTGAAGGGTGAGATTCA
CCAAGCTTTGAAGCTTAAAGATGGAGGTCACTACTTGGTTGAGTTCAAGACCATTTACATGGCTAAGAAACCAGTTCAACTTCCTGGATACTATTACGTTGACACT
AAGCTGGACATTACCTCTCACAACGAAGACTACACCATCGTTGAGCAATACGAGAGATCCGAAGGTAGACACCACTTGTTCTTGTACGGTATGGACGAGCTTTAT
AAGTAAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTT
TGTCATTTTGTTTCTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTT
TCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCaCGCGGGTCACCCGGCCAGCGACA
TGGAGGCCCAGATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACATCCCCATGTATAATCATTT
GCATCCATACATTTTGATGGCCGCGACGGCGCAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCAGCAACGCGTTGAATTC
TCCCCACGGCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACTC
ACATCCGAACATAAACAAAAATGGGTAAGGAAAAGACTCACGTTAACCTCGAGATGGGAAAAGACTATACACCACTAGTTGCTACCATCGATATTGGTACTACCT
CCACCAGAGCTATTCTTTTTGACTACCACGGTCAGGAAGTGGCCAAGCACCAGATCGAGTACTCTACCTCTGCTCAGGATGATATCAAAAGAAAGCGTTCTCAGA
TCATCTCTTCCGAAGGTATTTCCCTGACAGTTTCTGACGACTTGGAAGTTGAGTCCGTTGACAATAAGGCTGGTCCAACTTTGCAATTTCCTCAGCCAGGCTGGGT
TGAATGTCGTCCAAGTCACATCTTGGCCAACGCCGTTCAGTGTCTTGCTGCTTGTTTAGTCACCATGGAGAACAAGAACTTGGACCGAGATGAAAAAAATAAATA
CAAGCTCATATCTATCGGTGTTGCCAACATGAGAGAGACCACGGTTGTTTGGTCCAAGAAGACAGGAAAGCCTCTTTACAACGGTATTGTGTGGAACGATACCA
GAAACAACGATATTGTTGACGAGTACACCGCCAAGTACTCTGAGAAGGAGAGAGAGGAAATGCGAACCTTGTGTGGTTGTCCAATCTCCACCTACTTTTCTGCAA
CCAAGTTCAGATGGTTACTGAAGCATGTTCCTGAGGTCAAACAGGCCTATGACAATGCTGATGGGGATCTAATGTTTGGTACTATTGACTCTTGGTTGATTTACCA
CTTGACTAACGAAAAATCCCACGTCACTGATGTTACCAATGCCTCCAGAACCAACTTCATGAACATTGAAACCAACAAATATGACGACAGACTTTTGAAATTCTGG
GACGTCGATACTTCCAAAGTCATCCTTCCAGAAATCAGATCTTCCGCAGAAGTCTACGGACACTTCAAGGTCCCACACTTGGAGTCTATTGGATATGTTGAGTCTT
ACCTTACTGATGACGCCCTTGCTCTTCTGGAAACTATTGAAGGTGCTCCTTTGGCTGGATGTCTGGGTGACCAATCTGCCTCTTTGGTTGGACAGTTGGCTGTCAG
AAAGGGTGATGCCAAATGTACATATGGTACCGGTGCTTTCTTGCTGTACAACACTGGTGATCAGACTTTGATTTCTGAGCACGGTGCTTTGACCACTGTGGGATA
TTGGTTCCCAGGTTTGGATGAGTCCGAAGATGGCAAACACTCTTCTAAGCCACAGTATGCTTTGGAGGGATCGATTGCTGTCGCTGGATCTGTCGTGCAATGGTT
AAGAGATAACCTTCGTTTGATTTCCAAGGCTCAGGACGTCGGACCATTGGCTTCTCAAGTTGACAACTCTGGAGGTGTGGTATTTGTTCCAGCATTTTCAGGATTG
TTTGCCCCTTACTGGGATTCCAACTCCAGAGGAACCATTTTCGGTCTGACCCAATACACCTCAGCTTCTCATATTGCTAGAGCTGCTTTGGAAGGTGTCTGTTTCCA
AACTAGAGCCATTTTGAAGGCCATGATCAGCGATGCAGGAGCTTCTGCTGACTTTTTGGAGGAATCATCCAAGGCCACTGGCCACAACCCTCTGTCAGTTCTTGC
CGTGGACGGAGGTATGTCCAAATCAGACGAGATGATGCAGATCCAAGCTGATATTTTGGGTCCATGTGTCACTGTTAGACGTTCCATCAACCCTGAATGTACTGC
ACTGGGAGCTGCCATTGCTGCCGGTTTTGGTGTCCCTAAGGAAGATAGAATTTGGGGTTCCTTGAAGGAATGTACCGAGGCCATTCTTGAGGGTAACAAGATGT
ACTTGGCTGCAGGGAACACTTCTTTGGACTTCAAGGCCACATTGAGCGACGAGGTCAGAAGAAAGGAATGGAGATTGTGGGAAAATGCCATTGCAAAGGCAAA
GGGCTGGCTTAAGGACACTGCTTAAGGTCGACAAGGGGTAAGAAATAGAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTATAGTTTTTTT
ATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGT
CAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAACGCCGCCCCAGTGTCTACCTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGATAGAA
TATGCTCGAACTGTTGAAGTCCATCAACAAGGATAACCCATATGCTCTATCGGCGGAGAAAACGTTCAGAGCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCT
AGATGAGAAGTACGGGGTTTAGTGTTTCCAGGCCTCGTAAATGCCGCAATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTC
AAGGACCGCCTTTTCATATATGGCCACCATCAATATAGAGCCAGCGAACGATGATGCATTCTGGTATATTGAACGCGAGGATCCCTGCAGGTACCACTGAGCGTC
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCA
AGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGG
GAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGGTACCC

4.3 ADDITIONAL *P. PASTORIS* PROMOTERS TESTED



4.4 REFERENCES

1. Kumar, R., Mustafiz, A. & Kumar, K. Functional screening of cDNA library from a salt tolerant rice genotype Pokkali identifies mannose-1-phosphate guanyl transferase gene (OsMPG1) as a key member of salinity stress response. 555–568 (2012). doi:10.1007/s11103-012-9928-8
2. Wan, D. *et al.* Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc Natl Acad Sci* **101,** 15724–15729 (2004).
3. Rabausch, U. *et al.* Functional Screening of Metagenome and Genome Libraries for Detection of Novel Flavonoid-Modifying Enzymes. **79,** 4551–4563 (2013).
4. Mattéotti, C. *et al.* New glucosidase activities identified by functional screening of a genomic DNA library from the gut microbiota of the termite Reticulitermes santonensis. **166,** (2011).
5. Ogura, A. *et al.* Effective gene collection from the metatranscriptome of marine microorganisms. **12,** (2011).
6. Baltimore, D. & Smoler, D. Primer requirement and template specificity of the DNA polymerase of RNA tumor viruses. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 1507–11 (1971).
7. Baltimore, D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226,** 1209–1211 (1970).
8. Britain, G. *et al.* The Biology of Hepadnaviruses. 1215–1235 (2015).
9. Modrow, S., Falke, D., Truyen, U. & Schätzl, H. in *Molekulare Virologie* 409–430 (2010).
10. Krug, M. S. & Berger, S. L. First-strand cDNA synthesis primws with oligo(dT). *Methods Enzymol.* **152,** 316–325 (1987).
11. Lim, D. *et al.* Crystal structure of the moloney murine leukemia virus RNase H domain. *J. Virol.* **80,** 8379–8389 (2006).
12. Hook, B., Bratz, M. and Schagat, T. *M-MLV Reverse Transcriptase RNase H Minus Point Mutant: A Cost-Effective Alternative to SuperScript® II.* (2012).
13. Roth, M. J., Tanese, N. & Goff, S. P. Purification and characterization of murine retroviral reverse transcriptase expressed in Escherichia coli. *J. Biol. Chem.* **260,** 9326–35 (1985).
14. Lambert, K. N. & Williamson, V. M. cDNA library construction from small amounts of RNA using paramagnetic beads and PCR. *Nucleic Acids Res.* **21,** 775–776 (1993).
15. Raineri, I., Moroni, C. & Senn, H. P. Improved efficiency for single-sided PCR by creating a reusable pool of first-strand cDNA coupled to a solid phase. *Nucleic Acids Res.* **19,** 4010 (1991).
16. Walker, J. M. & Ying, S.-Y. *Generation of cDNA Libraries*. (2003).

17. Clark, J. M. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16,** 9677–9686 (1988).

18. Zajac, P., Islam, S., Hochgerner, H., Lönnerberg, P. & Linnarsson, S. Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS One* **8,** e85270 (2013).

19. Kulpa, D., Topping, R. & Telesnitsky, A. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors. *EMBO J.* **16,** 856–865 (1997).

20. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: A SMART??? approach for full-length cDNA library construction. *Biotechniques* **30,** 892–897 (2001).

21. Cereghino, J. L. & Cregg, J. M. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol. Rev.* **24,** 45–66 (2000).

22. Cregg, J. M., Cereghino, J. L., Shi, J. & Higgins, D. R. Recombinant Protein Expression in Pichia pastoris. *Mol. Biotechnol.* **16,** 23–52 (2000).

23. Wegner, G. H. Emerging applications of the methylotrophic yeasts. *FEMS Microbiol. Lett.* **87,** 279–283 (1990).

24. Ellis, S. B. *et al.* Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. *Mol. Cell. Biol.* **5,** 1111–21 (1985).

25. Cregg, J. M., Barringer, K. J., Hessler, A. Y. & Madden, K. R. Pichia pastoris as a host system for transformations. *Mol. Cell. Biol.* **5,** 3376–85 (1985).

26. Cregg, J. M. & Russell, K. a. *Pichia Protocols*. *Methods in molecular biology (Clifton, N.J.)* **103,** (1998).

27. Watts, F., Castle, C. & Beggs, J. Aberrant splicing of Drosophila alcohol dehydrogenase transcripts in Saccharomyces cerevisiae. *EMBO J* **2,** 2085–2091 (1983).

28. Langford, C., Nellen, W., Niessing, J. & Gallwitz, D. Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 1496–500 (1983).

29. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. *Methods Enzymol.* **194,** 795–823 (1991).

30. Giga-Hama, Y. in *Foreign Gene Expression in Fission Yeast: Schizosaccharomyces pombe* 3–28 (1997).

31. European Molecular Biology Laboratory. PomBase. (2016). Available at: http://www.pombase.org/status/statistics. (Accessed: 8th April 2016)

32. Potashkin, J., Naik, K. & WentzHunter, K. U2AF homolog required for splicing in vivo. *Science\r262* 573–575 (1993).

33. Schwarzhans, J. P. *et al.* Integration event induced changes in recombinant protein productivity in Pichia pastoris discovered by whole genome sequencing and derived vector optimization. *Microb. Cell Fact.* 1–15 (2016). doi:10.1186/s12934-016-0486-7

34. Keeney, J. B. & Boeke, J. D. Efficient Targeted Integration at leu1-32 and ura4-294 in Schizosaccharomyces pombe. *Genetics* **136,** 849–856 (1994).

35. Clyne, R. K. & Kelly, T. J. Identification of autonomously replicating sequence (ARS) elements in eukaryotic cells. *Methods* **13,** 221–33 (1997).

36. Clyne, R. K. & Kelly, T. J. Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe. *EMBO J.* **14,** 6348–57 (1995).

37. Parent, S., Fenimore, C. & Bostian, K. Vector systems for the expression, analysis and cloning of DNA sequences in S. cerevisiae. *Yeast* **1,** 83–138 (1985).

38. Sikorski, R. S. & Hieter, P. A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in. (1989).

39. Steinborn, G. *et al.* Application of a wide-range yeast vector (CoMed) system to recombinant protein production in dimorphic Arxula adeninivorans, methylotrophic Hansenula polymorpha and other yeasts. *Microb. Cell Fact.* **5,** 33 (2006).

40. Felder, Y. Analysis of Biological Networks : Transcriptional Networks - Promoter Sequence Analysis. *Gene* 1–15 (2007).

41. Yano, T., Takigami, E., Yurimoto, H. & Sakai, Y. Yap1-Regulated Glutathione Redox System Curtails Accumulation of Formaldehyde and Reactive Oxygen Species in Methanol Metabolism of Pichia pastoris □ †. **8,** 540–549 (2009).

42. Vogl, T. *et al.* A toolbox of diverse promoters related to methanol utilization – functionally verified parts for heterologous pathway expression in Pichia pastoris. *ACS Synth. Biol.* acssynbio.5b00199 (2015). doi:10.1021/acssynbio.5b00199

43. Cohen, R., Yokoi, T., Holland, J. P., Pepper, a E. & Holland, M. J. Transcription of the constitutively expressed yeast enolase gene ENO1 is mediated by positive and negative cis-acting regulatory sequences. *Mol. Cell. Biol.* **7,** 2753–61 (1987).

44. Drocourt, D., Calmels, T., Reynes, J. P., Baron, M. & Tiraby, G. Cassettes of the Streptoalloteichus hindustanus ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Res.* **18,** 4009 (1990).

45. Gatignol, A., Baron, M. & Tiraby, G. Phleomycin resistance encoded by the ble gene from transposon Tn 5 as a dominant selectable marker in Saccharomyces cerevisiae. *MGG Mol. Gen. Genet.* **207,** 342–348 (1987).

46. Hentges, P., Van Driessche, B., Tafforeau, L., Vandenhaute, J. & Carr, A. M. Three novel antibiotic marker cassettes for gene disruption and marker switching inSchizosaccharomyces pombe. *Yeast* **22,** 1013–1019 (2005).

47. Benko, Z. & Zhao, R. Zeocin for selection of bleMX6 resistance in fission yeast. *Biotechniques* **51,** 1–4 (2011).

48. Lin-Cereghino, J. *et al.* Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *Biotechniques* **38,** 44–48 (2005).

49. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–5 (2009).

50. Suga, M. & Hatakeyama, T. A rapid and simple procedure for high-efficiency lithium acetate transformation of cryopreserved Schizosaccharomyces pombe cells. *Yeast* **22,** 799–804 (2005).

51. Nurse, P. *Fission Yeast Handbook: Welcome to the lab.* (2014).

52. Soden, D. M., O'Callaghan, J. & Dobson, a. D. W. Molecular cloning of a laccase isozyme gene from Pleurotus sajor-caju and expression in the heterologous Pichia pastoris host. *Microbiology* **148,** 4003–4014 (2002).

53. Kooiker, M. & Xue, G. Cereal Genomics. **1099,** 29–40 (2014).

54. Harju, S., Fedosyuk, H. & Peterson, K. R. Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnol.* **4,** 8 (2004).

55. Ying, S.-Y. *Generation of cDNA Libraries*. *Generation of cDNA Libraries* **221,** (2003).

56. Nam, D. K. *et al.* Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6152–6156 (2002).

57. Marrs, R. H. & Watt, A. S. Biological flora of the British Isles: Pteridium aquilinum (L.) Kuhn. *J. Ecol.* **94,** 1272–1321 (2006).

58. Moelling, K. Characterization of reverse transcriptase and RNase H from friend-murine leukemia virus. *Virology* **62,** 46–59 (1974).

59. Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res.* **16,** 265–77 (1988).

60. Xu, L. *et al.* Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Mol. Biol. Evol.* **23,** 1107–1108 (2006).

61. Berger, S., Wallace, D., Puskas, R. & Eschenfeldt, W. Reverse transcriptase and its associated ribonuclease H: interplay of two enzyme activities controls the yield of single-stranded complementary deoxyribonucleic acid. *Biochemistry* **22,** 2365–2372 (1983).

62. Gerard, G. F. *et al.* The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res.* **30,** 3118–29 (2002).

63. Lévesque, V., Fayad, T., Ndiaye, K., Diouf, M. N. & Lussier, J. G. Size-selection of cDNA libraries for the cloning of cDNAs after suppression subtractive hybridization. *Biotechniques* **35,** 72–78 (2003).

64. Näätsaari, L. *et al.* Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* **7,** (2012).

65. Embleton, M. L., Vologodskii, A. V. & Halford, S. E. Dynamics of DNA loop capture by the SfiI restriction endonuclease on supercoiled and relaxed DNA. *J. Mol. Biol.* **339,** 53–66 (2004).

66. Clarke, L. & Carbon, J. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287,** 504–509 (1980).

67. Vogl, T. & Glieder, A. Regulation of Pichia pastoris promoters and its consequences for protein production. *N. Biotechnol.* **30,** 385–404 (2013).

68. Wang, H. *et al.* Identification and refinement of two strong constitutive promoters for gene expression system of Schizosaccharomyces pombe. *World J. Microbiol. Biotechnol.* **30,** 1809–1817 (2014).

69. Steiner, S., Philippsen, P., Gieben, U. & Str, F. Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus Ashby gossypii. 263–271 (1994).

70. Kurth, R., Philippsen, P., Steiner, S. & Wright, M. C. Isolated promoter and terminator of elongation factor 1-α. (1994).

71. Scharff, L. B., Childs, L., Walther, D. & Bock, R. Local absence of secondary structure permits translation of mrnas that lack ribosome-binding sites. *PLoS Genet.* **7,** 1–11 (2011).

72. Project Lab Report, F. *Jasmin Elgin Fischer, Novel Expression strategies in P. pastoris*. (2014).

73. Kjærulff, S. & Jensen, M. R. Comparison of different signal peptides for secretion of heterologous

proteins in fission yeast. *Biochem. Biophys. Res. Commun.* **336,** 974–982 (2005).

74. Geier, M., Fauland, P., Vogl, T. & Glieder, A. Compact multi-enzyme pathways in P. pastoris. *Chem. Commun.* **51,** 1643–1646 (2015).

75. Tekere, M., Mswaka, a Y., Zvauya, R. & Read, J. S. Growth, dye degradation and ligninolytic activity studies on Zimbabwean white rot fungi. *Enzyme Microb. Technol.* **28,** 420–426 (2001).

76. Solovyev, V. V. Solovyev, A Salamov (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), Nova Science Publishers, p.61-78. *Nov. Sci. Publ.* 61–78 (2011).

4.5 LIST OF TABLES

4.6 LIST OF FIGURES

4.7 ACIB PROTOCOL

ACIB Protocol

cDNA library generation from total RNA by template switching mechanism

Author: Jasmin Elgin Fischer

Jasmin.fischer@acib.at

Creation date: 2015/10/20 Restrictions: no Validity: until revoked/cancelled This ACIB protocol replaces the version from: Developed in project/working group: 22061/AG09		
Notice of modification:		
Written by: Jasmin Elgin Fischer		
Date/signature:		
Checked:	by:	Date/Signature
Approved by:		
Released by		

1 Purpose and Field of Application

This method was developed to generate full-length cDNA libraries from total RNA of eukaryotic cells for sequencing and cDNA expression library generation.

2 Principle

The method is based on the template switching mechanism of the M-MLV reverse transcriptase during first strand cDNA generation. M-MLV adds with ~80% probability a triple cytidine to the cDNA first strand as it reaches the 5' end. By using a primer harboring three RNA molecules (guanosines) at the 3' end and a vector overhang at the 5' end, the M-MLV RT recognizes the guanosines as extended template and keeps on synthesizing the complement to the vector overhang. Two normal DNA primers are then used to generate the second strand and amplify the generated cDNA.

3 Key Words, Definitions & Abbreviations

cDNA library, library generation, template switching, RNA, reverse transcription

4 METHODOLOGY

4.1 REAGENTS

Name	Formula	MW	Purity	Supplier	CAS No.	Comments
DTT	$C_4H_{10}O_2S_2$	154,25	≥99%	Roth	6908.2	
Hydrogen chloride	HCl	36,46	37%	Roth	4625.2	
Magnesium chloride	$MgCl_2$	95,21	≥99%	Roth	KK36.3	
Potassium chloride	KCl	74,55	≥99%	Roth	P017.1	
Tris-Base	$C_4H_{11}NO_3$	121,14		Roth	4855.2	
Sodium dodecyl sulfate	$C_{12}H_{25}NaO_4S$	288,4		Roth	CN30.4	

4.2 SOLUTIONS

SOC Medium

Name	Concentration
SOB	30.7 g/L
glucose	3.6 g/L

Poly-d(T) Primer Mix:

Contains 12 different poly-d(T) primers, endconcentration 10 µM

5X RT Buffer pH 8.3

Name	Concentration
DTT	10 mM
Tris	250 mM
$MgCl_2$	30 mM
KCl	375 mM

Adjust pH with HCl conc. to 8.3

5X ISO reaction buffer (for assembly cloning master mix)

Name	Concentration
PEG-8000	25% (w/v)
Tris-HCl pH 7.5	500 mM
$MgCl_2$	50 mM
DTT	50 mM
dNTPs	1 mM of each dNTP
NAD	5 mM

Assembly Cloning Master Mix

Name	volume
5X ISO reaction buffer	320 µL
T5 exonuclease (10 U/µL)	0.64 µL
Phusion Polymerase (2 U/µL)	20 µL
Taq DNA Ligase (40 U/µL)	160 µL
ddH_2O	Up to 1.2 mL

4.3 MATERIALS

Name	Supplier	Order No.
1 µg total RNA	Self-isolated	
Adapter Primer fwd (10 µM)	IDT	
Adapter Primer rev (10 µM)	IDT	
CHROMA SPIN-TE 1000 columns	Clontech	636079
dNTP mix 10 mM (RNase free)	Thermo Fisher Scientific	R0192
FastDigest *Bcu*I (*Spe*I)	Thermo Fisher Scientific	FD1253
FastDigest *Not*I	Thermo Fisher Scientific	FD0594
GeneJET Gel Extraction Kit	Thermo Fisher Scientific	K0692
Microcentrifuge tubes (RNase-free)	Greiner bio-one	227261
Millipore MF-Millipore™ Mixed Cellulose Ester Membranes: 0.025µm Pore Size	Merck Millipore	VSWP01300
M-MuLV Reverse Transcriptase (100 U/µL)	NEB	M170A
NAD	Sigma-Aldrich	000000010127965001
PCR tubes (RNase-free)	Any supplier	
PEG-8000	Roth	0263.2
Phusion High-Fidelity DNA Polymerase (2 U/µL)	NEB	M0530L
Plant/Fungi RNA Purification Kit	Norgen Biotek Corporation	25800
Poly-d(T) Primer Mix 10 µM	IDT	
Ribo-Primer 12 µM	IDT	

RNase H (2 U/µL)	Thermo Fisher Scientific	EN0201
RNase out	Invitrogen	10777019
RNase-free water	Thermo Fisher Scientific	R0581
SOB Medium	Carl Roth	AE27.1
T5 Exonuclease (10 U/µL)	NEB	M0363S
Taq DNA Ligase	NEB	M0208L

4.4 Apparatus

Name	Supplier
2720 Thermal Cycler	Applied Biosystems
Sub-Cell® GT Cell	Bio-Rad Laboratories
PowerPac Basic power supply	Bio-Rad Laboratories
MicroPulser™ Electroporator	Bio-Rad Laboratories

4.5 Procedure

Primer Design:

- Poly-d(T) primer mix (the “reverse” primers)

  Design the primer which is used to bind to the poly-A tail of the mRNA with appropriate overhangs to terminator on the vector either with restriction sites for conventional cloning or Assembly Cloning. Use 17 Ts at the 3’ end followed by each possible combination of A, C and G (anchor nucleotides), which results in 12 primers in total. For example:

  TTCTGACATCCTCTTGAGCGGCCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTXX

  Vector overhang + *Not*I restriction site anchor nucleotides

- Ribo-Primer (the “forward” primer)

  The primer should contain an appropriate overhang to the promoter by which the library expression should be driven because it is used as “forward” primer with 3 ribonucleotide G at the 3’ end and a melting temperature about 58°C. For example:

CTCTAGTCAAGACTTACAATTAAAACTAGACTAGTTCGAAACGr(GGG)

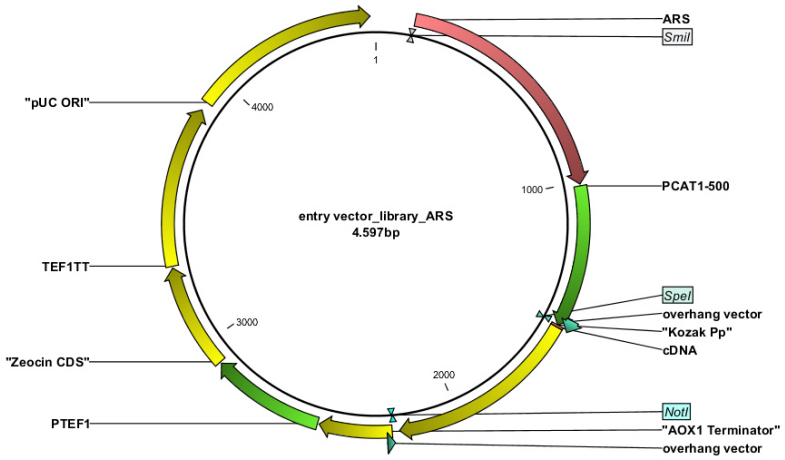
Vector overhang *Spe*I restriction site ribonucleotide G

- Primer for 2nd strand synthesis primers

  The forward primer for the second strand synthesis and amplification of the cDNA should be the same as the ribo primer but without the ribonucleotides, the reverse primer should be the same as the poly-d(T) primer lacking the T-tail. (see the yellow parts above)

  These primers can also be used for colony-PCR afterwards to evaluate the length of the inserts, as well as for sequencing the library.

Vector for example:



RNA Isolation

For plant RNA about 1-2 g of freeze-dried tissue are necessary as starting material and purified Plant/Fungi RNA Purification Kit.

1. About 1-2 g of freeze-dried plant tissue is doused with liquid nitrogen in a mortar and pestled until the tissue is completely pulverized.
2. From 150 mg of this tissue total RNA is purified according to the kits manual.
3. Concentration of the RNA after purification as well as 260 nm/280 nm and 230 nm/260 nm values were determined by spectrophotometric measurements with nanodrop. The 230/260 value should exceed 2.2, the 260/280 value should exceed 2.1 to make sure that the RNA quality is sufficient for successful cDNA generation.

Vector preparation:

1. Digest at least 2 µg of the plasmid "entry vector_library_ARS" with *Spe*I and *Not*I to remove the stuffer fragment
2. Prepare the following reaction:

Reagent	volume
2 µg entry vector	X µL
Fast Digest *Not*I	2 µL
Fast Digest *Spe*I	2 µL
10X Fast Digest Buffer	4 µL
ddH_2O	Up to 40 µL

3. Incubate the reaction at 37°C for 1 hour.
4. Load the reaction mixture and an appropriate DNA ladder on a 1% agarose gel. Electrophoresis is done for 40 minutes at 110 V.
5. Excise the correct sized backbone (3870 bp)
6. Purify via GeneJET Gel Extraction Kit
7. Determine concentration photometrically or check on a control gel.

First strand synthesis:

1. RNase-free PCR tube with the following contents (always on ice!)

Reagent	volume
Total mRNA	1 µg
Poly-d(T) primer mix 10 µM	1 µL
RNase-free water	Up to 4 µL

Incubate the mixture at 72°C for 2 minutes, then at 42°C for 3 minutes in a thermocycler. Place on ice afterwards.

2. In the meantime prepare the following mixture in a RNase-free PCR tube

Reagent	volume

5X RT Buffer	2 µL
dNTP mix RNase free (10 mM)	1 µL
Ribo primer (12 µM)	1 µL
RNase out	0.25 µL
RNase free water	0.75 µL
M-MLV reverse transcriptase	1 µL

Add the mixture to the PCR tube of step 1, mix well by pipetting. Incubate for 90 minutes at 42°C, followed by 10 minutes at 70°C. Place on ice.

3. Add 1 µL RNase H and incubate at 37°C for 20 minutes, then heat to 85°C for 5 minutes to inactivate the enzymes. Place on ice again.

Second strand synthesis:

1. Prepare PCR reaction solution containing

Reagent	**volume**
5X HF Buffer	20 µL
dNTP mix (2 mM)	10 µL
Anchor primer fwd (10 µM)	4 µL
Anchor primer rev (10 µM)	4 µL
cDNA (from first strand sysnthesis)	10 µL
Phusion polymerase (2 U/µL)	0.6 µL
ddH_2O	71.4 µL

Use the following program for the 2nd strand synthesis

Step	**time**	**temperature**
Initial denaturation	180 s	98°C

Denaturation	30 s	98°C
Annealing	30 s	58°C
Extension	150 s	72°C
Final extension	420 s	72°C
storage	∞	4°C

Size selection of cDNA

1. Possibility 1
    a. Pour a 1% (w/v) agarose gel containing 0.5 µg/mL Ethidiumbromide
    b. Add 20 µL 6x loading dye + SDS to the PCR reactions and mix well
    c. Apply the whole sample on a 1% (w/v) agarose gel, next to Gene Ruler 1 kb DNA ladder and run the gel for 40 min at 120V.
    d. Cut out the area between 2000 and 10000 bp and purify via GeneJET Gel Extraction Kit.
    e. Measure concentration photometrically or check on control gel.
2. Possibility 2
    a. Add 20 µL 6x loading dye + SDS to your PCR reactions and mix well
    b. Purify via CHROMA SPIN-TE 1000 columns according to manufacturer's manual.

Cloning

1. Use 15 µL assembly cloning master mix for 1 µL appropriate digested vector (about 30 ng) and 4 µL double stranded cDNA. Incubate for 1 h at 50°C. The master mix contains T5 Exonuclease for generating sticky ends, *Taq* DNA ligase for ligating insert and vector and Phusion polymerase for filling up the gaps.
2. Desalt the Assembly reaction by pipetting it on a 0.025 µm membrane filter and dialyzing for 20 minutes.
3. Transform the whole reaction in 100 µL *E. coli* K12 Top10 F' and resuspend the cells in 900 µL SOC medium. For using of ampicillin resistance as selection marker 10 minutes of regeneration time are required, for other antibiotics I would recommend 45-60 minutes.

4. Plate 100 µL on an appropriate antibiotic plate and inoculate 10 mL selective medium containing the antibiotic with the residual 900 µL culture. Incubate o/n at 37°C shaking.

5. Make Minipreps of the liquid culture (2-3 Minis)

5 SAFETY PRECAUTIONS

Please follow instructions described in "acib-Mitarbeiterleitfaden Gefahrstoff- und Laborordnung"

6 DOCUMENTATION

All observations, protocol deviations and calculations are recorded in the lab book (lab book I, page 58).

7 REFERENCES

Farrell, Robert E. Jr., 'cDNA Synthesis', in RNA Methodologies, 2010, pp. 376–84 <http://dx.doi.org/10.1016/B978-0-12-374727-3.00017-6>

Gerard, G F, D K Fox, M Nathan, and J M D'Alessio, 'Reverse Transcriptase. The Use of Cloned Moloney Murine Leukemia Virus Reverse Transcriptase to Synthesize DNA from RNA.', Molecular Biotechnology, 8 (1997), 61–77 <http://dx.doi.org/10.1007/BF02762340>

Gerard, Gary F, R Jason Potter, Michael D Smith, Kim Rosenthal, Gulshan Dhariwal, Jun Lee, and others, 'The Role of Template-Primer in Protection of Reverse Transcriptase from Thermal Inactivation.', Nucleic Acids Research, 30 (2002), 3118–29 <http://dx.doi.org/10.1093/nar/gkf417>

Gibson, Daniel G, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde a Hutchison, and Hamilton O Smith, 'Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases.', Nature Methods, 6 (2009), 343–45 <http://dx.doi.org/10.1038/nmeth.1318>

Ying, S.-Y, Generation of cDNA Libraries, Generation of cDNA Libraries, 2003, ccxxi <http://dx.doi.org/10.1385/1592593593>

Zajac, Pawel, Saiful Islam, Hannah Hochgerner, Peter Lönnerberg, and Sten Linnarsson, 'Base Preferences in Non-Templated Nucleotide Incorporation by MMLV-Derived Reverse Transcriptases.', PloS One, 8 (2013), e85270 http://dx.doi.org/10.1371/journal.pone.0085270

Hook, B., Bratz, M. and Schagat, T., M-MLV Reverse Transcriptase RNase H Minus Point Mutant: A Cost-Effective Alternative to SuperScript® II., 2012 <http://at.promega.com/resources/pubhub/mmlv-reverse-transcriptase-rnase-h-minus-point-mutant-a-cost-effective-alternative/>

Kotewicz, Michael L, Craig M Sampson, James M D'Alessio, and Gary F Gerard, 'Isolation of Cloned Moloney Murine Leukemia Virus Reverse Transcriptase Lacking Ribonuclease H Activity', Nucleic Acids Research, 16 (1988), 265–77 <http://dx.doi.org/10.1093/nar/16.1.265>

Lim, David, G Glenn Gregorio, Craig Bingman, Erik Martinez-Hackert, Wayne A Hendrickson, and Stephen P Goff, 'Crystal Structure of the Moloney Murine Leukemia Virus RNase H Domain', Journal of Virology, 80 (2006), 8379–89 <http://dx.doi.org/10.1128/JVI.00750-06>

Nam, Douglas Kyung, Sanggyu Lee, Guolin Zhou, Xiaohong Cao, Clarence Wang, Terry Clark, and others, 'Oligo(dT) Primer Generates a High Frequency of Truncated cDNAs through Internal poly(A) Priming during Reverse Transcription.', Proceedings of the National Academy of Sciences of the United States of America, 99 (2002), 6152–56 <http://dx.doi.org/10.1073/pnas.092140899>

Roth, M J, Naoko Tanese, and Stephen P Goff, 'Purification and Characterization of Murine Retroviral Reverse Transcriptase Expressed in Escherichia Coli.', The Journal of Biological Chemistry, 260 (1985), 9326–35 http://dx.doi.org/10.1111/j.1462-5822.2011.01744.x

Sequence Vector:

>entry vector_library_ARS

AGATCCAATTCCCGCTTTGACTGCCTGAAATCTCCATCGCCTACAATGATGACATTTGGATTTGGTTGACTCATGTTGGTATTGTGAAATA
GACGCAGATCGGGAACACTGAAAAATACACAGTTATTATTCATTTAAATGACAGAATGGGATACAAGGGCATCGTCTATAATTATAGCTA
AAAATTGTATTTTAATTTGTATTTTTTGTAATTTTATTTCTCATTGTTTTACTTAAAATGAATGCGAATTAGAGAAAACTTCAACGAAATGTC
AAAATAAGCTCAGCAAAATATACAATTTTAGGGAAAGCGATCAGCAACTTATTCTCCGTTGTAGATTCTTTTTCTTCATAAATTCCAACAAC
GTAACTGCCAATTGATTTTTCTCTAGATGCTAATAAATTGAAACTAATGAGTATAATAACACAGGTTTTGATAAATTGTAAACTCTTACTCA
AAAAAAGAATGTTAAAACTAATTTAACCTTTTTATAAAGGCTCATGGCGTTAAAGAGGATAGTAAAAATTCTACAATTTTTTAATTTAATG
TCGCAATTGTCTTGCGAATATAATTCTATTAATTTTCGCTAACTGCTGTTTTAGTACTATTACTTTCAATATGCCTTATATGCAACTTTAATTC
ATGAAGAATGCAATATTTTTGCAATCTTTCTTGTACTGAGCGCTTTTTATCATTTAGTACTCATTATTTAATTTTTCAGTAAAGGGCAATAAA
TAAATTTGTAGAAGATGCAATGTAATCTCCTCTATCCTTTTGCTCATATATGTTTATGAGTATACCTAGTCTAGAAAGGCTTGTATTTAAAA
TATTATTCAATAAAAATTCACAATTTTTACGACATGTGCTACCATTCACCTAACTTCCTGATTATAAAATTGGTTCGTTTATACTAATTACTT
AAGTACCTTTAACTAAACAAAATGCCTATATATATATTAATTTACAATGAGTGTCAGATAAGTCACTATGTCCGAGTGGTTAAGGAGAGAT
CTGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTAAATAAACTCATGGCCTGACATCTGTACACGTTCTTATTG
GTCTTTTAGCAATCTTGAAGTCTTTCTATTGTTCCGGTCGGCATTACCTAATAAATTCGAATCGAGATTGCTAGTACCTGATATCATATGAA
GTAATCATCACATGCAAGTTCCATGATACCCTCTACTAATGGAATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATACTATGC
ACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTGGCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTA
AGTTTCATTATCCCAGCGGGGTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGCACTTCAATTACTTGAA
ATTCACCATAACACTTGCTCTAGTCAAGACTTACAATTAAAACTAGACTAGTTCGAAACGATGGTTTCTAAGGGTGAGGAAGTTATCAAG
GAGTTCATGCGTTTCAAGGTCAGAATGGAAGGTTCTATGAACGGTCATGAGTTCGAGATTGAAGGAGAAGGTGAAGGAAGACCATATG
AGGGTACTCAAACCGCAAAGTTGAAGGTTACTAAAGGAGGTCCTTTACCATTCGCTTGGGATATCCTGTCTCCACAATTCATGTATGGTTC
TAAGGCATACGTTAAGCATCCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGAGGGTTTCAAGTGGGAAAGAGTCATGAACTT

CGAAGACGGTGGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGATGCGTGGTACCAACTT
CCCACCAGATGGTCCTGTTATGCAGAAAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATCCAAGAGATGGTGTTTTGAAGG
GTGAGATTCACCAAGCTTTGAAGCTTAAAGATGGAGGTCACTACTTGGTTGAGTTCAAGACCATTTACATGGCTAAGAAACCAGTTCAAC
TTCCTGGATACTATTACGTTGACACTAAGCTGGACATTACCTCTCACAACGAAGACTACACCATCGTTGAGCAATACGAGAGATCCGAAG
GTAGACACCACTTGTTCTTGTACGGTATGGACGAGCTTTATAAGTAAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTGCCTGAGAG
ATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCTCCT
GATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACTCC
TCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCAAGCTTGCCTCGTCCCACGCGGGTCACCCGGCCAGCGACATGG
AGGCCCAGATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCACGGGGCATGATGTGACTGTCGCCCGTACATTTACCCATACATCCCCAT
GTATAATCATTTGCATCCATACATTTTGATGGCCGCGACGGCGCAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAAC
GCTCCCCTCAGCAACGCGTTGAATTCTCCCCACGGCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCACTGAGGTTCTTCTTTCA
TATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACTCACATCCGAACATAAACAAAAATGGGTAAGGAAAAGACTCACGTTAACCT
CGAGATGGCTAAACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTGGACTGATAGACTCGGTTTC
TCCCGTGACTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCAGGACCAGGTTGTGCCA
GACAACACTCTGGCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACTTCCGTGATGC
ATCTGGTCCAGCTATGACCGAGATCGGTGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTGCGTGCATTTCGT
CGCAGAAGAGCAGGACTAAGTCGACAAGGGGTAAGAAATAGAGTAACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTAT
AGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTGCCTCGACATCATCTGCCCAGATGCGAAGTTAA
GTGCGCAGAAAGTAATATCATGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAACGCCGCCCCAGTGTCTAC
CTGTCAAATTTGCCAGCGTCAAATGCCTCCAGGATAGAATATGCTCGAACTGTTGAAGTCCATCAACAAGGATAACCCATATGCTCTATCG
GCGGAGAAAACGTTCAGAGCCGCTTCCTTCCGCAGACGTGCCCCTTCCACTGCTAGATGAGAAGTACGGGGTTTAGTGTTTCCAGGCCTC
GTAAATGCCGCAATAAATGCTTCCTTGGGTTCGCTACGCCCTCAGGCAGACGAGTTTCTACAAAACTTCAAGGACCGCCTTTTCATATATG
GCCACCATCAATATAGAGCCAGCGAACGATGATGCATTCTGGTATATTGAACGCGAGGATCCCTGCAGGTACCACTGAGCGTCAGACCC
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG
GTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGT
AGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGG
CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACA
CAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAA
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATA
GTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC
GCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGGTACCC