Imperial College London

# Viral dsRNA and Killer System in *Saccharomyces Paradoxus*

**Mahsan Nematbakhsh**

A thesis submitted for the degree of Doctor of Philosophy of Imperial College London

Department of Life Sciences
Faculty of Natural Sciences
Imperial College London

June 2016

# DECLARATION

I certify that this thesis represents my own work, unless otherwise stated. All external contributions and any information derived from other sources have been acknowledged accordingly.

MAHSAN NEMATBAKHSH

*In loving memory of my late grandfathers, Mirza Mahmoud Amin and Fatolah Nematbakhsh…*

*Dedicated to my loving husband Meysam, my wonderful son Ali and my darling daughter Layli, who are all so full of life; without them my PhD journey would not have been possible.*

*Dedicated to my mom & dad Sima and Masoud, the most precious gift in my life: for their endless support, love and encouragement throughout my life.*

# ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Professor Austin Burt for accepting me onto his research team and for all his patience and valuable inputs to successfully drive this project forward. I would like also to thank my co-supervisor, Professor Tony Cass and Dr. Vassiliki Koufopanou for their continued support and guidance throughout the project.

Thank you to colleagues, past and present, particularly Dr. Muthu and Ms. Susan Lomas for their continued support and encouragement throughout my tenure.

Finally, a very special thank you to my family members, relatives and friends for being there for me. Last but not least, to the Almighty God for giving me the strength and perseverance to complete my studies within the stipulated time frame and for keeping me surrounded by these wonderful people.

# Abstract

Some yeast strains (killer strains) release killer toxins which kill other strains of the same or different species of yeast but not the strain producing the toxin itself. In *Saccharomyces cerevisiae,* killer toxins are encoded by either dsRNA viruses or nuclear genes.

This study aims to characterise the genetic basis of the killer toxin synthesised in *Saccharomyces paradoxus*, a wild non-domesticated relative of *S. cerevisiae*. One hundred and nine stains of *S. paradoxus* gathered from Silwood Park, Continental Europe, Far East, and North America were screened for the killer-immune phenotype against killer and sensitive yeast strains. In order to find whether the killer toxin is encoded by dsRNA or a nuclear gene, the killer strains were treated by cyclohexamide that removes the dsRNA which carries the killer toxin gene. The strains were also screened for presence and absence of dsRNA by directly visualising the dsRNA on the agarose gel and also by Next Generation Sequencing (NGS).

Several strains were found to have the killer phenotype (30% of the screened strains). In the majority of killer strains the toxins were encoded by dsRNA except the strains from Canada and one strain from continental Europe, which seems to have other genetic bases. Sixteen full sequences of large dsRNA variants (L-A; with the length of about 4.5 kb) composed of L-A-Q, L-A-D1, L-A-C, L-A-P1.1, L-A-P1.2, L-A-P1.3, L-A-P1.4, L-A-P1.5, L-A-P1.6, L-A-P1.7, L-A-P2.1, L-A-P2.2, L-A-P2.3, L-A-P2.4, L-A-P2.5, and L-A-P2.6, and seven new types of medium size dsRNA (M; with the length between 0.8 and 2kb) composed of MQ, MC, M-P1G1, M-P1G2, M-P1G3-1, M-P1G3-2, M-P1G5, and M-P1SG, have been identified in this study. M28 was the only M dsRNA in *S. paradoxus* which was nearly identical to the M28 from *S. cerevisiae*; therefore likely to have been transferred between the two species.

To test whether the killer toxins were encoded from the new types of M dsRNAs, the MQ ORF sequence from *S. paradoxus* was cloned and expressed into *S. cerevisiae* and then tested for the killer phenotype. The *S. paradoxus* MQ protein was successfully expressed in the *S. cerevisiae* strain and a positive killer response in the killer assay confirmed its function as a killer toxin.

# List of Tables

## List of Figures

# Table of Contents

**Chapter one**

# 1 Introduction

The killer phenotype was first reported in *Saccharomyces cerevisiae* in 1963 by Makower and Bevan (cited in Wickner et al. 2002). Since then, the yeast-killer phenomenon has been discovered in numerous yeast genera and species (Liu et al., 2015a). Yeast killer strains release a toxin which is lethal to the strains of the same or related yeast species that do not have the same toxin gene as the killer strains (sensitive strains). The killer toxin is not lethal to the killer stains itself (Woods and Bevan, 1968). The killer toxin in *S. cerevisiae* is encoded by either viral double strand RNA (dsRNA) or nuclear genes (Dignard et al., 1991, Schmitt and Tipper, 1990, Wickner, 1983, Rodríguez-Cousiño et al., 2011, Goto et al., 1990a, Goto et al., 1990b). The four well-known killer toxins in *S. cerevisiae,* K1, K2, K28, and K-lus, are encoded by four types of the medium-size dsRNA (M dsRNA): M1, M2, M28, and M-lus. The M dsRNA is the satellite of a large-size dsRNA mycovirus (L-A dsRNA) and is completely dependent on the enzymes of L-A dsRNA for replication and encapsidation. The killer toxins encoded by the genome, KHS and KHR, are less known. Their killer activity is weaker than the viral toxin (Goto et al., 1991, Goto et al., 1990b). One of the yeast species with killer activity, which seems to be related to dsRNA, is *Saccharomyces paradoxus* (Naumov et al., 2005, Pieczynska et al., 2013b).

## 1.1 Mycovirus

Fungal viruses, which are believed to be of ancient origin, are ubiquitous in the fungi kingdom and are called mycoviruses (Bruenn, 1993, Ghabrial, 1998). They usually exist in the cytoplasm and occasionally have been found in mitochondria (Polashock and Hillman, 1994, Varga et al., 2003). The genomes of most of the mycoviruses are dsRNA or single strand RNA (ssRNA); an exceptional single strand circular DNA mycovirus, *Sclerotinia sclerotiorum*, hypovirulence associated DNA virus 1 (SsHADV-1), has been reported (Yu et al., 2010). They do not have an extracellular phase in their life cycle and are transmitted intracellularly during cell division, sporogenesis and cell fusion (Ghabrial, 1998). The range of phenotypic changes induced by these viruses seems to be varied. It can be from severe to symptomless infection (Ghabrial and Suzuki, 2009).

The gene that is common to all mycoviruses is RNA-dependent RNA polymerase (RDRP). They have several conserved motifs within the gene. Based on RDRP sequence comparisons, the mycoviruses have the same origin (Bruenn, 1991, Bruenn, 1993, Ghabrial, 1998). The presence of ancestral mycoviruses might trace back to a time prior to the separation of protozoa and fungi. Therefore, mycoviruses are widespread in the fungal kingdom (Bruenn, 1993, Pearson et al., 2009). Ten microbial families are listed by ICTV on virus taxonomy (Fauquet et al., 2005). Table 1-1 explains some of the characteristics of five important families of dsRNA viruses. The viral dsRNA widespread in the yeast genera are categorised in the *Totiviridae* family (Bruenn, 1991).

**Table 1-1**. Characterisation of dsRNA mycovirus families (Fauquet et al., 2005)

| **dsRNA family** | **Size of the genome** | **Number of genomic segment** | **Morphology of virus particles** |
|---|---|---|---|
| ***Totiviridae*** | 4.6 - 7.0 kb | 1 packed | 30 - 40 nm diameter icosahedral, icosahedral capsid protein |
| ***Chrysoviridae*** | 2.4 - 3.6 kb | 4 packed separately | 30 - 40 nm diameter icosahedral, icosahedral capsid protein, multiple component |
| ***Reoviridae*** | 0.7 - 5.0 kb | 9, 10, 11, or 12 packed | 70 - 90 nm diameter icosahedral, one, two, or three layered capsid protein |
| ***Hypoviridae*** | 9.0 - 13.0 kb | 1 unpacked | 50 - 80 nm diameter, pleomorphic vesicle |
| ***Partitiviridae*** | 1.3 - 2.3 kb | 2 packed separately | 30 - 40 nm diameter icosahedral, icosahedral capsid protein |

## 1.2 Totiviridea family

Viruses belonging to the Totiviridea family have non-segmented dsRNA genomes that are between 4.6 and 7 kb in size. They are encapsidated in isometric particles that are 40 nm in diameter. The genome organisation and expression strategy of the viruses are similar. The genome encompasses two large open reading frames (ORF); 5’ proximal encodes the coat protein (CP) (Gag) and 3’ proximal encodes an RDRP (Pol). Except for one of the viruses in this family, LRV2-1, the ORF of the *gag* and *pol* has overlap. They express either Gag protein by finishing translation at the end of *gag* gene or a fusion protein, Gag-Pol, with -1 or +1 translational frameshifting (Fauquet et al., 2005, Ghabrial, 1998, Icho and Wickner, 1989b). The predicted amino acid sequences of the RDRP have eight conserved motifs and relatively significant sequence similarity (Bruenn, 1993). There are three genuses in this family: *Totivirus, Giardavirus* and *Leishmaniavirus*. The dsRNAs found in *S. cerevisiae* belong to the *Totivirus* genus (Bruenn, 1991, Fauquet et al., 2005). The species of the *Totivirus* genus is listed in Table 1-2.

**Table 1-2**. The characterisation of the species in *Totivirus* genus (Dinman et al., 1991, Kang et al., 2001, Kondo et al., 2016, Fauquet et al., 2005, Li et al., 2011).

| **Species** | **Length (bp)** | **Genes** | **Expression** |
|---|---|---|---|
| *Helminthosporium victoriae virus 190S* | 5200 | Two large overlapping ORFs; Capsid (Gag) and RNA dependent RNA polymerase (Pol) | Express Gag by stopping at stop codon of *gag* gene and Gag-Pol fusion by -1 frameshifting |
| *Saccharomyces cerevisiae virus L-A* | 4580 | | |
| *Saccharomyces cerevisiae virus L-BC* | 4615 | | |
| *Ustilago maydis virus H1* | 6000 | Three ORFs; a capsid, a putative protease and an RNA-dependent RNA polymerase | The whole genome is expressed as a single polyprotein, then it will be cleaved into these three proteins |

## 1.3 DsRNA viruses and killer systems in *S. cerevisiae*

As previously mentioned, in *S. cerevisiae,* the killer phenotype is encoded by ether viral dsRNA or nuclear genes. The expression of killer phenotype in the *S. cerevisiae* strains infected with the virus requires the presence of two different dsRNAs: the large-size dsRNA, L, and the toxin-coding medium-size dsRNA, M (Table 1-3). In vivo, both dsRNAs are separately encapsidated in virus-like particles (VLP) (Hopper et al., 1977a, Rodríguez-Cousiño et al., 2011). M dsRNA is a satellite dsRNA of the L, meaning that L encodes the coat protein (Gag) which encapsidates M and L, as well as the RNA-dependent RNA polymerase (Pol), to replicate them both (Figure 1-1) (Hopper et al., 1977a, Icho and Wickner, 1989a). Other killers produce toxins that are encoded by nuclear genes, called KHS and KHR (Goto et al., 1991, Goto et al., 1990a, Goto et al., 1990c). In addition to the L and M dsRNAs, two other groups – W and T dsRNAs – have also been reported which do not have any effect on the killer phenotype (Wesolowski and Wickner, 1984).

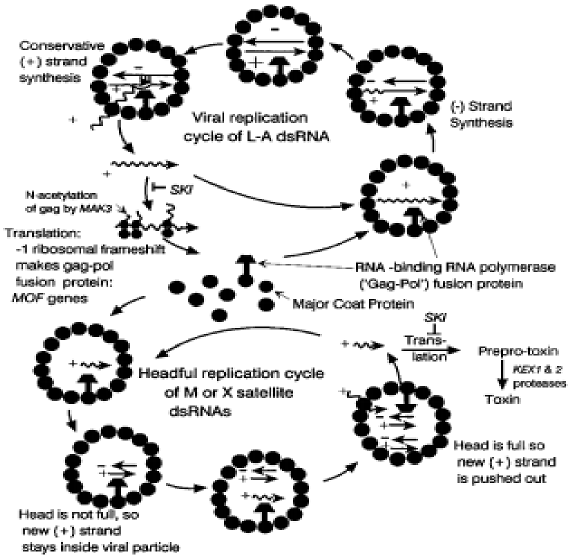


**Figure 1-1**. Replication cycle of L-A and its satellites (Wickner, 1996).

### 1.3.1 L dsRNA

L dsRNA, which is a *Totivirus*, is composed of two dsRNAs: L-A and L-BC (Bruenn, 1991). L-A is an unsegmented genome, 4.6 kb, packed inside 39 nm icosahedral capsids. It encodes two virion proteins: a 76 kD major structural protein, Gag, and a 180 kD Gag-Pol fusion protein, which is composed of an N-terminal Gag domain and a C-terminal Pol domain (Esteban and Wickner, 1986a, Icho and Wickner, 1989a). Each capsid is formed from 60 asymmetric Gag dimers and one or two Gag-Pol molecules (Cheng et al., 1994, Ribas and Wickner, 1998). Similar to other viruses, the genome has two overlapped ORFs; the 5′ ORF encodes Gag while the 3′ ORF has an amino acid sequence of Pol. Pol is only expressed as a fusion protein with Gag using the -1 ribosomal frameshifting regions that exist in the overlapped area of Gag and Pol ORFs (Figure 1-2).

In the (+) strand of L-A dsRNA, a *cis* signal has been demonstrated to start replication at the 3′ UTR (Esteban et al., 1989). There is no experimental evidence for the presence of a signal at the 5′ end of L-A-L1. However, in X dsRNA, which is a deletion mutant of L-A-L1, the 25 nt at the 5′ is identical to the 5′ of L-A-L1. This dsRNA is maintained stable by L-A-L1. It appears that the *cis* signal for transcription exists in this 25 nt. Similar to other dsRNA viruses, this region is AU-rich and seemes to facilitate the melting of the dsRNA for conservative transcription (Rodríguez-Cousiño et al., 2013, Esteban and Wickner, 1988). There is a packing site near the end of the dsRNA, nucleotides 4,180 - 4,203, which contains the encapsidation signal (Figure 1-2) (Wickner et al., 1995).

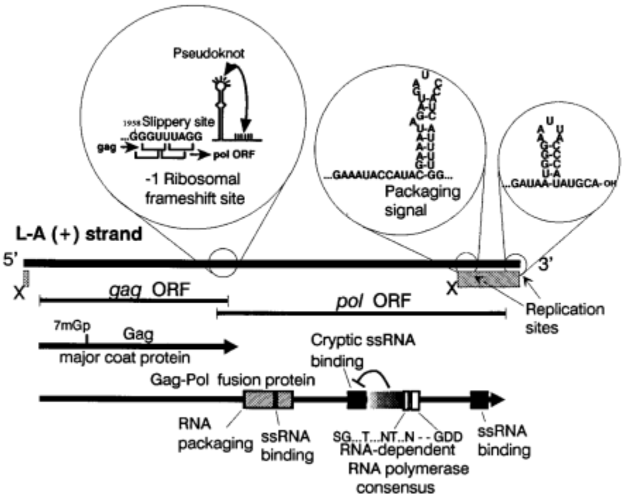


**Figure 1-2.** L-A dsRNA genome structure (Wickner, 1996). The genome contains two overlapped ORFs with a frameshift site in the middle. The Gag-Pol fusion protein is expressed using this site. There are two replication sites at the 3' end.

There are four types of L-A in *S. cerevisiae*: L-A-L1, L-A-2, L-A-lus, and L28. The sequence of all L-A dsRNAs is available, except L28. The sequence identity between the three known L-A dsRNAs is between 73% and 75%. Three subtypes of L-A-lus have been found in nature. The conservation between them with respect to nucleotide level is between 83% and 85%, which is higher than the conservation between different types of L-A. The encoded proteins of these subtypes are almost identical at 97% to 98%. In nature, L-A-lus and L-A-2 is more widely geographically distributed than L-A-L1 and they are shown to be more stable inside the cell in difficult situations, e.g. when exposed to high temperatures (Rodríguez-Cousiño et al., 2013).

L-A may coexist in the same cell with L-BC. L-BC dsRNA has around 25% identity with L-A dsRNA and is encapsidated in particles whose major protein is different from that of L-A and M VLPs. The copy number of this particle is substantially lower than that of L-A (Sommer and Wickner, 1982).

### 1.3.2 Satellite M dsRNAs

The presence of a satellite M dsRNA in cells co-infected with the L dsRNA virus is responsible for the killer-immune phenotype observed in the *S. cerevisiae* killer strains. The four killer toxins, K1, K2, K28, and K-lus, are encoded by four different M dsRNAs, M1, M2, M28, and M-lus, differing in sizes (1.5 - 2.5 kb) and showing similar organisation, despite not having any significant sequence similarity (Dignard et al., 1991, Schmitt and Tipper, 1990, Wickner, 1983). However, the genome structure is similar in all M dsRNAs. The genome starts with a 5' terminal coding region, followed by a poly A sequence. After the poly A sequence, there is a 3' non-coding region (Figure 1-3). M dsRNAs have the same *cis* signal as L-As at the 3'-terminal region, which is essential for packaging and replication. Similar to L-A dsRNA, it seems that transcription initiation exists in the 25 nt at the 5'-terminal sequences (Rodríguez-Cousiño et al., 2011).

Although the preprotoxins have no sequence relationship to each other or to other known proteins, they share conserved patterns of potential processing sites. At the amino termini they contain a stretch of hydrophobic amino acid which follows one or more signal cleavage sites. The signal cleavage sites are used for entering into the endoplasmic reticulum (ER). Moreover, they have potential Kex2p/Kex1p processing sites, which divide the preprotoxin into three subunits: α, β, and γ and also three potential sites for N-glycosylation. Once the preprotoxins are synthesized, they undergo post-translational modifications via the ER, Golgi apparatus, and secretory vesicles. This results in the secretion of mature active toxins that are composed of α and β subunits with a disulfide bond (Magliani et al., 1997b, Rodríguez-Cousiño et al., 2011, SCHMITT and TIPPER, 1995).

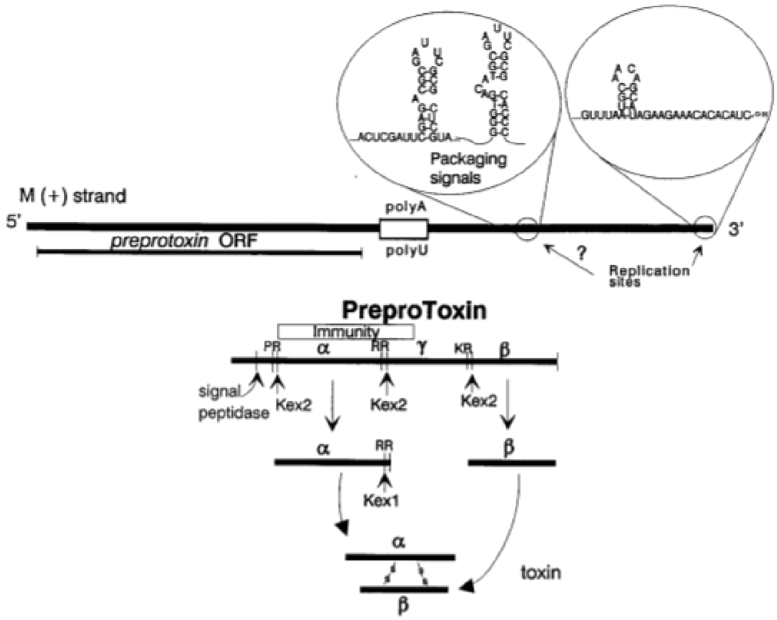


**Figure 1-3.** The structure of M1 dsRNA and K1 preprotoxin, and processing of K1 preprotoxin (Wickner, 1996). The M1 dsRNA sequences start with the killer toxin ORF followed poly A and long non-coding sequences. K1 preprotoxins, using their signal cleavage site, enter to ER and Golgi for post-translational modifications. They break into four subunits using three Kex2/Kex1 cleavage sites. The α and β subunits will be attached to make the active killer toxin.

Most of the infected cells with M dsRNAs show killer phenotypes (Killing the other yeast strains by secreting killer toxins)($K^+$) and are resistant to their own toxin ($R^+$). However, some infected strains with $K^+R^-$ or $K^-R^+$ phenotypes have also been reported. These phenotypes arise from the mutations in chromosomal or non-chromosomal genes (Wickner, 1974). The presence of more than one M in a single yeast cell does not occur, as M genomes are mutually exclusive at the replicative level (Schmitt and Tipper, 1992).

### 1.3.3 Other dsRNAs

In addition to the L and M dsRNAs, two other groups – W and T dsRNAs – have also been reported as having no effect on killer phenotypes (Wesolowski and Wickner, 1984). Both are cytoplasmically inherited and have their own RNA-dependent RNA polymerase. Based on Northern blot hybridisation, they show no homology with each other or other dsRNAs. On average, the proteins of these two dsRNAs share 22.5% of identical amino acids, but there are several regions with highly conserved sequences in a region that includes the consensus sequences for viral RNA-dependent RNA polymerase, suggesting that T and W are evolutionarily related (Esteban et al., 1992a).

### 1.3.4 Nuclear killer genes

Two types of killer yeasts were discovered by Kitano et al. whose killer phenotypes were weaker than K1 and K2, and were not cured by cycloheximide or high temperatures (KITANO et al., 1984). They were classified into two groups, which differed in their optimum pH, thermostability, designated KHS (killer of heat susceptible), and KHR (killer of heat resistance). Later, their genes were found in the right arm of chromosome V and the left arm of chromosome IX respectively (Goto et al., 1990a, Goto et al., 1991).

The nucleotide sequence of the KHS gene, which is near telomeres, was cloned and expressed by Goto et al. The ORF of the gene consists of 2,124 nucleotides and its protein was estimated to be 79 kD. However, the molecular size of the killer toxin was about 75 kD, which suggested some protein processing occurred during maturation of the killer toxin. The expressed protein was monomeric (Goto et al., 1991).

Frank and Wolf in 2009 studied the sequence of KHS in other *S. cerevisiae* genome sequences. They believe that there is an inversion in the sequence that ends at a restriction site which was used in the cloning. As a result, they considered the SCY_1690 ORF in the genome of *S. cerevisiae* YJM789 as the KHS encoding gene. SCY_1690, which is 1,057 bp, is intact and codes a 350-amino acid protein in *S. cerevisiae* strains YJM789 (protein name SCY_1690), M22, and most of the strains sequenced by the *Saccharomyces* Genome Resequencing Project (the project is about getting the picture of evolution by analysing the sequences of *S. cerevisiae and S. paradoxus* strains; https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html) (Doniger et al., 2008, Liti et al., 2009, Ruderfer et al., 2006, Wei et al., 2007). Its coding region from strains YJM789 and M22 has greater similarity to the *S. paradoxus* genome sequence, 99% identity, than the null alleles in *S. cerevisiae,* 89% identity, which suggests horizontal exchange of this gene between two species. They also introduce SCY_1690 as a NUPAV (nuclear sequences of plasmid and viral origin) related to

M2 killer virus double-stranded RNA because of the similarity they found between their sequences (Frank and Wolfe, 2009).

KHR gene is 888 bp and the toxin is a simple monomeric protein that goes through the processing steps of maturation of the killer toxin. The KHR toxin has special characters not previously reported such as: wide pH stability, high temperature tolerance, wide killer spectrum, and toxicity to its toxin-producing cells (Goto et al., 1990a, Goto et al., 1990b). In the transformed cells, because of increasing the copy number of the gene in the cells, the killer phenotype and sensitivity of the KHR producing strains to its own toxin was clearly observed. However, in nature, since the killer strains have one or two copies of the gene, their killer activity is very weak (Goto et al., 1990b).

**Table 1-3.** The genetic basis of the killer toxins in *S. cerevisiae* (Goto et al., 1991, Goto et al., 1990a, Rodríguez-Cousiño et al., 2011, SCHMITT and TIPPER, 1995, Magliani et al., 1997b)

| Genetic basis | | | | Toxin | Replication |
|---|---|---|---|---|---|
| dsRNA<br>(Linear) | L dsRNA<br>(*Totivirus*) | L-A | L-A-L1 | - | Use their own RNA-dependent RNA polymerase (RDRP) |
| | | | L-A-2 | - | |
| | | | L-A-lus | - | |
| | | | L28 | - | |
| | | | | | |
| | M dsRNA | M1 | | K1 | Using L-A-L1 RDRP |
| | | M2 | | K2 | Using L-A-2 RDRP |
| | | M28 | | K28 | Using L28 RDRP |
| | | M-lus | | K-lus | Using L-A-lus RDRP |
| Nuclear genes | *khs* | | | KHS | Nuclear polymerase |
| | *khr* | | | KHS | |

## 1.4 Killer systems in other yeasts

Killer systems in various yeast genera are controlled by dsRNA, linear DNA plasmids, or nuclear genes (Table 1-4). Viral dsRNAs are responsible for the killer phenotype in *Hanseniaspora uvarum* and *Zygosaccharomyces bailii*. L and M dsRNAs similar to *S. cerevisiae* were detected in both yeasts. However, an additional Z dsRNA (2.8 kb) was only present in the wild-type *Z. bailii* killer strain (Schmitt and Neuhausen, 1994). Furthermore, four dsRNAs associated with virus-like particles in *Phaffia rhodozyma* are said to encode a killer system (Castillo and Cifuentes, 1994).

The killer phenomenon in *Kluyveromyces lactis* and two species of *Pichia, Pichia inositovora* and *Pichia acacia*, are controlled by DNA linear plasmids. In *Kluyveromyces lactis*, killer strains always contain one of the two cytoplasmically inherited linear plasmids designated pGKL1 (K1) and pGKL2 (K2). There are similar plasmids in both the functional and structural organisation of *P. acaciae,* designated p*Pac*1-1 and p*Pac*1-2, which produce killer toxins. The presence of three linear dsDNA plasmids has been reported in *P. inositovora*. Only two of them (pPin1-1 and pPin 1-2) seem to be associated with the killer phenotype (Magliani et al., 1997a).

Other killer phenotypes associated with chromosomal genes have been found in two other species of *Pichia (Pichia kluyveri* and *Pichia farinose*), *Williopsis marakii* and some of the pathogenic yeasts. The killer toxins in *Pichia kluyveri* and *Pichia farinos* functionally resemble the *S. cerevisiae* K1 toxin. However, the cell-wall receptor in both, as well as in *W. marakii*'s toxin, is different. In almost all genera of pathogenic yeasts, such as species of *Candida* and *Torlopsis*, the killer systems are controlled by nuclear genes, even though it is unlikely that killer toxins are significant as virulence factors (Magliani et al., 1997a).

**Table 1-4.** The genetic basis of killer phenotype in yeast species

| Species | Genetic base | Killer toxin |
|---|---|---|
| ***Ustilago maydis*** | dsRNA | KP1, KP4, KP6 |
| ***Hanseniaspora uvarium*** | dsRNA | Similar to *S. cerevisiae* |
| ***Zygosaccharomyces baili*** | dsRNA | Similar to *S. cerevisiae* |
| ***Kluyveromyces lactis*** | DNA linear plasmid | K1 & K2 |
| ***Pichia inositovora*** | DNA linear plasmid | Killer toxin |
| ***Pichia acacia*** | DNA linear plasmid | K1 & K2 (Similar to *K. lactis)* |
| ***Pichia kluyveri*** | Nuclear gene | Similar to K1 in *S. cerevisiae* |
| ***Pichia farinose*** | Nuclear gene | Similar to K1 in *S. cerevisiae* |
| ***Williopsis marakii*** | Nuclear gene | HMK & K-500 |
| ***Candida*** | Nuclear gene | I & II |
| ***Cryptococcous humicola*** | Nuclear gene | Mycocin & microcin |
| ***Torulopsis*** | Nuclear gene | Killer toxin |

Of six different species of *Saccharomyces* genus, dsRNAs have been reported in five of them: *S. cerevisiae, S. paradoxus, S. kudriavzevii, S. mikatae,* and *S. bayanus* (Ivannikova et al., 2007, Naumov et al., 2005, Naumov et al., 2009). Table 1-5 highlights the reported dsRNAs and killer activity in these species. Despite the presence of M dsRNA in *S. kudriavzevii*, *S. mikatae*, and *S. bayanus*, none of them show killer activity (Ivannikova et al., 2007, Naumov et al., 2009).

**Table 1-5.** Reported dsRNA and killer activity in *Saccharomyces* species (Ivannikova et al., 2007, Naumov et al., 2005, Naumov et al., 2009)

| Species | dsRNA | Killer activity |
|---|---|---|
| ***S. cerevisiae*** | L dsRNA, M dsRNA, W and T dsRNA | Killer |
| ***S. paradoxus*** | L dsRNA, M dsRNA | Killer |
| ***S. mikatae*** | L dsRNA, M dsRNA | Non-killer |
| ***S. bayanus*** | L dsRNA, M dsRNA | Non-killer |
| ***S. kudriavzevii*** | L dsRNA, M dsRNA | Non-killer |

The killer phenotype is very rare in the population of *S. paradoxus* (Chang et al., 2015, Pieczynska et al., 2013b). Two types of M dsRNA, M1 and M28, have been found in this species. These dsRNA sequences are very close to M1 and M28 in *S. cerevisiae.* M1 has just one nucleotide difference, in the α subunit of preprotoxin, from M1 of *S. cerevisiae.* However, one of its carrier strains has the potential to kill the K1 strains of *S. cerevisiae*. Variation in the sequence of M28 between the two species is higher than that of M1; they are 90% identical. However, the killer phenotype of most of the *S. paradoxus* strains infected with M28 is similar to that of *S. cerevisiae* infected strains. In addition to K1 and K28, there are some unknown killer toxins in the *S. paradoxus* population (Chang et al., 2015).

Some of the killers that carry these dsRNAs showed different killer-immune reactions not only to the *S. cerevisiae* strains but also to the *S. paradoxus* strains that have the same dsRNA. These indicate the effect of the host-genome background on the expression of the phenotype and co-evolution of the viruses and hosts' genomes in different populations (Chang et al., 2015). Regarding the genome phylogeny tree of the *S. paradoxus* strains, the strains (on the tree) which show the killer phenotype seem to be related to each other more than those which show the non-killer phenotype (Pieczynska et al., 2013b). Moreover, Chang et al. suggest killer-virus infections in *S. paradoxus* seem to be a more ancient event than in *S. cerevisiae* because killer toxin resistance in *S. paradoxus* populations is more wildly widespread than *S. cerevisiae* or *S. eubayanus* populations. In addition, genetic analyses indicate that the immunity in *S. paradoxus* often arises from dominant alleles that have independently evolved in different populations (Chang et al., 2015).

## 1.5 Viral killer system, survival, replication, expression and immunity in *S. cerevisiae*

### 1.5.1 Viral dsRNA survival

In order to defend against viruses, most of Eukaryotic cells are equipped with RNA interference (RNAi). RNAi is a conserved biological response that mediates resistance to both foreign nucleic acids such as viruses, and mobile segments of the host genome (transposons and retroelements). (Ghildiyal and Zamore, 2009, Malone and Hannon, 2009). It is an ancient mechanism which exists in plants, animals, and most fungi. However, during the evolution of budding yeasts it has been lost in most of the species including *S. cerevisiae* (Moazed, 2009, Drinnenberg et al., 2009). It seems that the absence of the RNAi in this species has led to the development of the condition in which viral dsRNA can thrive. Reconstituting RNAi in this species causes loss of the dsRNA (Drinnenberg et al., 2011).

### 1.5.2 Replication cycle

Similar to the other viral dsRNAs, the polymerisation of L and M dsRNA takes place in the viral particles. The two strands of L-A dsRNA ((+) and (-)) are synthesised sequentially at distinct stages of the replication cycle (Figure 1-1). The (+) strands are synthesised using the RNA polymerisation activity of Gag-Pol fusion protein in the mature virions and released into the host cytoplasm. In the cytoplasm, they are either translated into Gag proteins (by stopping at stop codon of *gag*) and Gag-Pol fusion proteins (using -1 frameshifting) or encapsidated by the coat proteins. Once the (+) ssRNAs are encapsidated, they are converted into dsRNA in the particles, using the polymerisation activity of the virions (Fujimura and Wickner, 1987).

In terms of M and X, Both are satellites of L-A and they use its replication system. As a result, their replication cycles are essentially the same as that of L-A dsRNA (Figure 1-1). Since the coat proteins are designed to encapsidate L-A and the size of these two dsRNAs are significantly less than L-A dsRNA, the M particles contain two molecules of M dsRNA and the X particles are found to have up to eight X dsRNA molecules (Esteban and Wickner, 1988, Esteban and Wickner, 1986b).

### 1.5.3 Frameshifting

In L-A dsRNA of *S. cerevisiae,* the Gag and Pol ORFs have 130 nucleotide overlap. The frameshifting region located in this area is composed of a seven-nucleotide slippery site, followed by an RNA pseudoknot. At the slippery site, the tRNA in the A and P sites in the ribosome slips back one nucleotide on the mRNA. The RNA pseudoknot promotes frameshifting by blocking the forward proration of the ribosome. The efficiency of ribosomal frameshifting indicates the ratio of Gag-Pol fusion protein to Gag protein in the cell. This ratio is critical to propagation of the M1 satellite dsRNA

(Dinman and Wickner 1992). Increasing or decreasing the efficiency of frameshifting results in reducing the copy number of M1 in the cell. This is probably because of a requirement for a specific ratio of Gag-Pol to Gag for the assembly of viral particles.

### 1.5.4 Killer toxin precursor processing and secretion

The killer toxin maturation is similar in the known killer toxins. The killer toxin transcript is translated into a preprotoxin. The preprotoxin goes through the secretory pathway for post-translational modification and secretion. The N-terminal hydrophobic signal of the preprotoxin is responsible for introducing the protein into the endoplasmic reticulum (ER), the location of protein folding and maturation. In the ER, the γ sequence becomes N-glycosylated and a disulphide bond is generated between the α and β sequences, which are the subunits of the mature killer toxin. Then, the preprotoxin goes to the Golgi using the second signal and the γ sequence is removed by Kex2p/Kex1p. The mature α/β heterodimer is secreted into the environment (Schmitt and Breinig, 2006, Magliani et al., 1997b).

### 1.5.5 The killer toxin reaction

The sensitive cells are killed in two different ways, depending on the concentration of the killer toxin. At high concentrations of toxin, susceptible yeasts are killed by necrosis whereas, at low concentrations, apoptosis is triggered in sensitive yeasts (Sommer and Wickner, 1982).

The initial stage of necrosis, in terms of the interaction with receptors of the cell wall and cytoplasmic membrane, is identical across the various toxins. In the following stage, K2 displays very similar toxin activity to that of K1, despite having a different structure. K28, however, acts differently (Magliani et al., 1997a, Schmitt and Tipper, 1990).

Initially, all secreted mature toxins bind to their specific cell-wall receptors on the susceptible cells. This binding is strongly pH dependent. For K1 and K2, this receptor is β-1,6-D-glucan, whereas α-1,3-mannoprotein is the receptor for K28 (Bussey et al., 1979a, Schmitt and Radler, 1988). They then interact with cytoplasmic membrane receptors. The receptors for K1 and K28 are Kre1p and probably Erd2P respectively (Bussey et al., 1979a, Schmitt and Breinig, 2006). Susceptible strains can become resistant by chromosomal mutation in a set of genes that encode the proteins which are involved in the structure and assembly of these receptors (Boone et al., 1990, Schmitt and Breinig, 2006).

Following interaction with receptors, the K1 transferred into the cytoplasmic membrane acts as an ionophore and causes the membrane to become permeable to proton and potassium ions. As a result, permeability increases and higher molecular mass, such as ATP, leaks out. Two strongly hydrophobic regions near the C terminus of the K1 α-subunit have an α-helical structure separated by a short, highly hydrophilic segment. This may act as a membrane-spanning domain responsible for channel formation (Sturley et al., 1986, Ahmed et al., 1999).

K28 affects the cell cycle in a different way. After endocytotic uptake and retrograde transport through the Golgi and ER, the toxin enters the cytosol. Yeast mutants that are blocked in the early stages of both fluid-phase endocytosis and receptor-mediated endocytosis, or at any stage in retrograde transport, are toxin resistant. In the cytosol, the β-subunit is ubiquitinated and proteasomally degraded, whereas the α-subunit enters the nucleus and interacts with cell-cycle control progression proteins, as well as initiating DNA synthesis proteins in the early S phase. Therefore, since K28 targets essential and evolutionarily conserved host proteins with basic cellular functions, it allows its toxin to develop an ‘intelligent’ strategy to effectively penetrate and kill its target cell (Schmitt and Breinig, 2006).

### 1.5.6 The immunity of *S. cerevisiae* strains to the toxins

The exact mechanism of immunity against K1 is not clearly understood. However, the α and γ component of preprotoxin seem to be required for the expression of immunity (Zhu et al., 1993). They probably act as a competitive inhibitor of mature toxins by saturating or eliminating the plasma-membrane receptors that normally make toxicity (Schmitt and Breinig, 2006). Given that the γ-component is essential in the maturation of the toxins it postulates, it might not only act as an intramolecular chaperone – ensuring proper preprotoxin processing – but might also provide some sort of masking function. This can be achieved by protecting the membrane of toxin-producing cells against damage caused by the hydrophobic α-subunit (Bostian et al., 1984).

Recently, the mechanism of K28 immunity has been elucidated at the molecular level. This toxin enters K28-infected cells by endocytosis and, after passing through Golgy and ER, is transported into the cytosol. Before K28 enters the nucleus to start its killer activity, it forms a complex with a preprotoxin (produced by an internal virus). In this complex, the K28 heterodimer is selectively ubiquitinated and protosomally degraded. In this way, the preprotoxin is released and can either be imported into ER or form a complex with a new internalized K28 heterodimer. Interestingly, the amount of cytosolic ubiquitin plays a crucial role in toxin immunity (Breinig et al., 2006, Schmitt and Breinig, 2006).

In addition, susceptible strains can become resistant to a killer toxin by mutation in the genes which encode receptors or proteins that killer toxins use to perform their killer activity (Schmitt and Breinig, 2006). Only around 30% of the genes involved in the immunity or sensitivity of the cells against K1, K2, and K28 are the same (Servienė et al., 2012).

### 1.5.7 The host cell and virus interaction

The host cell plays a critical role in the maintenance and expression of killer phenotypes. Two groups of host genes have been detected which have an impact on viral dsRNA. The superkiller (*ski*) genes (the phenotype of mutants) are a group of genes that repress the copy number of M, L-A, and L-BC. These genes appear to constitute a host antiviral system that is essential to the cell only for repressing viral replication and propagation (Magliani et al., 1997a).

Maintenance of killer (*mak*) host genes is essential for cell growth and necessary for maintenance and propagation of viral genomes. Of more than 30 chromosomal MAK genes identified, only three are required for maintenance of both L-A and M dsRNAs. All other MAK genes are only necessary for maintaining each of the three known M satellite dsRNAs (Schmitt and Tipper, 1992, Magliani et al., 1997a).

## 1.6 The Ecology of the killer system

Alleopathy is a type of interference competition in which toxin compounds are produced that kill or suppress the growth of competitors like killer systems in yeasts (Starmer et al., 1987). The failure or success of this reaction is dependent on the frequency of toxin producer, environmental structure, the cost associated with toxin production, the effect of the toxin on competitor growth, and the relative importance of interference competition and resource competition (Frank, 1994). The first two factors are critical. Their importance suggests that density is likely to affect the cost and benefit of alleopathy (Levin et al., 1988).

These two factors have been widely researched in yeast-killer systems. The results show that there is a high value associated with producing toxin when the producers are dense, but this advantage is proportional to the level of toxin concentration. Additionally, the growth rate of the killer strains is lower than that of sensitive strains, probably due to toxin production. As a result, in competition this is a disadvantage for the killer strains (Greig and Travisano, 2008b).

## 1.7 Potential applications of the killer system

Killer yeasts and their toxins may have a range of applications. They can be used in the taxonomy of yeasts. Given the fact that yeasts make up a highly heterogeneous group of unicellular organisms, it is important to be able to dissect yeast toxins through a simple, cheap, and easy-to-perform test. Fermentation is widely used in industry. Since the killer activity plays a critical role in the optimisation of the fermentation reaction, it could well be beneficial in the improvement of the quality of industrial products. In addition, recent interest in the development of killer toxin as a food preservative has increased. A further application is in medicine where it has potential as an antimicrobial agent. It is also important in pathogenic yeast diseases. Furthermore, it can be used in transgenic plants to improve their resistance. Finally, it can be used as a model to study the interaction between host and virus as well as in modelling post-translational activity (Schmitt and Breinig, 2002, Marquina et al., 2002).

## 1.8 Research aim

*S. cerevisiae* has a number of characteristics that makes it an ideal model for the study of molecular biology as well as population and evolutionary genetics (Zeyl, 2000). The long-term association with artificial, man-made environments is, however, a disadvantage of these species (Vaughan-Martini and Martini, 1995). *S. paradoxus*, on the other hand, which is a close relative of *S. cerevisiae* and has never been domesticated (Goddard and Burt, 1999), contains a variety of viral dsRNAs and also shows the killer phenomenon. Therefore, it seems to be an appropriate model to study the nature, ecology and evolution of yeast-killer systems. To date, there has been very little research on the killer systems in *S. paradoxus*. My research, therefore, aims to:

a) Find and characterise the different groups of killer systems in *S. paradoxus*;
b) Determine and categorise the various viral dsRNAs which are related to the killer phenotypes in this species;
c) Find the relationship between the killer systems in *S. paradoxus* and *S. cerevisiae.*

**Chapter two**

# 2 Detection and characterisation of the killer-immune phenotype

## 2.1 Introduction

The killer toxin in *S. cerevisiae* is encoded by either viral dsRNA, M1, M2, M28, and M-lus, or nuclear genes, KHS and KHR. In all killer types, K1, K2, K28, and K-lus, of which the genetic basis is viral dsRNA, a single open reading frame encodes the toxin as a single polypeptide, preprotoxin. The preprotoxin comprises larger hydrophobic amino termini than are usually found on secreted proteins, potential Kex2/Kex1 cleavage and N-linked glycosylation sites. They all have similar overall structures. They are composed of δ, α, β and γ sequences. Once synthesized, they undergo post-translational modifications via the endoplasmic reticulum (ER), Golgi apparatus, and secretory vesicles. This results in the secretion of the mature active toxins that are composed of α and β subunits with disulfide bonds. Mature secreted proteins are active in low pH (Magliani et al., 1997a).

Most of the infected cells show both killer phenotype ($K^+$) and resistance to their own toxin ($R^+$). However, some infected strains with $K^+R^-$ or $K^-R^+$ phenotypes have also been reported. These phenotypes arise from the mutations in chromosomal or non-chromosomal genes (Wickner, 1974).

The killer toxins encoded by the genome, KHS and KHR, are less known. They are monomeric and their killer activity is weaker than the viral toxin (Goto et al., 1991, Goto et al., 1990b). Apart from the genetic basis of the killer toxins or the structure, their killer activity mostly depends on environmental conditions like pH, temperature, concentration of salt, and so on (Liu et al., 2015a).

Different types of M viruses use different pathways for killing sensitive cells. For instance, K1 killer toxin binds to the cell membrane receptors of susceptible cells and triggers potassium homoeostasis perturbation, which results in cell death (Sturley et al., 1986, Ahmed et al., 1999), whereas K28 toxin enters the cell nucleus using endocytosis and blocks DNA synthesis in the early S phase (Schmitt and Breinig, 2006, SCHMITT and TIPPER, 1995). Yeast mutants that are are blocked in binding membrane receptors, in the early stage of both fluid-phase endocytosis and receptor-mediated endocytosis, or at any stage in retrograde transport, are toxin resistant (Schmitt and Breinig, 2006). In addition to the immunity that arises from mutations, each type of killer is immune to its own toxin. The γ subunit in K1 and the preprotoxin in K28 play an active role in this immunity.

In the majority of infected yeasts, the killer toxins are active at acidic pH values, ranging between pH 3.5 and pH 5, and at low temperatures (Bussey et al., 1979b, Liu et al., 2015b, McBride et al., 2013, PALFREE and BUSSEY, 1979, Tipper and Bostian, 1984). Within the pH and temperature range, toxin production can benefit toxin-producing yeast. However, the advantage of having the killer toxin is lost at pH and temperatures out of this range (McBride et al., 2013, Greig and Travisano, 2008a, McBride et al., 2008). The size range of killer toxins in yeast is from 8 kDa to 156.5 kDa and made of one to three subunits (Liu et al., 2015a).

Killer-immune reactions between *S. paradoxus* and *S. cerevisiae* strains were studied by Chang et al. (2015). Against *S. cerevisiae* strains, nearly all the *S. paradoxus* strains in this study were killer toxin resistant. As a result, they believe that the killer system first evolved in *S. paradoxus* then transferred to *S. cerevisiae* (Chang et al., 2015). The killer phenotype is very rare in populations of both species (Chang et al., 2015, Pieczynska et al., 2013b). Two types of M dsRNA, M1 and M28, have been found in *S. paradoxus.* Some of the killer strains that carry these dsRNAs not only showed different killer-immune reactions to the *S. cerevisiae* strains but also to the *S. paradoxus* strains that have the same dsRNA. For instance, even though Q74.4 and N-45 both carry an M1 virus with one different amino acid from *S. cerevisiae's* M1, some of the *S. cerevisiae* M1 killer strains were sensitive to these strains. Or, although T21.4 and DBVPG4650 both have M28 virus, only T21.4 killed DBVPG6040. These indicate the effect of the host-genome background on the expression of the phenotype and co-evolution of the viruses and host genomes in different populations. In addition to K1 and K28, there are some unknown killer toxins in the *S. paradoxus* population (Chang et al., 2015).

The killer-immune reaction between strains is usually studied using Methylene Blue (MB) diffusion assay. In this method, to check for killer or non-killer states, the yeast strains (killer testers) are put as a spot on an MB agar plate, which has been seeded with a strain that is being tested for immunity (immunity tester). Because most of the killer toxins are active in acidic pH, the plates are normally prepared with low pH. The killer strains, by secreting killer toxins in the plate, either do not allow the sensitive strains to grow around them, or, after growing, kill them. As a consequence, the Methylene Blue can pass through the cell membrane and the grown sensitive strains become blue (Lopes and Sangorrín, 2010).

In this chapter, using the MB agar diffusion assay, 109 *S. paradoxus* strains were screened to detect the killer strains. Then the killer-immune reactions between the *S. paradoxus* and *S. cerevisiae* strains were studied. Since the strain Q62.5 showed the strongest killer phenotype among all strains of both species, further studies were performed on the killer toxin of this strain.

## 2.2 Methodology

### 2.2.1 Strains and media

*S. cerevisiae* strain M894 (lys11 clv3 len2) (sensitive strain), which is sensitive to all killer strains in *S. cerevisiae*, was used for testing the killer phenotype of the other strains. The following *S. cerevisiae* strains were used as killer-positive strains: K1617 (K1), K1963 (K1), K2618 (K2), and MS300 (K28). The killer activities of 109 *S. paradoxus* strains from the Burt group's collection were tested. The list of strains and media used is in the Appendix (Table 8-1).

### 2.2.2 Concentrating the medium containing the killer toxin

Concentrating the medium was done in two different ways. In the first method, after growing the yeasts, the cells were removed using centrifugation and filtration. Then, the medium was concentrated using an Amicon ultra centrifugal filter (Millipore), Vivacell 250 (Sartorious) or Vivaflow 200 (Sartorious), depending on the amount of the sample. Four different cut-offs were tried: 10K, 30K, 50K, and 100K, and the concentration performed at 4°C or 25°C. To remove the salts in the medium, the concentrate was dialysed with the filter using 1 or 10 mM citrate-phosphate buffer or the buffers used in different tests.

The second method, following Santos and Marquina (Santos and Marquina, 2004), was carried out using both concentrating devices and precipitating with ethanol. After centrifugation and filtration of the medium in which the yeast was grown, it was adjusted to a final glycerol concentration of 15% (v/v) and concentrated 40x. Cold ethanol was added to a final concentration of 45%. Following 30 minutes incubation at 0-4˚C and removal of the precipitate, the proteins were precipitated using additional cold ethanol up to a final concentration of 50%, 70%, 75%, or 80% (Ciani and Fatichenti, 2001, Santos et al., 2004, Ha et al., 1997). The precipitated proteins were dissolved in 1 or 10 mM sodium citrate-phosphate buffer (pH 4.5).

### 2.2.3 Killer assay

**Methylene Blue (MB) agar diffusion assay**

This killer assay was performed at different pH levels (3.5, 4.2, 4.5, 4.7, and 5.3). To check for killer or non-killer states, 5 µl of each yeast strain was grown in YPD (Yeast Extract Peptone Dextrose) liquid medium (Appendix; Table 8-2) and was put as a spot on an MB agar plate (Appendix; Table 8-2) (killer tester). In testing for immunity to this sample, the same or a different strain was sprayed on the plate (immunity tester). The immunity tester was sprayed on same day as the killer tester was put on the plate, Same-Day-Test (SDT), or two days after the development of the killer tester, Different-Day-Test (DDT). After spraying, the plate was kept at 22˚C for 48 hours. If the killer tester produces a toxin to which the immune tester is not immune, the immune-tester cells cannot grow around the killer tester or their numbers decrease. In some strains, they grow but their colour changes to blue, which results from the entrance of the Methylene Blue from cell membranes into the dead immune-tester cells.

**MB agar toxin assay**

In this assay, after growing the yeast strains in the YPD liquid medium, either 100 to 200 µl of the YPD medium was filtered and put in a hollow made in the MB agar plate, or 5 µl to 10 µl of the YPD liquid medium, which was concentrated, was spotted on the MB agar plate. After the liquid medium was absorbed by the MB agar, the tester strain was sprayed on the plate. The result was checked after two days growing at 22.5°C.

**Microtiter plate killer assay**

The concentrated media were added to the wells of a microtiter plate containing the toxin-sensitive strain (1 x $10^4$ or 1 x $10^5$ cells $ml^{-1}$) in YPD pH 4.5 and, following that, optical density at 600 nm was measured after incubation at 22.5˚C for one day.

### 2.2.4 Measuring the growth rate of strains

The growth rates of four yeast strains, M894, A33, Q62.5, and Q14.4, were measured using a microtiter plate reader in the absence and presence of the killer toxin of the strongest *S. paradoxus* killer strain, Q62.5. The growth rates were measured in the presence of 0, 0.5, 1, and 5 folds of the concentrated medium contained the Q62.5 killer toxin in 62 hours. To determine the effects of the number of cells at the start of the test, two different concentrations of the strains were used, $10^4$ and $10^5$ cells.

### 2.2.5 Ethanol treatment

To test the effect of ethanol on the killer toxins, the killer strains were cultured on YPG (Yeast Extract Peptone Glycine) or YPD liquid medium (Appendix; Table 8-2). Following filtration and concentration, the medium was put on the MB agar plates containing 0%, 6%, 12%, and 14% ethanol. After incubation at 22.5°C for 48 hours, it was sprayed with the sensitive strain M894 and kept at 22.5°C. The results were documented on the second and fourth days. The test was repeated by replacing the toxin with the three killer strains, Q62.5, T21.4, and Y8.5.

### 2.2.6 Testing the stability of the toxin at pH 7

The pH of Q62.5's concentrated medium was increased through dialysis with citrate-phosphate buffer pH 7. Half of the medium was kept in a fridge for 16 hours and the other half for 32 hours. Following that, they were dialysed with a citrate-phosphate buffer pH 4.5 and their killer activities were tested using the MB agar toxin assay.

### 2.2.7 Gel electrophoresis

Four types of polyacrylamide gel were used: low pH polyacrylamide gel (Bollag et al., 1996), isoelectric point focusing (IEF) gel (Novex® pH 3–7 IEF Gel (Invitrogen)), Native PAGE and SDS PAGE (NuPAGE® Novex® 4-12% Bis-Tris Midi Protein Gels (Invitrogen)).

In the low pH polyacrylamide gels, the pH of the gel was 4.3 and the pH of the running buffer was 4.5 (Bollag et al., 1996). Because the pH of the gel was low, riboflavin was used instead of ammonium persulfate. Polymerisation by riboflavin is considerably slower than with ammonium persulfate and is triggered by exposure to light, to which it is very sensitive. In trying to achieve the polymerisation of the gel, different concentrations of riboflavin and TEMED, various levels of light, degasification and temperature were tested. The media were loaded on 10%, 12%, and 15% gels in different concentrations and the gels were run at a range of voltages between 40 and 250 V and from two hours to overnight with normal and reverse electrode polarity.

In all the polyacrylamide gels, except SDS PAGE, each sample was run twice on the gels; half of the gels were stained using SimplyBlue (Invitrogen) and the other half were laid on the seeded MB agar for bio assay. The SDS PAGE gel was stained only with SimplyBlue.

In the bioassay, the gel was soaked for 10 minutes at room temperature in citrate-phosphate buffer pH 4.5. It was then laid on an MB plate seeded with the sensitive strain and was kept at 22˚C for 48 hours. The sensitive cells did not grow around any band containing the toxin.

### 2.2.8 Chromatography

**Ion-exchange chromatography**

The concentrated and desalted supernatant was injected into a column (4 x 1 cm) of DEAE. After washing with 20 mM Bis-Tris buffer (pH 6.5), a NaCl gradient from 0 to 1 M was applied and fractions of 500 µl were collected. The activity of fractions was tested using a microtiter plate killer assay.

**Gel filtration**

Superdex 75 (GE Healthcare) and Sephadex G25 (GE Healthcare) were packed, tested, and applied for purification of the killer toxin based on the manufacturer's instructions.

### 2.2.9 Measuring the amount of protein

The amount of the protein was measured using Bradford Protein assay (Bio-Rad).

### 2.2.10 Trichloroacetic acid (TCA) protein purification

200 µl of TCA was added to 200 µl of the concentrated medium. Following 20 minutes incubation on ice, it was centrifuged at 8,000 x g. Then, the precipitated proteins were air dried and dissolved in 1 or 10 mM sodium citrate-phosphate buffer (pH 4.5).

## 2.3 Results and discussion

### 2.3.1 Killer-immune phenotype

The killer-immune phenotype in *S. paradoxus,* similar to other yeasts (Frank, 1994, Greig and Travisano, 2008a, Liu et al., 2015a)*,* is very complex and is sensitive to small changes in the environment. There is a range of killer phenotypes from very poor killer to strong killer in this species (Figure 2-1). The expression level of this phenotype in each strain depends on various environmental factors, such as the amount of killer tester and immune tester, pH and temperature. Even the thickness of the medium has an effect on the phenotype. Like most of the yeast killer toxins, the killer toxins in the tested strains were stable and acted in acidic pH and at low temperatures (Liu et al., 2015a). Within the range of pH tested using the MB agar diffusion method, pH 3.5, 4.2, 4.5, 4.7, and 5.3, none of the samples showed the killer phenotype at pH 5.3. Regarding temperature, all killer samples were inactive at 28˚C and above.

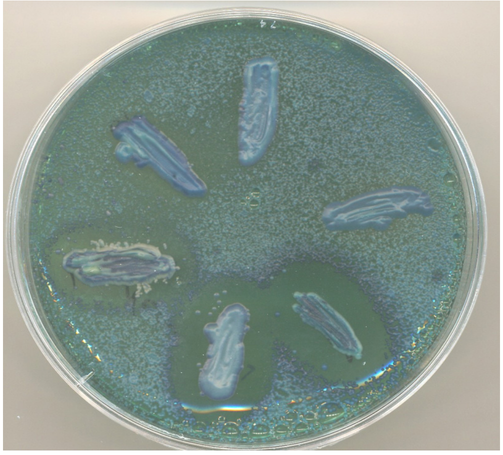


**Figure 2-1.** A range of killer phenotypes in *S. paradoxus* strains. All the strains on the plate are killer strains. Three strains on the bottom, from right to left, T4b, Q62.5, and T21,4, showed the killer phenotype more strongly than the three killer strains on the top, from left to right, Q74.4, Q14.4, and Q43.5. In the strong killers, a clear halo is visible around the yeast strain, whereas in the poor killer strains only one or two layers of the sensitive strain cannot grow around the killer strain. The blue clones around the halo are the dead cells that Methylene Blue passed through their membranes.

Killer activity was present in 39 out of 109 *S. paradoxus* strains (36%) against the *S. cerevisiae* sensitive strain, M894, on MB agar. In all, 27% (17 out of 63) of the strains from Silwood Park, 33% (5 out of 15) from continental Europe, 25% (3 out of 12) from the Far East, 67% (10 out of 15) from North America and 100% (2 out of 2) from South America were killer. The frequency of the phenotype in the studied population was about twice (Chang et al., 2015) and three times (Pieczynska et al., 2013b) more than previous reports.

In this study, the sensitive strain was sprayed on the plates either on the same day as the killer tester was put onto the medium (Same-Day-Test), or two days after growing the killer tester (Different-Day-Test). This strategy provided a new condition for expression of the killer phenotype and helped to find the killer phenotype in four strains which previously reported as non-killer strains, Q59.1, UFR50816, A12 (Chang et al., 2015, Pieczynska et al., 2013b), and DBVPG4650 (Pieczynska et al., 2013b). The killer phenotype in all these strains was not strong. In general, depending on the time of spraying the sensitive strain, most of the tested killer strains showed different levels of expression of the phenotype. The majority of the strains from Silwood Park, Continental Europe, the Far East and South America expressed the phenotype more strongly in Different-Day-Test. Two strains from the UK, Q43.5 and Q59.1, and three strains from Continental Europe, DPVPG4650, C10 (4/2SW2), and C15 (CECT10176) did not show the killer phenotype when the tester was sprayed on the same day on which these killer strains were put on the medium. In contrast, all North American killer strains, including A12, showed their killer phenotype more clearly on Same-Day-Test than Different-Day-Test.

The difference in the phenotype expression may be as a result of various reactions of the sensitive strain against the different concentration of the killer strains' toxins. In the *S. cerevisiae* strains*,* depending on the concentration of the killer toxin, the sensitive cells are killed in two different ways; at high concentrations of the toxin, susceptible yeasts are killed by necrosis whereas, at low concentrations, apoptosis is triggered in sensitive yeasts (Sommer and Wickner, 1982). Perhaps, in *S. paradoxus* strains, there are also different modes of actions in various killer toxins which trigger different killing and immunity pathways in the sensitive strains. As a result, different strains behave differently in Same-Day-Test and Different-Day-Test.

Overall, the killer phenotype in North American strains looks different from the other killer strains. The phenotype in these strains is very weak; there is no hollow around the killer strain. Only one or two layers of the sensitive strain's clones grown around the killer strain become blue, which results from Methylene Blue entering into the dead cells through their membrane. Moreover, the frequency of the killer phenotype in the strains in this region is higher than Silwood Park, Europe and the Far

East. Looking at the location of the strains showed that all the killer strains in North America belong to Canada. Of the strains from this area, 10 out of 12 (83%) are killer. The results suggest that the killer toxins in the strains from this region are different from the others. Despite finding the phenotype in 100% of the South American strains, as there are only two strains from this area, the number is not reliable.

In order to establish whether the killer toxins in *S. paradoxus* and *S. cerevisiae* are similar or different, the killer-immune reaction between the *S. cerevisiae* killer strains, K1617 (K1), K2618 (K2), and MS300 (K28), and 16 strains of *S. paradoxus* were studied (Table 2-1). In this test, the *S. cerevisiae* killer strains were sprayed on the medium as an immune tester and all the other strains were put on the medium as killer testers. As can be seen in the table below, some strains again showed different phenotypes when the immune tester was sprayed on the medium on the same day as the killer tester, or two days after the development of the killer tester. Of all the *S. paradoxus* strains that were tested, only three, Y8.5, Q74.4, and T68.3, showed the same killer-immune pattern as the K2 killer phenotype. They were killer against the K1 and K28, whereas K2 was immune to their toxin. The optimum pH for these three strains was also the same as that for K2 (pH 4.2). Nevertheless, previous studies showed that Q74.4 contains K1 ORF with one amino acid variation from the *S. cerevisiae* K1 ORF, whereas Y8.5 has a different killer toxin from the *S. cerevisiae* reported toxin (Chang et al., 2015). Although Q74.4 has the K1 killer toxin, it also showed a different killer pattern from the K1 killer toxin in *S. cerevisiae* in the previous study. In addition, two strains, T21.4 and DBVPG4650, which were expected to have the K28 killer toxin (Chang et al., 2015), could not express their killer phenotype very well against K1 and K2 killer strains. Overall, it seems that the killer-immune reaction between the two species is different from that previously reported in *S. cerevisiae*. This might be as a result of the different killer toxins that exist in these strains, the different immunity pathways that they use, the mutation in the genes that is necessary for expression, or the different genome background of each species.

**Table 2-1.** The killer-immune phenotype of yeast strains against the *S. cerevisiae* sensitive strain, M894, *S. cerevisiae* killer strains, K1 (K1 617), K2 (K2 618), and K28 (MS 300), and the *S. paradoxus* non-killer strains, CBS 8437, T18.2, A33. SDT and DDT represent Same-Day-Test and Different-Day-Test spraying of the immune tester and K, PK, VPK and (-) represent Killer, Poor Killer, Very poor and non-killer, respectively.

| | | *S. cerevisiae* Immune tester | | | | | | | | *S. paradoxus* Immune tester | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Species** | **Strains** | **Sensitive** | | **K1** | | **K2** | | **K28** | | **CBS 8437** | | **T18.2** | | **A33** | |
| | | **SDT** | **DDT** | **SDT** | **DDT** | **SDT** | **DDT** | **SDT** | **DDT** | **SDT** | **DDT** | **SDT** | **DDT** | **SDT** | **DDT** |
| K1 *S. cerevisiae* | K1 617 | K | K | - | - | K | K | K | K | PK | PK | PK | PK | PK | VPK |
| K2 *S. cerevisiae* | K2 618 | K | K | K | K | - | - | K | K | PK | K/VPK | K | PK | - | - |
| K1 *S. cerevisiae* | K1 963 | K | K | - | - | K | K | K | K | PK | VPK | K/PK | PK | PK/- | - |
| K28 *S. cerevisiae* | MS 300 | K | K | K | K | PK | K/VPK | - | - | PK | K | VPK | VPK | VPK | PK |
| Tester *S. cerevisiae* | M894 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| *S. paradoxus* | DPVPG4650 | - | K | K | - | PK | - | - | - | PK | PK | - | - | VPK | - |
| *S. paradoxus* | Q43.5 | - | VPK | VPK | - | - | - | - | - | PK | PK | K/PK | K/VPK | - | - |
| *S. paradoxus* | Q59.1 | - | VPK | - | - | - | - | - | - | - | - | - | - | VPK/- | - |
| *S. paradoxus* | Q62.5 | K | K | PK | PK | K | K | K | K | PK | VPK | PK | PK | PVPK/- | - |
| *S. paradoxus* | Q14.4 | - | PK | - | - | - | - | - | - | PK | VPK | - | - | PK | PK/- |
| *S. paradoxus* | Q74.4 | PK | PK | K | K | - | - | PK | PK | K | PK | K | K | PK | PK/- |
| *S. paradoxus* | T4b | PK | K | PK/- | - | K/- | PK | K/- | K | - | - | - | - | - | - |
| *S. paradoxus* | T21.4 | PK | K | - | - | - | - | PK | - | VPK | - | K | K/- | PK | PK/- |
| *S. paradoxus* | T68.3 | PK | PK | K | K | - | - | PK | PK | K | PK | K | PK | PK | PK/- |
| *S. paradoxus* | T18.2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| *S. paradoxus* | Y8.5 | PK/K | PK | K | K | - | - | PK | PK | K | K | K | PK | K | PK/- |
| *S. paradoxus* | Y10 | PK/- | - | - | - | - | - | - | - | - | - | - | - | - | - |
| *S. paradoxus* | CBS 8444 | PK | PK | - | - | - | - | - | - | - | - | - | - | - | - |
| *S. paradoxus* | CBS8435 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| *S. paradoxus* | A24 | PK | VPK | - | - | - | - | - | - | - | - | PVPK | - | - | - |
| *S. paradoxus* | A33 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |

The killer-immune tests were repeated using three non-killer *S. paradoxus* strains as the immune tester, A33, T18.2, and CBS8437, in order to get a better picture of the killer-immune reaction between the two yeast species. The results of the Killer-immune tests of the yeast strains (killer tester) against the *S. cerevisiae,* killer and non-killer, strains and the *S. paradoxus* non-killer strains (immune tester) were compared (Table 2-1). Although the *S. cerevisiae* killer strains expressed their killer phenotype against the majority of *S. paradoxus* strains, the level of expression against *S. paradoxus* strains is significantly less than that of the *S. cerevisiae* strains. This result suggests that an additional immunity exists in the *S. paradoxus* strains, which increases the tolerance of the strains against the *S. cerevisiae* killer strains, similar to that reported by Chang et al. (Chang et al., 2015). Of all the *S. paradoxus* strains, A33 showed the highest level of immunity against *S. cerevisiae* strains.

Interestingly, two *S. cerevisiae* strains, K1 617 and K1963, which have the same killer toxin, showed different reactions against this strain, which might be as a result of the killer-toxin expression level. A33 is also immune to K2 killer toxin.

Four *S. paradoxus* killer strains, Q59.1, Y10, CBS8444, and A24, which express a weak killer phenotype against the *S. cerevisiae* sensitive strain, could not show the phenotype against any of the *S. cerevisiae* killer strains and *S. paradoxus* non-killer strains. This suggests that these six strains are immune to these killer strains. The phenotype of T4b against the six strains showed that all the *S. paradoxus* non-killer strains are immune to this strain, whereas all the *S. cerevisiae* strains are killed by it. This demonstrates a type of immunity that seems to exist specifically in *S. paradoxus* strains. Q62.5 is the only strain that kills all of the *S. cerevisiae* and *S. paradoxus* strains. However, similar to the *S. cerevisiae* killer strains, its killer activity is reduced against the *S. paradoxus* strains. All of the immunity might result from mutation in the genes, which encodes receptors or proteins used by killer toxins to perform their killer activity (Schmitt and Breinig, 2006), or from some unknown pathway of immunity in the strains.

Overall, there is not any particular pattern of killer-immune reaction between two species. It mostly depends on the type of killer toxin and the genome of the strains that exist in one environment. Perhaps, in addition to the environmental factors, the reason that the killer phenotype was not recognised in previous studies in the four strains of *S. paradoxus* (Chang et al., 2015, Pieczynska et al., 2013b), or the reason that the killer phenotype in Q74.4 and Y8.5, which was reported to be the strongest amongst *S. paradoxus* killer strains (Chang et al., 2015) compared to this study in which Q62.5 expressed the strongest phenotype, is that the genome of the immune testers was different in these projects.

### 2.3.2 Characterisation of the killer toxins

Given that Q62.5 expressed the strongest killer phenotype amongst the *S. paradoxus* strains and appears to have a new type of killer toxin, its killer toxin was used for the further studies. Given that the killer toxin produced by the killer strains is accumulated in the medium, the medium was used for the characterisation of the killer toxin. After removing the cells that were grown on the liquid medium using centrifugation and filtration, the medium was concentrated using concentrator filters or ethanol.

When concentrating using the concentrator filters, in order to reduce the number of proteins in the secretome, various cut-offs, 10K, 30K, 50K, and 100K, were tested using Amicon ultra centrifugal filter. In this concentrator, the filter is fitted on the top of a centrifugal tube. The concentrated medium was collected from the filter and its killer activity was tested using MB agar toxin assay. The result of the killer assay indicated that the concentrated supernatants with 10K and 30K have the same killer activity, whereas the killer activity in that of 50K is reduced, and the concentrated supernatant from the 100K filter is not active. In the case of the 50K filter, if the supernatant is dialyzed twice, it does not show any killer activity. These results indicate that the molecular weight of the toxins is more than 30 kDa and less than 50 kDa.

In concentrating the medium using ethanol, different amounts of cold ethanol, 50%, 70%, 75%, and 80%, were applied into the 40x concentrated media in order to precipitate the proteins and the killer toxin in them (Ciani and Fatichenti, 2001, Santos et al., 2004, Ha et al., 1997). After washing the pellets, which contains the killer toxin, and dissolving them in Citrate-Phosphate buffer pH 4.5, the activities of the concentrated  toxins were tested using MB agar toxin assay. Only the killer toxins in the media that were concentrated using 50% and 70% ethanol were still active. The killer activity of the concentrated medium using 50% ethanol was greater than that of 70%. The results indicated that the killer toxin is not stable in relation to ethanol.

**The effect of ethanol on the killer activity of killer toxins**

The result of concentration using ethanol encouraged us to study the effect of the ethanol on the killer activity of the killer toxins. Since the maximum ethanol in the fermentation reaction is 14%, four types of MB agar containing 0%, 6%, 12%, and 14% ethanol were prepared. The killer phenotype of the three killer strains, Q62.5, T21.4, and Y8.5, and the killer activity of the killer toxin in the Q62.5 concentrated medium were tested on these media. Although Saccharomyces strains are resistant to ethanol, the results of the treatment showed that the expression of the killer phenotype in the two of strong killer strains, Q62.5 and T21.4, reduces by the increased concentration of ethanol and the killer toxin in strain Y8.5, which is a weaker killer strain, and killer toxin in the Q62.5's concentrated medium becomes completely inactive with all the concentrations of ethanol (Table 2-2). However, the effect is not permanent. If the killer toxin concentrated medium treated by the three concentrations of ethanol for two days, is put on a normal MB agar seeded with the sensitive strain, it becomes active again. These results suggest that the nature of producing ethanol by the yeast strain decreases the efficiency of the killer toxin of the yeast-killer strains. Nevertheless, this is not true for all the yeast-killer toxins. In the case of the KHR killer toxin, increasing the ethanol in fermentation improves the efficiency of the killer toxin (de Ullivarri et al., 2014).

**Table 2-2.** The effect of ethanol on the killer activity of the three *S. paradoxus* strains, O62.5, T21.4, and Y8.5, and the concentrated medium that contains Q62.5 killer toxin; K (Killer), PK (Poor Killer), VPK (Very Poor Killer), VVPK (Very Very Poor Killer) and (-) non-killer.

| Sample | 0% | 6% | 12% | 14% |
|---|---|---|---|---|
| **Q62.5 strain** | K | PK | VPK | VVPK |
| **T21.4 strain** | K | PK | VPK | - |
| **Y8.5 strain** | PK | - | - | - |
| **Q62.5 concentrated medium** | K | - | - | - |

**Effect of the pH on the activity and stability of Q62.5 killer toxin**

A comparison of the killer phenotype of Q62.5 on MB agar diffusion assay (comparing the halo and the blue cells around the strain) at three different pH, 4.2, 4.5, and 4.7, showed that at pH 4.5 the strain expresses the highest level of the killer phenotype. The level of the phenotype expression at pH 4.7 is less than the level at 4.2. The amount of toxin production of this strain in YPAD broth at pH 4.5 is also materially higher than at pH 4.2; putting the same amount of both media (YPAD broth pH 4.5 and pH 4.2 which were cultivated with Q62.5) on MB agar pH 4.2 and pH 4.7 showed that the killer activity of the YPAD medium with pH 4.5 is about three times stronger than that of the medium with pH 4.2 on both MB agar. This suggests that the optimum pH of the Q62.5 killer toxin is 4.5. Testing the stability of the Q62.5 toxin at pH 7 showed that, despite being inactive at this pH, reducing the pH below 5 activates the toxin again. However, its activity is slightly lower than non-treated samples. Increasing the time of treatment from 16 hours to 32 hours did not make a difference in the killer activity.

Testing the pH of the medium after two days of growing the strain showed a reduction in the medium's pH. The pH after growing was between 3.2 and 3.5. The pH of the medium that was shaken reduced more than that which was not shaken.

### 2.3.3 Growth rate of *S. paradoxus* and *S. cerevisiae* strains in the presence of Q62.5 killer toxin

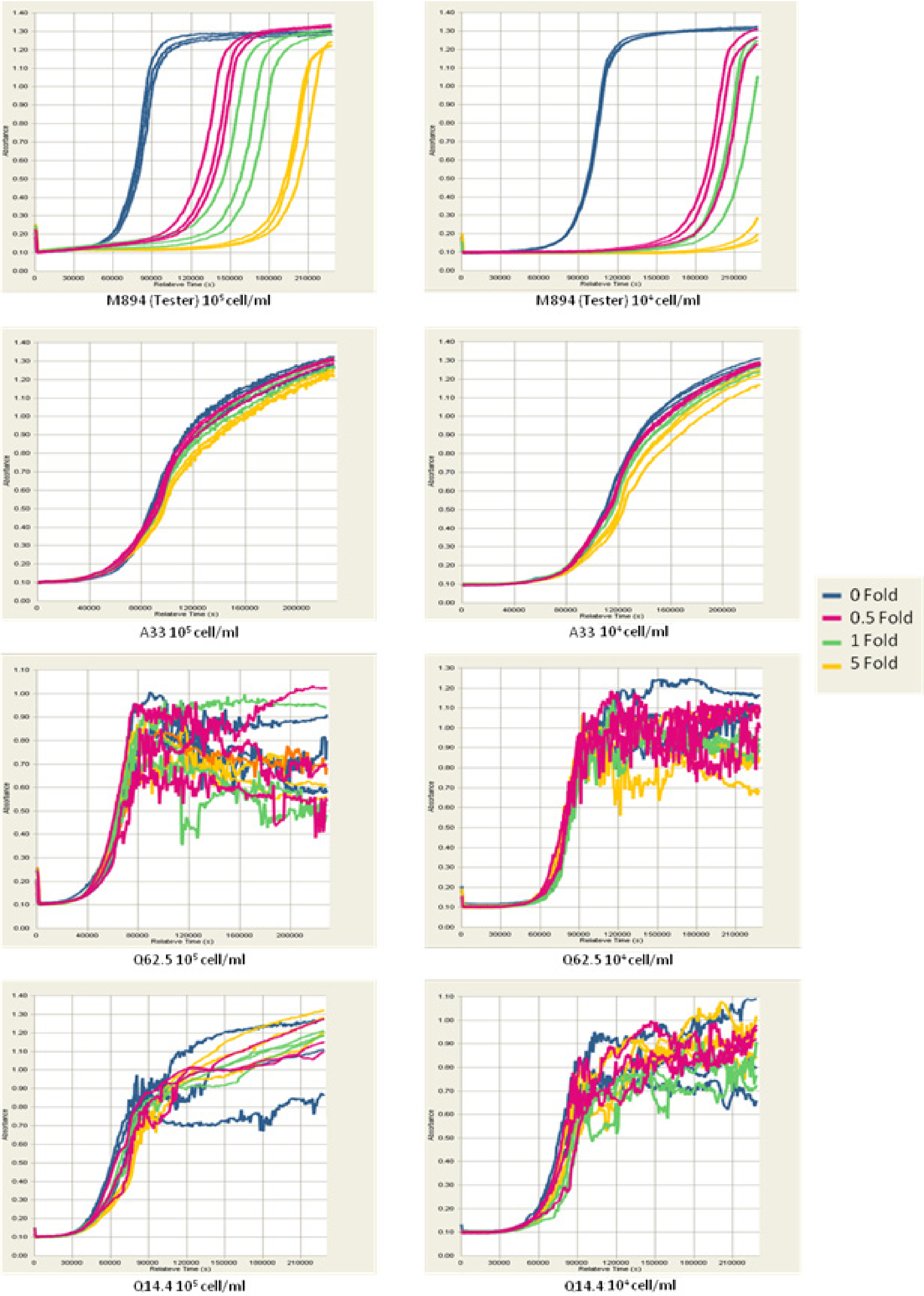
In order to study the interaction between the killer toxin of Q62.5 and the other strains, the growth rate of Q62.5, M894 (the *S. cerevisiae* sensitive strain), Q14.4 (*S. paradoxus* killer strain) and A33 (the non-killer *S. paradoxus* that showed the highest amount of immunity to Q62.5 killer toxin) was measured using a microtiter plate reader in the absence of Q62.5's concentrated medium and in the presence of 0.5, 1, and 5 folds of that in 62 hours. To determine the effect of the number of cells at the start of the test, two different amounts of the strain were used: $10^4$ and $10^5$ cells.

The results unexpectedly showed that the killer toxin does not kill all the sensitive cells. Even in the high concentration of the toxin (i.e. five times more than normal concentration in the medium) and low concentration of sensitive cells, there is only a delay in the growth of the cells (Graph 2-1). The higher the concentration of the toxin and the lower the number of sensitive cells at the start, the greater the delay in growth. It seems that, during this period of delay, the killer toxin is rendered inactive by the sensitive cells. In contrast, the filtered concentrated medium containing the killer toxin stays active, when kept at the same temperature and for the same period of time without any killer or sensitive cells.

Reducing the killer activity of the killer toxins were also observed during the culturing of Q62.5. By increasing the age of the culture, the killer activity of the killer toxins accumulated in the liquid medium decreases (Graph 2-2). Given the fact that, by ageing the culture, the amount of the ethanol accumulated in the medium increases and the pH of the medium decreases, and based on the previous results which indicated that both changes have negative effects on the killer activity of the killer toxins, perhaps the increase in ethanol and the decrease in pH levels are themselves the two factors that negatively affect the killer activity of the killer toxins in both tests. During culturing strains, in addition to numerous factors that are changing in the media, the strains release their secretome and metabolites in the media that may affect the killer activity of the killer toxin.

A comparison between the graphs of the two non-killer strains (Graph 2-1), A33 and M894, indicated that the delay in the growth of A33 (which had higher immunity to the killer toxin compared to M894 (Table 2-1)), is significantly less than M894. These probably arise from the genetic background of the strains and the different immunity pathways within them. It seems that, in addition to pH and ethanol which decrease the activity of the toxin, all the strains, even the sensitive strain, are equipped with extra immunity which can protect them against killer toxin in the environment in order to increase their chance of survival. This may explain why killer strains of *S. cerevisiae* have not been dominant at the end of fermentation in some of the previous studies (Heard and Fleet, 1987) and why the killer strains have a lower frequency than the non-killer strains in the environment (Chang et al., 2015, Pieczynska et al., 2013b). Perhaps it is the strategy that is used in the environment to make a balance between the killer and sensitive strains and prevent the extinction of the strains.

The growth rate of killer strains Q62.5 and Q14.4 fluctuated significantly more than that of non-killer strains M984 and A33, which might result from their toxin production.



M894 (Tester) $10^5$ cell/ml

M894 (Tester) $10^4$ cell/ml

A33 $10^5$ cell/ml

A33 $10^4$ cell/ml

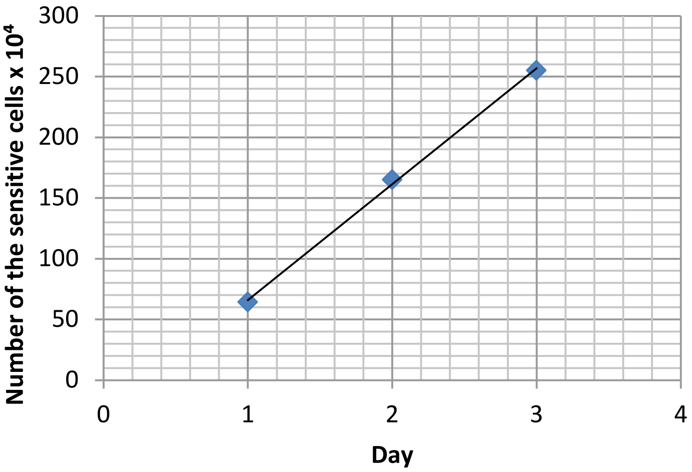
Q62.5 $10^5$ cell/ml

Q62.5 $10^4$ cell/ml

Q14.4 $10^5$ cell/ml

Q14.4 $10^4$ cell/ml

**Graph 2-1.** Comparison of the growth rate of four different strains, M894 (tester), A33, Q62.5, and Q14.4, in the presence of the 0 fold, 0.5 fold, 1 fold and 5 folds of Q62.5's concentrated medium that contains Q62.6 killer toxin,. The tests were conducted with two different concentrations of the cells, $10^4$ cell/ml and $10^5$ cell/ml and each test was repeated three times.



**Graph 2-2**. Decreasing the killer activity of the Q62.5 killer toxin by aging the culture. Q62.5 were cultured for three days and the killer activity of its secreted killer toxin were measured in the first, second and third days, using microtiter plate killer assay. The number of the sensitive cells after finishing the killer assay is shown on the graph. By increasing the age of the culture, the killer activity of the killer toxin decreased. As a result the number of the sensitive cells in the killer assay increased.

### 2.3.4 Characterisation of the killer toxins using gel electrophoresis

The concentrated medium was used to study the killer toxin on the gel electrophoresis. YPAD buffered with citrate phosphate (pH 4.5), which is a common medium for growing yeast and producing toxins, was initially used for toxin production. However, because it has a high amount of protein due to the presence of peptone and yeast extract, it caused problems in separating the bands in polyacrylamide gel, and hence detecting the toxin bands. As a result, minimal media with and without amino acids, buffered with citrate-phosphate buffer, were tried. The results showed that both forms were suitable for producing toxins. However, the amount of toxin in the medium with amino acid exceeded that of the medium without. In addition, supplementing this medium with the non-ionic detergent Brij-58 (Santos and Marquina, 2004), resulted in the highest killer activity in the supernatant.

On the low pH gels, IEF gels and native gels, because the killer toxins stay active, each sample was run on the gel in two different lanes. After running, the gel was cut into two pieces. In one part of the gel all proteins were visualized using Coomassie Blue, and in the other part the toxin was detected using the bio assay. In the bio assay, the gel was laid on the MB agar and the sensitive strain sprayed on the plate.

**Low pH polyacrylamide gel**

The pH of the gel made in the lab was 4.3 and the pH of the running buffer was 4.5. With normal electrode polarity, there were some faint bands at the top of the gel that did not move down with increasing voltage and running time. This suggests that perhaps the isoelectric point of the toxin is near to the pH of the gel and, as a result, it did not have enough charge to move in the gel. To solve this issue, IEF gel was substituted.

**IEF gel**

An IEF gel with pH 3-7 from Invitrogen was used. In order to detect the killer toxin band and compare it with the known killer toxins, the concentrated media of K1 618 (K1 killer), Q62.5 and M894 (non-killer) were run on the gel. The concentrated media from both concentration methods, ethanol precipitation and concentrator filter, were loaded on the gel in different lanes. Both concentrated media were stained with Coomassie Blue but a bioassay was done only on the lane which was concentrated using the concentrators. The results showed that the proteins in the medium concentrated using ethanol were separated considerably more sharply than when using the concentrators. It seems that the unremoved salts in the first method prevented clear separation of the bands.

The results of the IEF gel indicated that the isoelectric point of the majority of the proteins, which are secreted by both *S. cerevisiae* and *S. paradoxus*, is between 3 and 4.5 (Figure 2-2); this explains why the proteins did not move very well on the low pH gel. After doing the bio assay, a clear zone was detected on the gel. It indicated that the isoelectric point of the toxin is approximately 4.5–5.2. In addition to this part, there was a zone with a lower number of the sensitive clones around pH 3.5–4.2. This might arise from the poor separation of the proteins. Since most of the secreted proteins are flocked where the toxin band is detected, it was not possible to detect the toxin bands among other bands by comparing the results of the Coomassie Blue staining and bioassay, and comparing the bands of the Q62.5 with the non-killer strain M894. As a consequence, we tried to purify the killer toxin for further studies.

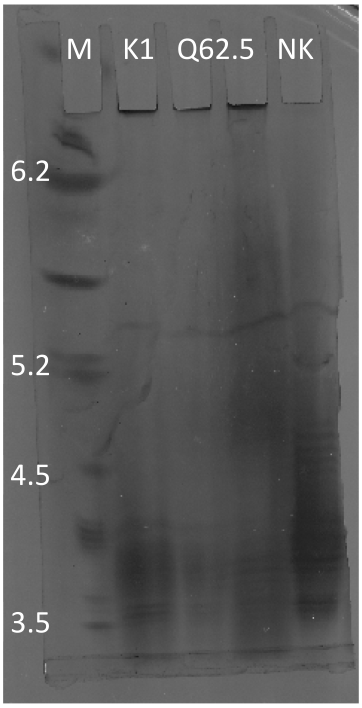


**Figure 2-2.** The IEF gel electrophoresis. The first lane is the marker (M). The second lane is the concentrated medium of K1 617 (K1). The third lane is the Q62.5's medium concentrated using ethanol precipitation. The fourth lane is Q62.5's medium concentrated with the concentrator filter and the fifth lane is that of M894, which is not killer (NK). A comparison of second, third, and fourth lanes, which contain the medium of the killer strains, with the fifth lane, that of the non-killer strain, does not show any specific band for the killer toxins. The isoelectric point of the majority of the proteins secreted by the yeasts in the medium is between 3 and 4.5.

### 2.3.5 Purification of killer toxin

Different strategies were tried to purify the killer toxin. Since the killer activity of the killer toxin reduced during the concentration of the medium using ethanol, the medium was concentrated using the concentrator devices, although this took longer time. Because the size of the killer toxin in Q62.5 was between 30 kDa and 50 kDa, the medium was concentrated using 30 K cut-off concentrator devices in order to remove the proteins smaller than 30 kDa toxin.

After growing Q62.5 in a minimal medium containing Brij-58 (pH 4.5), removing the cells and concentrating the supernatant, ion exchange chromatography using the DEAE column was performed. Then, the fractions were concentrated, their buffer was changed with citrate-phosphate buffer (pH 4.5), and their killer activity tested by a microtiter plate-reader killer assay. Finally, the active fractions were studied on the native and SDS PAGEs.

The purification of the killer toxin was started with 250 ml of medium but, because the amount of protein was low, it was increased to 1,000 ml. Testing the killer activity of the fractions showed that 6 out of 40 fractions of the chromatography was active, fractions 6 to 11. The peak of the elution period includes the active fractions.

On the native gel, the bands did not separate from each other very well in either staining or bioassay; as a result, a specific band for the toxin was not detectable (Figure 2-3). On the SDS PAGE there were several bands in each fraction which show the toxin did not purify completely. It was not possible to detect the toxin bands from comparison of active and inactive fractions. Given that the killer activity in killer fractions is high, it shows that the killer toxins in this strain are highly active; as a result, the low amount of protein which is not detectable on the gel produces a highly active phenotype.

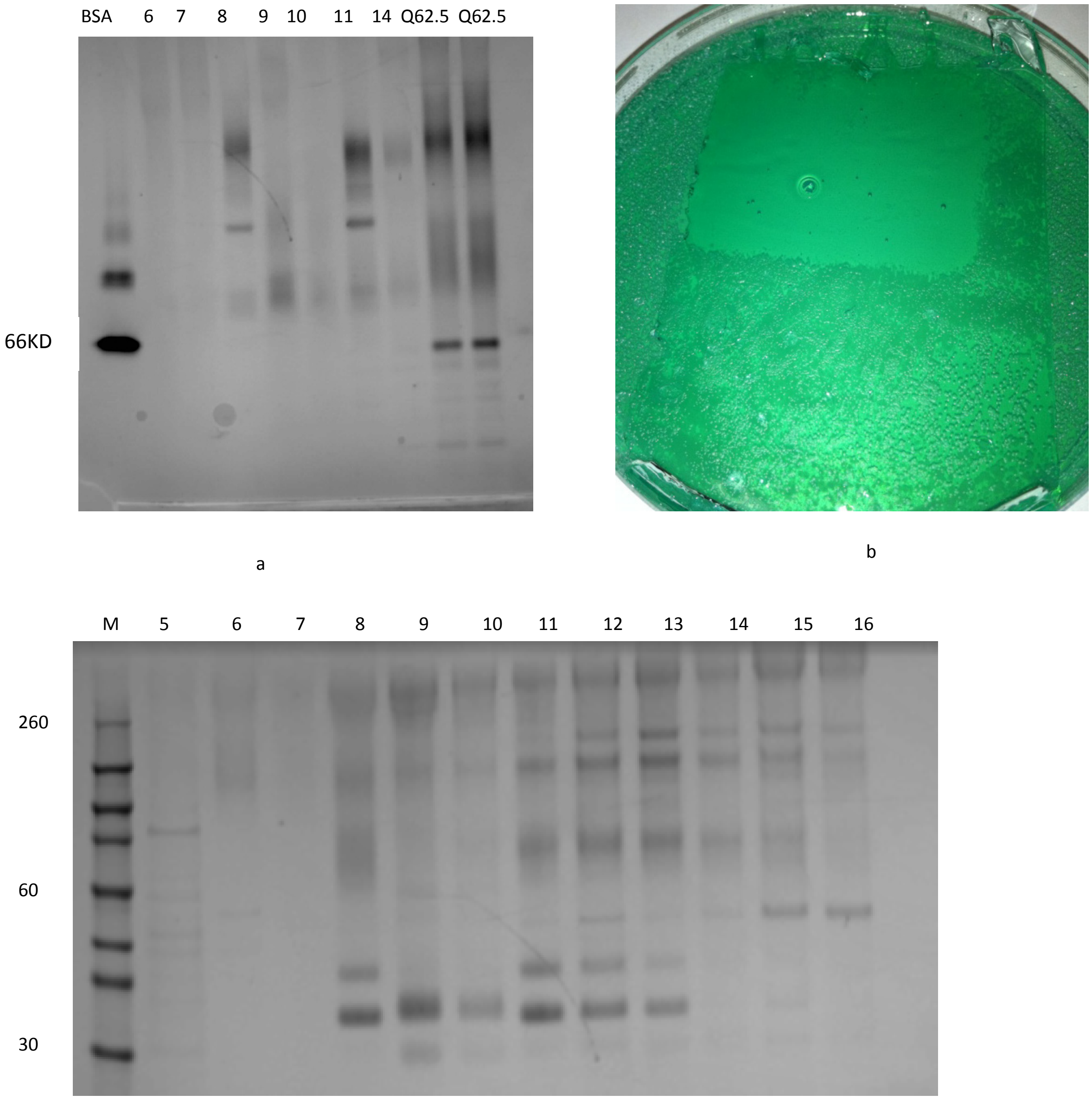

BSA 6 7 8 9 10 11 14 Q62.5 Q62.5
66KD
a
b
M 5 6 7 8 9 10 11 12 13 14 15 16
260
60
30

**Figure 2-3.** The native and SDS-PAGE analysis of the fractions eluted from the DEAE column. (A) The half of native gel that was strained with Coomassie Blue. The first lane is bovine serum albumin (BSA) loaded as a marker (approximately 66 KD). The six active fractions, 6, 7, 8, 9, 10, and 11, and one inactive fraction, 14, and Q62.5 concentrated medium were run on the gel to compare. The bands on the gel did not separate very well. (B) The other half of the native gel laid down on MB agar in order to perform the bioassay. The sensitive strain could not grow on half of the gel, which is as a result of poor separation. A comparison of A and B shows that the killer toxin protein is a highly active protein; as a result, the low amount of protein, which is not detectable on the stained gel, produces a highly active phenotype (C) The SDS PAGE stained with Commassie Blue. The active fractions, 6, 7, 8, 9, 10, and 11 were compared with inactive fractions, 5, 14, 15, and 16. No specific band was found for the killer toxin.

To overcome this problem, the amount of medium was increased firstly to 10 litres then to 20 litres. Concentrating 20 L of the medium increased the viscosity of the solution. Since the amount of protein was not very high in the concentrated medium, it seems that the metabolites that exist in the medium increased the viscosity. Washing the concentrated medium with a greater amount of citrate-phosphate buffer did not help decrease the viscosity of the solution. It seems that the contaminants cannot pass through the 30 K cut-off filter. As a result, to remove the contaminants in the medium and the secreted proteins, which were bigger than the killer toxin, the concentrated samples were passed through the gel filtration column, Superdex 75. However, after passing the medium through the column, the viscosity of the medium did not change. The bigger proteins were also not removed from the active fractions. Since the column was packed in the lab, firstly, it seemed that the problem arose from packing the column. However, testing the column using protein markers showed that packing was fine. It seems that the speed of the proteins and the metabolite contaminant in the column is similar. Since the viscosity of the solution is high, it does not allow the proteins to separate based on their size. Using Sephadex, the G25 column was not successful either.

In order to eliminate the contaminants, the protein purification using TCA was tested. However, after purification, part of the precipitate did not dissolve in the buffer. Moreover, the killer activity of the toxin also decreased. Since the purification of the protein was complicated, took a long time and did not produce a good result, we decided to study the killer system in this species through its genetic base.

### 2.3.6 Conclusion

The killer phenotype was detected in 36% of the *S. paradoxus* strains. It was very sensitive to the environmental conditions. The killer activity of the strains was tested in four different PH, 3.5, 4.2, 4.5, and 5.3 (the result not shown). None of them were active at a pH more than 5 and at a temperature more than 28°C.

The frequency of the phenotype in North and South America was more than other regions. In general, depending on the time of spraying the sensitive strains, most of the tested killer strains showed different levels of expression of the phenotype. The majority of the strains from Silwood Park, Continental Europe, the Far East and South America express the phenotype more strongly when the sensitive strain was sprayed on the medium on the different day as the killer testers were applied. Two strains from Silwood Park and three strains from Continental Europe did not show the killer phenotype when the tester was sprayed on the same day. In contrast, all North American killer strains showed their killer phenotype more clearly on Same-Day-Test than on Different-Day-Test.

The killer phenotype in the North American strains looks different and they are very weak killers. The different level of the phenotype expression in Same-Day-Test and Different-Day-Test may be as a result of various reactions of the sensitive strain against the different concentration of the killer toxin.

None of the *S. paradoxus* killer strains showed the same killer-immune pattern as *S. cerevisiae* strains, except three strains from Silwood Park. The level of the killer-phenotype expression of *S. cerevisiae* killer strains against *S. paradoxus* strains is significantly less than that of *S. cerevisiae* strains. This shows an additional immunity that exists in the *S. paradoxus* strains.

The killer toxin of Q62.5, which was the strongest of *S. paradoxus* killer strains, was characterised. Its molecular weight is between 30 kDa and 50 kDa, its isoelectric point is between 4.5 and 5.2, and its optimum pH is 4.5. Although it is not active at pH 7, it is stable at this pH. The killer activity of the toxin decreases by increasing the amount of ethanol in the medium. Measuring the growth rate of four *S. paradoxus* strains in the presence of the killer toxin indicated that none of the strains, even the sensitive one, was killed completely by the killer toxin; there is only a delay in the growth of the strains. The higher the concentration of the toxin and the lower the number of sensitive cells, the higher the delay in the growth of the cells. In the strains that are more sensitive to the killer toxin, this delay increases. Overall, results suggest that all the strains, even the sensitive strains, are equipped with some immunity for their survival. Nevertheless, this immunity in the sensitive strains is significantly less than in the resistant strains. In addition to this immunity, increasing the ethanol level and decreasing the pH during fermentation increases the chance of survival for the sensitive strains. Moreover, these might be the reasons for the killer's inability to be dominant at the end of all fermentation and to be the dominant strains in nature.

**Chapter three**

# 3 Determination of the genetic basis of the killer toxin in *S. paradoxus*

## 3.1 Introduction

There are two genetic bases for the killer phenotype that have been discovered in *S. cerevisiae*, viral dsRNA (Rodríguez-Cousiño et al., 2011, Hopper et al., 1977b, Magliani et al., 1997b) and chromosomal genes (Goto et al., 1991, Goto et al., 1990b). In the killer yeasts that were infected with viral dsRNA, two types of dsRNA, L-A and M dsRNA, coexist in the cell. L-A dsRNA, with a size of 4.5 kb, produces Gag and Pol protein to replicate and encapsidate itself, and M dsRNA. M dsRNA, with a size of between 1.5 kb and 2.5 kb, has just one gene to produce the killer toxins. Various types of L-A, L-A-L1, L-A-2, L-A-28, and L-A-lus, and M dsRNA, M1, M2, M28, and M-lus have been detected, which co-evolved with each other respectively. The M dsRNAs encode four well-known killer toxins: K1, K2, K28 and K-lus respectively (Rodríguez-Cousiño and Esteban, 2017, Rodríguez-Cousiño et al., 2013, Rodríguez-Cousiño et al., 2011, SCHMITT and TIPPER, 1995, Wickner, 1996). L-A can exist alone but it does not produce any phenotype in the cell (Hopper et al., 1977b, Wickner et al., 1995). One L-A dsRNA and at least seven different M dsRNAs, based on size, were detected in *S. paradoxus* (Naumov et al., 2004) and the killer gene of M1 has been sequenced in two strains of these species, Q74.4 and N-45, and the M28 killer gene has been found in DBVPG4650 and T21.4 (Chang et al., 2015).

In addition to L-A and M dsRNA, there are three other dsRNAs, L-BC (4.5 kb), W (2.25) and T (2.7 kb). They produce their own RNA-dependent RNA polymerase and do not make any phenotype in the infected cells (Esteban et al., 1992b, Rodriguez-Cousiño et al., 1991, Sommer and Wickner, 1982).

The identified killer chromosomal genes in *S. cerevisiae* are composed of KHR and KHS. They are encoded by the right arm of chromosome V and the left arm of chromosome IX respectively, with no homology with each other or another killer gene (Goto et al., 1991, Goto et al., 1990b). The sequence of KHS was studied by Frank and Wolfe (2009) in other *S. cerevisiae* genome sequences.

They believe that there is an inversion in the sequence that ends at a restriction site that was used in the cloning of the gene before its sequencing. As a result, they considered the SCY_1690 ORF in the genome of *S. cerevisiae* YJM789 as the KHS encoding gene. SCY_1690, which is 1,057 bp, codes a 350-amino acid protein in *S. cerevisiae* strains YJM789 (protein name, SCY_1690), M22, and most of the strains sequenced by the *Saccharomyces* Genome Resequencing Project. Its coding region from strains YJM789 and M22 has greater similarity with the *S. paradoxus* genome sequence, 99% identity, than the null alleles in *S. cerevisiae,* 89% identity, which suggests a horizontal exchange of this gene between two species. Also, they introduce SCY-1690 as an NUPAV (nuclear sequences of plasmid and viral origin) related to M2 killer virus double-stranded RNA because of the similarity which they found between their sequences (Frank and Wolfe, 2009).

Cycloheximide treatment is the routine method for identifying the genetic basis of killer phenotypes. This antibiotic, which acts on the large ribosomal subunit protein L29 (Fried and Warner, 1982, Stöcklein and Piepersberg, 1980), at low concentration, removes M dsRNA from the infected cell. As a result, if the killer toxin is encoded by M dsRNA, the phenotype changes to non-killer (Carroll and Wickner, 1995, Fink and Styles, 1972).

In this chapter, various types of dsRNA were detected in the *S. paradoxus* strains. The results were compared with the killer phenotype of each strain. Then, to identify the genetic basis of the killer strains, these strains were treated with a low concentration of cycloheximide. Finally, the presence of KHS and KHR sequences were studied in the strains in which their genome sequences were available.

## 3.2 Materials and methods

### 3.2.1 dsRNA extraction

For the extraction of dsRNA from yeast cells, five different methods were tried. (1) the method described by Fried and Fink with some changes (Fried and Fink, 1978); (2) the method used by Wickner and Leibowitz (Wickner and Leibowitz, 1976); (3) the method used by Coenen and colleagues (Coenen et al., 1997); (4) RiboPure Yeast kit (Amibion); and (5) RNeasy mini kit (QIAGEN).

Of the five methods, the first is not only the cheapest but also enables the extraction of the greatest amount of dsRNA. Also, the concentration of DNA obtained is lower than that from the three other methods used from the papers, but the purity of dsRNA after extraction is not as high as the purity of extracted dsRNA when using the kits. This purity is satisfactory for detecting dsRNAs on the agarose gel. In cases where highly purified dsRNA is needed, an extra treatment with lithium chloride is performed on the dsRNA in order to remove the carbohydrates (Salzman et al., 1999).

333 µl of 8 M LiCl was added to the sample in which the volume was made to 1 ml with water. Then, the RNA was precipitated at 4 ◦C for at least 3 h (or overnight). After centrifugation (20 min at 12,000 × g at 4 ◦C) and washing the pellet with 400 µl 80% Ethanol, the pellet was resuspended with water. Although this treatment reduces the amount of dsRNA, it produces much higher purity (Figure 3-2).

Using this method, the cells grown to stationary phase in YPAD broth were washed with 50 mM Na2EDTA (pH 7.0). Then they were incubated for 15 min in 50 mM Tris-H2SO4 (pH 9.3) (Sigma) containing 2.5% 2-mercaptoethanol (Sigma) and, following that, stirred for one hour at room temperature with 0.1 M NaCl (Sigma), 10 mM Tris-HCl (pH 7.5) (Sigma), 10 mM Na2EDTA (Sigma), 0.2% sodium dodecyl sulfate and an equal volume of phenol-chloroform-isoamyl alcohol. After washing the aqueous phase with an equal volume of chloroform-isoamyl alcohol (Sigma), the nucleic acid was precipitated with two volumes of cold ethanol and incubated overnight at -20˚C. After centrifugation and washing the nucleic acid with 70% ethanol, it was dissolved in water. All the centrifugation from phenol-chloroform-isoamyl alcohol steps were performed at 16,000 x *g*.

### 3.2.2 ssRNA and DNA digestion

To digest ssRNA, RNase A was tried initially. RNase A in high salt concentration digests only ssRNA, whereas, in low salt concentration it digests both ssRNA and dsRNA. However, this digestive process is very sensitive to the concentration of salt, and minor changes affect the results. As a consequence, RNase A was replaced with S1 nuclease, an enzyme which digests all single-strand nucleic acids.

Since S1 nuclease is more sensitive to changes in salt concentration than DNase, treatment with S1 nuclease was performed before DNase treatment. 1, 1.5, 2 or 3 units of the enzyme were used for each µg of extracted nucleic acid, and the samples were incubated at 37°C or 67°C for half an hour. After stopping the reaction of S1 nuclease with 15 mM Na2EDTA (pH 7.5), the remaining DNA was treated with DNaseI (Sigma); it was then washed using the same volume of phenol-chloroform-isoamyl alcohol and, following that, with the same volume of chloroform-isoamyl alcohol. The dsRNA was precipitated with two volumes of cold ethanol (-20°C). After washing with 70% ethanol, it was dissolved in RNase-free water.

### 3.2.3 Agarose gel electrophoresis

Samples were run overnight on a 1% agarose gel prepared using a TAE buffer and GelRed. After running, the bands were visualized with UV.

### 3.2.4 Cycloheximide treatment

YPD medium containing 2% (20 mg/ml), 6% (60 mg/ml), and 10% (100 mg/ml) of cycloheximide were tested for the treatment. The killer strains were cultured on the media and incubated at 30°C until the clones appeared. Three clones from each strain were cultured on YPD broth overnight. Then, their killer activity was tested on MB agar diffusion assay (Section 2.2.3.). All the strains were tested on Same-Day-Test and Different-Day-Test.

### 3.2.5 Bioinformatic analysis of KHS and KHR genes in *S. paradoxus*

The sequences of KHS and KHR genes, as well as SCY_1690 ORF, were taken from NCBI and studied in the genomes of 37 *S. paradoxus* strains using Geneious. In order to find the position of KHS ORF and SCY_1690, their sequences were firstly aligned with each other then they were aligned with the genomes of all the *S. paradoxus* strains.

## 3.3 Results and discussion

### 3.3.1 Viral dsRNA in *S. paradoxus* strains

dsRNA extraction was performed on 68 strains of *S. paradoxus*, 30 strains from Silwood Park, 15 strains from continental Europe, eight strains from the Far East and 15 strains from Canada (Table 3-1). The result of the gel electrophoresis of the dsRNA extractions indicated that there are two categories of dsRNA in these strains; a large band of approximately 4,000 bp in size, compatible with that of L dsRNA in *S. cerevisiae*, and a medium-sized band of 1,000 - 2,500 bp, similar to that of M dsRNA in *S. cerevisiae*. The size of the large-sized band was similar among all the strains, whereas the medium-sized bands varied in size (Figure 3-2). In total, 25 strains contained both L and M dsRNA. Three of the killer strains from Silwood Park, T21.4, T4b, and Y10, initially showed only one large-sized dsRNA. However, increasing the concentration of dsRNA revealed a medium-sized band, which was still not as clear as the similar band in the other strains. Perhaps it arose from the lower copy number of the dsRNA in these strains compared to the other strains. Testing the bands with DNase and S1 nuclease proved that both bands in all the strains are dsRNA (Figure 3-2).

The S1 nuclease was tried with two incubation temperatures, 37°C and 67°C, and four different concentrations (1, 1.5, 2, and 3 units of enzyme per µg of the nucleic acid). M dsRNA showed different behaviour in the various incubation temperatures. At 67°C the M dsRNA was cut into two pieces, similar to that reported by Welsh and Leibowitz (Welsh and Leibowitz, 1982) (Figure 3-1). Perhaps at higher temperatures, the poly A region of M dsRNA, which is located in the middle of the coding and noncoding part of the RNA (Figure 1-3), is melted and digested by S1 nuclease. Although in the instruction of the enzyme it mentioned that, at high concentration of the enzyme, dsRNA can be digested, increasing the concentration did not show any visible effects on the dsRNA when it was run on the 1% agarose gel. Since the sensitivity of the gel is not very high, despite not seeing any differences on the gel, there is the possibility of the effect of S1 nuclease on the lower amount of dsRNA or the hairpin structures that exist at the ends of dsRNA.

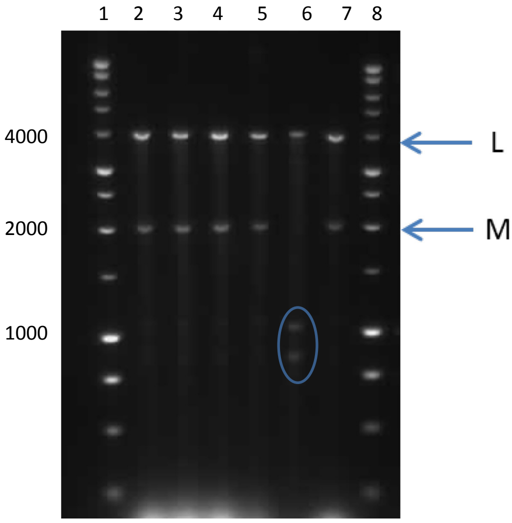


**Figure 3-1.** A comparison of dsRNA S1 nuclease treatments with different concentrations of enzyme and incubation temperatures. The dsRNA extracted from strain Q43.5 was treated with different concentrations of the S1 nuclease: 1, 1.5, 2 and 3 units of enzyme per 1µg of nucleic acid; and incubation temperatures: 37°C and 67°C. All the samples were treated with DNase. Lanes 1 and 8 are 1 kb DNA ladder. The condition of treatment of each Lane is: Lane 2: 1 unit of S1 nuclease per 1µg of nucleic acid, incubated at 37°C; Lane 3: 1.5 units of S1 nuclease per 1µg of nucleic acid, incubated at 37°C; Lane 4: 2 units of S1 nuclease per 1µg of nucleic acid, incubated at 37°C; Lane 5: 3 units of S1 nuclease per 1µg of nucleic acid, incubated at 37°C; Lane 6: 1 unit of the enzyme per 1µg of nucleic acid and incubated at 67°C; Lane 7: the samples were just treated with DNase. The concentration of the enzyme had no effect on the cleavage of the dsRNA, Lanes 2-5, whereas increasing the temperature from 37°C to 67°C promoted the cleavage, Lane 6.

The various types of M dsRNA in *S. paradoxus* strains, as previously reported, are of different sizes (Naumov et al., 2004). Moreover, the size of the digested fragments, as well as the pattern of digestion, was different in various strains. Based on the size and digestion pattern with S1 nuclease, there are at least eight different medium-sized dsRNAs in the Silwood Park strains (Figure 3-2). The M dsRNA in T68.2, T68.3, Q43.5, Y8.5 and Q74.4 seems to be the same. Among all the S1 nuclease treated samples, the M dsRNA in Y2.8 strain did not show digestion on the gel and Q95.3 and Y1 showed partial digestion. These patterns of digestion suggest that the dsRNA in these strains has a more stable structure, probably shorter poly A in the middle. The M dsRNAs of Q95.3 and Q16.1 are shortest at 1 kb and 1.3 kb, respectively.

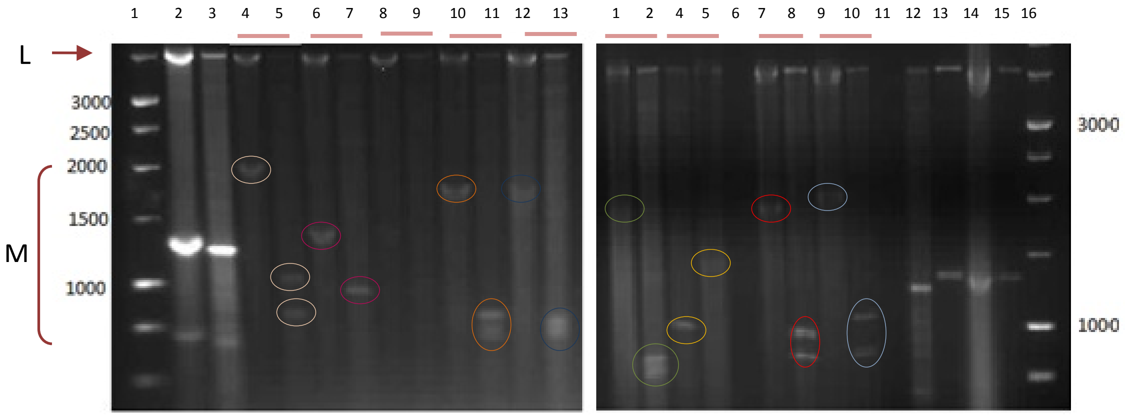


**Figure 3-2.** Various dsRNAs in *S. paradoxus* strains in Silwood Park. Lanes 2 and 3 on the left are a *S. cerevisiae* positive control without and with S1 nuclease digestion, respectively. Other pairs of samples, 4-13 on the left and 1-10 on the right image, represent different strains of *S. paradoxus* from Silwood Park, without and with S1 nuclease digestion. *S. paradoxus* strains in the left image from left to right are Y8.5, Q16.1, T21.4, Q14.4, Q62.5, and Q95.3; and in the right image, Q62.5, Q16.1 (Lane 4 is S1 treated and 5 without S1 treatment), T26.3, and Q74.4. All the M dsRNAs were broken with S1 nuclease. Based on the size of the dsRNA and S1 digestion pattern, there are at least eight different M dsRNAs in the Silwood Park strains. Lane 12 is a *S. cerevisiae* positive control, extracted by a RiboPure kit. Lanes 13, 14 and 15 are *S. cerevisiae* positive controls, which were extracted using the Fink and Frank method. Lanes 13 and 15 were treated with lithium chloride.

Compared to those from Silwood Park strains, the size of M dsRNA in the strains from Europe and the Far East is similar and does not show variation (Figure 3-3a). DBVPG4650 from Italy has three additional bands in the cell (Figure 3-3b). Two of them, which were smaller than M dsRNA, because of low copy number, were visible when the agarose gel was run for one hour. Running the gel for a longer time revealed an additional band that was integrated with M dsRNA in the previous running. However, the first two bands become unclear. It seems that there are two medium-sized dsRNAs in this strain. They may be different dsRNAs of similar size or a dsRNA that is similar to L or M, which has a deletion in the sequence similar to X dsRNA in *S. cerevisiae* (Esteban and Wickner, 1988).

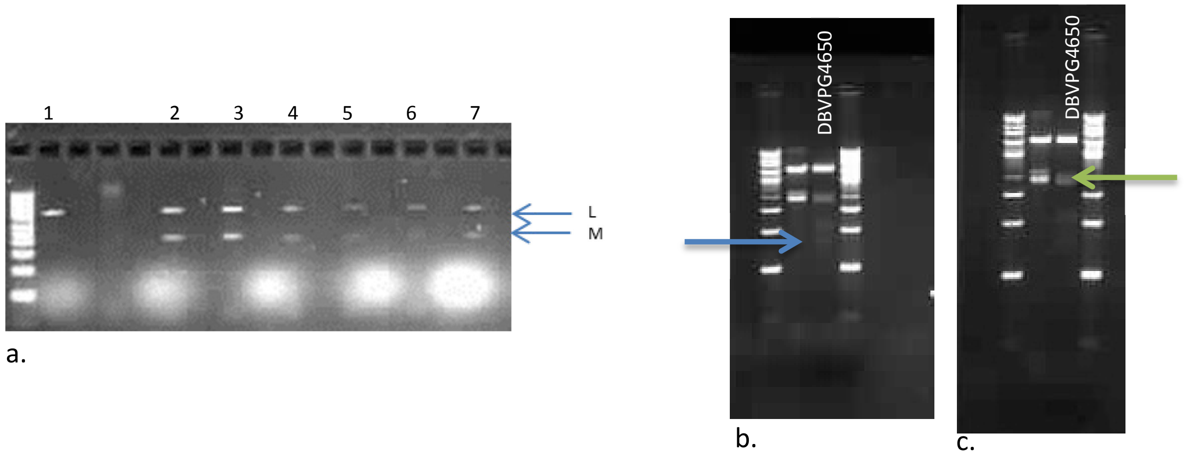


**Figure 3-3**. The gel electrophoresis of viral dsRNA in the continental Europe and the Far East *S. paradoxus* strains. (a) The extracted dsRNA from the continental Europe and Far East strains; Lane 1. CBS8439, Lane 2. N17, Lane 3. CBC8444, Lane 4. C02, Lane 5. C05, Lane 6. C10 and Lane 7. C07. Both LA and M dsRNA in all the continental Europe and the Far East strains have a similar size. (b) and (c) The third lane in both gels is DBVPG4650. The gel b was run for one hour whereas gel c is the same gel as gel b, which was run for two hours. In the gel b, two tiny bands with the size of around 1.5 kb are visible under the M dsRNA band (showed with the blue arrow). Running the gel for a longer time in the gel c made the smaller bands disappear and revealed an additional band that was integrated with M dsRNA in the gel b (showed with the green arrow).

In the infected cells, the L and M dsRNAs co-exist together and no L dsRNAs were found alone in this species. This result suggests that the presence of M dsRNA has some advantages for the L dsRNA, including increased chance of survival in competitive conditions. In addition to the strains which were previously reported as the infected strains, Q74.4, Y8.5, DBVPG4650, T21.4, and Q62.5, the dsRNAs were found in the three strains reported as the non-infected strains, Q59.1, Q95.3, and N17 (Chang et al., 2015, Pieczynska et al., 2013b).

**Table 3-1.** Summary of the results of killer assay, dsRNA extraction of *S. paradoxus* strains and S1 nuclease treatment of the dsRNA. ”K” is the abbreviation of killer and “-“ means it is not killer.

| **Strains** | **Location** | **Large dsRNA** | **Medium dsRNA** | **Digestion status of M dsRNA with S1 nuclease in S1 nuclease treatment** | **Killer phenotype** |
|---|---|---|---|---|---|
| **Q59.1** | Silwood Park | √ | √ | √ | K |
| **Q16.1** | Silwood Park | √ | √ | √ | - |
| **Y2.8** | Silwood Park | √ | √ | X | - |
| **Q43.5** | Silwood Park | √ | √ | √ | K |
| **Q74.4** | Silwood Park | √ | √ | √ | K |
| **T68.2** | Silwood Park | √ | √ | √ | K |
| **T68.3** | Silwood Park | √ | √ | √ | K |
| **Y8.5** | Silwood Park | √ | √ | √ | K |
| **Q14.4** | Silwood Park | √ | √ | √ | K |
| **Q62.5** | Silwood Park | √ | √ | √ | K |
| **T26.3** | Silwood Park | √ | √ | √ | - |

| Strains | Location | Large dsRNA | Medium dsRNA | Digestion status of M dsRNA with S1 nuclease in S1 nuclease treatment | Killer phenotype |
|---|---|---|---|---|---|
| **Q95.3** | Silwood Park | √ | √ | Partly digested | - |
| **Y1** | Silwood Park | √ | √ | Partly digested | - |
| **T8.1(C19)** | Silwood Park | √ | √ | Not tested | K |
| **T4b** | Silwood Park | √ | √ | Not tested | K |
| **T21.4** | Silwood Park | √ | √ | Not tested | K |
| **Y10** | Silwood Park | √ | √ | Not tested | K |
| **T76.2** | Silwood Park | X | X | X | - |
| **T18.2** | Silwood Park | X | X | X | - |
| **Q6.1** | Silwood Park | X | X | X | - |
| **Y7** | Silwood Park | X | X | X | - |
| **Y9.6** | Silwood Park | X | X | X | - |
| **Y6.5** | Silwood Park | X | X | X | - |
| **Y2.6** | Silwood Park | X | X | X | - |
| **S36.7** | Silwood Park | X | X | X | - |
| **W7** | Silwood Park | X | X | X | - |
| **T76.6 (C16)** | Silwood Park | X | X | X | - |
| **Z3 (C18)** | Silwood Park | X | X | X | - |
| **Y9.6** | Silwood Park | X | X | X | - |
| **Z1.1** | Silwood Park | X | X | X | - |
| **DBVPG4650 (C14)** | Continental Europe | √ | √ | Not tested | K |
| **OS20 (C02)** | Continental Europe | √ | √ | Not tested | K |
| **OS11,12W (C05)** | Continental Europe | √ | √ | Not tested | K |
| **4/2 02 (C10)** | Continental Europe | √ | √ | Not tested | K |
| **N17** | Continental Europe | √ | √ | Not tested | - |
| **CECT10176 (C15)** | Continental Europe | X | X | X | K |
| **N44** | Continental Europe | X | X | X | - |
| **KNP3828** | Continental Europe | X | X | X | - |
| **CBS5829** | Continental Europe | X | X | X | - |
| **YPS3** | Continental Europe | X | X | X | - |
| **YPS138** | Continental Europe | X | X | X | Not tested |
| **SIG1 (C13)** | Continental Europe | X | X | X | - |
| **STOC3** | Continental Europe | X | X | X | - |
| **OS5,6W (C03)** | Continental Europe | X | X | X | - |
| **O14-3,4W (C06)** | Continental Europe | X | X | X | - |
| **CBS8436 (C20)** | Far East Asia | X | X | X | - |

| Strains | Location | Large dsRNA | Medium dsRNA | Digestion status of M dsRNA with S1 nuclease in S1 nuclease treatment | Killer phenotype |
|---|---|---|---|---|---|
| **CBS8439 (C23)** | Far East Asia | √ | √ | Not tested | K |
| **CBS8441 (C25)** | Far East Asia | √ | √ | Not tested | K |
| **CBS8444 (C27)** | Far East Asia | √ | √ | Not tested | K |
| **CBS8437 (C21)** | Far East Asia | X | X | X | - |
| **CBS8438 (C22)** | Far East Asia | X | X | X | - |
| **CBS8440 (C24)** | Far East Asia | X | X | X | - |
| **CBS8442 (C26)** | Far East Asia | X | X | X | - |
| **A3** | Canada | X | X | X | K |
| **A8** | Canada | X | X | X | - |
| **A11** | Canada | X | X | X | - |
| **A12** | Canada | X | X | X | K |
| **A17** | Canada | X | X | X | K |
| **A19** | Canada | X | X | X | K |
| **A21** | Canada | X | X | X | K |
| **A22** | Canada | X | X | X | K |
| **A23** | Canada | X | X | X | K |
| **A24** | Canada | X | X | X | K |
| **A25** | Canada | X | X | X | K |
| **A27** | Canada | X | X | X | K |
| **A28** | Canada | X | X | X | K |
| **A29** | Canada | X | X | X | K |
| **A33** | **Canada** | **X** | **X** | **X** | **-** |

Based on the *S. paradoxus* samples collected from Silwood Park (UK), Europe not including UK, the Far East and Canada, it seemes that these dsRNAs are probably widespread throughout the world, except in Canada; 17 out of 30 (57%) from Silwood Park, five out of the 15 strains (33%) from the rest of Europe and three out of eight strains (37.5%) from the Far East are infected by viruses (Table 3-2). However, there is a possibility that the samples from Canada are not good representatives of the Canadian strains.

### 3.3.2 The relationship between the killer phenotype and viral infection

A comparison of the results of the killer assay against dsRNA extraction demonstrates that, although M dsRNA is present in most of the killer strains, there is one killer strain, CECT10176 from Europe, and 12 strains from Canada, which do not have any dsRNA (Table 3-1 and Table 3-2). Increasing the start volume of dsRNA extraction did not change the result. There is a possibility that the killer toxins in these strains are encoded by different genetic resources such as chromosomal genes, which were reported in *S. cerevisiae* (Goto et al., 1991, Goto et al., 1990b). As mentioned before, the killer phenotype in the Canadian strains was different from the other killers. The killer phenotype in these strains was very mild; only one or two layers of sensitive strain clones around the killer strain became blue (which was as a result of passing Methylene Blue from the membrane of the dead cells). In contrast to other killer strains from other geographical areas, the killer phenotype was expressed stronger when the sensitive strain was sprayed on the medium on the  same day as the killer strain was spotted, compared to when it was sprayed two days later. These data suggest that there is a possibility that the killer phenotype in the *S. paradoxus* strains from Canada evolved separately, and its genetic base is different from the strains from other regions. The killer phenotype in CECT10176, unlike that from the Canadian strains, is strong and expresses stronger when the sensitive strain is sprayed on the medium after two days of growing the killer strain. No killer phenotype was seen in Same-Day-Test.

There are five strains in Silwood Park, Q16.1, Q95.3, Y2.8, T26.3, and Y1, and one strain in Europe, N17 which contained dsRNA but did not show the killer phenotype. As mentioned before, the M dsRNA in Q95.5 and Q16.1 has the smallest size compared to the M dsRNA in reported in *S. cerevisiae* and *S. paradoxus* which has the minimum size of 1.5 Kb (Hopper et al., 1977b, Wickner et al., 1995, Naumov et al., 2004). Moreover, just part of the M dsRNA in Q95.3 was digested by S1 nuclease, and the size of the longer digested fragment in this strain is about 600 bp, which is shorter than the shortest toxin ORF; M-lus (729 bp) (Rodríguez-Cousiño et al., 2011) reported in these two species. These suggest that maybe there is a deletion in the M dsRNA of these strains. As a consequence, they do not have the full sequence to express the killer toxin. Changing the pattern of digestion in M dsRNA by S1 nuclease of Q95.5 and Y1, which were partly digested, and Y2.8, which was not digested, maybe arose from changing the structure of the dsRNA , which decreased the amount of the digestion or eliminated the digestable sequences. If the digestion happened as a result of melting the poly A in the middle of M dsRNA, decreasing the length or deletion of this part would reduce or prevent the digestion of the dsRNA. Since M dsRNA uses the expression system of yeast, the presence of poly A at the end of the mRNA of the killer toxin is essential for toxin expression. It is also possible that these dsRNAs are not M and they have just similar size to M dsRNA. However, in previous studies, many strains were reported to have had M dsRNA but, as a result of a mutation in the promoter of the killer toxin or in the host genes which are essential for the toxin expression, they cannot express the toxin (Wickner, 1974).

**Table 3-2.** The distribution of dsRNAs and killer phenotypes in *S. paradoxus* strains from four different regions. The columns show: the total number of strains, dsRNA$^+$; the number of strains containing dsRNA, dsRNA$^+$ K$^+$; the number of strains carrying dsRNA and showing the killer phenotype, dsRNA$^+$ K$^-$ the number of strains carrying dsRNA and not showing the killer phenotype and dsRNA$^-$ K$^+$ the number of strains that do not have dsRNA but show the killer phenotype.

| | Total No. strains | dsRNA$^+$ | dsRNA$^+$ K$^+$ | dsRNA$^+$ K$^-$ | dsRNA$^-$ K$^+$ |
|---|---|---|---|---|---|
| **Silwood Park (UK)** | 30 | 17 | 12 | 5 | 0 |
| **Europe without UK** | 15 | 5 | 4 | 1 | 1 |
| **Far East** | 8 | 4 | 3 | 0 | 0 |
| **Canada** | 15 | 0 | 0 | 0 | 12 |
| **Total strains** | 68 | 25 | 19 | 6 | 13 |

The rest of the strains containing dsRNA are killer. Interestingly, Q74.4, Q62.3, and Y8.5, which showed similar killer patterns against the *S. cerevisiae* killer strains, seem to have the same type of M dsRNA based on the size and S1 nuclease digestion. However, the results of the previous study, which was based on RT-PCR with specific primers for *S. cerevisiae* M dsRNA, indicated that Q74.4 carries M1, whereas Y8.5 has an unknown M dsRNA (Chang et al., 2015).

### 3.3.3 Cycloheximide treatment of the killer strains

In order to find out whether, in the killer strains, the killer phenotype expresses from the dsRNA or from the host genome, the killer strains were treated with cycloheximide. As reported previously, cycloheximide can remove M dsRNA from the infected cells completely from all the clones or from some of the cells that were treated. As a result, the killer toxin cannot be expressed if it was encoded by the M dsRNA (Carroll and Wickner, 1995, Fink and Styles, 1972).

In order to find the best concentration of cycloheximide in this treatment Q62.5 was cultured on YPD medium containing 2%, 6%, and 10% of cycloheximide and incubated at 30°C until the clones appeared. Since the strain did not grow at 6% and 10% of cycloheximide, the concentration of 2% was selected to continue for the rest of the killer strains.

The growth rate of various strains on the YPD media that contained 2% cycloheximide was different. All the Canadian strains were sensitive to 2% cycloheximide and, except for A12 and A24, the others did not grow on the medium. The clones of these two strains appeared after three weeks. However, after treatment with cycloheximide they became weak and could not grow very well on the MB agar. As a result, the killer assay test was not successful. The other strains could survive on 2% cycloheximide. The results of the treatment of the rest of the strains are summarised in Table 3-3 .

Table 3-3. The  cyclohexamide treatment of *S.paradoxus* killer strains. Both the Same-Day-Test (SDT) and Different-Day-Test (DDT)were performed to test the killer phenotype of the strains before and after the cycloheximide treatment.  The presence of  L and M dsRNA checked by dsRNA extraction and running on the gel. C1, C2, and C3 are the three clones from each cyclohexamide treated strain which their killer phenotype was tested.  Only CECT10176 strain, which does not carry any dsRNA, did not lose its killer phenotype, after cycloheximide treatment. In the rest of the killer strains, minimum one clone became non – Killer after cyclohexamide treatment. dsRNA extraction of the non – killer clones indicated that, M dsRNA is removed from all the non-killer clones. K, NK, PK and VPK are the abbreviations of Killer, Non-killer, Poor Killer and Very poor respectively. “+” and “-“indicate the presence and absent of the mentioned dsRNA in each column in each strain.

| Strains | Location | Original strain | | | | Strains after cyclohexamide treatment | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | L | M | Killer assay | | L* | M* | Killer assay | | | | | |
| | | | | SDT | DDT | | | SDT | | | DDT | | |
| | | | | | | | | C1 | C2 | C3 | C1 | C2 | C3 |
| Q59.1 | Silwood Park | + | + | NK | VPK | + | - | NK | NK | NK | NK | NK | NK |
| Q43.5 | Silwood Park | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| Q74.4 | Silwood Park | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| T68.2 | Silwood Park | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| T68.3 | Silwood Park | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| Y8.5 | Silwood Park | + | + | VPK | PK | + | - | VPK | NK | NK | NK | NK | NK |
| Q14.4 | Silwood Park | + | + | NK | PK | + | - | NK | NK | NK | NK | NK | NK |
| Q62.5 | Silwood Park | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| T8.1 (C19) | Silwood Park | + | + | K | K | + | - | NK | NK | NK | PK | PK | NK |
| T4b | Silwood Park | + | + | PK | K | + | - | NK | NK | NK | VPK | VPK | NK |
| T21.4 | Silwood Park | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| Y10 | Silwood Park | + | + | K | K | + | - | NK | VPK | VPK | NK | NK | VPK |
| DBVPG4650 (C14) | Continental Europe | + | + | VPK | K | + | - | NK | NK | NK | NK | NK | NK |
| OS20 (C02) | Continental Europe | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| OS11,12W (C05) | Continental Europe | + | + | K | K | + | - | NK | NK | NK | NK | NK | NK |
| 4/2 02 (C10) | Continental Europe | + | + | NK | K | + | - | NK | NK | NK | NK | NK | NK |
| CECT10176 (C15) | Continental Europe | - | - | PK | K | - | - | PK | PK | PK | K | K | K |
| CBS8439 (C23) | Far East Asia | + | + | PK | PK | + | - | - | - | - | - | - | - |
| CBS8441 (C25) | Far East Asia | + | + | PK | PK | + | - | - | - | - | - | VPK | VPK |
| CBS8444 (C27) | Far East Asia | + | + | PK | PK | + | - | - | - | - | - | - | - |

*the dsRNA extraction was done only on the clones that their phenotype changed from Killer to non-killer after cyclohexamide treatment.

Most of the strains that contained dsRNA, after the treatment did not show the killer phenotype in either of the clones, neither in SDT nor in DDT assay. Some of the clones in five strains, Y8.5, T8.1, T4b, Y10, and CBS 8441, were still killer, which could be as a result of not removing the dsRNA completely from all the cells, similar to Fink and Styles’ result (Fink and Styles, 1972). The gel electrophoresis of dsRNA extractions from all the clones that changed their phenotype from killer to non-killer confirmed that the M dsRNAs were eliminated from all the cells while the L dsRNAs still exist in the cell. These results indicate that in these strains the medium-sized dsRNA is M and it encodes the killer toxin in the cell. Moreover, these killer strains do not have any active genomic genes to express the killer toxin.

In the case of CECT10176 from continental Europe, which showed a strong killer phenotype without having the dsRNA, the phenotype of the cell did not change after treatment with cycloheximide. This result suggests that the genetic basis of the killer toxin might be in the genome. To find out if the same killer genes as *S. cerevisiae* exist in the genome of this strain and other strains from *S. paradoxus*, the sequence of the killer genes was analysed in the strains in which their genome sequence is available.

### 3.3.4 Analysis of the KHS and KHR sequences in *S. paradoxus* strains

In order to study the presence of KHS gene in *S. paradoxus* strains, the original sequence of the KHS gene (Goto et al., 1991) and its hypothetical ORF sequence in the other *S. cerevisiae* strains, SCY-1690 (Frank and Wolfe, 2009),  were aligned with the genome sequences of 35 *S. paradoxus* strains. The result indicated that the sequences exist in all the *S. paradoxus* strains. In the alignment, as Frank and Wolf reported, there is an inversion in the original KHS sequence (with the length of 1.4 bp) this inversion ends at a restriction site which was used in the cloning (Frank and Wolfe, 2009).

SCY_1690 ORF covers about half of the KHS ORF (Figure 3-4). The inverted site in KHS sequence is located after the SCY_1690 ORF.  In the sequence of the *S. paradoxus* strains, about one-quarter of residues of SCY_1690 and one-third of residues of KHS ORF are ambiguity nucleotides ("N" and "?") , which might the result of being close to the telomeres. There is no ambiguity nucleotide at the 5' end of the SCY_1690, in any of the strains. The sequence of this part of SCY_1690 is very close that to *S. paradoxus* strains (Frank and Wolfe, 2009), especially the strains from Silwood Park and continental Europe, 99%-100% identity, which showed more similarity than the Far East strains, 88% identity. The sequence of this ORF in CECT10176 (C15) is identical to Q95.3, Y7, T76.6, Q31.4, Y8.5, SIG1 and Q59.1 strains. Q95.3, Y1, Q76.6, Q31.4 and SIG1 are not killers, and the killer phenotype in Y8.5 was cured after treating with cycloheximide. As a result, it does not seem that the killer phenotype in these species arose from this ORF unless there is some mutation in the controlling region of the other strains.

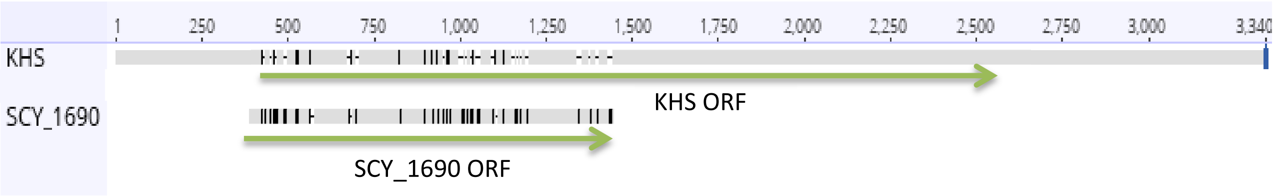


**Figure 3-4.** The alignment of KHS and SCY_1690 sequences. The SCY_1690 ORF covers about half of the KHS ORF sequence.

The identity of the KHS to *S. paradoxus* strains sequence is less than that of the SCY_1690 (81.5% to 84,5%). It does not seem that this sequence encodes a killer toxin because lots of stop codons have formed in the sequence of the strains as a result of the variation that exists in these sequences. In CECT10176 (C15) strains there are 18 stop codons in the sequence of this part of the ORF. Therefore, even if the sequence of the ORF is right, this part cannot be the encoding sequence of the killer toxin in this strain.

There is no sequence similar to KHR in the *S. paradoxus* strains. The results suggest that in CECT10176 (C15) an unknown gene encodes the killer toxin. This gene may be on the chromosomal DNA or on a plasmid DNA.

## 3.4 Conclusion

Viral dsRNAs were found in the 26 strains of the *S. paradoxus* strains. They were found in all of the strains from all around the world except Canada. L and M dsRNA co-exist in all the strains. The size of L dsRNA is similar in all the strains, whereas that of M dsRNA varies in different strains. Also, most of the M dsRNAs in Silwood Park strains were digested with S1 nuclease. Just two strains were digested partly and one of them was not digested. All of the killer strains were infected by the viral dsRNA except CECT10176 (C15) and the killer strains from Canada. The killer phenotype in Canadian strains was different from the other strains. It was significantly weaker than most of the killer strains and against the other strains expressed better in the Same-Day-Test.

Killer phenotype was not expressed in six strains Q95.3, Y2.8, Y1, T26.3, Q16.1, and N17 that contained M dsRNA. In the first three strains the dsRNAs were not digested with S1 nuclease very well.

The killer phenotype was cured in all killer strains which were infected by the viral dsRNA. In CECT10176 (C15) the killer phenotype was still stable in the strain after the treatment. A similar sequence to 1.2 kb at the end of the KHS ORF was found in all the killer strains of which their genome is available including CECT10176 (C15). However, there are lots of stop codons in the *S. paradoxus* strains. No KHS sequences were found in the *S. paradoxus* strains. It seems that an unknown chromosomal gene or DNA plasmid gene in CECT10176 (C15) produces the killer toxin. The

Canadian strains were sensitive to cycloheximide and did not grow. The genome sequence of these strains was not available.

In summary, the results suggest that the killer phenotype in all the killer *S. paradoxus* strains except CECT10176 (C15) from Continental Europe and the Canadian strains is encoded by M dsRNA. The genetic basis of the killer phenotype of CECT10176 (C15) and the Canadian strains may be chromosomal genes or a DNA plasmid.

**Chapter four**

# 4 Finding dsRNA viruses in *S. paradoxus* through sequencing

## 4.1 Introduction

Two types of viral dsRNA, a large-size (approximately 4,000 bp) and a medium-size dsRNA (1,000-2,500 bp), have been found in *S. paradoxus* in this study, using gel electrophoresis (Table 3-1). The size of the large band is similar to L dsRNA and the size of the medium band is similar to M dsRNA in *S. cerevisiae*.

In *S. cerevisiae*, which is close a relative of *S. paradoxus,* there are four types of dsRNA: L dsRNA, M dsRNA, T dsRNA and W dsRNA. Two types of L dsRNA has been reported; L-A and L-BC. In killer cells, L-A co-exists with M dsRNA and, by producing the capsid proteins and RDRP, it encapsidates and replicates itself and acts as a helper to replicate M dsRNA. There are four types of L-A in *S. cerevisiae*: L-A-L1, L-A-2, L-A-lus and L28. The sequence of all L-A dsRNAs is available, except L28. The sequence identity between the three known L-A dsRNAs is between 73% and 75%. The genomic organisation of L-A is composed of two overlapped ORFs of the *gag* and *pol*. A conserved sequence in the overlapping area of L-A-L1 and L-A-lus, with 100% identity, programmes a -1 ribosomal frameshift event in this region to synthesise the Gag-Pol fusion protein. About 1.8% of the ribosomes, which translate Gag, shift to the -1 frame at this site and continue translation to make the Gag-Pol fusion protein (Dinman et al., 1991, Wickner et al., 1995, Rodríguez-Cousiño et al., 2013).

In the (+) strand of L-A-L1 dsRNA, two *cis* signals have been demonstrated to be essential in starting replication. One is 30 nucleotide at the 3’ terminal containing an essential stem-loop structure and the other *cis* site is 400 nucleotide away from 3’ end called Internal Replication Enhancer (IRE) (Figure 1-2) (Esteban et al., 1989). The IRE overlaps with the (+) ssRNA binding site in the empty particle. Probably, binding the replicas to IRE increase the chance of the 3’ end site seen by replicase (Fujimura and Wickner, 1992). There is no experimental evidence for the presence of a signal at the 5’ end of L-A-L1. However, in X dsRNA, which is a deletion mutant of L-A-L1, the 25 nt at the 5’ is identical to the 5’ of L-A-L1. This dsRNA is maintained stable by L-A-L1. It appears that the *cis* signal for transcription exists in this 25 nt. Similar to other dsRNA viruses, this region is AU-rich, which seems to facilitate the melting of the dsRNA for conservative transcription (Rodríguez-Cousiño et al., 2013).

At the protein level, a comparison between the predicted Gag and Gag-Pol proteins of L-A-lus and that of L-A-L1 indicates that the identity rises to 86%. The identity increased above 95% in the four conserved motifs in RdRps (Rodríguez-Cousiño et al., 2013).

Three subtypes of L-A-lus have been found in nature. The conservation between them with respect to nucleotide level is between 83% and 85%, which is higher than the conservation between different types of L-A. The encoded proteins of these subtypes are almost identical at 97% - 98%. The distribution of L-A-lus and L-A-2 in nature is greater than those of L-A-L1 and they are shown to be more stable inside the cell in difficult situations, e.g. when exposed to high temperatures (Rodríguez-Cousiño et al., 2013).

There are four well-characterised M dsRNAs in *S. cerevisiae*: M1, M2, M28 and M-lus. Although the identity between different types of M is very low, the genome organisation is similar in all M dsRNAs. The genome starts with a 5’ terminal coding region, followed by a poly A sequence. After the poly A sequence, there is a 3’ non-coding region. Although the preprotoxins have no sequence relationship to each other or to other known proteins, they share conserved patterns of potential processing sites. In terms of the relationship between L-A dsRNA and M dsRNA during evolution, Rodríguez-Cousiño suggests that each population of L-A virus has evolved specifically with distinct types of satellite RNA (Rodríguez-Cousiño et al., 2011).

L-BC dsRNA is a 4.5 kb linear dsRNA, unrelated to L-A. It is encapsidated in particles whose major protein is different from that of L-A and M VLPs. However, it has the same genome organisation and mode of expression as L-A and can coexist with it in the same cell. The copy number of this particle is substantially lower than L-A (Sommer and Wickner, 1982). It has no helper activity and does not produce any phenotype in its host (PARK et al., 1996). It shows a lesser degree of variation compared to L-A (Rodríguez-Cousiño et al., 2013).

T and W dsRNAs, with the approximate molecular sizes of 2.7 kb and 2.25 kb respectively, are both cytoplasmically inherited and were not encapsidated into viral particles. Similar to L dsRNA, they do not produce any phenotype in the infected cells. They have their own RDRP. On average, the proteins of these two dsRNAs share 22.5% of identical amino acids, but there are several regions with highly conserved sequences in a region that includes the consensus sequences for viral RNA-dependent RNA polymerase, suggesting that T and W are evolutionarily related (Esteban et al., 1992b).

Although several dsRNAs are identified in *S. paradoxus* with the size of L dsRNA and M dsRNA (Naumov et al., 2004, Pieczynska et al., 2013b), recently just two types of M dsRNA, M1 and M28, have been sequenced in *S. paradoxus*. These dsRNA sequences are very close to M1 and M28 in *S. cerevisiae.* M1, which was found in strains Q74.4 and N-45, has just one nucleotide difference, in the β subunit of preprotoxin, from M1 of *S. cerevisiae.* However, Q74.4 has the potential to kill the K1 strains of *S. cerevisiae*. Variation in the sequence of M28 between the two species, *S. cerevisiae* and *S. paradoxus* is higher than that of M1, 90% identical. This M dsRNA was found in DBVPG4650 and T21.4. The killer phenotype of these strains is similar to that of *S. cerevisiae* (Chang et al., 2015).

Several methods of sequencing dsRNA have been developed so far. The three currently mostly used methods are: single-primer amplification sequence-independent dsRNA technique (SPAT) (Potgieter et al., 2002, Lambden et al., 1992, Attoui et al., 2000); Full-length Amplification of cDNA (FLAC) (Maan et al., 2007); and the method that was published by Potgieter in 2009 (Potgieter et al., 2009).

In this study, firstly four strains, Q62.5, Y8.6, T26.4 and T4b, were sent to the Pirbright Institute lab, where the FLAC method was invented, to see whether sequencing is possible using FLAC. However, it was not successful. As a result, instead of going through the whole process with the two other methods, we decided to sequence the dsRNA directly using NGS. Among the various commercial methods available, MiSeq 300 had the longest reads. Therefore, it was selected for sequencing the dsRNA detected in the *S. paradoxus* strains. In this chapter, we report viral dsRNA sequences from 27 strains of *S. paradoxus* from Silwood Park, Europe and the Far East.

## 4.2 Methodology

### 4.2.1 Strains and cDNA libraries

The dsRNA of the 27 strains that had previously been identified as having viral dsRNA (Table 3-1) was purified and sent to GATC Biotech to sequence. Since the library preparation of each strain is very expensive, three strains from the three different regions (Q62.5 from Silwood Park, DBVPG4650 from Continental Europe and CBS8441 from the Far East), containing both dsRNA and showing killer phenotypes, were selected to be sequenced separately. In addition, two mixtures of the dsRNA extractions, Pool1 and Pool2, were prepared. Pool1 included 16 strains from Silwood Park and Pool2 four strains from Continental Europe and three from the Far East (Appendix; Table 8-3). To have a control for analysing the sequence of the strains in the Pools, Q62.5 was added to Pool1. The library preparation of Q62.5, DBVPG4650 and Pool2, performed by GATC, was successful. However, the two other samples failed several times. As a result, the cDNA of the dsRNA was prepared and sent to the company to facilitate their work.

### 4.2.2 dsRNA extraction and purification

Since the amount of extracted dsRNA in the method described by Freid and Fink was higher than the other methods (Chapter 3; section 3.2.1), we firstly tried to purify the dsRNA using this method, followed by a lithium chloride treatment to remove the ssRNA, as explained below, and a DNase treatment to remove the DNA genomic contamination. The high amount of digested DNA in the extraction made it impossible to concentrate the dsRNA. The purity of dsRNA was also not as high as the dsRNA extracted by kits.

In the extraction using the kits, since the RNA is purified by binding to a column, the amount of DNA contamination is significantly less than other methods. Of all the kits, RiboPure™ Yeast Kit (Amicon) extracted the higher amount of dsRNA. The extracted RNAs were treated with lithium chloride and DNAse. After that, to remove the DNase and digested nucleic acids, it was cleaned up using RNeasy MinElute or MidiElute Cleanup Kits (Qiagen). Then, the purity of dsRNA was measured by NanoDrop and the amount of the remaining DNA and digested nucleic acid were studied using 1% agarose gel electrophoresis. If the DNA had not been completely digested, the DNase treatment and RNeasy Cleanup were repeated. If the digested nucleic acid band was still visible on the gel, the clean-up stage was performed again.

In dsRNA extraction using the RiboPure™ Yeast Kit, the amount of samples used in one extraction was increased to five times higher than the amount suggested by the kit instructions. Consequently, the amount of elution buffer used to elute the RNA from the column was doubled. Increasing the amount of the sample increased the amount of both dsRNA and DNA contamination. However, the ratio of 260/280 measured by NanoDrop did not go over the range.

In the lithium chloride treatment, for 500 µl of the extracted RNA, the same amount of 4M lithium chloride (Sigma) was added and the samples were incubated at 4°C overnight. After 30 minutes centrifugation at 16,000 x *g*, the supernatant was transferred into a new tube and 200 µl isopropanol (Sigma) and 50 µl of 7.5 M ammonium acetate were added. The samples were incubated at -20°C for two hours. They were then centrifuged for 10 minutes at 16,000 x *g* and their precipitates were washed with 70% ethanol. After drying, they were dissolved in RNeasy free water.

The result of running the lithium-chloride-treated samples indicated that, in addition to DNA, there are still some ssRNA in the samples. Since there is a possibility of digesting some parts of the dsRNA that have a secondary structure with S1 nuclease, it was not used to remove the ssRNA.

We tried three different types of DNase I: DNase I (Sigma), DNase I Amplification Grade (Sigma) and DNase I provided by the RiboPure kit  (Figure 4-1). The DNase I from Sigma has 6% RNase activity. The result of running the gel showed that this RNase activity is enough to remove the remaining ssRNA. As a result, this DNase was selected. The result of the next generation sequencing of the three samples that were first sequenced, Q62.5, DBVPG4650 and Pool2, showed that, despite not seeing any DNA band after treating with DNase in the gel, there was still DNA contamination in these samples. As a result, the two other samples, CBS8441 and Pool1, which were prepared later, were treated with TURBO DNase (Ambion) after treating with DNase I from Sigma. Companies' instructions were followed in all the enzyme treatments.

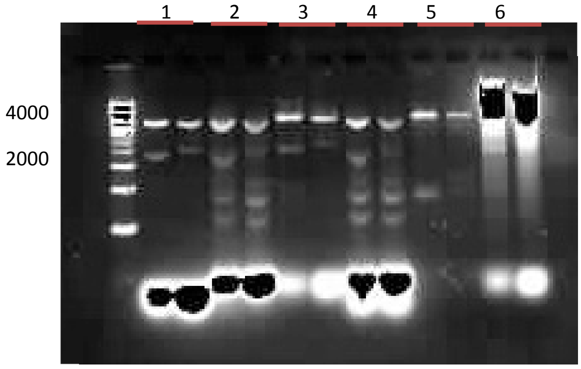


**Figure 4-1**. DNase and S1 Nuclease treatment. The first lane from left is 1 kb DNA ladder; after the ladder every two lanes are Q62.5 and Y8.5 treated with different DNase and S1 Nuclease. The treatment of each two lanes is: 1) DNase I (Sigma); 2) DNase I Amplification Grade (Sigma); 3) DNase I Amplification Grade (Sigma) and S1 Nuclease (Promega) (S1 nuclease incubated at 37°C); 4) RiboPure DNase I; 5) DNase I Amplification Grade (Sigma) and S1 Nuclease (Promega) (S1 nuclease incubated at 67°C); and 6) S1 Nuclease (Promega) (incubated at 37°C). A comparison between the treatments indicated that the first treatment prepares the cleanest dsRNA. DNase I from Sigma has 6% RNase activity. It appears that this RNase activity is enough to remove the remaining ssRNA. The third treatment, DNase I Amplification Grade and S1 Nuclease, showed the best result after the first treatment. In this treatment, although the S1 Nuclease did not cut the dsRNA at 37°C, since viral dsRNAs usually have a secondary structure, digestion of some parts is still possible. S1 Nuclease at 67°C, in the fifth treatment, digested the dsRNA. In the second and fourth treatments, in which there is no S1 nuclease and the DNases have no ribonuclease activity, the ssRNA was not removed. In the last treatment, which has just S1 Nuclease, there is DNA contamination.

### 4.2.3 cDNA preparation

Two methods were tested to prepare the cDNA of CBS8441 and Pool1. In the first method, described by Froussard (Froussard, 1992), the first and second strand cDNA were made using the 26 nt forward primers with six random nucleotides at the 3’ end. Then a PCR was done using the reverse primer, which is complementary to the 20 nt beginning of the forward primer. The methods were used with some modifications. In order to find out the effect of DMSO on the denaturation of the dsRNA, the dsRNA was denatured for 5 minutes at 99°C, either with or without 15% DMSO (Sigma) and immediately snap frozen on -80°C ethanol. The first strand DNA was made using Reverse Transcription System (Promega) or Super Script III First-Strand Synthesis System (Invitrogen), and the second strand by DNA Polymerase I, Large (Klenow) Fragment (BioLab) using the primers mentioned in the paper. The primers were then washed with QIAquick PCR Purification Kit (Qiagen). The amplification of the randomly synthesized double-strand cDNA was performed using Platinum *Pfx* DNA Polymerase (Invitrogen) or OneTaq Polymerase (Promega). The samples were denaturated at 94°C for 2 minutes then subjected to 40 cycles of amplification: 94°C – 30 seconds, 60°C – 30 seconds, and 72°C – 4 minutes and 30 seconds. The PCR product was run on 1% agarose gel.

The results of the gel electrophoresis showed that the product smears started from either 10 kb or the wells, which were at least twice the size of the longest dsRNA. To optimise the cDNA preparation, many changes were made. Treating the samples with Turbo DNase, adding DMSO in the denaturing stage, decreasing the amount of double-strand cDNA at the PCR reaction, using different kits in the first-strand synthesis and polymerases in the PCR reaction, and changing the time and temperature of PCR reaction, did not improve the performance.

In the second method, the dsRNA with random hexamer primers and dNTP were denatured at 99°C for 5 minutes and snap frozen on -80°C ethanol. The first strand was synthesized using Super Script III First-Strand Synthesis System (Invitrogen) and the second strand was made using NEBNext® mRNA Second Strand Synthesis Module (BioLab). The cDNA were washed with the QIAquick PCR Purification Kit (Qiagen). Firstly, the cDNA were run on the 3% agarose gel. Given the low efficiency of the method, the smear of the product was not clear. There were two bands on the gel the same size as the dsRNAs and a background smear started from the longer band (Figure 4-2). To be sure of the result, the PCR product was treated with DNase. The smear decreased with the DNase treatment. To guarantee production of the cDNA, the dsDNA was measured by Qubit® dsDNA HS Assay Kit before starting and after finishing the reactions. As there was not a Qubit® Fluorometer in the lab, four dsRNA standards, 10 ng/µl, 5 ng/µl, 2.5 ng/µl and 0 ng/µl, were prepared and, using a standard graph, the concentration of the dsDNA was measured using a normal fluorometer.

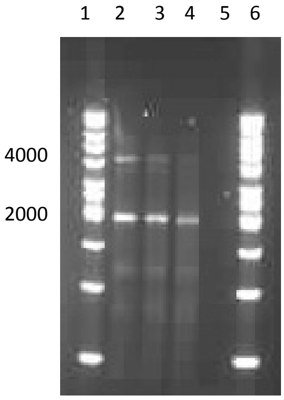


**Figure 4-2.** CDNA preparation using the second method. Lane 1:  1 kb DNA ladder; lane 2: cDNA1, dNTP and primer were added in the denature stage; lane 3: cDNA2, dNTP and primer were added in the master mix; lane 4: cDNA3 dNTP and primer were added to the master mix and the dsRNA were denatured in 15% of DMSO; lane 5: negative control. There were two bands in all of the reactions, except negative control, with the same size as the dsRNAs and a background smear started from the longer band. It seems that the bands are the dsRNAs and the smear is the cDNA.

Even though the cDNA produced in the second method was significantly less than the first method, the PCR artifact seemed to be significantly less than the first one. As a consequence, the cDNA were prepared with this method using high concentrations of dsRNA at the start of the reaction.

### 4.2.4 Bioinformatics analyses

The NGS data was analysed using Geneious software. FastQC was also used to measure the quality of the data. To identify the sequences of the viral dsRNAs, both map-to-reference and *de-novo* assemblies were used. The known dsRNA sequences in *S. cerevisiae* and two reported M dsRNAs in *S. paradoxus* (Table 4-2) were used as references in map-to-reference assemblies. The preliminary analyses of the data and the viral dsRNA assemblies are explained in the Appendix; section 8.2.1.

The prediction of the killer toxin ORF, translation of the preprotoxin, the signal cleavage sites and the Kex2p/Kex1p cleavage sites were also performed in the Geneious software. Since the cleavage sites of Kex2p/Kex1p are KR and RR this sequences were simply searched in the amino acid sequences of the preprotoxins. Only the N-glycosylation sites were predicted in the NetNGlyc website (http://www.cbs.dtu.dk/services/NetOGlyc/) from Technical university of Denmark. The threshold of the program for N-Glycosylation was 0.5. As one of the N-Glycosylation sites has potential of 0.4, in K28 in *S. cerevisiae,* which has a well-known structure, all potential sites were considered. However, those that have potential less than 0.5 were mentioned.

## 4.3 Result and discussion

### 4.3.1 Viral dsRNAs found in *S. paradoxus* strains

The viral dsRNAs found in *S. paradoxus* are listed in Table 4-2: 16 full sequences of L-A dsRNA (L-A-Q, L-A-D1, L-A-C, L-A-P1.1, L-A-P1.2, L-A-P1.3, L-A-P1.4, L-A-P1.5, L-A-P1.6, L-A-P1.7, L-A-P2.1, L-A-P2.2, L-A-P2.3, L-A-P2.4, L-A-P2.5 and L-A-P2.6); one sequence of only the *gag* gene of L-A, L-A-P1g1; one sequence of only *pol*, L-A-P1p2; seven types of M dsRNA composed of two full sequences of M dsRNA (MQ and MC); five sequence of killer toxin genes (M-P1G1, M-P1G2, M-P1G3-1, M-P1G3-2, M-P1G5 and M-P1SG); and three sequences of the non-coding part of M dsRNA (M-P1NC1, M-P1NC2 and M-P1NC3). The gag and *pol* sequences (L-A-P1g1 and L-A-P1p2) were found in Pool1. As this sample is a mixture of 17 different strains and the sequence between these two genes is highly similar to different types of L-A dsRNA, it was not possible to ascertain whether they are from one L dsRNA or if they are satellites of an L dsRNA with a deletion in the sequence of an L-A dsRNA, similar to X dsRNA in *S. cerevisiae* (Esteban and Wickner, 1988). The sequences of the coding and non-coding parts of M dsRNA in Pool1 were also sequenced separately. This was as a result of the poly A in the middle of M dsRNA.

Both map-to-reference and *de-novo* assemblies were used to find the sequences of L dsRNAs. In Q62.5 and CBC8441, which contained only one type of L dsRNA, the consensus sequences of both assemblies were identical, whereas, in Pool1 and Pool2, which were a mixture of the strains, as well as DBVPG4650, which had two L-As, their consensus sequences were different. This arises from high similarity between different types of L-A dsRNA in *S. paradoxus* (73.5% – 96.6% identity). The normal setting of map-to-reference and *de-novo* assembly was not able to categorise the reads of each type of dsRNA. As a result, the sequences of the Ls were found using the metagenomics settings optimised for the LQ, which was as the control in Pool1. In terms of M dsRNA, only M28 was sequenced using both assemblies. Since the reference sequences of the other Ms were not available, their sequences were found using the *de-novo* assembly. Because the similarity between M dsRNAs is low (15% – 35%), the normal *de-novo* setting was able to assemble their sequences. However, the accuracy of the sequences of Ms in Pool1 was also checked with the metagenomics settings. In Pool2, as a result of the high amount of the host genome and low amount of the M dsRNA reads, the *de-novo* assembly was not successful. As a consequence, the Pool2 reads were mapped to all Ms of *S. cerevisiae* and the Ms found in other libraries using normal and metagenomics settings. Since the software cannot find the accurate number of poly A, 50 residues were considered for this part.

In two of the single sequenced strains, Q62.5 and CBS8441, which had two dsRNA bands on the gel electrophoresis (Table 4-1), the large-size and the medium-size bands, the size of the bands was comparable with the L-A and the M dsRNA found. The predicted molecular weight of the killer toxin of Q62.5 and its isoelectric point is also similar to those measured in the lab. The predicted molecular weight is 30.422 and the isoelectric point is 4.72. It seems that MQ is the sequence of the M in the Q62.5 that encodes the new killer toxin in this strain, which was stronger than the other killer toxins.

Table 4-1. The results of killer assay, the gel electrophoresis of dsRNA extraction and dsRNA sequencing of Q62.5, DBVPG4650 and CBS8441.

| Strain | Killer phenotype | dsRNA bands on the gel | | | Sequenced dsRNA | | | |
|---|---|---|---|---|---|---|---|---|
| | | Number of the large bands (~4.5 kb) | Number of the medium band ( ~2 kb) | Number of the small bands (~0.8 kb) | L-A | | M | |
| | | | | | Name | Size (bp) | Name | Size (bp) |
| Q62.5 | K | 1 | 1 | - | L-A-Q | 4,580 | MQ | 1933 |
| DBPVG4650 | K | 1 | 2 | 2 | L-A-D1<br>L-A-L1<br>L-BC | 4,580<br>4,580<br>4615 | M28 | 2012 |
| CBS8441 | K | 1 | 1 | - | L-A-C | 4580 | MC | 1973 |

Finding the dsRNA of each band in the gel electrophoresis of DBVPG4650 was more complicated compared with the two other strains. Five dsRNA bands were detected in the gel electrophoresis of this strain (Figure 3-3b and Table 4-1). One dsRNA band with the same size as L, two bands in the range of M dsRNA, and two smaller than M dsRNA were detected in this strain. Three types of L, L-A-L1, L-A-D and L-BC, and one type of M, M28, were found in this strain. The size of the full sequence of three Ls and the M is similar to the large- and medium-size bands on the gel. However, in the mapping alignment of the *de-novo* contigs to L-A-D1, there are several contigs identical to the reference with lengths between 502 bp and 1,363 bp. This was not true in the other two single strains. The presence of the different size L-A-D1 contigs in each *de-novo* assembly suggests that there are various L-A-D1 with different sizes in this strain. This might result from the deletion in the original sequence of L-A-D1, similar to X dsRNA in *S. cerevisiae*, which is derivative of L-A-L1 (Sommer and Wickner, 1982).

Table 4-2. The viral dsRNAs reported in *S. cerevisiae* (Icho and Wickner, 1989b, Rodríguez-Cousiño et al., 2013, Russell et al., 1997, Meskauskas, 1990, SCHMITT and TIPPER, 1995, Rodríguez-Cousiño et al., 2011, PARK et al., 1996, Esteban et al., 1992b, Rodriguez-Cousiño et al., 1991) and *S. paradoxus* (Chang et al., 2015), and those found in *S. paradoxus* in this study. (C) is the abbreviation of Coding and (NC) is the abbreviation of Non-Coding parts of M. In terms of L, all the L-A found in this study contain full sequences of L-A except L-A-P1-g1 and L-A-P1-p1, which contained only *gag* and *pol* genes.

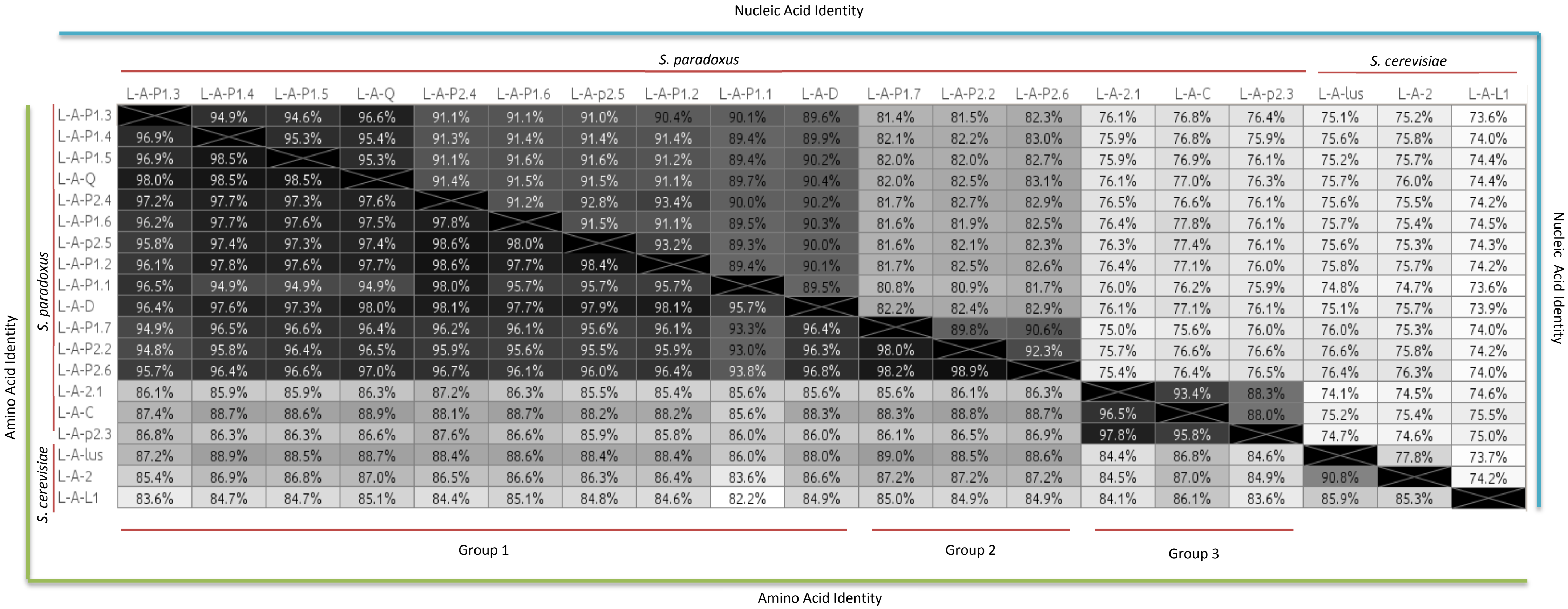
| Type of dsRNA | | dsRNA reported in previous studies in *S.cerevisiae* | dsRNA reported in previous studies in *S.paradoxus* | Length (kb) | dsRNA in *S. paradoxus* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Q62.5 | | DBVPG4650 | | CBS8441 | | Pool1<br>(T4b, T8.1,T21.4, T26.3, T68.2, T68.3, Q14.4, Q16.1, Q43.5, Q59.1, Q74.4, Q95.3, Y1, Y2.8, Y8.5, Y10, Q62.6) | | Pool2<br>(OS20, OS11.12W. OS3.4Wa,4/25W2, N17, CBS8439, CBS8444) | |
| | | | | | dsRNA | Assembly | dsRNA | Assembly | dsRNA | Assembly | dsRNA | Assembly | dsRNA | Assembly |
| L | L-A | L-A-L1<br>L-A-2<br>L-A-lus<br>L28 | - | 4.5 | L-A-Q | Map-to-reference and *de-novo* assembly with Normal setting | L-A-D1<br>L-A-L1 | Map-to-reference and *de-novo* assembly with metagenomics settings | L-A-C | Map-to-reference and *de-novo* assembly with Normal setting | L-A-P1.1<br>L-A-P1.2<br>L-A-P1.3<br>L-A-P1.4<br>L-A-P1.5<br>L-A-P1.6<br>L-A-P1.7<br>L-A-P1-g4(gag)<br>L-A-P1-p1(pol) | *De-novo* assembly with metagenomics settings | L-A-P2.1<br>L-A-P2.2<br>L-A-P2.3<br>L-A-P2.4<br>L-A-P2.5<br>L-A-P2.6 | *De-novo* assembly with metagenomics settings |
| | L-B | L-BC | - | 4.6 | - | - | L-BC | Map-to-reference and *de-novo* assembly with Normal setting | - | - | - | - | - | - |
| M | | M1<br>M2<br>M28<br>M-Lus | M1<br>M28 | 1.0 - 2.5 | MQ | *De-novo* assembly with Normal setting | M28 | *De-novo* assembly with Normal setting | MC | *De-novo* assembly w *de-novo* with Normal setting | M-P1G1(C)<br>M-P1G2(C)<br>M-P1G3-1(C)<br>M-P1G3-2(C)<br>M-P1G5(C)<br>M-P1SG(C) | *De-novo* assembly with Normal and metagenomics settings | M28<br>MC(F)<br>MQ<br>M-P1G2 | Map-to-reference assembly with Normal and metagenomics settings |
| | | | | | M28 | Map-to-reference and *de-novo* assembly with Normal setting | | | | | M-P1NC1 (NC)<br>M-P1NC2 (NC)<br>M-P1NC3 (NC) | | | |
| W | | W | | | - | - | - | - | - | - | - | - | - | - |
| T | | T | | | - | - | - | - | - | - | - | - | - | - |

### 4.3.2 L dsRNA

**a) Nucleotide sequence**

Of all three strains that were sequenced separately, Q62.6, DBVPG4650 and CBS8441, only one strain, DBVPG4650, had more than one L dsRNA: L-A-D1, L-A-L1 and L-BC. Although the number of reads in L-A-L1 and L-BC is significantly less than L-A-D1 (1,253 and 406 reads respectively), they were assembled into both Ls quite specifically, and running the map-to-reference using more restricted settings did not change the result. Since none of the strains in this study contained L-A-L1 or L-BC, it would be unlikely that their reads come from contamination. The presence of low copy numbers of L-BC dsRNA with L-A has been reported previously (Sommer and Wickner, 1982). However, there are no reports with respect to the presence of two types of L-A in one strain found in nature. In addition, the tests that were done to put two types of L-A dsRNA, L-A-L1 and L-A-lus, in one cell show that L-A-lus excludes L-A-L1 from the cell (Rodríguez-Cousiño et al., 2013). There is the possibility that, instead of excluding, they affect the copy number of each other. Primarily, L-A-L1 is a weaker viral dsRNA in nature when compared to the other types of L-A dsRNA (Rodríguez-Cousiño et al., 2013). Both L-A-L1 and L-BC have been reported in *S. cerevisiae* and were just seen in this strain of S. *paradoxus*. The L-A-L1 is identical with the L-A-L1 of *S. cerevisiae*, whereas L-BC has 89% identity with the same L in *S. cerevisiae*. In addition to these Ls, the M dsRNA in this strain is the only M, M28, in *S. paradoxus* that was reported in *S. cerevisiae*. The identity of this dsRNA to the M28 of *S. cerevisiae* is 87.3%. Given that in *S. cerevisiae* each type of L evolves with specific M (Rodríguez-Cousiño and Esteban, 2017, Rodríguez-Cousiño et al., 2013), there is a possibility that L-A-D1 is the L28 in *S. cerevisiae*, which its sequence has not yet reported. The identity between this dsRNA and the L-As in *S. paradoxus* is higher than that of *S. cerevisiae* (Table 4-3). This result supports the theory of the evolution of viral dsRNA starting in *S. paradoxus* and then transferring into *S. cerevisiae* (Chang et al., 2015).

**Table 4-3.** The nucleic acid and amino identities between the different types of L-A dsRNA in *S. cerevisiae* and *S. paradoxus*. The numbers on the top that are marked with the blue lines are nucleic acid identities and the numbers at the bottom marked with green lines are amino acid identities.  Based on the identities between different L-A and their phylogeny tree, there are three groups of dsRNA in the *S. paradoxus* strains.



Nucleic Acid Identity

| | | *S. paradoxus* | | | | | | | | | | | | | | | | *S. cerevisiae* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | L-A-P1.3 | L-A-P1.4 | L-A-P1.5 | L-A-Q | L-A-P2.4 | L-A-P1.6 | L-A-p2.5 | L-A-P1.2 | L-A-P1.1 | L-A-D | L-A-P1.7 | L-A-P2.2 | L-A-P2.6 | L-A-2.1 | L-A-C | L-A-p2.3 | L-A-lus | L-A-2 | L-A-L1 |
| *S. paradoxus* | L-A-P1.3 | | 94.9% | 94.6% | 96.6% | 91.1% | 91.1% | 91.0% | 90.4% | 90.1% | 89.6% | 81.4% | 81.5% | 82.3% | 76.1% | 76.8% | 76.4% | 75.1% | 75.2% | 73.6% |
| | L-A-P1.4 | 96.9% | | 95.3% | 95.4% | 91.3% | 91.4% | 91.4% | 91.4% | 89.4% | 89.9% | 82.1% | 82.2% | 83.0% | 75.9% | 76.8% | 75.9% | 75.6% | 75.8% | 74.0% |
| | L-A-P1.5 | 96.9% | 98.5% | | 95.3% | 91.1% | 91.6% | 91.6% | 91.2% | 89.4% | 90.2% | 82.0% | 82.0% | 82.7% | 75.9% | 76.9% | 76.1% | 75.2% | 75.7% | 74.4% |
| | L-A-Q | 98.0% | 98.5% | 98.5% | | 91.4% | 91.5% | 91.5% | 91.1% | 89.7% | 90.4% | 82.0% | 82.5% | 83.1% | 76.1% | 77.0% | 76.3% | 75.7% | 76.0% | 74.4% |
| | L-A-P2.4 | 97.2% | 97.7% | 97.3% | 97.6% | | 91.2% | 92.8% | 93.4% | 90.0% | 90.2% | 81.7% | 82.7% | 82.9% | 76.5% | 76.6% | 76.1% | 75.6% | 75.5% | 74.2% |
| | L-A-P1.6 | 96.2% | 97.7% | 97.6% | 97.5% | 97.8% | | 91.5% | 91.1% | 89.5% | 90.3% | 81.6% | 81.9% | 82.5% | 76.4% | 77.8% | 76.1% | 75.7% | 75.4% | 74.5% |
| | L-A-p2.5 | 95.8% | 97.4% | 97.3% | 97.4% | 98.6% | 98.0% | | 93.2% | 89.3% | 90.0% | 81.6% | 82.1% | 82.3% | 76.3% | 77.4% | 76.1% | 75.6% | 75.3% | 74.3% |
| | L-A-P1.2 | 96.1% | 97.8% | 97.6% | 97.7% | 98.6% | 97.7% | 98.4% | | 89.4% | 90.1% | 81.7% | 82.5% | 82.6% | 76.4% | 77.1% | 76.0% | 75.8% | 75.7% | 74.2% |
| | L-A-P1.1 | 96.5% | 94.9% | 94.9% | 94.9% | 98.0% | 95.7% | 95.7% | 95.7% | | 89.5% | 80.8% | 80.9% | 81.7% | 76.0% | 76.2% | 75.9% | 74.8% | 74.7% | 73.6% |
| | L-A-D | 96.4% | 97.6% | 97.3% | 98.0% | 98.1% | 97.7% | 97.9% | 98.1% | 95.7% | | 82.2% | 82.4% | 82.9% | 76.1% | 77.1% | 76.1% | 75.1% | 75.7% | 73.9% |
| | L-A-P1.7 | 94.9% | 96.5% | 96.6% | 96.4% | 96.2% | 96.1% | 95.6% | 96.1% | 93.3% | 96.4% | | 89.8% | 90.6% | 75.0% | 75.6% | 76.0% | 76.0% | 75.3% | 74.0% |
| | L-A-P2.2 | 94.8% | 95.8% | 96.4% | 96.5% | 95.9% | 95.6% | 95.5% | 95.9% | 93.0% | 96.3% | 98.0% | | 92.3% | 75.7% | 76.6% | 76.6% | 76.6% | 75.8% | 74.2% |
| | L-A-P2.6 | 95.7% | 96.4% | 96.6% | 97.0% | 96.7% | 96.1% | 96.0% | 96.4% | 93.8% | 96.8% | 98.2% | 98.9% | | 75.4% | 76.4% | 76.5% | 76.4% | 76.3% | 74.0% |
| | L-A-2.1 | 86.1% | 85.9% | 85.9% | 86.3% | 87.2% | 86.3% | 85.5% | 85.4% | 85.6% | 85.6% | 85.6% | 86.1% | 86.3% | | 93.4% | 88.3% | 74.1% | 74.5% | 74.6% |
| | L-A-C | 87.4% | 88.7% | 88.6% | 88.9% | 88.1% | 88.7% | 88.2% | 88.2% | 85.6% | 88.3% | 88.3% | 88.8% | 88.7% | 96.5% | | 88.0% | 75.2% | 75.4% | 75.5% |
| | L-A-p2.3 | 86.8% | 86.3% | 86.3% | 86.6% | 87.6% | 86.6% | 85.9% | 85.8% | 86.0% | 86.0% | 86.1% | 86.5% | 86.9% | 97.8% | 95.8% | | 74.7% | 74.6% | 75.0% |
| *S. cerevisiae* | L-A-lus | 87.2% | 88.9% | 88.5% | 88.7% | 88.4% | 88.6% | 88.4% | 88.4% | 86.0% | 88.0% | 89.0% | 88.5% | 88.6% | 84.4% | 86.8% | 84.6% | | 77.8% | 73.7% |
| | L-A-2 | 85.4% | 86.9% | 86.8% | 87.0% | 86.5% | 86.6% | 86.3% | 86.4% | 83.6% | 86.6% | 87.2% | 87.2% | 87.2% | 84.5% | 87.0% | 84.9% | 90.8% | | 74.2% |
| | L-A-L1 | 83.6% | 84.7% | 84.7% | 85.1% | 84.4% | 85.1% | 84.8% | 84.6% | 82.2% | 84.9% | 85.0% | 84.9% | 84.9% | 84.1% | 86.1% | 83.6% | 85.9% | 85.3% | |
| | | Group 1 | | | | | | | | | | Group 2 | | | Group 3 | | | | | |

Amino Acid Identity

Nucleic Acid Identity

All the sequenced L-A dsRNA in *S. paradoxus* had an identity between 73.6% and 76.6% with the L dsRNAs of *S. cerevisiae* (Table 4-3). The genome organisation of all L-A dsRNAs found in *S. paradoxus* strains is the same as that of *S. cerevisiae* (Figure 4-3A). 25 nucleotides at the 5' of L-A dsRNA in *S. cerevisiae* seem to have *cis* signal for transcription. X dsRNA, which is a deletion mutation of L-A, has only this 25 bp at 5' and is stably maintained as a satellite of L-A in the infected cells (Esteban and Wickner, 1988). The full sequence of this 25 bp has assembled completely in three L-As from *S. paradoxus* strains: L-A-Q, L-A-D1, and L-A-P2.6, and partly in the other L-A dsRNA (Figure 4-3B). These 25 nucleotides are identical in all the L-A dsRNAs from *S. paradoxus* whereas there are variations in the sequence of this part of L-A in *S. cerevisiae* (Figure 4-3B). Similar to other viruses, this region is AU-rich and seems to facilitate the melting of the dsRNA in order to create access for *gag* (Rodríguez-Cousiño et al., 2011). The six conserved nucleotides at 5' in all *S. paradoxus* L-As is GAAUAA, which is identical to L-A-2. This sequence in the other two dsRNAs in *S. cerevisiae* is GAAAAA.

The first ORF, *gag,* starts at nucleotide 30 and extends to 2,102. Stop codons in three of the L-A: L-A-P1.1, L-A-P1.4, and L-A-P2.3 dsRNA from UAA changed to UGA. The *pol* ORF runs from nucleotide 1,961 to almost the end of the molecule at nucleotide 4,607. The whole sequence of *pol* ORF was assembled in eight of the sequences: L-A-Q, L-A-D1, L-A-P2.2, L-A-P2.5, L-A-P1.4, L-A-P1.5, L-A-P1.7, and L-P1.2. The complete sequence at the 3' end after Pol ORF was assembled only in two dsRNAs: L-A-Q and L-A-D1. In L-A-P2.2, L-A-P2.5, L-A-P1.7, and L-P1.2, part of the sequence was assembled. Similarity between this part of the dsRNA in *S. paradoxus* L-As is greater than between the *S. cerevisiae* L-As (Figure 4-3B). The two ORFs have 130 nucleotides overlap. In the *S. paradoxus* L-A, similar to the *S. cerevisiae* L-A, the *gag* and *pol* ORF probably express as a Gag-Pol fusion protein using the -1 ribosomal frameshifting region that exists in the overlapped area of the two ORFs. This region contained a stem-loop structure that is involved in frameshifting, nucleotides 1,988-2,034, adjacent to the slippery site, 1988GGGUUUA1995 (Dinman et al., 1991, Dinman and Wickner, 1992). A comparison between *S. cerevisiae* and *S. paradoxus* L-A dsRNAs indicates that the regions are almost identical in all L-A dsRNAs. There is only one residue between the stem-loop structure and the slippery site, nucleotide 1,997, which is different in four of the *S. paradoxus* L-A dsRNAs: L-A-P2.4, L-A-P2.1, L-A-C, and L-A-P2.3, and one residue in the stem-loop structure of L-A-L1, which shows variation compared to the other L-A dsRNAs (Figure 4-3C).

The other region that is 100% identical in all the dsRNAs from both species is 24-nucleotide stem-loop structure in (+) strand of L-A. This region is responsible for binding to the Gag-Pol fusion protein and encapsidation of dsRNA (Figure 4-3D).

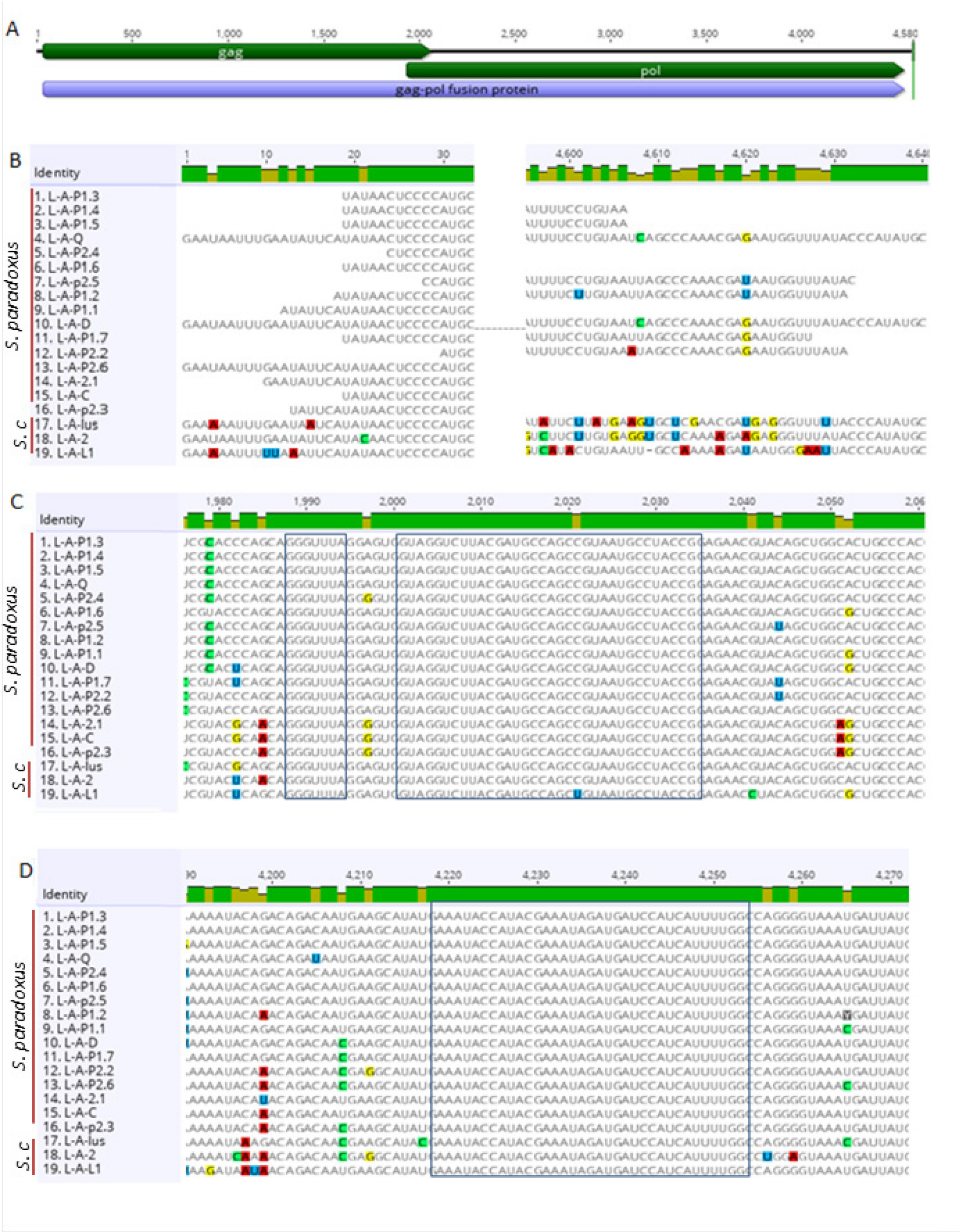
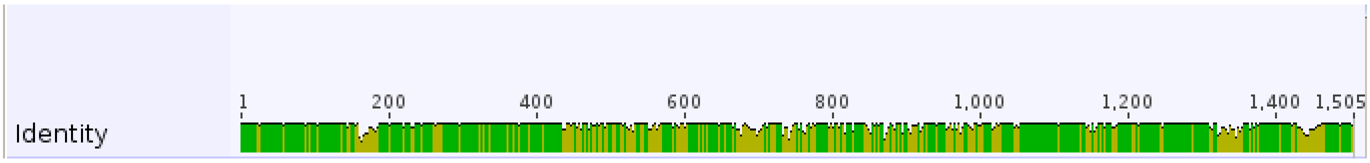


Figure 4-3. The structure and conserved sequences of L-A in *S. paradoxus* and *S. cerevisiae.* (A) The structure of L-A. L-A is composed of a conserved sequence at 5’ followed by *gag* and *pol* ORFs, which have overlap in the middle of the dsRNA, the frameshifting sequence in the overlap area, which is responsible for the expression of Gag-Pol fusion protein and conserved sequence at 3’. (B) The alignment of the sequence of the beginning and end of the L-A dsRNAs. The 25 nucleotide at the beginning of L-A contains *cis* transcription signal and AU rich region. It is identical in all *S. paradoxus* strains*.* However, there is variation between the sequences of this region in both species (C). The frameshifting region in L-As. The left rectangle shows the slippery site and the right rectangle indicates the stem-loop structure. Except for one nucleotide in the stem-loop structure, the rest of the sequences are identical. (D) Encapsidation stem-loop structure. 24 nucleotide near the 3’ end is responsible for binding to Gag protein for encapsidation. This region is identical in all Ls. S.C stands for *S. cerevisiae.*

### b) Amino acid sequence

The identity between the Gag-Pol fusion proteins encoded by L-As of both species is between 82.2% and 98.9%. They would be expected to have the same three-dimensional and spatial organisation and mode of action. In the alignment of all Gag-Pol fusion proteins there are five regions that show higher amounts of identity: residues 1-160; residues 186-436; residues 1,029-1,310; residues 1,350-1,428; and residues 1,464-1,503. The four motifs conserved in RDRP are in the third region from nucleotides 1,039 to 1,320; the identity in this region between the L-As is higher than 97.4%. In all *S. paradoxus* L-As, except L-A-C, L-A-2.1, L-A-P2.3, and L-A-P1.4, the identity of this part is 100%.



**Graph 4-1.** The identity between the amino acid sequence of the Gag-Pol fusion protein of *S. paradoxus* and *S. cerevisiae* L-As. The dark green shows the identical residues and the light green represents the lower identity between the residues. The numbers on the top are amino acid residues.

### c) Phylogeny analysis

A comparison between the tree of the *gag* and *pol* genes shows that they evolve almost at the same speed (Figure 4-4). The only difference between the trees is that, in the *gag* tree, L-A-L1 from *S. cerevisiae* is grouped with L-A-P2.1, L-A-P2.3 and L-A-C from *S. paradoxus*. However, the bootstrap of their branch is 83. This is also true for the tree of the whole sequence of the L-A. The bootstrap in this tree is less than that of the *pol* tree. Since the trees of both genes are similar, the study of the phylogeny of the L-A was done on the whole sequence of the dsRNA.

Based on the tree aligning all L-As from both species (Figure 4-4) and the identity between them (Table 4-3), there are three groups of L-As in *S. paradoxus* strains. Group 1 contained seven L-As from Silwood Park: L-A-Q, L-A-P1.1, L-A-P1.2, L-A-P1.3, L-A-P1.4, L-A-P1.5, and L-A-P1.6; the L-A of DBVPG4650, L-A-D1, from continental Europe; and two L-As from Pool2, which is a mixture of the strains from continental Europe and the Far East. The identity between the L-As in this group is between 89.2% and 96.6%. The closest L-As in this group are L-A-Q, L-A-P1.3, L-A-P1.4, and L-A-P1.5. The identity between these L-As is between 94.6% and 96.6%.

Group 2 of L-As is composed of L-A-P1.7 from Silwood Park together with L-A-P2.2 and L-A-P2.6 from Pool2. The identity between the L-As in this group is between 89.5% and 92.2%. This group is closer to the first group than to the other group; the identities between this group and the other two groups, group 1 and group 3, are around 81% and 76% respectively.

Group 3 of L-As includes two L-As from Pool2, L-A-P2.1 and L-A-P2.3, and the L-A from CBS 8,441, which is from the Far East. The identity in this group is from 88% to 93.4% and between this group and the other groups is around 75%. The results suggest that, firstly, since the two Pool2 L-As in this group are closer to L-A-C than L-A-Q and L-A-D1, they are from the Far East. Secondly, the two L-As in group one, L-A-P2.5 and L-A-P2.4, which are categorised with L-A-Q and L-A-D1, are from Continental Europe.

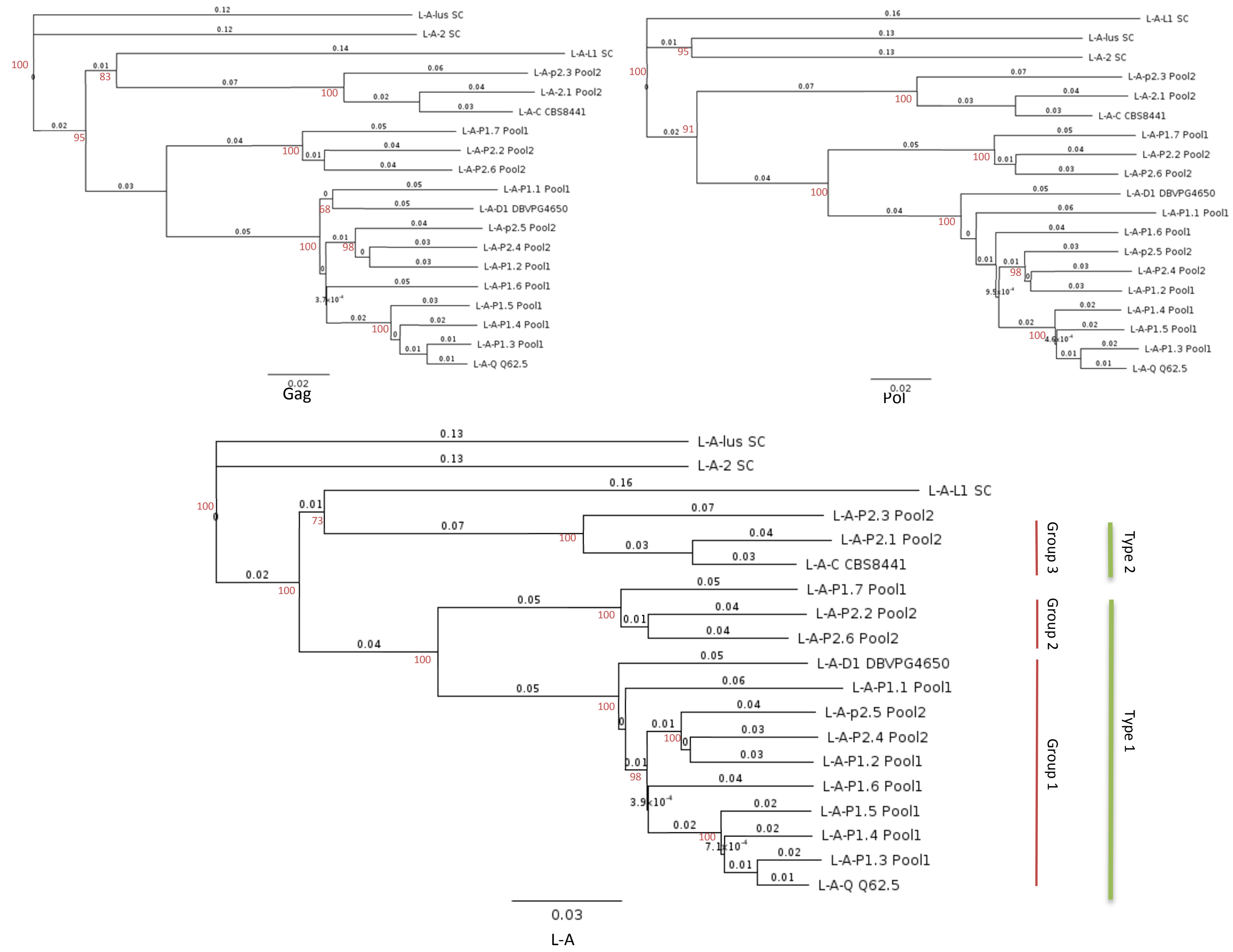


**Figure 4-4**. The neighbour joining phylogeny tree of *gag, pol* ORFs and the whole sequence of L-A dsRNA in *S. cerevisiae* and *S. paradoxus.* Based on the tree, there are three groups of L-A in *S. paradoxus*. A comparison between the identity of the three groups and L-A of *S. cerevisiae* shows that two types of L-A were found in this study.

Group 1 and group 2 are closer together than these two groups are with group 3. The identity between the first two groups is between 80.6% and 82.7%, whereas the identity between these two groups and group 3 is between 73.3% and 77.8%. Since the identity between the different types of L-A dsRNA in *S. cerevisiae* is 73.7% and 77.8%, it seems that two types of L-A were identified in *S. paradoxus*, one of which has two subtypes.

### 4.3.3 M dsRNA

Seven new types of M dsRNA were identified in *S. paradoxus* in this study (Figure 4-5). Of all M dsRNAs reported in *S. cerevisiae*, only the sequence of M28 was found in Q62.5, DBVPG4650 and Pool2. M1 was previously reported in the *S. paradoxus* strain Q74.4 (Chang et al., 2015), which was sequenced in Pool1 in this study. However, even by map-to-reference, the sequence of the dsRNA did not assemble. The information about all M dsRNAs found in *S. paradoxus* can be seen in Table 4-4.

**Table 4-4**. The result of the prediction of the structure of the new M dsRNAs and their preprotoxins

| Name M dsRNA and the host strain | | Length | Conserved sequence at 5’ | AU-rich after 5’ conserved sequence | Start codon residue | ORF length (bp) | 3’ Non-coding sequence length | Killer toxin name | preprotoxin length (aa) | Preprotoxin starts from | Hydrophobic aa at N-terminal of preprotoxin | Signal cleavage sites | Kex2 cleavage site | N-glycosylation site |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MQ | Q62.5 | 1933 | GAAAAAAUUUGA | + | 36 | 819 | 917 | KQ | 272 | Start codon | + | 15-27, 19-31 | 112 | 49, 66, 79, 95, 105*, 230** |
| | Pool2 | 1943 | GAAAAAAUUUGA | + | 36 | 846 | 917 | | 272 | Start codon | + | -*** | 112 | 49, 66, 79, 95, 106, 230** |
| MC | CBS-8441 | 2012 | - | + | 59 | 1008 | 818 | KC | 355 | Beginning of RNA | - | 5-17, 92-104 | 40, 81, 163, 251 | 44, 67, 186, 199, 231, 259** |
| | Pool2 | 2069 | GAAAAAAUGAAG | + | 7 | 1113 | 783 | | 370 | Start codon | + | 15-27,21-33 109-121 | 56,97,179,267 | 5, 60, 83, 202, 215, 247, 275** |
| M28 | DBVPG-4650 | 1793 | GAAAAAAUUUAA | + | 55 | 996 | 554 | K28 | 331 | Start codon | + | 10-22, 107-119 | 178, 231 | 147, 168*, 197, 210, 224** |
| | Pool2 | 1790 | - | + | 55 | 996 | 600 | | 331 | Start codon | + | 10-22, 107-119 | 178, 231 | 147, 194,210**, 224 |
| M-P1G1 Pool1 | | 1207 | - | + | 26 | 1002 | NA | K-P1G1 | 333 | Start codon | + | - | 110, 171, 220 | 132**, 150,159, 199, 213 |
| M-P1G2 | Pool1 | 1118 | - | + | 26 | 678 | NA | K-P1G2 | 233 | Beginning of RNA | short | 4-16, 8-20 | 161 | 43, 50, 100, 122**, 154, |
| | Pool2 | 1096 | - | + | 26 | 876 | NA | | 299 | Beginning of RNA | short | 3-15, 8-20 | 33, 161 | 43,50, 100, 122, 154 |
| M-P1G3-1 Pool1 (M-P1G3) | | 1235 | GAAAAAAUACUA- | + | 76 | 972 | NA | K-P1G3-1 | 323 | Start codon | + | 8-20, 9-21, 10-22, 11-23, 12-24, 51-63 | 36, 158, 216 | 118, 126**,168, 208 |
| M-P1G3-2 Pool1 (M-P1G4) | | 1222 | GAAAAAAUGCUA- | + | 7 | 1041 | NA | K-P1G3-2 | 346 | Start codon | + | 23-35,24-36, 25-37, 27-39, 31-43, 32-44, 33-45, 34-46, 35-47, 74-86 | 59, 181, 239 | 141, 149**, 231, 315 |
| M-P1G5 Pool1 | | 947 | - | Shorter | 7 | 780 | NA | K-P1G5 | 259 | Beginning of RNA | + | - | 71, 128 | 15, 36, 59, 121, 233 |
| M-P1SG Pool1 | | 784 | GAAAAAAUAAUC- | + | 51 | 420 | NA | K-P1SG | 139 | Start codon | short | - | 112 | 35, 131 |

*N-glycosylation potential is 0.3 **N-glycosylation potential is 0.4 ***Did not detect the signal because of the ambiguity nucleotide which exists in the signal cleavage site

### a) The structure of M dsRNA

Although the identity between different types of M dsRNA in both species is low, 15% – 35% the structure of the dsRNA is the same in all. The sequences start with a conserved sequence followed by an AU-rich region and the killer toxin ORF with direction of 5’ to 3’. There is a possibility that the AU region is facilitating the melting of the dsRNA for transcription. After the ORF, there are poly A and long non-coding sequences respectively (Rodríguez-Cousiño et al., 2011) (Figure 4-5A). Because of the poly A in the middle of the M dsRNAs, it was not possible to discern which coding part belongs to which non-coding region in Pool1 M dsRNA. As a result, their sequences were analysed separately.

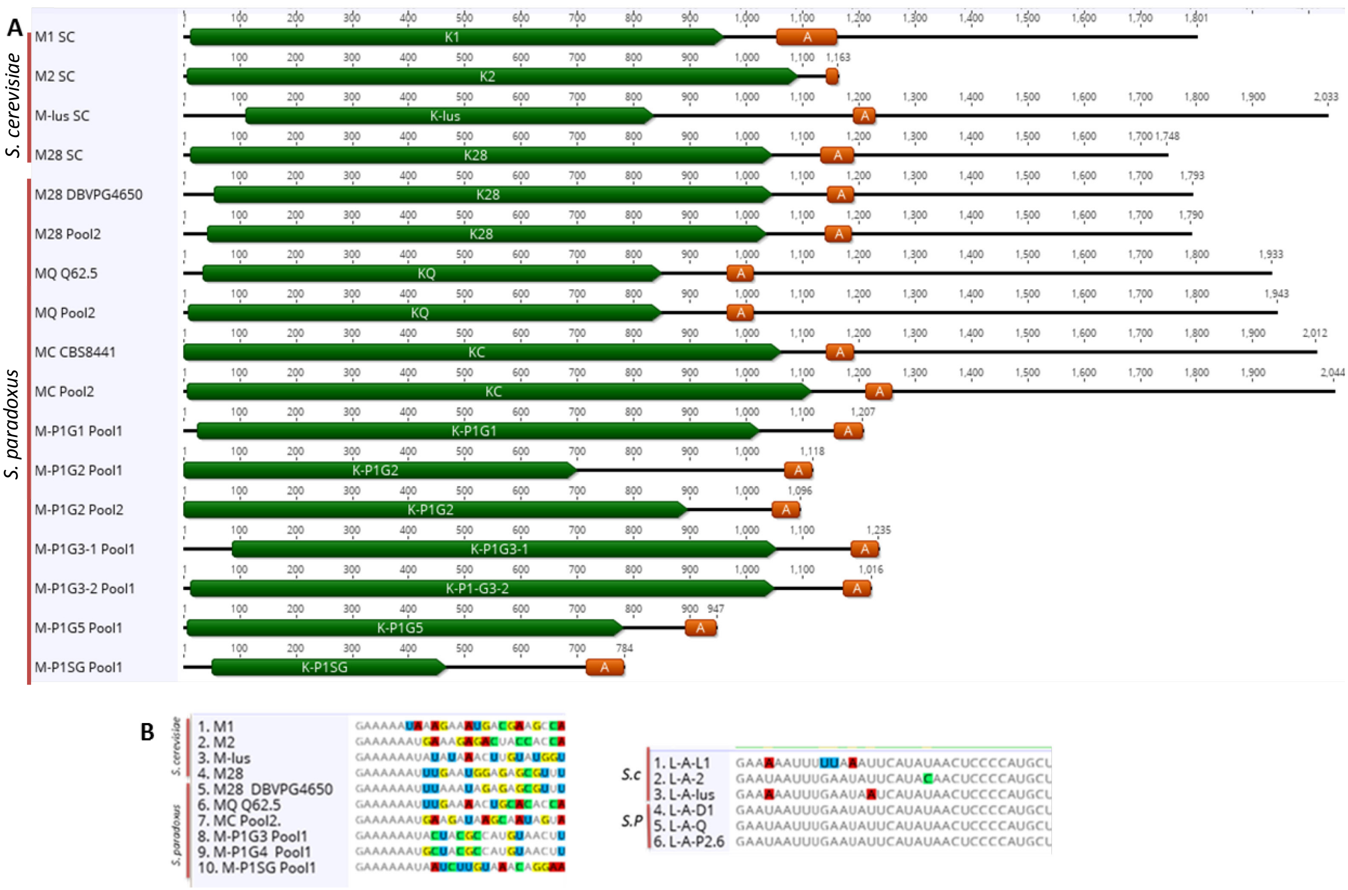


**Figure 4-5**. (A) M dsRNA structures in *S. cerevisiae* and *S. paradoxus.* All sequences have a killer toxin ORF (420-1113 bp), followed by poly A (since the software was not able to count the accurate number of A, all were considered 50 bp) and a non-coding area (600-917 bp). The green arrows are the killer toxin ORFs and the orange areas are the poly A sequences. Since it was not possible to match the coding and non-coding region of Ms in Pools due to the poly A in the middle of the dsRNAs, the dsRNAs belong to the Pools do not have the non-coding region. SC is the abbreviation of *S. cerevisiae.* (B) The conserved sequence at 5’ of Ms and L-As. The 5’ end has assembled in six of the Ms in *S. paradoxus.*

The conserved sequence at 5’ is an essential recognition element in transcription initiation. The sequence of GAAAAA- was reported for *S. cerevisiae’s* dsRNA (Rodríguez-Cousiño et al., 2011). The 5’ end of all M dsRNAs sequenced in this study formed, except M-P1G1, M-P1G2 and M-P1G5. As can be seen in Figure 4-5B, the eight residues at 5’ end of all M dsRNA of both species, except M1 from *S cerevisiae*, are identical: GAAAAAAU-, which seems to be the conserved sequence of M of saccharomyces strains. In M1, the sequence changed to GAAAAAUA-. A comparison between the L and M 5’ conserved sequences shows that this sequence in L either changed to GAAAAAUU- (in L-A-L1 and L-A-lus) or GAAUAAUU- (in L-A-2, L-A-D1, L-A-Q, and L-A-P2.6). The L helper of M1 and M-lus (L-A-L1 and L-A-lus) has the identical sequence in the six residues at 5’ end (Rodríguez-Cousiño et al., 2013). However, this is not true for M2 in *S. cerevisiae* and MQ in *S. paradoxus* (their helpers are L-A-2 and L-A-Q respectively). In DBVPG4650, which contains two helpers for M28, L-A-L1 and L-A-D1, although the copy number of L-A-L1 is less than L-A-D1, the conserved sequence is closer to L-A-L1. The rest of 25 bp at 5’ end of M, which has *cis* signal for transcription in L-A (Esteban and Wickner, 1988), not only is not conserved in Ms but is also different from Ls. This is true in both species, which suggest another unknown signal in M.

The AU-rich region is formed in all the Ms. It only has a shorter length in M-P1G5, which does not have the conserved sequence either. It seems that the missing part at the 5’ end of this RNA is longer than that of M-P1G1 and M-M-P1G2. The length of ORF in both species is between 420 bp and 1,113 bp. M-P1SG has the shortest length of ORF.

**b) Protein structure**

Unlike L, in M dsRNA the identity of the amino acid sequence is less than that of the nucleotide sequence. The nucleotide identity between all the coding sequences of all Ms is between 12% and 30.2%, whereas the identity between all preprotoxins is between 5.3% and 16% (including gaps). However, it seems that they all have a structure similar to the known preprotoxins in *S. cerevisiae,* K1 and K28 (SCHMITT and TIPPER, 1995) (Figure 4-6). Apart from K-P1G1 and K-P1G5, in which the beginning of the dsRNA is missing, and also K-P1SG, the rest of the preprotoxins start with a stretch of hydrophobic amino acids and signal cleavage sites. All preprotoxins have between one and four Kex2p/Kex1p cleavage sites and between two and six N-glycosylation sites. All these results suggest that in *S. paradoxus* as in *S. cerevisiae*, the preprotoxin uses the post-translational modification system in ER and Golgi to mature.

In most of *S. paradoxus's* preprotoxins, similar to the K1, K28 and K-lus in *S. cerevisiae,* there is more than one signal cleavage site. In terms of the Kex2p/Kex1p cleavage sites, KC with four cleavage sites has the closest structure to K1. Although K-P1G1 and K-P1G3-1 contained three cleavage sites, K-P1G5 contained two cleavage sites, and KQ, K-P1G2 and K-P1SG has only one Kex2p/Kex1p cleavage site, there is still the possibility of having unknown cleavage sites in the preprotoxins similar to K28 to break them to four subunits (SCHMITT and TIPPER, 1995). Further lab protein studies are needed to find out the number of the preprotoxins and mature toxin subunits.

A comparison between the N-glycosylation sites in *S. paradoxus* and *S. cerevisiae* indicates that the number of the N-glycosylation sites in *S. paradoxus* is higher than that in *S. cerevisiae.* Even in K28, which exists in both species, the number increased from three N-glycosylation sites to five N-glycosylation sites. Evolutionarily, N-glycosylation is an advantage for proteins; it stabilizes the proteins against proteolysis and denaturation, increases solubility, facilitates orientation of proteins relative to a membrane, confers structural rigidity to proteins, facilitates orientation of proteins relative to a membrane, regulates protein turnover and fine-tunes the charge and isoelectric point of proteins (Helenius and Aebi, 2004).

M-P1SG, with a length of 784 bp, has the shortest preprotoxin, K-P1SG (139 aa). The length of the preprotoxin is about half that of the other preprotoxins. Since it has both a conserved sequence and an AU-rich region at 5' and poly A in 3' of the dsRNA, as well as the hydrophobic amino acid at amino terminal of the protein, it seems that no part of the dsRNA is missed in sequencing. However, the start codon in the dsRNA is TTG instead of AUG. There is a possibility of having a deletion in the sequence of this dsRNA. Maybe, this is the M dsRNA in Q95.3 and Q16.1, which do not produce any active killer toxins. Both of these strains had the shortest M dsRNA on the gel and did not show the killer phenotype in the killer assay.

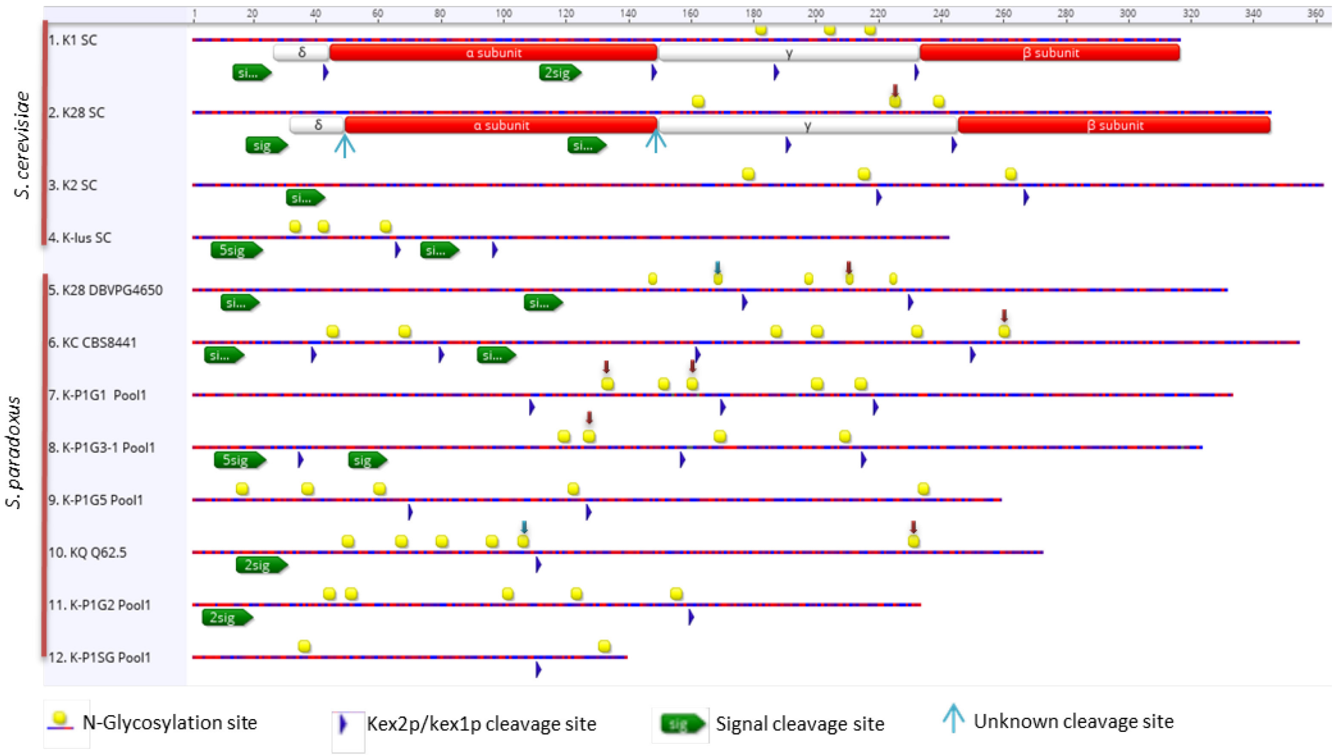


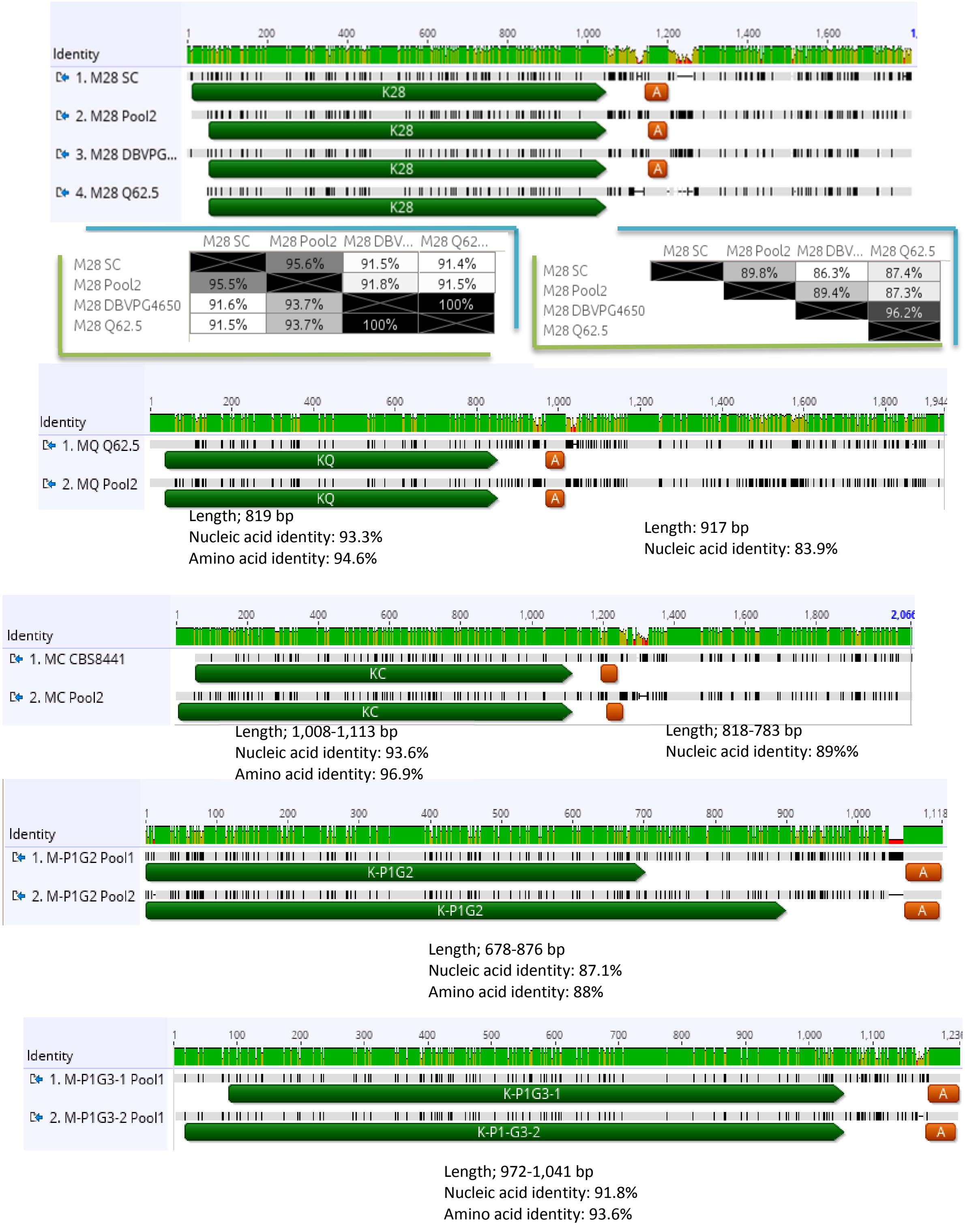
**Figure 4-6**. The structure of preprotoxins. The first two preprotoxins are K1 and K28, the structure of which is well-known. They start with a stretch of hydrophobic amino acid and signal cleavage sites to enter to ER and Golgi for post- translational modifications. They break into four subunits using three Kex2/Kex1 cleavage sites in K1, and one Kex 2/Kex1 cleavage site and two unknown cleavage sites in K28, δ,α,γ and β. There are three N-glycosylation sites on the γ subunit of both preprotoxins. Using two disulphide bond the α and β subunits will be attached and make the active killer toxin. A comparison between the preprotoxins of *S. cerevisiae* and *S. paradoxus* suggests that the preprotoxins in *S. paradoxus* go through the same pathway that *S. cerevisiae*'s preprotoxins use to become mature. They have between one and six signal cleavage sites (except K-P1G5 and K-P1SG), one and three Kex2p/Kex1p cleavage sites, and four and six N-glycosylation sites. Only K1P1SG has two N-glycosylation sites, which seems to be as a result of deletion in this sequence. The red arrows on N-glycosylation sites indicate that the site has the potential of 0.4 and the blue arrows indicate that the site has the potential of 0.3.

### a) Polymorphism in M dsRNA of *S. cerevisiae* and *S. paradoxus*

Three sequences of K28 (in Q62.5, DBVPG4650 and Pool2) and two sequences of MQ (in Q62.5 and Pool2), MC (in CBS8441 and Pool2), M-P1G2 (in Pool1 and Pool2) and M-P1G3 (in M-P1G31 and M-P1G3-2 in Pool1) were found in this study. The sequences of each type of M were aligned separately. The alignments are in the Appendix (Section 8.2.8) and the results are summarised in Figure 4-7.

A comparison between the alignments shows that in all the dsRNA the identity of the coding sequences is higher than that of the non-coding sequences, which shows the higher conservation in coding sequences compared to the non-coding sequences. Of all, only the size of the ORF of MQ has not changed in the alignment. The start codon in the M28 alignment and M-P1G3-1 and M-P1G3-2 alignment is changed. Moreover, there is a stop codon in the sequence of the M-P1G2 from Pool1. In terms of signal cleavage site, Kex2p/Kex1p cleavage site and N-glycosylation sites are almost conserved in the proteins (Table 4-4). The only differences are: the Kex2p/Kex1p site at residue 33 of the M-p1G2 of Pool2 is changed; MC from Pool2 has one N-glycosylation site more than MC from CBS8441; and the number of N-glycosylation sites in M28 of the *S. paradoxus* strains, DBVPG4650 and Pool2, is more than that in *S. cerevisiae*, it is five, four and three respectively.

As mentioned before, K28 is the only M that is found in both species and it has very low reads in Q62.5. The coding sequence of this dsRNA is identical to that of DBVPG4650. On the one hand, it could result from contamination in Q62.5. On the other hand, this strain showed the strongest killer phenotype among all tested *S. cerevisiae* and *S. paradoxus* strains. It was the only strain in which none of the *S. cerevisiae* and *S. paradoxus* immune tester strains were immune to its killer toxin (Table 2-1). In addition, in the bio assay of the IEF gel, there was one clear zone of sensitive cells in the area with a pH of 4.5 – 5.2 and a zone with the lower number of sensitive clones in the area with a pH of 3.5 – 4.2 (Chapter 2), which might result from the two present M dsRNAs in this strain. In order to be sure about the presence of M28 in this strain, RT-PCR with specific primes is needed.



| | M28 SC | M28 Pool2 | M28 DBV... | M28 Q62... |
|---|---|---|---|---|
| M28 SC | | 95.6% | 91.5% | 91.4% |
| M28 Pool2 | 95.5% | | 91.8% | 91.5% |
| M28 DBVPG4650 | 91.6% | 93.7% | | 100% |
| M28 Q62.5 | 91.5% | 93.7% | 100% | |

| | M28 SC | M28 Pool2 | M28 DBV... | M28 Q62.5 |
|---|---|---|---|---|
| M28 SC | | 89.8% | 86.3% | 87.4% |
| M28 Pool2 | | | 89.4% | 87.3% |
| M28 DBVPG4650 | | | | 96.2% |
| M28 Q62.5 | | | | |

Length; 819 bp
Nucleic acid identity: 93.3%
Amino acid identity: 94.6%

Length: 917 bp
Nucleic acid identity: 83.9%

Length; 1,008-1,113 bp
Nucleic acid identity: 93.6%
Amino acid identity: 96.9%

Length; 818-783 bp
Nucleic acid identity: 89%%

Length; 678-876 bp
Nucleic acid identity: 87.1%
Amino acid identity: 88%

Length; 972-1,041 bp
Nucleic acid identity: 91.8%
Amino acid identity: 93.6%

**Figure 4-7**. The alignment, the nucleic acid identity and amino acid identity of different sequences of each M dsRNA. For M28, which has four sequences, the nucleic acid and amino acid identity of the coding sequence is shown in the left table and nucleic acid identity of non-coding sequence is shown in the right table. The green lines on the tables show the amino acid identity and the blue lines show the nucleic acid identity. For the rest of the alignments, the length, the nucleic acid and amino acid identities of coding sequence and the length, the nucleic acid identity of the non-coding sequence showed under each part.

The ORF of M28 in DBVPG4650 was previously reported (Chang et al., 2015). The sequence of the M28 ORF had two nucleotide differences with the previous report (Chang et al., 2015). The first different residue, nucleotide 15, is critical because it is the third nucleotide at the start of the ORF. At this position, nucleotide G is changed to A. The coverage of the residue is 12,114 reads, of which 97% are A and only 0.8% are G. The result of this change is that, instead of starting the ORF from nucleotide 13, it starts from nucleotide 55, codon 15. The expression of the killer phenotype from codon 15 was proved in *S. cerevisiae*. Nevertheless, the expression from the first codon is four-fold more efficient than the expression from codon 15 (SCHMITT and TIPPER, 1995). That is probably why this M28 killer strain, against the *S. cerevisiae* and *S. paradoxus* M28 killer strains MS300 and T21.4, expresses its killer phenotype only in Same-Day-Test. Furthermore, it was killed by T21.4, which is a M28 killer (Chang et al., 2015) and overall has a weaker killer phenotype compared to MS300 and T21.4 (Table 2-1). In all *S. paradoxus* M28, the ORF starts from codon 15. Although the length of the hydrophobic amino acids decreased, the sequence of the signal cleavage did not change.

For M-P1G3-1, mutation in its first start codon changed the AUG to AUA. Of the nucleotides aligned in the third position of the codon in M-P1G3-1, 94.3% are A. As a result, instead of the ORF starting from nucleotide 7, it starts from nucleotide 76. Therefore, the preprotoxin from M-P1G3-1 has 23 amino acids and five signal cleavage sites less than M-P1G3-2. However, there are still five signal cleavage sites in its sequence to enter the ER system. Since the activity of K28 killer toxins in the M28 *S. cerevisiae* strains, that the killer toxin starts from the second start codon, is approved (SCHMITT and TIPPER, 1995), there is the possibility that the killer toxin of M-P1G3-1 is active, too.

In M-P1G2 of Pool1, a mutation in residue 703 changed the codon UGG to UAG, which is a stop codon. As a result of this change, the ORF of this dsRNA has 198 nucleic acids; its protein contained 66 amino acids less than those of Pool2. In addition, as mentioned before, as a result of a mutation, one of its Kex2p/Kex1p cleavage sites was also removed. These results increase the chance of producing an inactive killer toxin by this dsRNA. Previous results in this study showed that, although strains Y2.8, T26.3 and Y1, which exist in Pool1, contain M dsRNA, they do not produce any killer toxins.

### 4.3.4 Evolution of dsRNA in *S. cerevisiae* and *S. paradoxus*

From the evolutionary perspective, M is the satellite of the L-A virus which, during evolution, becomes compatible with the replication system of the L-A to encapsidate and replicate itself (Rodríguez-Cousiño and Esteban, 2017). The killer toxin of the M has benefit for both virus and yeast; at the same time, it has costs for both (Greig and Travisano, 2008b). It seems that each M has specifically evolved with one variant of L-A in *S. cerevisiae* (Rodríguez-Cousiño and Esteban, 2017, Rodríguez-Cousiño et al., 2013). There are *S. cerevisiae* strains that have only L-A dsRNA. However, in this study all infected strains contained both L and M. The result of cycloheximide treatment (Table 3-3) proved that L-A dsRNA is stable alone in the *S. paradoxus* strains, too. It seems that in *S. paradoxus* the advantage of M for the L-A virus is more than its costs. As a result, during natural selection almost all the L-A viruses are equipped with M satellite in *S. paradoxus.*

A comparison between the nucleic acid identity and amino acid identity of L-A dsRNAs (Table 4-3) and each type of M dsRNA (Figure 4-7) shows that in both of them the sequence of amino acids is more conserved than the sequence of nucleic acid level. However, the difference between the nucleic acid and amino acid identity in L-A is much higher than the Ms. This arises from the critical role of Gag and Gag-Pol fusion protein in the survival of the virus.

The presence of M28 and L-BC in both species, as well as the L-A dsRNA with maximum 26% difference, indicate that without any doubt the viral dsRNA has the same origin in both species and transferred between them. On the one hand, the presence of wider types of M in *S. paradoxus*, particularly finding five new types of M just in Silwood Park, more glycosylation sites in the preprotoxin of the *S. paradoxus* and wider immunity to killer toxins in *S. paradoxus* strains, all support the theory of the evolution of dsRNA in *S. paradoxus* and transfer to *S. cerevisiae* (Chang et al., 2015). In addition, L-A-D1, which seems to be the M28 helper, is closer to the *S. paradoxus* Ls than the *S. cerevisiae* Ls (Table 4-3). On the other hand, the L-A helper of each type M is more diverse in *S. cerevisiae* than in *S. paradoxus*. For the three types of M, for which their L has been reported in *S. cerevisiae*, the identity between the Ls is around 75%; whereas in *S. paradoxus*, for eight types of M, only two variants of L with that much diversity were detected. This can be seen from two points of view: firstly, as result of the longer evolution of L and M in *S. cerevisiae*, they become more specified. Or secondly, because the evolution of the viruses in *S. paradoxus* was longer than that in S*. cerevisiae*, each type of L has been equipped with different types of M.

Since DBVPG4650 is the only strain that has the same dsRNAs as *S. cerevisiae*, L-A-L1, L-BC and M28, it is of evolutionary significance for further studies.

## 4.4 Conclusion

Sixteen full sequences of L-A dsRNA (L-A-Q, L-A-D1, L-A-C, L-A-P1.1, L-A-P1.2, L-A-P1.3, L-A-P1.4, L-A-P1.5, L-A-P1.6, L-A-P1.7, L-A-P2.1, L-A-P2.2, L-A-P2.3, L-A-P2.4, L-A-P2.5, and L-A-P2.6), one sequence of *gag* gene of L-A, L-A-P1g1, one sequence of *pol* gene, L-A-P1p2, and seven new types of M dsRNA, which in turn were composed of two full sequences of M dsRNA (MQ and MC), five of killer toxin genes (M-P1G1, M-P1G2, M-P1G3, M-P1G4, and M-P1G5) and three sequences of non-coding parts of M dsRNA (M-P1NC1, M-P1NC2, and M-P1NC3) were found in the *S. paradoxus* strains.

The sequence of L-As shows homology with the L-A in *S. cerevisiae*. The sequence of the *gag* and *pol* fusion protein is almost similar. There are two variants of L-A that have been identified, based on the tree that is formed from aligning L-As of both species and the identity between them. Although the identity between the sequences of the M dsRNAs is very low, the organisation of the dsRNA and preprotoxin is similar between both species. The presence of a wider range of M, more glycosylation sites in the preprotoxin, wider immunity to killer toxins in *S. paradoxus* strains compared to *S. cerevisiae*, all support the theory of the evolution of the viral dsRNA in *S. paradoxus* and transfer to *S. cerevisiae*.

**Chapter Five**

# 5 Expression of killer toxin

## 5.1 Introduction

Q62.5 was the strongest killer strain amongst all the *S. paradoxus* strains that were studied in this project. It expresses its killer phenotype in both Same-Day-Test and Different-Day-Test. Its killer-immune phenotype was not similar to any of the well-known killers in *S. cerevisiae*, K1, K2 and K28, when it was tested against them. It was the only strain which killed both *S. cerevisiae* and *S. paradoxus* strains. However, its killer activity is reduced against *S. paradoxus* strains. The killer phenotype in this strain was cured using cycloheximide treatment.

Two dsRNAs, L and M, were detected in this strain. The M dsRNAs were removed from the cells after treatment with cycloheximide. All the results suggest that there is a new type of M dsRNA in this species that encodes an unknown killer toxin in this strain. The detected viral dsRNAs, L and M, were sequenced using MiSeq 300 and their sequences were assembled using Geneious software. A new type of L (LQ) and M (MQ) dsRNA were detected in this strain. To find out, firstly, whether the sequence assembled in Geneious software is reliable, the ORF of the MQ was amplified using the designed complementary primers. Secondly, the amplified ORF was cloned into a vector and expressed into a *S. cerevisiae* strain to become certain that the MQ ORF encodes the killer toxin in this strain.

pYES2.1/V5-His-TOPO vector was selected for the expression. This plasmid is linearized with a single thymidine (T), which is overhung at the 3’ end of the vector for TA cloning, and topoisomerase, which is covalently bounded to the vector. Single T nucleotide allows the PCR products that have a single A at the 3’ end to ligate efficiently with the vector, and the topoisomerases facilitate the ligation. The promoter in this vector is *GAL1*.

## 5.2 Material and method

### 5.2.1 MQ ORF amplification

**Designing primer**

The MQ forward primer (5'-AACTGCACACCACTCGATAGTT-3') and the MQ reverse primer (5'-CATTAGCTGCACCGACAGTT-3') were designed and tested for the MQ ORF using Geneious software and Primer3 online software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi).

**RT-PCR reaction**

The reverse transcription reaction was performed using Super Script III First-Strand Synthesis System (Invitrogen) with some modifications. The Q62.5 purified dsRNA (method 4.2.2) with primers and dNTP were melted at 99°C for five minutes and immediately were put in -80°C ethanol for two minutes. Then, the instructions of the kit were followed for the first strand cDNA synthesizing. For DNA amplification, 2 µl of the cDNA were used.

The cDNA amplification was tried with Pfx (Invitrogen), AmpliTaq Gold® 360 DNA Polymerase (Invitrogen) and One*Taq* DNA Polymerase. The PCR reaction was performed at 94°C for 5 minutes and 40 cycles of 94°C for 15 seconds, 55°C for 30 seconds, 72°C (One*Taq* and AmpliTaq Gold® 360) or 68°C (Pfx) for 50 seconds.

The result of RT-PCR was checked on 3% agarose gel using GelRed.

### 5.2.2 Cloning the MQ ORF

Cloning was performed using pYES2.1 TOPO®TA Expression Kit (Invitrogen). In the cloning reaction, 3 µl of the PCR product was used. The transformation was performed using TOP10 One Shot® Chemical Transformation (Invitrogen). 10 µl and 50 µl of the transformed cells were spread on the selective media, LB Agar (Sigma) containing 100 µg/ml Ampicillin (Sigma), and they were incubated overnight at 37°C. Of the colonies, 20 were picked and resuspended in 20 µl nuclease-free water. 10 µl of the water were used for culturing the transformed cells, and the remaining suspended cells were used in the PCR reactions to analyse the positive clones. Two PCR reactions were performed to find the positive transformants. In the first reaction, for which MQ primers were used, the presence of the PCR product was tested. The second reaction was performed with the MQ forward primer and TOPO reverse primer to choose the cloned vector with the right direction of the MQ ORF. This reaction was incubated at 94°C for 10 minutes, followed by 40 cycles of 94°C for 15 seconds, 59°C for 30 seconds, 72°C for 1 minute and the final extension step of 72°C for 10 minutes. The result was checked on 3% agarose gel and the positive clones were selected. In the next stage, the plasmids were extracted from the selected clones using the QIAGEN Plasmid Mini Kit, and the extraction amount was measured with NanoDrop.

The purified plasmid were transformed into the INV*Sc*1, *S. cerevisiae* strain (Invitrogen) with the genotype *MATa his3D1 leu2 trp1-289 ura3-52 MA*T *his3D1 leu2 trp1-289 ura3-52*. The competent cells were prepared using the S.c. EasyComp™ Transformation Kit (Invitrogen) and we followed its instructions for transformation. The transformed cells were cultured on the selective medium, SC-U (0.67% yeast nitrogen base, 2% glucose, 0.01% [adenine, arginine, cysteine, leucine, lysine, threonine, tryptophan, uracil], 0.005% [aspartic acid, histidine, isoleucine, methionine, phenylalanine, proline, serine, tyrosine, valine] and 2% agar). The cells that are not transformed cannot grow on this medium. The plates were incubated at 30°C.

In order to express the MQ ORF and test the killer activity of the expressed protein, the killer activity of the transformed clones was tested on a modified MB agar (pH 4.5) (Appendix; Table 8-2). In this medium, instead of the 2% glucose, 2% Galactose and 1% Raffinose were added. The MB agar diffusion assay was performed as described in section 2.2.3. They were incubated at 22.5°C.

## 5.3 Results and discussion

### 5.3.1 MQ ORF amplification

The RT-PCR amplification of the MQ ORF was performed successfully with Pfx and One*Taq* Polymerase. There was a band the same size as the ORF of the MQ, 800 bp (Figure 5-1). An additional band of around 920 bp was visualised when 5 µl of the product of the Pfx enzyme were run on the gel. Since the One*Taq* Polymerase has the 5' to 3' nuclease activity and adds a single A at the end of the amplified DNA, it was chosen for the amplification of the ORF for cloning. There was no additional band when the One*Taq* product was run with higher concentration. No band was seen when the RT-PCR was performed using AmpliTaq Gold® 360 DNA Polymerase.

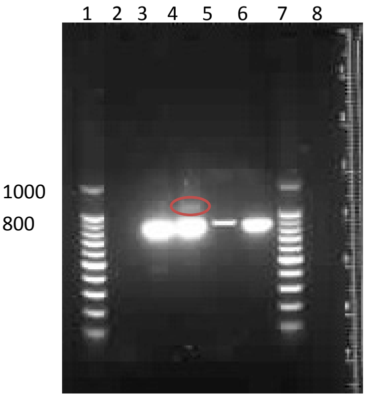


**Figure 5-1.** RT-PCR amplification of MQ ORF. Lanes from left to right: Lane 1, 100 bp DNA ladder; Lane 2, Pfx negative control; Lane 3, Pfx product (2.5 µl of the product were run); Lane 4, Pfx product (5 µl of the product were run); Lane 5, One*Taq* product (2.5 µl of the product were run); Lane 6, One*Taq* product (5 µl of the product were run); Lane 7, 100 bp DNA ladder; and Lane 8, One*Taq* negative control. The RT-PCR amplification of the MQ ORF, 800 bp band, was performed successfully with Pfx and One*Taq* Polymerase. An additional band, around 920 bp, was visualised when 5 µl of the product of the Pfx enzyme was run on the gel.

### 5.3.2 Cloning the MQ ORF

The MQ ORF was successfully cloned in the vector. The PCR of the insert fragment using the MQ primers was positive for the 20 clones that were tested. On the agarose gel of the PCR amplification, all the clones, except clone numbers 1 and 12, have a sharp band whereas, in these two clones, the MQ band is thicker and there is an additional bigger band on the gel (Figure 5-2a). The result of the PCR using the MQ forward primer and the TOPO reverse primer, which was done to find the direction of the inserted fragment, showed that a). in clones numbers 2, 3, 7, 8, 11, 18, and 19, there is a sharp band with the same size as MQ ORF; b). there is no band in  clones numbers 1 and 12; and c). in the other clones there are two unclear bands with a size of 900 bp and about 300 bp (Figure 5-2b). This suggests that, firstly, clone numbers 2, 3, 7, 8, 11, 18 have the vector with the right direction of amplified ORF; secondly, in clone numbers 1 and 12 the direction of the ORF is opposite; and thirdly, perhaps the 900 bp band in the rest of the clones is the unclear band that was seen in the RT-PCR using Pfx enzyme. Probably, the concentration of this band in the RT-PCR using One*Taq* was low. As a result, the amplfied fragment, which was not visible on the gel, was inserted into the vector during cloning.

Based on the aforementioned results, the clone number 2 was selected to continue the cloning. The plasmid was extracted from this clone using the QIAGEN Plasmid Mini Kit, and the purified plasmid was prepared for transformation.

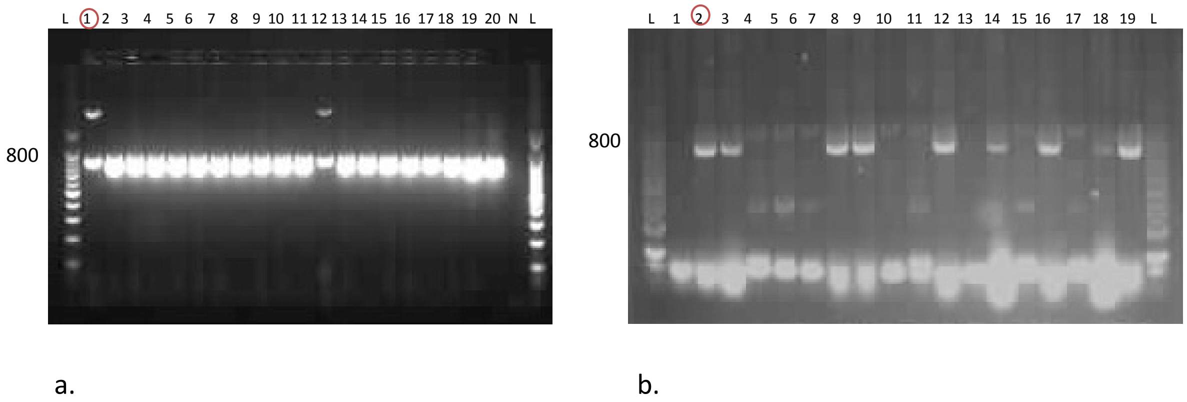

L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 N L
800
a.
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 L
800
b.

**Figure 5-2.** Amplification of the inserted fragment to find the positive transformants with the right direction of the fragment. a) Amplification using MQ primers. L is the abbreviation of ladder (100 bp DNA ladder), N is the abbreviation of the negative control and the numbers on the top of the gel are the name of the clones. The 800 bp fragment was amplified in all clones. Clone numbers 1 and 12 had an additional band larger than 1000 bp. The 800 bp band was thinner than the other clones. b) Amplification using the MQ forward primer and the TOPO reverse primer to check the direction of the fragments. L is the abbreviation of ladder (100 bp DNA ladder) and the numbers on the top of the gel are the name of the clones. There is a 800 bp sharp band in clone numbers 2, 3, 7, 8, 11, 18, and 19, no band in clone numbers 1 and 12 and two unclear bands with a size of 900 bp and about 300 bp in the rest of the clones. Clone number 2, which seems to have the right fragment with right direction, was selected for further study.

Since the KQ killer toxin (the killer toxin produced by MQ) seems to go through post transcriptional modification (section 4.3.3b), the selected strain for transformation had to be Eukaryote. In addition, during the maturation of the well-known killer toxins in *S. cerevisiae,* the host genome affects the process (Schmitt and Breinig, 2006). In order to increase the efficiency of the expression in the transformed cells, an *S. cerevisiae* strain (INV*Sc*1) was selected. As mentioned previously, *S. cerevisiea* is a close relative of *S. paradoxus* and the killer system in both species seems to have the same origin (section 4.3.4). The selected strain does not have any killer toxin and the selective media for the transformed cells is SC-U.

After transforming, the cells were cultured on the selective medium. The number of cells grown on the medium was significantly higher than normal. It seemed that the purity of the material which was used in the medium preparation was not very high. As a result, Uracil was inserted into the medium, and the medium was not selective. To save time, before preparing a new medium, 20 clones from the SC-U medium were selected for killer assay (Methylene Blue (MB) agar diffusion assay, section 2.2.3) to find the transformed cells. In order to activate the *GAL1* promoter, a modified MB agar medium was prepared. Instead of the glucose, Galactose and Raffinose were added to the medium as the carbon source. The test was done at the optimum pH and optimum temperature where Q62.5 killer toxin is most active (pH: 4.5 and temperature: 22°C). After growing the clones on the seeded medium, the result showed that the killer toxin, KQ, is expressed in two of the clones, 14 and 19. Nevertheless, the killer phenotype of these two clones was not as strong as that of the original strain. Possibly, it arises from the medium which was not optimised for the expression of the *GAL4* promoter. Alternatively, it arises from the effect of the differences between the genome of *S. paradoxus* and *S.cerevisiae (*which has influence on the post translational modification of preprotoxin).

Expression of KQ in the transformed cells proved the result of dsRNA sequencing. It also indicated that the MQ dsRNA which was founded in Q62.5 encodes the killer toxin in this strain.

## 5.4 Conclusion

The MQ dsRNA is a new type of dsRNA which was detected in Q62.5 in this project. The result of the killer-immune assays and cycloheximide treatment suggested that the killer phenotype in these strains is encoded by this dsRNA. In order to be certain about the sequence of this dsRNA, and prove that the ORF of this dsRNA encodes the killer toxin, the ORF was firstly amplified using specified primers and then cloned in a *S. cerevisiae* strain. The result of the killer assay of the transformed cells indicated that the killer phenotype is expressed in the transformed cells. Nevetheless, the expression of the phenotype in the transformed strain was significantly less than that of the original strain. Perhaps this arises from the medium which was not optimised for the expression of the *GAL4* promoter, or from the effect of the new host genome in the expression of the killer toxin in the post-translational stage.

**Chapter six**

# 6 Discussion

## 6.1 Introduction

The killer phenomenon was first reported in *Saccharomyces cerevisiae* (Wickner et al. 2002). Killer strains release a toxin that is lethal to sensitive yeast strains, but not to the killer itself (Woods and Bevan, 1968). Since then, the yeast-killer phenomenon has been found in numerous yeast genera and species. However, most of the studies on killer system in yeasts were performed on *S. cerevisiae.* Two genetic bases for the killer phenotype have been discovered in *S. cerevisiae*, viral dsRNA (Rodríguez-Cousiño et al., 2011, Hopper et al., 1977b, Magliani et al., 1997b) and chromosomal genes (Goto et al., 1991, Goto et al., 1990b). In the killer yeasts that were infected with viral dsRNA, There are two types of dsRNA, L-A and M dsRNA. L-A dsRNA produces Gag and Pol protein to replicate and encapsidate itself, and M dsRNA. M dsRNA has just one gene which encodes the killer toxin.

*S. paradoxus* is a wild non-domesticated close relative of *S. cerevisiae* and is a good model to study for its ecology and evolution. In this project, the killer system and its genetic bases were studied in this species.

## 6.2 Killer-immune phenotype and its genetic basis in *S. paradoxus* strains

The killer-immune phenotype in *S. paradoxus,* similar to other yeasts (Frank, 1994, Greig and Travisano, 2008a, Liu et al., 2015a)*,* is very complex and is sensitive to small changes in the environment. Like most of the yeast killer toxins, the killer toxins are active in acidic pH and at low temperatures (Liu et al., 2015a); None of the killer strains showed the killer phenotype at pH 5.3 and all killer samples were inactive at 28˚C and above.

In this study 35% of the strains showed the killer phenotype, which was about twice (Chang et al., 2015) or three times (Pieczynska et al., 2013b) greater than in previous reports. The killer strains are widespread across the world, with the highest frequency in American strains (70%). We report a different killer-immune reaction between the killer and immune tester based on the time of spraying the sensitive strains on the medium – labelled ‘Same-Day-Test’ and ‘Different-Day-Test’ (Chapter 2;

section 2.2.3). This protocol provides a condition that the killer phenotype expresses in the four strains that were previously reported as non-killer strains: Q59.1, UFR50816, A12 (Chang et al., 2015, Pieczynska et al., 2013b), and DBVPG4650 (Pieczynska et al., 2013b). The killer phenotype in all of these strains is weak. Q59.1, UFR50816 and DBVPG4650 showed their killer phenotype just in Different-Day-Test and A12 expressed it more clearly in the Same-Day-Test.  The majority of strains from Silwood Park, Continental Europe, the Far East, and South America expressed the phenotype more strongly in Different-Day-Test. Two strains from the Silwood Park, Q43.5 and Q59.1, and three strains from Continental Europe, DPVPG4650, C10 (4/2SW2), and C15 (CECT10176) did not express the killer phenotype in the Same-Day-Test. By contrast, all Canadian killer strains showed their killer phenotype more clearly in Same-Day-Test. Perhaps, in different concentrations of the killer toxins, there are different modes of actions in various killer toxins which trigger different killing and immunity pathways in the sensitive strains. As a consequence, different strains behave differently in Same-Day-Test and Different-Day-Test.

The results of the cycloheximide treatment and visualisation of extracted dsRNA from killer strains indicate that in all of *S. paradoxus* strains the killer phenotype is encoded by M dsRNA in the cells, except one strain from Continental Europe, C15 (CECT10176), and the Canadian killer strains. In C15 and Canadian killer strains; the killer phenotype is probably encoded by the nuclear genome or DNA plasmid. In general, the killer phenotype in the Canadian strains looks different from the other killer strains. The killer phenotype is expressed more clearly in the Same-Day-Test, it is very weak, there is no hollow around the killer strain, and only one or two layers of the sensitive strain’s clones grown around the killer strain become blue. In addition, the frequency of the phenotype in this region is higher than in the other regions; 83% of the strains from this region are killer. All the results suggest that the killer system in the Canadian strains evolves separately from the other regions.

The killer phenotype in C15 is also weak, but the phenotype is different from the Canadian strain. Moreover, it does not express the killer phenotype in Same-Day-Test. It seems that different genes are involved in the killer phenotype of this strain.

There are six *S. paradoxus* strains, Q16.1, Q95.3, Y2.8, T26.3, Y1, and N17, which contain M dsRNA but do not express the killer phenotype. A changed pattern of digestion of M dsRNA using S1 nuclease in three of the strains, Y2.8, Q95.3, and Y1, and a smaller sized M dsRNA in two of the strains, Q95.3 and Q 16.1, suggested the presence of mutations that alter the structure of the dsRNA, such as deletions similar to X dsRNA in *S. cerevisiae* (Esteban and Wickner, 1988); mutations in the sequence of the M dsRNA which inactivate the killer toxin or prevent its expression; or, mutation in the nuclear genes, which is essential for the expression of the killer toxin (Wickner, 1974).

None of the *S. paradoxus* killer strains showed the same killer-immune pattern as the *S. cerevisiae* strains, except three strains from Silwood Park: Y8.5, Q74.4, and T68.3. Nevertheless, Q74.4 was reported to have M1 dsRNA, whereas Y8.5 has an unknown M dsRNA. Although Q74.4 is expected to have the K1 killer toxin, it also showed a different killer pattern from the K1 killer toxin in *S. cerevisiae* in a previous study. They also showed that Q74.4, in line with our results, can kill the *S. cerevisiae* K1 killer strains (Chang et al., 2015).

Chang et al. (2015) suggested that the killer-immune system in *S. paradoxus* might be a more ancient than in *S. cerevisiae*. One of their reasons was that nearly all *S. paradoxus* populations in their study were immune to killer toxins of both (Chang et al., 2015). Our result, however, was slightly different from their report. We could not see an immunity in the *S. paradoxus* tested strains against the killer strains from both species that stop the expression of their killer phenotype. Nevertheless, the level of killer-phenotype expression against *S. paradoxus* strains was significantly less than that of the *S. cerevisiae* strains, which still suggests an additional immunity in the *S. paradoxus* strains. The difference in the results might be caused by the difference between the strains selected for the tests and the environmental conditions.

## 6.3 The effect of ethanol on the killer activity of killer toxins

The killer phenotype of the three killer strains, Q62.5, T21.4, and Y8.5, and the killer activity of the Q62.5 killer toxin were tested on MB agar containing 0%, 6%, 12%, and 14% ethanol. Although Saccharomyces strains are resistant to ethanol, the results of the treatment showed that the expression of the killer phenotype of the killer strains reduces by the increased concentration of ethanol. In addition, Q62.5's concentrated killer toxin, which was put on the same media, becomes completely inactive with all the concentrations of ethanol (Table 2-2). However, the effect is not permanent; if the killer-concentrated medium treated by the three concentrations of ethanol for two days is put on a normal MB agar seeded with the sensitive strain, it becomes active again. These results suggest that the nature of producing ethanol by the yeast strains decreases the efficiency of the killer toxin of the yeast-killer strains, which in turn can increase the chance of survival of the yeast species in the environment. Nevertheless, this is not true for all the yeast-killer toxins. In the case of the KHR killer toxin, increasing the ethanol in fermentation improves the efficiency of the killer toxin (de Ullivarri et al., 2014).

## 6.4 Characterisation of the killer system in Q62.5

Q62.5 expressed the strongest killer phenotype amongst the *S. paradoxus* strains. It is the only strain that kills all of the *S. cerevisiae* and *S. paradoxus* strains and appears to have a new type of killer toxin. As a result, it was used for further studies. The results showed that it has a killer toxin with a molecular weight of between 30 kDa and 50 kDa, and an isoelectric point between 4.5 and 5.2. The optimum pH of its killer toxin is 4.5. Although it is not active at pH 7, it is stable at this pH. In contrast to KHS killer strains, whose killer activity increases by raising the ethanol in the environment (de Ullivarri et al., 2014), the killer activity of this killer toxin decreases with increasing ethanol.

A new type of M dsRNA was detected in this strain: MQ. The structure of the M dsRNA and its preprotoxin were similar to the known M dsRNA. The predicted molecular weight and isoelectric point of the protein were in the range measured for the killer toxin of this strain.

In addition to MQ, a sequence of M28 with a very low number of reads was found in Q62.5 (56 reads) in this study. The coding sequence of this dsRNA is identical to that of DBVPG4650. On the one hand, it might result from contamination in Q62.5. On the other hand, this strain showed the strongest killer phenotype among all tested *S. cerevisiae* and *S. paradoxus* strains. None of the *S. cerevisiae* and *S. paradoxus* immune-tester strains was immune to its killer toxin (Table 2-1). In addition, in the bio assay of the IEF gel, there was one clear zone of sensitive cells in the area with the pH of 4.5–5.2 and a zone with the lower number of sensitive clones in the area with the pH of 3.5–4.2, which might result from the presence of two killer toxin in this strain. In order to be sure about the presence of M28 in this strain, RT-PCR with specific primers is needed.

## 6.5 Growth rate of yeast strains in the presence of Q62.5 killer toxin

Measuring the growth rate of four *S. paradoxus* strains: Q62.5, M894 (the *S. cerevisiae* sensitive strain), Q14.4 (*S. paradoxus* killer strain), and A33 (the non-killer *S. paradoxus* that showed the highest amount of immunity to Q62.5 killer toxin), in the presence of the Q62.5 killer toxin, surprisingly indicated that none of the strains, not even the sensitive one, was killed completely by the killer toxin. There is only a delay in the growth of the strains. The higher the concentration of the toxin and the lower the number of sensitive cells, the higher the delay in the growth of the cells. In the strains with greater sensitivity to the killer toxin, this delay increases (Graph 2-1). Overall, results suggest that all the strains, even the sensitive strains, are equipped with some immunity for their survival. Nevertheless, this immunity in the sensitive strains is significantly less than in the resistant strains. In addition to this immunity, increasing the ethanol level and decreasing the pH during fermentation, which also decreases the activity of the killer toxin, increases the chance of survival for the sensitive strains. Moreover, these might be the reasons for the killers' inability to be dominant always at the end of all fermentation and to be the dominant strains in nature (Heard and Fleet, 1987, Chang et al., 2015, Pieczynska et al., 2013b).

## 6.6 Viral dsRNA in *S. paradoxus* strains

Viral dsRNA of 27 *S. paradoxus* strains was sequenced in this study using MiSeq 300. The samples were categorised in five libraries: three strains from the three different regions (Q62.5 from Silwood Park, DBVPG4650 from Continental Europe and CBS8441 from the Far East) and two mixtures of the dsRNA extractions (Pool1 included 16 strains from Silwood Park and Pool2 four strains from Continental Europe and three from the Far East) (Appendix; Table 8-3). The result of the sequencing is summerised in Table 4-2.

In two of the single sequenced strains, Q62.5 and CBS8441, which had two dsRNA bands on the gel electrophoresis (Table 4-1), the large-size and the medium-size bands, the size of the bands was comparable with the L-A and the M dsRNA found. However, finding the dsRNA of each band in the gel electrophoresis of DBVPG4650 was more complicated compared with the two other strains. Five dsRNA bands were detected in the gel electrophoresis of this strain (Figure 3-3b and Table 4-1). One dsRNA band with the same size as L, two bands in the range of M dsRNA, and two smaller than M dsRNA were detected in this strain. Three types of L, L-A-L1, L-A-D and L-BC, and one type of M, M28, were found in this strain. The size of the full sequence of three Ls and the M is similar to the large- and medium-size bands on the gel. However, in the mapping alignment of the *de-novo* contigs to L-A-D1, there are several contigs identical to the reference with lengths between 502 bp and 1,363 bp. This was not true in the other two single strains. The presence of the different size L-A-D1 contigs in each *de-novo* assembly suggests that there are various L-A-D1 with different sizes in this strain. This might result from the deletion in the original sequence of L-A-D1, similar to X dsRNA in *S. cerevisiae*, which is derivative of L-A-L1 (Sommer and Wickner, 1982).

### 6.6.1 M dsRNA

In addition to MQ and M28, six new M dsRNAs were found in the *S. paradoxus* killer strains. For one of the dsRNAs, MC, we have the complete genome; for the others, just the coding sequences before the poly A repeat: M-P1G1, Mp1G2, M-P1G3-1, Mp1G3-2, M-P1G5, and M-P1SG. Moreover, three sequences of the non-coding part of the M dsRNA: M-P1NC1, M-P1NC2, M-P1NC3, which is located after the poly A, were found (Table 4-2). Since there is poly A in the middle of all dsRNA, it was not possible to find out which coding part belonged to which non-coding part in Pool1 and Pool2.

The only *S. cerevisiae* M dsRNA found in this study in the *S. paradoxus* strain was M28 from DBVPG4650, as previously reported (Chang et al., 2015). This dsRNA, which has an identity of 92.2% with that of *S. cerevisiae*, has a mutation in the first AUG start codon and seems to express the killer toxin from the second start codon. Expressing the K28 from the second start codon with lower killer activity has been reported before in *S. cerevisiae* (SCHMITT and TIPPER, 1995). In DBVPG4650, the killer toxin is also active but the level of expression is less than in *S. cerevisiae* M28, which was used in the killer-immune test (Table 2-1). This killer strain, which shows its killer phenotype only in Different-Day-Test, is killer when it is tested with the K1 *S. cerevisiae* strain and poor killer when it is tested with the K2 *S. cerevisiae* strain, only in the Same-Day-Test. It cannot kill the K28 *S. cerevisiae* strain. In the same test, the K28 *S. cerevisiae* strain shows the same pattern of killing but showed the killer phenotype much more strongly in both Same-Day-Test and Different-Day-Test.

In addition to M28,  M1 was previously reported in the *S. paradoxus* strain, Q74.4 (Chang et al., 2015).  This strain was sequenced in Pool 1 in this study. However, no M1 sequence has been found in the sequencing.  Even by map-to-reference, the sequence of the dsRNA did not assemble. As mentioned before, the killer immune reaction of this strain, in both studies, was different from K1 killer assay pattern. This strain can kill the *S. cerevisiae* K1 killer strains (Chang et al., 2015). Also, results in this project indicated that the K2 *S. cerevisiae* strain, which was killed by the *S. cerevisiae* K1 killer strain, was immune to Q74.4 killer toxin, whereas, the K28 *S. cerevisiae* strain is killed by Q74.4. This pattern in the killer assay suggests the presence of K2 toxin in this strain. However, No M2 has also been identified in this study. The M1 sequence reported by Chang et al has just one nucleotide difference with *S. cerevisiae's* M1 which causes one amino acid difference in the α subunit of M1 (Chang et al., 2015). Aligning the Q74.4's M1 sequence reported by them (GenBank: KJ796681.1) with three *S. cerevisiae's* M1 sequences (GenBank; NC_001782, SCU78817 and SEG_DQ0171595) shows a difference in the results; instead of changing Ala to Thr at residue 123, Ile changed to Ser at amino acid 103. A comparison between the sequences of Q74.4's M1 in the Chang et al paper and Genbank indicated that there are two nucleotide differences between the sequences. Cloning and expressing the Q74.4's M1 sequence and *S. cerevisiae*'s M1 sequence indicated that the Q74.4's M1 encodes a stronger killer toxin which seems to arise from the nucleotide difference (Chang et al., 2015). Nevertheless, Q74.4 could not kill the K2 *S. cerevisiae* strain in this study. There is a possibility that the differences in the genome background of the strains that was used in the killer assays and the mutations in them caused the changes in the killer pattern of this strain. However, resequencing the M dsRNA in this strain using specific primers can help to clarify this further.

The other M dsRNA found in the *S. paradoxus* strains was entirely different from *S. cerevisiae* M dsRNA. This explains why the killer-immune reaction in *S. paradoxus* was different from *S. cerevisiae* strains in this and in previous studies (Pieczynska et al., 2013b, Chang et al., 2015). Despite low identity (15% – 35%), in the sequence of M dsRNAs, both in the M dsRNAs from *S. paradoxus* and those of *S. cerevisiae*, the organisation of the Ms and their preprotoxins in all the *S. paradoxus* strains was similar to that of *S. cerevisiae*. The dsRNA started with a six-nucleotide conserved sequence, GAAAAA, followed by an AU-rich region and preprotoxin ORF. There is a poly A after the ORF connected to a long non-coding sequence (Figure 4-5 A). In the preprotoxins there is a stretch of hydrophobic amino acids followed by at least one signal cleavage and N-glycosylation sites. Between one and three Kex2p/Kex1p processing sites were detected in the sequence of the preprotoxins (Figure 4-6).

In most of the *S. paradoxus'* preprotoxins, similar to the K1, K28, and K-lus in *S. cerevisiae,* there is more than one signal cleavage site. In terms of the Kex2p/Kex1p cleavage sites, KC, with four cleavage sites, has the closest structure to K1. Even though K-P1G1 and K-P1G3-1 have three cleavage sites, K-P1G5 has two cleavage sites, and KQ, K-P1G2 and K-P1SG contain only one Kex2p/Kex1p cleavage site, there is still the possibility of having unknown cleavage sites in the preprotoxins, similar to K28, to break them into four subunits (SCHMITT and TIPPER, 1995). A comparison between the N-glycosylation sites in *S. paradoxus* and *S. cerevisiae* indicates that the number of N-glycosylation sites in *S. paradoxus* is higher than that in *S. cerevisiae.* Even in K28, which exists in both species, the number increased from three N-glycosylation sites to five N-glycosylation sites. N-glycosylation is evolutionarily an advantage for proteins (Helenius and Aebi, 2004).

M-P1SG (784 bp) from Pool 1, has the shortest ORF. The length of the preprotoxin (139 aa) is about half that of the other preprotoxins. However, it has both a conserved sequence and an AU-rich region at 5' and poly A at 3' of the dsRNA, as well as the hydrophobic amino acids at amino terminal of the protein. It seems that no part of the dsRNA is missed in sequencing. Nevertheless, the start codon in the dsRNA is TTG instead of AUG. There is a possibility of having a deletion in the sequence of this dsRNA. It may be that this is the M dsRNA in Q95.3 and Q16.1, which do not produce any active killer toxins. Both of these strains had the shortest M dsRNA on the gel, did not show the killer phenotype in the killer assay, and were sequenced in Pool 1.

A comparison between the alignments of similar M dsRNA found in this study (K28 in Q62.5; DBVPG4650 and Pool2; MQ in Q62.5 and Pool2; MC in CBS8441 and Pool2; M-P1G2 in Pool1 and Pool2; M-P1G3 in M-P1G31; and M-P1G3-2 in Pool1) shows that in all the dsRNA the identity of the coding sequences is higher than that of the non-coding sequences, which indicates the higher conservation in coding sequences compared to the non-coding sequences. In all, only the size of the ORF of MQ has not changed in the alignment. In terms of signal cleavage site, Kex2p/Kex1p cleavage site and N-glycosylation sites, they are almost conserved in the proteins (Table 4-4). The only differences are: the Kex2p/Kex1p site at residue 33 of the M-p1G2 of Pool2 is changed; MC from Pool2 has one N-glycosylation site more than MC from CBS8441; and the number of N-glycosylation sites in M28 of the *S. paradoxus* strains, DBVPG4650 and Pool2, is more than that in *S. cerevisiae*: it is five, four and three, respectively.

For M-P1G3-1, mutation in its first start codon changed the AUG to AUA. Of the nucleotides aligned in the third position of the codon in M-P1G3-1, 94.3% are A. As a result, instead of the ORF starting from nucleotide 7, it starts from nucleotide 76. Therefore, the preprotoxin from M-P1G3-1 has 23 amino acids and five signal cleavage sites less than M-P1G3-2. However, there are still five signal cleavage sites in its sequence to enter the ER system. Since the activity of K28 killer toxins in the M28 *S. cerevisiae* strains, that the killer toxin starts from the second start codon, is approved (SCHMITT and TIPPER, 1995), there is the possibility that the killer toxin of M-P1G3-1 is active, too.

In M-P1G2 of Pool1, a mutation in residue 703 changed the codon UGG to UAG, which is a stop codon. As a result of this change, the ORF of this dsRNA has 198 nucleic acids; its protein contained 66 amino acids less than those of Pool2. In addition, as mentioned before, as a result of a mutation, one of its Kex2p/Kex1p cleavage sites was also removed. These results increase the chance of producing an inactive killer toxin by this dsRNA. Previous results in this study showed that, although strains Y2.8, T26.3 and Y1, which exist in Pool1, contain M dsRNA with similar size to M-P1G2, they do not produce any killer toxins.

### 6.6.2 L dsRNA

In terms of L dsRNA, the following were detected in the *S. paradoxus* strains: 16 full sequences of L-A dsRNA (L-A-Q, L-A-D1, L-A-C, L-A-P1.1, L-A-P1.2, L-A-P1.3, L-A-P1.4, L-A-P1.5, L-A-P1.6, L-A-P1.7, L-A-P2.1, L-A-P2.2, L-A-P2.3, L-A-P2.4, L-A-P2.5, and L-A-P2.6); one sequence of a Gag gene of an L-A (L-A-P1g1); and one sequence of a Pol gene (L-A-P1p2). They have about 75% identity with the L-A dsRNA in *S. cerevisiae* and between 75% and 96.6% identity with each other. At the protein level, the identity between the L-As in the two yeast species increases to 92% (Table 4-3). The conserved sequence at the 5’ end of all L dsRNAs in *S. paradoxus* is GAAUAA. The genome organisation of the L-A dsRNA in *S. paradoxus* are similar to that of *S. cerevisiae*. There are two ORFs in L-As, *gag* and *pol* with 130 nucleotides overlap. The *gag* ORF starts at nucleotide 30 and extends to nucleotide 2102, and the *pol* ORF runs from nucleotide 1961 to almost the end of the molecule at nucleotide 4607. Stop codons of the *gag* ORF in three of the L-A: L-A-P1.1, L-A-P1.4, and L-A-P2.3 dsRNA from UAA changed to UGA. In the *S. paradoxus* L-A, similar to the *S. cerevisiae* L-A, the *gag* and *pol* ORF probably express as a Gag-Pol fusion protein using the -1 ribosomal frameshifting region that exists in the overlapped area of the two ORFs.

25 nucleotides at the 5' of L-A dsRNAs in *S. cerevisiae* seem to have a *cis* signal for transcription (Esteban and Wickner, 1988). The full sequence of this 25 bp has assembled completely in three L-As from *S. paradoxus* strains: L-A-Q, L-A-D1, and L-A-P2.6, and partly in the other L-A dsRNA of *S. paradoxus* strains (Figure 4-3B). This sequence in *S. paradoxus* dsRNAs is identical, whereas there are variations in this sequence of the L-As in *S. cerevisiae.* Similar to other viruses, this region is AU-rich and seems to facilitate the melting of the dsRNA in order to create access for *gag* (Rodríguez-Cousiño et al., 2011).

There are two regions in the sequence of L-A from both species that are almost identical; 130 nucleotides in overlapped region, and 24-nucleotide stem-loop structure toward the 3' end of L-A. The overlapped region contained a stem-loop structure that is involved in frameshifting, nucleotides 1988-2034, adjacent to the slippery site, 1988GGGUUUA1995 (Dinman et al., 1991, Dinman and Wickner, 1992). There is only one residue between the stem-loop structure and the slippery site, nucleotide 1997, which is different in four of the *S. paradoxus* L-A dsRNAs: L-A-P2.4, L-A-P2.1, L-A-C, and L-A-P2.3, and one residue in the stem-loop structure of L-A-L1, which shows variation compared to the other L-A dsRNAs (Figure 4-3C). The 24-nucleotide stem-loop structure in (+) strand of L-A, (from residue 4218 to 4254) is 100% identical in all the dsRNAs from both species. This region is responsible for binding to the Gag-Pol fusion protein and encapsidation of dsRNA (Figure 4-3D).

Based on the tree aligning all L-As from both species (Figure 4-4) and the identities between all L-A (Table 4-3), there are three groups of L-A in *S. paradoxus* strains. Group 1 and group 2 are closer together than these two groups with group 3. The identity between the first two groups is between 80.6% and 82.7%, whereas the identity between these two groups and group 3 is between 73.3% and 77.8%. Since the identity between the different types of L-A dsRNA in *S. cerevisiae* is 73.7% and 77.8%, it seems that two types of L-A were identified in *S. paradoxus*, one of which has two subtypes.

Of all three strains that were sequenced separately, only one strain, DBVPG4650, had more than one L dsRNA, L-A-D1, L-A-L1, and L-BC. Although the number of the reads in L-A-L1` and L-BC is very low (1,253 and 406 reads respectively), they were assembled into both Ls quite specifically, and running the map-to-reference using more restricted settings did not change the result. Since none of the strains in this study contained L-A-L1 or L-BC, it would be unlikely that their reads come from contamination. The presence of low copy numbers of L-BC dsRNA with L-A was reported before (Sommer and Wickner, 1982). However, there are no reports with respect to the presence of two types of L-A in one strain found in nature. In addition, previous research showed that, in the case of two L-A present in one cell, one of them excludes the other from the cell (Rodríguez-Cousiño et al., 2013, Rodríguez-Cousiño and Esteban, 2017). There is a possibility that, instead of excluding, they affect the copy number of each other. Primarily, L-A-L1 is a weaker viral dsRNA in nature when compared to the other types of L-A dsRNA (Rodríguez-Cousiño et al., 2013). Both L-A-L1 and L-BC have been reported in *S. cerevisiae* and was just seen in this strain of S. *paradoxus*. In addition to these Ls, the M dsRNA in this strain is the only M, M28, in *S. paradoxus* that was reported in *S. cerevisiae*. Given the fact that in *S. cerevisiae* each type of L specifically evolves with specific M (Rodríguez-Cousiño and Esteban, 2017, Rodríguez-Cousiño et al., 2013), there is a possibility that L-A-D1 is the L28 in *S. cerevisiae*, which its sequence has not yet reported. The identity between this dsRNA and the L-As in *S. paradoxus* is higher that of *S. cerevisiae* (Table 4-3).

There are *S. cerevisiae* strains that have only L-A dsRNA. However, in this study, all infected strains contained both L and M. The result of cycloheximide treatment (Table 3-3) proved that L-A dsRNA is stable alone in the *S. paradoxus* strains, too. It seems that, as a result of natural selection, almost all the L-A viruses are equipped with M satellites in *S. paradoxus.*

## 6.7 Cloning the MQ ORF and expressing the KQ killer toxin in *S. cerevisiae*

In order to be certain about the sequence of dsRNAs and to test whether the killer toxins were encoded from the new types of M dsRNAs, the MQ ORF sequence from *S. paradoxus* strain, Q62.5, was cloned and expressed into *S. cerevisiae* and then tested for the killer phenotype. As mentioned before Q62.5 was the strongest killer strain amongst all the *S. paradoxus* strains that were studied in this project. Its killer-immune phenotype was not similar to any of the well-known killers in *S. cerevisiae*, K1, K2 and K28, when it was tested against them. It was the only strain which killed both *S. cerevisiae* and *S. paradoxus* strains. The killer phenotype in this strain was cured using cycloheximide treatment, and MQ were removed from Q62.5 after the treatment. These results suggest MQ carries the gene that encodes the killer toxin.

The ORF was firstly amplified using specified primers and then cloned in a *S. cerevisiae* strain. The result of the killer assay of the transformed cells indicated that the killer phenotype is expressed in the transformed cells. Nevertheless, the expression of the phenotype in the transformed strain was significantly less than in that of the original strain. Perhaps this arises from the medium, which had not been optimised for the expression of the *GAL4* promoter, or from the effect of the new host genome in the expression of the killer toxin in the post-translational stage.

## 6.8 Evolution of dsRNA in *S. cerevisiae* and *S. paradoxus*

As mentioned before, *S. paradoxus* is a close relative of *S. cerevisiae* (Goddard and Burt, 1999). The presence of M28, L-A-L1 and L-BC in both species, as well as the L-A dsRNAs with maximum 26% difference and also 99% identity of SCY_1690 ORF in *S. paradoxus* with *khs gene* sequence in *S. cerevisiae* indicates that, without any doubt, the viral dsRNA has the same origin in both species and transferred between them. On the one hand, the presence of wider types of M in *S. paradoxus*, especially finding five new types of M just in Silwood Park, more glycosylation sites in the preprotoxin of the *S. paradoxus*, wider immunity to killer toxins in *S. paradoxus* strains, support the theory of the evolution of dsRNA in *S. paradoxus* and transfer to *S. cerevisiae* (Chang et al., 2015). In addition, L-A-D1, which seems to be the M28 helper, is closer to the *S. paradoxus* Ls than the *S. cerevisiae* Ls (Table 4-3). On the other hand, the L-A helper of each type M is more diverse in *S. cerevisiae* than in *S. paradoxus*. For the three types of M for which their L has been reported in *S. cerevisiae*, the identity between the Ls is around 75%, whereas in *S. paradoxus* for eight types of M only two variants of L with that much diversity were detected. This can be seen from two points of view: firstly, as result of the longer evolution of L and M in *S. cerevisiae*, they become more specified. Secondly, because the evolution of the viruses in *S. paradoxus* was longer than that in S. *cerevisiae*, each type of L has been equipped with different types of M and maybe this is the reason no L-A found alone in this species. Overall, most of the results in this study support the second point of view.

Horizontal gene transfer, in both *Totivirus* and yeast species and between their genomes (Taylor and Bruenn, 2009), and also introgression in yeast species especially between *S. paradoxus* and *S. cerevisiae* have been reported. However, Horizontal gene transfer is rare in yeast species (Liti and Louis, 2005). In nature, *S. paradoxus* and *S. cerevisiae* not only has being found together (Sniegowski et al., 2002) but also naturally occurring hybrid between them has been found (Liti et al., 2005). Since, the exogenous phase of mycoviruses has not reported, it is more likely that introgression of the yeast species is the cause of movement of the viral dsRNA between *S. paradoxus* and *S. cerevisiae strains.*

## 6.9 Potential applications of the killer system in *S. paradoxus*

As mentioned in chapter 1, killer yeasts and their toxins can have a range of applications. They can be used in the taxonomy of yeasts, optimising fermentation (which is widely used in industry), food preservatives, medicines, transgenic plants (to improve their resistance), and they can also be used as a model to study the interaction between host and virus as well as in modelling post-translational activity (Schmitt and Breinig, 2002, Marquina et al., 2002).

Identifying seven new types of M dsRNA and their killer toxins provides more opportunities for all of the mentioned applications. However, future characterisations of the killer toxins are needed to choose the best killer toxin for each application.  The effect of ethanol on the killer toxins studied in this study is one of the important factors that should be considered in most of the applications particularly optimisation of fermentation, food preservatives, and medicines. Choosing a killer toxin which is stable in ethanol or a toxin that its efficiency increase in presence of ethanol, like KHS (de Ullivarri et al., 2014), will have material impact on the optimization of fermentation and efficiency of the food preservatives. Moreover, knowing the effect of the killer toxin in the other strains' growth rate is the other key factor that should be consider in optimisation of fermentation, food preservatives medicine, and transgenic plants.

Prediction of signal cleavage sites, the kex2p/kex1p cleavage sites, and N-glycosylation sites in the founded preprotoxins suggested that the entire identified killer toxin in this study go through post translational modification. As a result, all these preprotoxins could be good candidates in modelling post-translational activity.

## 6.10 Conclusion

The killer system in *S. paradoxus* is very complex, as with other yeasts. Similar to *S. cerevisiae* killer strains, in the majority of *S. paradoxus* killer strains the viral dsRNAs play an active role in this system. However, there are some killer strains, Canadian strains and one strain from Continental Europe, C15 (CECT10176), for which other genetic resources encode the killer phenotype. Since the killer phenotype in the Canadian killer strains is different and no dsRNAs were found in the strains of this area, it seems that the killer system in the Canadian strains evolved separately. In the infected killer strains the dsRNAs show similarity to *S. cerevisiae* viruses. The sequence of the L-A dsRNA virus in this species has homology with that of *S. cerevisiae.* Regarding M dsRNA, only one type of the *S. cerevisiae* M dsRNA, M28, was found in the *S. paradoxus* strains. The other discovered M dsRNAs are new. Even though the sequence of the new types of M dsRNA have low similarity with the M dsRNAs in *S. cerevisiae*, all the Ms and their preprotoxins have the same structure as that of *S. cerevisiae.*

The killer-immune reaction between the *S. paradoxus* strains and both species' strains is also complex. This arises from, firstly, the different killer toxins that exist in this species; secondly, the different genomic background of each strain. Overall, it seems that there is an additional immunity in the *S. paradoxus* strains that decreases the level of expression of the killer phenotype in the killer strains from both species. In addition to all of the immunity reported so far, it seems that there is some mechanism in all of the strains, including the sensitive strains, which does not allow the killer toxin to kill all of the cells exposed to it. This mechanism, which is essential for the survival of the strains, might result from environmental changes during fermentation, such as increased ethanol and decreased pH, or some immunity expressed from the genome of the strains.

The presence of wider types of M in *S. paradoxus*, more glycosylation sites in the preprotoxin of *S. paradoxus*, wider immunity to killer toxins in *S. paradoxus* strains, support the theory of the evolution of dsRNA in *S. paradoxus* and transfer to *S. cerevisiae.*

## 6.11 Further perspectives

In order to get a better picture of the killer system in *S. paradoxus* and study the evolution of viral dsRNA in this species, some future work is needed:

1. Determine which dsRNA belongs to which strains for the dsRNA found in Pool1 and Pool2. This can be done using RT-PCR or dot blot analyses.
2. Determine the sequence of the ends of the dsRNAs.
3. Study the evolutionary genetics of the L and M dsRNAs in *S. paradoxus* and *S. cerevisiae.*
4. Investigate the relationship between the viral dsRNA and nuclear genome of the killer strain; the sequence of the genomes of most of the infected strains is available.
5. Express the killer toxin of the other M dsRNAs to make sure that they are the genetic basis of the killer phenotype in the killer strains.
6. Purify the expressed killer toxin in order to sequence and characterise the killer toxins.
7. Study the effect of the ethanol on the killer toxins.
8. Study the effect of each killer toxin on the growth rate of the other yeast strain and their mode of action.
9. Study the immunity pathways in the immune strains.

# 7 REFERENCES

AHMED, A., SESTI, F., ILAN, N., SHIH, T. M., STURLEY, S. L. & GOLDSTEIN, S. A. N. 1999. A Molecular Target for Viral Killer Toxin:: TOK1 Potassium Channels. *Cell,* 99**,** 283-291.

ATTOUI, H., BILLOIR, F., CANTALOUBE, J. F., BIAGINI, P., DE MICCO, P. & DE LAMBALLERIE, X. 2000. Strategies for the sequence determination of viral dsRNA genomes. *Journal of virological methods,* 89**,** 147-158.

BOLLAG, D., ROZYCKI, M. & EDELSTEIN, S. 1996. *Protein methods,* New York, John Wiley & Sons Inc.

BOONE, C., SOMMER, S. S., HENSEL, A. & BUSSEY, H. 1990. Yeast KRE genes provide evidence for a pathway of cell wall beta-glucan assembly. *The Journal of cell biology,* 110**,** 1833.

BOSTIAN, K. A., ELLIOTT, Q., BUSSEY, H., BUM, V., SMITH, A. & TIPPER, D. J. 1984. Sequence of the preprotoxin dsRNA gene of type I killer yeast: multiple processing events produce a two-component toxin. *Cell,* 36**,** 741-751.

BREINIG, F., SENDZIK, T., EISFELD, K. & SCHMITT, M. J. 2006. Dissecting toxin immunity in virus-infected killer yeast uncovers an intrinsic strategy of self-protection. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 3810.

BRUENN, J. A. 1991. Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic acids research,* 19**,** 217-226.

BRUENN, J. A. 1993. A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Research,* 21**,** 5667-5669.

BUSSEY, H., SAVILLE, D., HUTCHINS, K. & PALFREE, R. 1979a. Binding of yeast killer toxin to a cell wall receptor on sensitive Saccharomyces cerevisiae. *Journal of Bacteriology,* 140**,** 888.

BUSSEY, H., SAVILLE, D., HUTCHINS, K. & PALFREE, R. 1979b. Binding of yeast killer toxin to a cell wall receptor on sensitive Saccharomyces cerevisiae. *Journal of bacteriology,* 140**,** 888-892.

CARROLL, K. & WICKNER, R. B. 1995. Translation and M1 double-stranded RNA propagation: MAK18= RPL41B and cycloheximide curing. *Journal of bacteriology,* 177**,** 2887-2891.

CASTILLO, A. & CIFUENTES, V. 1994. Presence of double-stranded RNA and virus-like particles in Phaffia rhodozyma. *Current Genetics,* 26**,** 364-368.

CHANG, S. L., LEU, J. Y. & CHANG, T. H. 2015. A population study of killer viruses reveals different evolutionary histories of two closely related Saccharomyces sensu stricto yeasts. *Molecular ecology,* 24**,** 4312-4322.

CHENG, R. H., CASTON, J. R., WANG, G.-J., GU, F., SMITH, T. J., BAKER, T. S., BOZARTH, R. F., TRUS, B. L., CHENG, N. & WICKNER, R. B. 1994. Fungal virus capsids, cytoplasmic compartments for the replication of double-stranded RNA, formed as icosahedral shells of asymmetric Gag dimers. *Journal of molecular biology,* 244**,** 255-258.

CIANI, M. & FATICHENTI, F. 2001. Killer toxin of Kluyveromyces phaffii DBVPG 6076 as a biopreservative agent to control apiculate wine yeasts. *Applied and environmental microbiology,* 67**,** 3058.

COENEN, A., KEVEI, F. & HOEKSTRA, R. 1997. Factors affecting the spread of double-stranded RNA viruses in Aspergillus nidulans. *Genetics Research,* 69**,** 1-10.

DE ULLIVARRI, M. F., MENDOZA, L. M. & RAYA, R. R. 2014. Killer activity of Saccharomyces cerevisiae strains: partial characterization and strategies to improve the biocontrol efficacy in winemaking. *Antonie Van Leeuwenhoek,* 106**,** 865-78.

DIGNARD, D., WHITEWAY, M., GERMAIN, D., TESSIER, D. & THOMAS, D. 1991. Expression in yeast of a cDNA copy of the K2 killer toxin gene. *Molecular and General Genetics MGG,* 227**,** 127-136.

DINMAN, J. D., ICHO, T. & WICKNER, R. B. 1991. A-1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proceedings of the National Academy of Sciences,* 88**,** 174-178.

DINMAN, J. D. & WICKNER, R. B. 1992. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *Journal of Virology,* 66**,** 3669-3676.

DONIGER, S. W., KIM, H. S., SWAIN, D., CORCUERA, D., WILLIAMS, M., YANG, S.-P. & FAY, J. C. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS genetics,* 4**,** e1000183.

DRINNENBERG, I. A., FINK, G. R. & BARTEL, D. P. 2011. Compatibility with killer explains the rise of RNAi-deficient fungi. *Science,* 333**,** 1592-1592.

DRINNENBERG, I. A., WEINBERG, D. E., XIE, K. T., MOWER, J. P., WOLFE, K. H., FINK, G. R. & BARTEL, D. P. 2009. RNAi in budding yeast. *Science,* 326**,** 544-550.

ESTEBAN, L., RODRÍGUEZ-COUSIÑO, N. & ESTEBAN, R. 1992a. T double-stranded RNA (dsRNA) sequence reveals that T and W dsRNAs form a new RNA family in Saccharomyces cerevisiae. Identification of 23 S RNA as the single-stranded form of T dsRNA. *Journal of Biological Chemistry,* 267**,** 10874.

ESTEBAN, L. M., RODRIGUEZ-COUSIÑO, N. & ESTEBAN, R. 1992b. T double-stranded RNA (dsRNA) sequence reveals that T and W dsRNAs form a new RNA family in Saccharomyces cerevisiae. Identification of 23 S RNA as the single-stranded form of T dsRNA. *Journal of Biological Chemistry,* 267**,** 10874-10881.

ESTEBAN, R., FUJIMURA, T. & WICKNER, R. 1989. Internal and terminal cis-acting sites are necessary for in vitro replication of the LA double-stranded RNA virus of yeast. *The EMBO journal,* 8**,** 947.

ESTEBAN, R. & WICKNER, R. B. 1986a. Three different M1 RNA-containing viruslike particle types in Saccharomyces cerevisiae: in vitro M1 double-stranded RNA synthesis. *Molecular and cellular biology,* 6**,** 1552.

ESTEBAN, R. & WICKNER, R. B. 1986b. Three different M1 RNA-containing viruslike particle types in Saccharomyces cerevisiae: in vitro M1 double-stranded RNA synthesis. *Molecular and cellular biology,* 6**,** 1552-1561.

ESTEBAN, R. & WICKNER, R. B. 1988. A deletion mutant of LA double-stranded RNA replicates like M1 double-stranded RNA. *Journal of virology,* 62**,** 1278-1285.

FAUQUET, C. M., MAYO, M. A., MANILOFF, J., DESSELBERGER, U. & BALL, L. A. 2005. *Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses*, Academic Press.

FINK, G. R. & STYLES, C. A. 1972. Curing of a killer factor in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences,* 69**,** 2846-2849.

FRANK, A. C. & WOLFE, K. H. 2009. Evolutionary capture of viral and plasmid DNA by yeast nuclear chromosomes. *Eukaryotic cell,* 8**,** 1521-1531.

FRANK, S. A. 1994. Spatial polymorphism of bacteriocins and other allelopathic traits. *Evolutionary Ecology,* 8**,** 369-386.

FRIED, H. & FINK, G. 1978. Electron microscopic heteroduplex analysis of" killer" double-stranded RNA species from yeast. *Proceedings of the National Academy of Sciences of the United States of America,* 75**,** 4224.

FRIED, H. M. & WARNER, J. R. 1982. Molecular cloning and analysis of yeast gene for cycloheximide resistance and ribosomal protein L29. *Nucleic acids research,* 10**,** 3133-3148.

FROUSSARD, P. 1992. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic acids research,* 20**,** 2900-2900.

FUJIMURA, T. & WICKNER, R. B. 1987. LA double-stranded RNA viruslike particle replication cycle in Saccharomyces cerevisiae: particle maturation in vitro and effects of mak10 and pet18 mutations. *Molecular and cellular biology,* 7**,** 420-426.

FUJIMURA, T. & WICKNER, R. B. 1992. Interaction of two cis sites with the RNA replicase of the yeast LA virus. *Journal of Biological Chemistry,* 267**,** 2708-2713.

GHABRIAL, S. A. 1998. Origin, adaptation and evolutionary pathways of fungal viruses. *Virus genes,* 16**,** 119-131.

GHABRIAL, S. A. & SUZUKI, N. 2009. Viruses of plant pathogenic fungi. *Annual review of phytopathology,* 47**,** 353-384.

GHILDIYAL, M. & ZAMORE, P. D. 2009. Small silencing RNAs: an expanding universe. *Nature Reviews Genetics,* 10**,** 94.

GODDARD, M. R. & BURT, A. 1999. Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 13880.

GOTO, K., FUKUDA, H., KICHISE, K., KITANO, K. & HARA, S. 1991. Cloning and nucleotide sequence of the KHS killer gene of Saccharomyces cerevisiae. *Agricultural and biological chemistry,* 55**,** 1953-1958.

GOTO, K., IWASE, T., KICHISE, K., KITANO, K., TOTUKA, A., OBATA, T. & HARA, S. 1990a. Isolation and properties of a chromosome-dependent KHR killer toxin in Saccharomyces cerevisiae. *Agricultural and biological chemistry,* 54**,** 505-509.

GOTO, K., IWATUKI, Y., KITANO, K., OBATA, T. & HARA, S. 1990b. Cloning and nucleotide sequence of the KHR killer gene of Saccharomyces cerevisiae. *Agricultural and biological chemistry,* 54**,** 979-984.

GOTO, K., IWATUKI, Y., KITANO, K., OBATA, T. & HARA, S. 1990c. Cloning and nucleotide sequence of the KHR killer gene of Saccharomyces cerevisiae. *Agric Biol Chem,* 54**,** 979-84.

GREIG, D. & TRAVISANO, M. 2008a. DENSITY-DEPENDENT EFFECTS ON ALLELOPATHIC INTERACTIONS IN YEAST. *Evolution,* 62**,** 521-527.

GREIG, D. & TRAVISANO, M. 2008b. DENSITY DEPENDENT EFFECTS ON ALLELOPATHIC INTERACTIONS IN YEAST. *Evolution,* 62**,** 521-527.

HA, E. S., YIE, S. W. & CHOI, H. T. 1997. Biochemical Characteristics of a Killer Toxin Produced by Ustilago maydis Virus SH14 Isolated in Korea. *The Journal of Microbiology***,** 323-326.

HEARD, G. & FLEET, G. 1987. Occurrence and growth of killer yeasts during wine fermentation. *Applied and environmental microbiology,* 53**,** 2171.

HELENIUS, A. & AEBI, M. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annual review of biochemistry,* 73**,** 1019-1049.

HOPPER, J., BOSTIAN, K., ROWE, L. & TIPPER, D. 1977a. Translation of the L-species dsRNA genome of the killer-associated virus-like particles of Saccharomyces cerevisiae. *Journal of Biological Chemistry,* 252**,** 9010.

HOPPER, J. E., BOSTIAN, K., ROWE, L. & TIPPER, D. 1977b. Translation of the L-species dsRNA genome of the killer-associated virus-like particles of Saccharomyces cerevisiae. *Journal of Biological Chemistry,* 252**,** 9010-9017.

ICHO, T. & WICKNER, R. 1989a. The double-stranded RNA genome of yeast virus LA encodes its own putative RNA polymerase by fusing two open reading frames. *Journal of Biological Chemistry,* 264**,** 6716.

ICHO, T. & WICKNER, R. B. 1989b. The double-stranded RNA genome of yeast virus LA encodes its own putative RNA polymerase by fusing two open reading frames. *Journal of Biological Chemistry,* 264**,** 6716-6723.

IVANNIKOVA, Y. V., NAUMOVA, E. S. & NAUMOV, G. I. 2007. Viral dsRNA in the wine yeast Saccharomyces bayanus var. uvarum. *Research in microbiology,* 158**,** 638-643.

KANG, J.-G., WU, J.-C., BRUENN, J. A. & PARK, C.-M. 2001. The H1 double-stranded RNA genome of Ustilago maydis virus-H1 encodes a polyprotein that contains structural motifs for capsid polypeptide, papain-like protease, and RNA-dependent RNA polymerase. *Virus research,* 76**,** 183-189.

KITANO, K., SATO, M., SHIMAZAKI, T. & HARA, S. 1984. Occurrence of wild killer yeasts in Japanese wineries and their characteristics. *Journal of fermentation technology,* 62**,** 1-6.

KONDO, H., HISANO, S., CHIBA, S., MARUYAMA, K., ANDIKA, I. B., TOYODA, K., FUJIMORI, F. & SUZUKI, N. 2016. Sequence and phylogenetic analyses of novel totivirus-like double-stranded RNAs from field-collected powdery mildew fungi. *Virus research,* 213**,** 353-364.

LAMBDEN, P., COOKE, S., CAUL, E. & CLARKE, I. 1992. Cloning of noncultivatable human rotavirus by single primer amplification. *Journal of virology,* 66**,** 1817-1822.

LEVIN, B., ANTONOVICS, J. & SHARMA, H. 1988. Frequency-Dependent Selection in Bacterial Populations [and Discussion]. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences,* 319**,** 459.

LI, H., HAVENS, W. M., NIBERT, M. L. & GHABRIAL, S. A. 2011. RNA sequence determinants of a coupled termination-reinitiation strategy for downstream open reading frame translation in Helminthosporium victoriae virus 190S and other victoriviruses (Family Totiviridae). *Journal of virology,* 85**,** 7343-7352.

LITI, G., CARTER, D. M., MOSES, A. M., WARRINGER, J., PARTS, L., JAMES, S. A., DAVEY, R. P., ROBERTS, I. N., BURT, A. & KOUFOPANOU, V. 2009. Population genomics of domestic and wild yeasts. *Nature,* 458**,** 337.

LITI, G. & LOUIS, E. J. 2005. Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.,* 59**,** 135-153.

LITI, G., PERUFFO, A., JAMES, S. A., ROBERTS, I. N. & LOUIS, E. J. 2005. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the Saccharomyces sensu stricto complex. *Yeast,* 22**,** 177-192.

LIU, G.-L., CHI, Z., WANG, G.-Y., WANG, Z.-P., LI, Y. & CHI, Z.-M. 2015a. Yeast killer toxins, molecular mechanisms of their action and their applications. *Critical reviews in biotechnology,* 35**,** 222-234.

LIU, G. L., CHI, Z., WANG, G. Y., WANG, Z. P., LI, Y. & CHI, Z. M. 2015b. Yeast killer toxins, molecular mechanisms of their action and their applications. *Crit Rev Biotechnol,* 35**,** 222-34.

LOPES, C. & SANGORRÍN, M. 2010. Optimization of killer assays for yeast selection protocols. *Rev Argent Microbiol,* 42**,** 298-306.

MAAN, S., RAO, S., MAAN, N. S., ANTHONY, S. J., ATTOUI, H., SAMUEL, A. R. & MERTENS, P. P. C. 2007. Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. *Journal of virological methods,* 143**,** 132-139.

MAGLIANI, W., CONTI, S., GERLONI, M., BERTOLOTTI, D. & POLONELLI, L. 1997a. Yeast killer systems. *Clinical microbiology reviews,* 10**,** 369.

MAGLIANI, W., CONTI, S., GERLONI, M., BERTOLOTTI, D. & POLONELLI, L. 1997b. Yeast killer systems. *Clinical microbiology reviews,* 10**,** 369-400.

MALONE, C. D. & HANNON, G. J. 2009. Small RNAs as guardians of the genome. *Cell,* 136**,** 656-668.

MARQUINA, D., SANTOS, A. & PEINADO, J. 2002. Biology of killer yeasts. *International Microbiology,* 5**,** 65-71.

MCBRIDE, R., GREIG, D. & TRAVISANO, M. 2008. Fungal viral mutualism moderated by ploidy. *Evolution,* 62**,** 2372-2380.

MCBRIDE, R. C., BOUCHER, N., PARK, D. S., TURNER, P. E. & TOWNSEND, J. P. 2013. Yeast response to LA virus indicates coadapted global gene expression during mycoviral infection. *FEMS Yeast Res,* 13**,** 162-79.

MESKAUSKAS, A. 1990. Nucleotide sequence of cDNA to yeast M2-1 dsRNA segment. *Nucleic acids research,* 18**,** 6720.

MOAZED, D. 2009. Rejoice—RNAi for Yeast. *Science,* 326**,** 533-534.

NAUMOV, G., IVANNIKOIVA, I. & NAUMOVA, E. 2004. [Molecular polymorphism of viral dsRNA of yeast Saccharomyces paradoxus]. *Molekuliarnaia genetika, mikrobiologiia i virusologiia***,** 38-40.

NAUMOV, G., IVANNIKOIVA, I. V. & NAUMOVA, E. 2005. Molecular polymorphism of viral dsRNA of yeast Saccharomyces paradoxus]. *Molekuliarnaia genetika, mikrobiologiia i virusologiia***,** 38.

NAUMOV, G., IVANNIKOVA, Y. V., CHERNOV, I. Y. & NAUMOVA, E. 2009. Natural polymorphism of the plasmid double-stranded RNA of the Saccharomyces yeasts. *Microbiology,* 78**,** 208-213.

PALFREE, R. G. & BUSSEY, H. 1979. Yeast killer toxin: purification and characterisation of the protein toxin from Saccharomyces cerevisiae. *European Journal of Biochemistry,* 93**,** 487-493.

PARK, C.-M., LOPINSKI, J. D., MASUDA, J., TZENG, T.-H. & BRUENN, J. A. 1996. A second double-stranded RNA virus from yeast. *Virology,* 216**,** 451-454.

PEARSON, M. N., BEEVER, R. E., BOINE, B. & ARTHUR, K. 2009. Mycoviruses of filamentous fungi and their relevance to plant pathology. *Molecular Plant Pathology,* 10**,** 115-128.

PIECZYNSKA, M. D., DE VISSER, J. A. & KORONA, R. 2013a. Incidence of symbiotic dsRNA 'killer' viruses in wild and domesticated yeast. *FEMS Yeast Res,* 13**,** 856-9.

PIECZYNSKA, M. D., DE VISSER, J. A. G. & KORONA, R. 2013b. Incidence of symbiotic dsRNA 'killer'viruses in wild and domesticated yeast. *FEMS yeast research,* 13**,** 856-859.

POLASHOCK, J. J. & HILLMAN, B. I. 1994. A small mitochondrial double-stranded (ds) RNA element associated with a hypovirulent strain of the chestnut blight fungus and ancestrally related to yeast cytoplasmic T and W dsRNAs. *Proceedings of the National Academy of Sciences,* 91**,** 8680-8684.

POTGIETER, A., PAGE, N., LIEBENBERG, J., WRIGHT, I., LANDT, O. & VAN DIJK, A. 2009. Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *Journal of General Virology,* 90**,** 1423-1432.

POTGIETER, A., STEELE, A. & VAN DIJK, A. 2002. Cloning of complete genome sets of six dsRNA viruses using an improved cloning method for large dsRNA genes. *Journal of general virology,* 83**,** 2215-2223.

RIBAS, J. C. & WICKNER, R. B. 1998. The Gag Domain of the Gag-Pol Fusion Protein Directs Incorporation into the LA Double-stranded RNA Viral Particles inSaccharomyces cerevisiae. *Journal of Biological Chemistry,* 273**,** 9306-9311.

RODRIGUEZ-COUSIÑO, N., ESTEBAN, L. M. & ESTEBAN, R. 1991. Molecular cloning and characterization of W double-stranded RNA, a linear molecule present in Saccharomyces cerevisiae. Identification of its single-stranded RNA form as 20 S RNA. *Journal of Biological Chemistry,* 266**,** 12772-12778.

RODRÍGUEZ-COUSIÑO, N. & ESTEBAN, R. 2017. Relationships and evolution of double-stranded RNA totiviruses of yeasts inferred from analysis of LA-2 and L-BC variants in wine yeast strain populations. *Applied and environmental microbiology,* 83**,** e02991-16.

RODRÍGUEZ-COUSIÑO, N., GÓMEZ, P. & ESTEBAN, R. 2013. LA-lus, a new variant of the LA Totivirus found in wine yeasts with Klus killer toxin-encoding Mlus double-stranded RNA: possible role of killer toxin-encoding satellite RNAs in the evolution of their helper viruses. *Applied and environmental microbiology,* 79**,** 4661-4674.

RODRÍGUEZ-COUSIÑO, N., MAQUEDA, M., AMBRONA, J., ZAMORA, E., ESTEBAN, R. & RAMÍREZ, M. 2011. A new wine Saccharomyces cerevisiae killer toxin (Klus), encoded by a double-stranded RNA virus, with broad antifungal activity is evolutionarily related to a chromosomal host gene. *Applied and environmental microbiology,* 77**,** 1822-1832.

RUDERFER, D. M., PRATT, S. C., SEIDEL, H. S. & KRUGLYAK, L. 2006. Population genomic analysis of outcrossing and recombination in yeast. *Nature genetics,* 38**,** 1077.

RUSSELL, P. J., BENNETT, A. M., LOVE, Z. & BAGGOTT, D. M. 1997. Cloning, Sequencing and Expression of a Full-Length cDNA Copy of the M1 Double-Stranded RNA Virus from the Yeast, Saccharomyces cerevisiae. *Yeast,* 13**,** 829-836.

SALZMAN, R., FUJITA, T., ZHU-SALZMAN, K., HASEGAWA, P. & BRESSAN, R. 1999. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant Molecular Biology Reporter,* 17**,** 11-17.

SANTOS, A. & MARQUINA, D. 2004. Killer toxin of Pichia membranifaciens and its possible use as a biocontrol agent against grey mould disease of grapevine. *Microbiology,* 150**,** 2527.

SANTOS, A., SANCHEZ, A. & MARQUINA, D. 2004. Yeasts as biological agents to control Botrytis cinerea. *Microbiological Research,* 159**,** 331-338.

SCHMITT, M. & BREINIG, F. 2002. The viral killer system in yeast: from molecular biology to application. *FEMS microbiology reviews,* 26**,** 257-276.

SCHMITT, M. & RADLER, F. 1988. Molecular structure of the cell wall receptor for killer toxin KT28 in Saccharomyces cerevisiae. *Journal of Bacteriology,* 170**,** 2192.

SCHMITT, M. & TIPPER, D. 1990. K28, a unique double-stranded RNA killer virus of Saccharomyces cerevisiae. *Molecular and cellular biology,* 10**,** 4807.

SCHMITT, M. J. & BREINIG, F. 2006. Yeast viral killer toxins: lethality and self-protection. *Nature Reviews Microbiology,* 4**,** 212-221.

SCHMITT, M. J. & NEUHAUSEN, F. 1994. Killer toxin-secreting double-stranded RNA mycoviruses in the yeasts Hanseniaspora uvarum and Zygosaccharomyces bailii. *Journal of virology,* 68**,** 1765.

SCHMITT, M. J. & TIPPER, D. J. 1992. Genetic analysis of maintenance and expression of L and M double stranded RNAs from yeast killer virus K28. *Yeast,* 8**,** 373-384.

SCHMITT, M. J. & TIPPER, D. J. 1995. Sequence of the M28 dsRNA: preprotoxin is processed to an α/β heterodimeric protein toxin. *Virology,* 213**,** 341-351.

SERVIENĖ, E., LUKŠA, J., ORENTAITĖ, I., LAFONTAINE, D. L. & URBONAVIČIUS, J. 2012. Screening the budding yeast genome reveals unique factors affecting K2 toxin susceptibility. *PloS one,* 7**,** e50779.

SNIEGOWSKI, P. D., DOMBROWSKI, P. G. & FINGERMAN, E. 2002. Saccharomyces cerevisiae and Saccharomyces paradoxus coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS yeast research,* 1**,** 299-306.

SOMMER, S. S. & WICKNER, R. B. 1982. Yeast L dsRNA consists of at least three distinct RNAs; evidence that the non-Mendelian genes [HOK],[NEX] and [EXL] are on one of these dsRNAs. *Cell,* 31**,** 429-441.

STARMER, W., GANTER, P. & ABERDEEN, V. 1987. The ecological role of killer yeasts in natural communities of yeasts. *Canadian Journal of Microbiology/Revue Canadienne de Microbiologie,* 33**,** 783-796.

STÖCKLEIN, W. & PIEPERSBERG, W. 1980. Altered ribosomal protein L29 in a cycloheximide-resistant strain of Saccharomyces cerevisiae. *Current genetics,* 1**,** 177-183.

STURLEY, S. L., ELLIOT, Q., LEVITRE, J., TIPPER, D. J. & BOSTIAN, K. A. 1986. Mapping of functional domains within the Saccharomyces cerevisiae type 1 killer preprotoxin. *The EMBO Journal,* 5**,** 3381.

TAYLOR, D. J. & BRUENN, J. 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC biology,* 7**,** 88.

TIPPER, D. J. & BOSTIAN, K. 1984. Double-stranded ribonucleic acid killer systems in yeasts. *Microbiological reviews,* 48**,** 125.

VARGA, J., VÁGVÖLGYI, C. & TÓTH, B. 2003. Recent advances in mycovirus research. *Acta microbiologica et immunologica hungarica,* 50**,** 77-94.

VAUGHAN-MARTINI, A. & MARTINI, A. 1995. Facts, myths and legends on the prime industrial microorganism. *Journal of Industrial Microbiology and Biotechnology,* 14**,** 514-522.

WEI, W., MCCUSKER, J. H., HYMAN, R. W., JONES, T., NING, Y., CAO, Z., GU, Z., BRUNO, D., MIRANDA, M. & NGUYEN, M. 2007. Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. *Proceedings of the National Academy of Sciences,* 104**,** 12825-12830.

WELSH, J. D. & LEIBOWITZ, M. J. 1982. Localization of genes for the double-stranded RNA killer virus of yeast. *Proceedings of the National Academy of Sciences,* 79**,** 786-789.

WESOLOWSKI, M. & WICKNER, R. 1984. Two new double-stranded RNA molecules showing non-mendelian inheritance and heat inducibility in Saccharomyces cerevisiae. *Molecular and cellular biology,* 4**,** 181.

WICKNER, R. 1974. Chromosomal and nonchromosomal mutations affecting the" killer character" of Saccharomyces cerevisiae. *Genetics,* 76**,** 423.

WICKNER, R. 1983. Killer systems in Saccharomyces cerevisiae: three distinct modes of exclusion of M2 double-stranded RNA by three species of double-stranded RNA, M1, LAE, and LA-HN. *Molecular and cellular biology,* 3**,** 654.

WICKNER, R., BUSSEY, H., FUJIMURA, T. & ESTEBAN, R. 1995. Viral RNA and the killer phenomenon of Saccharomyces. *Genetics and Biotechnology.* Springer.

WICKNER, R. & LEIBOWITZ, M. 1976. Two chromosomal genes required for killing expression in killer strains of Saccharomyces cerevisiae. *Genetics,* 82**,** 429.

WICKNER, R. B. 1996. Double-stranded RNA viruses of Saccharomyces cerevisiae. *Microbiological reviews,* 60**,** 250.

WOODS, D. & BEVAN, E. 1968. Studies on the nature of the killer factor produced by Saccharomyces cerevisiae. *Microbiology,* 51**,** 115.

YU, X., LI, B., FU, Y., JIANG, D., GHABRIAL, S. A., LI, G., PENG, Y., XIE, J., CHENG, J. & HUANG, J. 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences,* 107**,** 8387-8392.

ZEYL, C. 2000. Budding yeast as a model organism for population genetics. *Yeast,* 16**,** 773-784.

ZHU, Y. S., KANE, J., ZHANG, X. Y., ZHANG, M. & TIPPER, D. J. 1993. Role of the component of preprotoxin in expression of the yeast K1 killer phenotype. *Yeast,* 9**,** 251-266.

# 8 Appendix

## 8.1 Chapter 2 appendix

**Table 8-1. The killer phenotype of *S. paradoxus* strains.**

| Location | Strain | Killer phenotype |
|---|---|---|
| Silwood | (C17) Q31.4 | NK |
| Silwood | Q32.3 | NK |
| Silwood | Q69.8 | NK |
| Silwood | Q70.8 | NK |
| Silwood | Q89.8 | NK |
| Silwood | T62.1 | NK |
| Silwood | Y2 | NK |
| Silwood | Y2.2 | NK |
| Silwood | Y3 | NK |
| Silwood | Y4 | NK |
| Silwood | Y4.5 | NK |
| Silwood | Y5 | NK |
| Silwood | Y5.1 | NK |
| Silwood | Y5.6 | NK |
| Silwood | Y5.8 | NK |
| Silwood | Y6 | NK |
| Silwood | Y6.2 | NK |
| Silwood | Y7 | NK |
| Silwood | Y7.1 | NK |
| Silwood | Y7.5 | NK |
| Silwood | Y8 | NK |
| Silwood | Y8.1 | NK |
| Silwood | Y8.4 | NK |
| Silwood | Y8.6 | NK |
| Silwood | Y8.8 | NK |
| Silwood | Y9 | NK |
| Silwood | Y6.5 | NK |
| Silwood | Y7 | NK |
| Continental Europe | (C02) OS20 | K |
| Continental Europe | (C05) OS11,12W | K |
| Continental Europe | (C07) OS3,4Wa | K |
| Continental Europe | (C10) 4/2S W2 | K |
| Continental Europe | (C15) CECT10176 | K |
| Continental Europe | (C03) OS5,6W(1) | NK |
| Continental Europe | (C06) O14-3,4W | NK |
| Continental Europe | (C18) Z3 | NK |
| Continental Europe | (C14) DBVPG4650 | K |
| Continental Europe | N17 | NK |
| Continental Europe | KNP3828 | NK |
| Continental Europe | CBS5829 | NK |
| Continental Europe | YPS3 | NK |

| Location | Strain | Killer phenotype |
|---|---|---|
| Continental Europe | (C13) SIG1 | NK |
| Continental Europe | STOC3 | NK |
| FAR EAST | (C23) CBS8439 | K |
| FAR EAST | (C25) CBS8441 | K |
| FAR EAST | (C27) CBS8444 | K |
| FAR EAST | (C20) CBS8436 | NK |
| FAR EAST | (C21) CBS8437 | NK |
| FAR EAST | (C22) CBS8438 | NK |
| FAR EAST | (C24) CBS8440 | NK |
| FAR EAST | (C26) CBS8442 | NK |
| FAR EAST | N43 | NK |
| FAR EAST | KPN3829 | NK |
| FAR EAST | IFO1804 | NK |
| FAR EAST | N44 | NK |
| North America (Canada) | A12 | K |
| North America (Canada) | A19 | K |
| North America (Canada) | A21 | K |
| North America (Canada) | A22 | K |
| North America (Canada) | A23 | K |
| North America (Canada) | A24 | K |
| North America (Canada) | A25 | K |
| North America (Canada) | A27 | K |
| North America (Canada) | A28 | K |
| North America (Canada) | A33 | NK |
| North America (Canada) | A4 | NK |
| North America (Canada) | A17 | K |
| North America (U.S.A) | YPS138 | NK |
| North America (U.S.A) | DBVPG6304 | NK |
| North America (U.S.A) | VWOPS91-917-1 | NK |
| South America | UFRJ50791 | K |
| South America | UFRJ50816 | K |

**Table 8-2. Media (1L):**

| | |
|---|---|
| YPD | Yeast extract 10 gr<br>Peptone 20 gr<br>Glucose 20 gr |
| YPD | Yeast extract 10 gr<br>Peptone 20 gr<br>Glucose 20 gr<br>Agar 18 gr |
| YPAD (pH4.5) | Yeast extract 10 gr<br>Peptone 20 gr<br>Glucose 20 gr<br>Brij 0.1gr |
| MB Agar | Yeast extract 10 gr<br>Peptone 20 gr<br>Glucose 20 gr<br>Agar 18 gr<br>Citric Acid 14.07 gr<br>K2Hpo4 18.96<br>Methylene blue 0.0003 |
| YPG | Yeast extract 10 gr<br>peptone 20 gr<br>Glycerol 30 ml |

## 8.2 Chapter 4

**Table 8-3. The list of the strains in each library of NGS.**

| Library | Strain | Location |
|---|---|---|
| Q62.5 | Q62.5 | Silwood Park |
| DBVPG4650 | DBVPG4650 | Continental Europe |
| CBS8441 | CBS8441 | Far East |
| Pool1 | T4b | Silwood Park |
| | T8.1 | |
| | T21.4 | |
| | T26.3 | |
| | T68.2 | |
| | T68.3 | |
| | Q14.4 | |
| | Q16.1 | |
| | Q43.5 | |
| | Q59.1 | |
| | Q74.4 | |
| | Q95.3 | |
| | Y1 | |
| | Y2.8 | |
| | Y8.5 | |
| | Y10 | |
| | Q62.5 | |
| Pool2 | OS20 (C02) | Continental Europe |
| | OS11,12W (C05) | Continental Europe |
| | OS3,4Wa (C7) | Continental Europe |
| | 4/2S W2 (C10) | Continental Europe |
| | N17 | Far East |
| | CBS8439 | Far East |
| | CBS8444 | Far East |

### 8.2.1 Preliminary bioinformatics analyses

### a) Analysing the NGS data

The NGS data was analysed using Geneious software. In addition to the variations that result from the lack of proofreading of RDRP, the dsRNA viruses have to go through the procedure of reverse-transcription reaction in sequencing. As a consequence, analysing the data is more complicated compared to DNA sequences. Therefore, before starting the data analyses, the accuracy of the Geneious algorithm for analysing the dsRNA data was verified with the Geneious team. They simply suggested that, in addition to using the *de-novo* setting with the default setting, we try to run the assembly by the turning off the option "*do not merge variant with coverage over a 6"*.

Since most of the studies on viral dsRNA were performed on *S. cerevisiae,* which is closely related to *S. paradoxus,* and two dsRNAs found in *S. paradoxus* are similar to that of *S. cerevisiae,* the sequence of the *S. cerevisia*e dsRNAs were used as references in the "mapping to the reference" assembly of the data. Moreover, the overall structure of each type of dsRNA was used to identify the dsRNA contigs in the *de-novo* assemblies.

### b) Insertion size and read numbers

Five cDNA libraries of *S. paradoxus* dsRNA, Q62.5, DBVPG4650, CBS8441, Pool1 and Pool2 were sequenced using MiSeq 300 (Table 8-3). The insertion size of the three libraries Q62.5, DBVPG4650 and Pool2, for which the cDNAs were prepared by the company, was less than that of CBS8441 and Pool1, for which the cDNAs were prepared in our lab (Table 8-4). The insertion size of Q62.5 and DBVPG4650 was even less than the length of the reads. Consequently it affected not only the quality of the nucleotides at the ends but it also devalued the purpose of the paired-end sequencing.

The number of reads per strain in the libraries containing a single strain was between 1.2 and 1.8 million reads, and the average number of reads per strain in Pool1 and Pool2 were 1.6 million and 1.3 million reads respectively (Table 8-4).

Table 8-4. The insertion size, number of reads after pairing the reads in each library and the percentage of reads mapped to nuclear and mitochondrial genome with and without considering MMQ200.

| Library | Insertion size | Number of reads after trimming and pairing | % of reads mapped to nuclear genome MQ200 | % of reads mapped to mitochondrial genome MQ200 |
|---|---|---|---|---|
| **Q62.5** | 199 | 1,170,856 | 7.6% | 0.7% |
| **DBVPG4650** | 260 | 1,819,200 | 14.2% | 0.2% |
| **CBS8441** | 772 | 1,656,910 | 0.03% | 0.03% |
| **Pool1*** | 684 | 27,643,010 | 0.03% | 0.03% |
| **Pool2**** | 354 | 10,604,622 | 5% | 0.2% |

* Pool 1 is composed of 17 strains from Silwood Park: T4b, T8.1, T21.4, T26.3, T68.2, T68.3, Q14.4, Q16.1, Q43.5, Q59.1, Q74.4, Q95.3, Y1, Y2.8, Y8.5, Y10, Q62.6.

**Pool 2 contains three strains from Continental Europe and three strains from the Far East: OS20, OS11.12W. OS3.4Wa,4/25W2, N17, CBS8439, CBS8444.

## c) Quality of the data

The quality of the data was studied by FastQC software. In Q62.5, DBVPG4650 and Pool2, in which the insertion size was low, the quality score of the nucleotides at the ends of the reads was less than 40. They also contained primer sequences. Around 180 bp in each read had high quality scores. The quality of the nucleotides increased in two other libraries that had higher insertion size. The full sequence of 92% of the reads in these libraries had quality scores above 40, meaning that the probability of incorrect base in their sequence is less than 1 base in 10,000 bases.

The results of Geneious confirmed the FastQC results. The primer sequences and the nucleotides with quality scores less than 40 were removed from the ends of the reads, and the single reads of each library were paired using Geneious (Table 8-4).

## d) The amount of nuclear and mitochondrial DNA in the libraries

The paired-end reads of each library were mapped to the *S. paradoxus* nuclear and mitochondrial genomes in order to measure the amount of the host-genome. Two mapping assemblies were run by the medium-sensitivity setting, as suggested by Geneious, with considering a minimum mapping quality of 200 (MMQ200). The results are shown in Table 8-4. The amount of the genomes' contamination in two samples, CBS8441 and Pool2, which were treated with Turbo DNase in addition to DNase I, was significantly less than that of the samples that were treated just with DNase I. Approximately two thirds of the data in Pool2, and one third of the data in Q62.5 and DBVPG4650 came from genomic DNA.

### e) Homology of viral dsRNA of *S. cerevisiae* with genomic DNA of *S. paradoxus*

A database was made from all known *S. cerevisiae* and *S. paradoxus* viral dsRNAs, and the genome of *S. paradoxus* was blasted to the database using Megablast. The result indicates that 593 bp of the genome were blasted to M28 dsRNA of *S. paradoxus* and *S. cerevisiae* with 87.2% and 89% identity respectively. The test was repeated by mapping the dsRNA references to the genomes. In the mapping alignment, the 593 nt is located in chromosome 11 of *S. paradoxus*. It is the reverse sequence of 593 nt at 5′ of M28, in the coding area. Part of the sequence after the homologous region, 59 residues before the poly A, the poly A and 122 residues after the poly A, are N in the sequence of chromosome 11. The percentage of identical sites in the rest of the alignment decreased to 39%. There was no homologous sequence on the mitochondrial DNA.

### 8.2.2 Viral assemblies

To identify the sequences of the viral dsRNAs, in the first stage, the reads of each library were mapped to the known references in *S. paradoxus* and *S. cerevisiae*. Since most of the references were from *S. cerevisiae,* the medium-low sensitivity setting was selected to run the assemblies. The assemblies were run with and without considering MMQ200. Then the consensus sequences of the contigs were made from the nucleotides with quality scores above 40 and were compared with each other and the reference sequences.

A series of *de-novo* assemblies with different settings were run in the second stage. To discover the effects of removing the nuclear and mitochondrial genomes, two groups of *de-novo* assemblies were run with and without removing the genomes. The reads that mapped to L-A-L1 were also removed to see if this has any effect on the assembly of M dsRNA. We assembled 25% and 100% of both data by turning ON and OFF the option suggested by the Geneious team: “*do not merge variant with coverage over a 6”.* In each *de-novo*, the consensus sequence of the contigs with more than 1,000 reads that had good coverage along their alignment was selected and mapped to the consensus sequences of L, which was assembled in the map-to-reference part. This mapping was done with an easy mapping custom sensitivity setting (MCS1) (Table 8-22). In this setting, since the maximum mismatch per read is 60% and it allows gaps with a maximum 60% per read, the sequences with different sequences in some parts or a high variation from the reference were still mapped to the reference and it can find similar sequences to the references. The best contigs of the mappings of all *de-novos* were aligned to compare. Primarily, the results of removing the genome and L-A in Q62.5 and DBVPG4650 showed that, although the genomes and L-A sequences were removed from the data, there were still enough reads to assemble 25s rRNA and L-A dsRNA. Also, turning ON and OFF the "*do not merge variant with coverage over approximately 6*"option did not show any effect on the assemblies' L contigs. However, the number of reads in the assemblies in which this option is ON is about half of the assemblies in which it is OFF.

In terms of M, the structure of the rest of the contigs was studied. Any contigs which had the same structure as M dsRNA in *S. cerevisiae* were selected. This structure is composed of a 5' to 3' ORF, followed by a poly A and a non-coding sequence. The nucleotide sequence, the translation of the ORF and three frames of translation of the RNA were blasted in nr and NCBI databases to see if any similar sequences were reported. The software could not make a complete sequence of M in most of the *de-novo* assemblies and the M contigs were usually broken in the poly A region. Polymorphisms only exist in the sequence of this area, which arises from the inability of the software to assemble the repeated sequences. Since the software cannot find the accurate number of poly A, 50 residues were considered for this part. The conserved sequence at 5' end of Ms, GAAAAA- (Rodríguez-Cousiño et al., 2011), was checked to see whether the 5' end of the molecule had assembled or not. After finding the sequence of the L and M in each strain, to study the assembly and coverage whole reads of each were mapped to the revealed M and L. Also, in the assembly of M, turning OFF and ON of the option *"do not merge variant with coverage over approximately 6"* had no effect on the assembly of the sequences in either area. However, the number of reads in the non-coding area was affected (Table 8-6 and Table 8-9).

In order to find out whether there are any other new sequences in all the *de-novo* assemblies, the contigs of each *de-novo* were mapped to the L sequence with the easy setting MCS1. Next, the unused contigs were mapped to the M with the same setting. Then, the original assembly of the unused contigs from mapping to the M was observed, and the good contigs were separated. After that, the consensus sequences of the good contigs were blasted in nr and NCBI databases to see if there are any non-detected sequences.

### Q62.5

### Mapping to the references

The result of the map-to-reference assemblies is summarised in Table 8-5. Aligning the consensus sequence of the mapping to the L-A dsRNAs shows that, apart from the uncovered area in MMQ200 contigs, the consensus sequences of the remaining sequences are identical. The identity of the consensus sequence of each contig with its own reference is between 71% and 75.6%. Since the identity between the contigs and their L dsRNA references is similar to the percentage between different L-A dsRNAs in *S. cerevisiae* (Rodríguez-Cousiño et al., 2013), this sequence may be a new L dsRNA in *S. paradoxus* and is therefore called L-A-Q dsRNA. In all of the L-A assemblies, the coverage of the sequences near to both ends was between 35 and 225 times higher than in the other areas. The coverage decreased in the middle, especially in the Gag-coding region in L-A-L1.

Table 8-5. Mapping the Q62.5 reads to *S. cerevisiae* and *S. paradoxus* references.

| **Reference** | **Host species** | **MMQ200** | **Number of reads** | **% reference coverage** | **% identity of each contig with its reference** |
|---|---|---|---|---|---|
| **L-A-L1** | *S. cerevisiae* | - | 225,397 | 97.6% | 71% |
| | | + | 136,985 | 83.8% | 71% |
| **L-A-2** | *S. cerevisiae* | - | 226,640 | 100% | 75.6% |
| | | + | 98,179 | 68.8% | 74.6% |
| **L-A-lus** | *S. cerevisiae* | - | 228,786 | 100% | 75.1% |
| | | + | 134,666 | 77.9% | 72.3% |
| **M28** | *S. cerevisiae* | - | 56 | 100% | 99.2% |
| | | + | 56 | 100% | 99.2% |
| **M28** | *S. paradoxus* | - | 56 | 100% | 87.6% |
| | | + | 56 | 100% | 87.6% |
| **M28 593bp from genome** | *S. paradoxus* | - | 37 | 100% | 87% |
| | | + | 37 | 100% | 89% |
| **M1** | *S. cerevisiae* | ± | - | - | - |
| **M2** | *S. cerevisiae* | ± | - | - | - |
| **M-lus** | *S. cerevisiae* | ± | - | - | - |
| **L-BC** | *S. cerevisiae* | ± | - | - | - |
| **W** | *S. cerevisiae* | ± | - | - | - |
| **T** | *S. cerevisiae* | ± | - | - | - |

The number of reads mapped to M28 dsRNA of both species is very low. The number was not shown to change when applying MMQ200. It showed that these reads were not random or an artefact of the library preparation. Because the number of reads was considerably low at 0.005% of total reads, there is a possibility that the reads came from the homologous part of M28 in the genome. Mapping the Q62.5 reads to this part of the genome, which is homologous with 593 bp at 5' of M28, shows that 37 reads were assembled. The consensus sequence of the mapping to M28 of both species' gene part and genome homologous sequence were aligned with the three references. The result shows that all three consensus sequences are identical and they are closer to the M28 of *S. paradoxus* than the homologous genome sequence (99.2% and 87.6% identity respectively).

### *De-novo* assemblies

The settings of the seven different *de-novo* assemblies that were run and the results of mapping the contigs of each *de-novo* to L-A-Q with MCS1 setting are in Table 8-6. Aligning the contigs that were mapped to L-A-Q indicates they are almost identical. The difference observed between the contigs arises from the low coverage areas.

Table 8-6. The settings and the results of the Q62.5 *de-novo* assemblies

| **Name of *de-novo*** | **Percentage of data** | **Removing L-A-L1 and host genomes** | **Do not merge variant with coverage over approximately 6** | **Long contigs mapped to L-A-Q with MSC1 setting** | **Number of reads** | **Length of contig** |
|---|---|---|---|---|---|---|
| **D1** | 25% | + | OFF | Contig 4 | 10,364 | 4,796 |
| **D2** | 25% | + | ON | Contig 4 | 5,768 | 4,658 |
| **D3** | 100% | + | OFF | Contig 3 | 48,094 | 5,535 |
| **D4** | 100% | + | ON | Contig 2 | 27,629 | 4,836 |
| **D5** | 25% | - | OFF | Contig 3 | 22,221 | 4,866 |
| **D6** | 25% | - | ON | Contig 3 | 14,129 | 4,725 |
| **D7** | 100% | - | ON | Contig 2 | 32,527 | 4,863 |

In terms of M, the contig 2 from D1 showed the same structure as M dsRNA and its sequence was not reported. The list of the best contigs in each *de-novo* is listed in Table 8-7. In the alignment of the best contigs mapped to MQ, the coding and non-coding regions of all contigs are identical with MQ. The numbers of the reads in non-coding contigs are significantly higher than the coding contigs.

Table 8-7. M dsRNA contigs in Q62.5 *de-novo* assemblies

| **Name of *de-novo*** | **MQ coding similar contig** | **Number of reads** | **MQ non-coding similar contig** | **Number of reads** |
|---|---|---|---|---|
| **D1** | 2 | 54,646 | 2 | 54,646 |
| **D2** | 13 | 1,195 | 2 | 9,017 |
| **D3** | 6 | 4,491 | 1 | 160,640 |
| **D4** | 13 | 4,703 | 3 | 25,834 |
| **D5** | 7 | 1,231 | 2 | 41,038 |
| **D6** | 10 | 1,176 | 2 | 20,303 |
| **D7** | 13 | 4,449 | 1 | 35,239 |

The 5’ end was assembled and its conserved sequence is identical with the conserved sequence of M28 dsRNA, GAAAAAATTTGA- (Rodríguez-Cousiño et al., 2011). In mapping the whole reads of the library to MQ, in total, 576,777 reads were assembled to MQ. Apart from the sequence of the 3’ end, which had low coverage, other parts of the 3' sequence did not change. The full sequence of MQ is available in the Appendix; section 8.2.7. Blasting of unused reads of L-A-Q and MQ in NCBI and nr did not find any new sequences.

To summarise, an L-A dsRNA named L-A-Q was found in Q62.5. It is identical to the L-A found in the mapping to the L-A references from *S. cerevisiae*. Also, a sequence was found that has the same structure as the other M dsRNAs but no detectable sequence in the databases; it was called MQ. It likewise has 56 reads with some similarity to M28, but probably derives from the nucleus genomes. L-A-Q and MQ are compatible with the two L and M bands on the gel electrophoresis.

### DBVPG4650

### Mapping to the references

The mapping results are summarized in Table 8-8. According to the results, it seems that there is an L-A dsRNA in BDVPG4650 with greater similarity to L-A-L1. The alignment of the consensus sequence of the mapping assemblies showed that the consensus sequence of mapping the reads to L-A-2 and L-A-lus is identical, and the identity percentage of their sequence with the consensus sequence of mapping to L-A-L1 without MMQ200 is 93.5% and with MMQ200 is 84%.

**Table 8-8.** The result of mapping DBVPG4650 reads to *S. cerevisiae* and *S. paradoxus* references

| Reference | Host species | MMQ200 | Number of reads | % reference coverage | % identity |
|---|---|---|---|---|---|
| **L-A-L1** | *S. cerevisiae* | - | 658,175 | 100% | 79.5% |
| | | + | 53,524 | 100% | 89.3% |
| **L-A-2** | *S. cerevisiae* | - | 935,615 | 100% | 75.1% |
| | | + | 76,573 | 91.9% | 74.3% |
| **L-A-lus** | *S. cerevisiae* | - | 849,043 | 100% | 74.6% |
| | | + | 190,926 | 82.1% | 73% |
| **L-BC** | *S. cerevisiae* | - | 406 | 98.6% | 89% |
| | | + | 406 | 98.6% | 89% |
| **M28** | *S. cerevisiae* | - | 739,170 | 100% | 87.3% |
| | | + | 559,355 | 100% | 87.3% |
| **M28** | *S. paradoxus* | - | 402,099 | 100% | 97.5 |
| | | + | 337,739 | 100% | 99.8% |
| **M1** | *S. cerevisiae* | ± | - | - | - |

| Reference | Host species | MMQ200 | Number of reads | % reference coverage | % identity |
|---|---|---|---|---|---|
| **M2** | *S. cerevisiae* | ± | - | - | - |
| **M-lus** | *S. cerevisiae* | ± | - | - | - |
| **W** | *S. cerevisiae* | ± | - | - | - |
| **T** | *S. cerevisiae* | ± | - | - | - |

The consensus sequence generated from mapping the reads to L-BC dsRNA is 99.2% identical with its reference. Although the number of reads assembled to the reference sequence is only 406 reads (0.02% of the data), they covered 98.6% of the L-BC sequence. The only areas that are not covered are 6 nucleotide at 5’, 49 bp from residue 1,819, and 8 nucleotide at the end. The number of reads did not change with MMQ200. It appears that the reads specifically belong to L-BC.

The DBVPG4650 reads were assembled to the M28 dsRNA sequence of both *S. cerevisiae* and *S. paradoxus*. The consensus sequence of both assemblies has only one nucleotide difference from the sequence reported by M. D. Pieczynska et al (Pieczynska et al., 2013a) in this strain when it was run with MMQ200. Applying MMQ200 had no effect on the coverage of either reference. Nevertheless, it increased the identity by 2.3% between the consensus sequence of mapping the reads to M28 from *S. paradoxus* and its reference. This might result from the reduction of the interference from M28 homologous sequence in the genome. However, the result of mapping M28 to the genome with the MCS1, did not show any similar sequences on the genome.

In summary, 658,178 and 935,615 reads were assembled to the three types of L-A dsRNA references. The consensus sequence of mapping to L-A-L1 was different from two other L-As. In this strain, in addition to L-A, the reads were also assembled to L-BC. However, the number of reads assembled to this dsRNA was very low at 403 reads. Of all the M dsRNAs, the reads were only assembled to M28. The consensus sequence of the assembly had only one nucleotide different from M28, as previously reported.

### *De-novo* assemblies

The result of the six *de-novo* assemblies that were run on this strain’s data is in . Since, in this strain, we could not find an identical sequence in mapping the reads to the three L-A references, and it seems that the L-A in this strain has greater similarity to L-A-L1 than the two other L-As, the consensus sequence of the *de-novo* assemblies was mapped to L-A-L1, using MCS1. The results were more complicated than that of Q62.5. There are many contigs longer than 1 kb, especially in the *de-novo* assemblies which were run with 100% of the data **(**Table 8-9). There are variations between the contigs, particularly in the D4, D5, D6 assemblies, where the L-A-L1 dsRNA sequence was not removed. In the D1 and D2 assemblies, there is just one contig longer than 4 kb with high coverage, whereas in the other assemblies this number increased.

Table 8-9. *De-novo* assembly settings in DVPG4650 and the L-A dsRNA contig that was formed in each assembly.

| **Name of *de-novo* assembly** | **Percentage of data** | **Removing genome contamination and L-A-L1** | **Do not merge variant with coverage over approximately 6** | **Long contigs mapped to L dsRNA** | **Type of dsRNA** | **Length** | **Number of reads** | **SNP** |
|---|---|---|---|---|---|---|---|---|
| **D1** | 25% | + | ON | 1 | L-A-D1 | 6,001 | 25,204 | 4 |
| **D2** | 25% | + | OFF | 1 | L-A-D1 | 5,045 | 68,146 | 2 |
| **D3** | 100% | + | ON | 1 | L-A-D1 | 4,556 | 114,636 | 2 |
| | | | | 23 | | 4,328 | 1,808 | 10 |
| | | | | 9 | | 4,900 | 4,584 | 15 |
| **D4** | 25% | - | ON | 1 | L-A-D1 | 5,264 | 48,841 | 1 |
| | | | | 70 | L-A-D2 | 2,338 | 150 | 2 |
| | | | | 49 | L-A-D2 | 2,200 | 189 | 0 |
| **D5** | 25% | - | OFF | 2 | L-A-D1 | 5,171 | 71,996 | 2 |
| | | | | 10 | L-A-D2 | 4,525 | 354 | 15 |
| **D6** | 100% | - | ON | 1 | L-A-D1 | 5,025 | 117,430 | 1 |
| | | | | 39 | L-A-D2 | 4,550 | 1,323 | 0 |

In the *de-novo* assemblies, in which the host genome and L-A-L1 were not removed from the data, two types of L-A dsRNA contigs were formed (Table 8-10). The first type of L-A contigs (named L-A-D1) were assembled in all *de-novo* assemblies, whereas the second type (L-A-D2) were formed just in the *de-novo* assemblies from which the host genome and L-A-L1 were not removed (Table 8-9). The complete sequences of the L-A-D2 were just assembled in D5 and D6. The sequence of these contigs was broken in 2 contigs in D4 assembly. Against the L-A-D1 contigs, there is just one L-A-D2 contig in each *de-novo*. The number of reads in this type of L-A is very low: 0.07% of the data (1,323 reads) in contig 39, which was run with 100% of the data. Although the number of reads in L-A-D2 dsRNA is significantly lower than that of L-A-D1 (Table 8-9), the identity between the contigs from different *de-novo* assemblies in this type is higher than that of the first type (Table 8-10). In order to be certain with respect to the sequence of L-A-D2, the reads of DBVPG4650 were mapped to contig 39 with a stringent custom sensitivity setting, MCS2 (Mapping custom Sensitivity 2) (Table 8-22). In this setting, the minimum mapping quality is 254, the minimum overlap increased to 200 bp, the maximum mismatch per reads decreased to 2%, and the minimum overlap identity increased to 98%. In this assembly, 610 reads were assembled to the reference. There is no polymorphic residue in the sequence and the sequence is identical with the reference. Since the insertion size of the reads in this strain is 260, selecting the minimum overlap of 200 reduced the number of reads mapped to the reference. Consequently, when the minimum overlap was reduced to 100 in MCS3 setting, the number of reads increased to 1,253 reads without changing the sequence. The percentage of the identity between these two types of L-A dsRNA is 62.2% to 75.2%, similar to that of different types of L-A dsRNA in *S. cerevisiae* (Rodríguez-Cousiño et al., 2013). Aligning the sequences with the L-A references and L-A-Q shows that the sequence of L-A-D2 is 100% identical to the sequence of L-A-L1. The only difference is that the contigs made in the *de-novo* assemblies do not cover 17 bp at the beginning and 12 bp at the end of L-A-L1 dsRNA. Mapping all the reads to L-A-L1 with MCS2 and MSC3 settings shows the same result; these sequences were not covered by any reads.

It seems that the presence of two L-As in this strain was the reason for not having identical contig in mapping the reads to the references. Perhaps the setting was not restricted enough to separate the reads of the two L-A dsRNAs.

Table 8-10. The identity between the L-A contigs in DBVPG4650 *de-novo* assemblies. There are two types of L-A in the *de-novo* assemblies based on identity: L-A-D1 and L-A-D2. L-A-D1 is composed of three contigs at the beginning of the table and L-A-D2 comprises the four contigs at the end of the table.

| | **D6** | **D5** | **D4** | **D4** | **D5** | **D4** | **D6** |
|---|---|---|---|---|---|---|---|
| | **C1** | **C2** | **C1** | **C40** | **C10** | **C70** | **C39** |
| **D6 C1** | | 99.9 | 99.8 | 71.7 | 73.7 | 75.2 | 73.8 |
| **D5 C2** | 99.9 | | 99.2 | 72.6 | 73.9 | 75.2 | 73.9 |
| **D4 C1** | 99.8 | 99.2 | | 72.7 | 73.9 | 75.2 | 73.9 |
| **D4 C40** | 71.7 | 72.6 | 72.7 | | 100 | 100 | 100 |
| **D5 C10** | 73.7 | 73.9 | 73.9 | 100 | | 99.7 | 99.9 |
| **D4 C70** | 75.2 | 75.2 | 75.2 | 100 | 99.7 | | 100 |
| **D6 C39** | 73.8 | 73.9 | 73.9 | 100 | 99.9 | 100 | |

Since the identity between these two types of L dsRNA is about 70% in both assemblies, to be certain about the sequence of the L-A sequences, two more *de-novo* assemblies were run: D7 (with 10% of data) and D8 (with 20% of data) with more restricted settings; *de-novo* custom sensitivity 11 (DCS11) and DCS11.50 (Table 8-14), respectively. In both assemblies, the full sequence of the L-A-D1 was formed. The contigs were identical to L-A-D1, except for a few residues in low coverage areas. As a result of the low number of reads of this dsRNA, its sequence was broken into several contigs. However, all contigs were identical to L-A-L1.

Aligning the sequence of L-A-D1 with the consensus sequence of mapping the reads to L-A-2 and L-A-lus shows that they have identical sequences. However, the consensus sequences generated from mapping to L-A-2 and L-A-lus do not have 13 residues at the beginning. Looking at L-A-D1 contigs in other *de-novo* assemblies shows that they also have low coverage at the beginning. It seems that during cDNA or library preparation, the sequence of the ends did not amplify very well.

In the alignment of mapping the consensus sequence of the D8 contigs to L-A-D1, in addition to contig 1, which has the full sequence of L-A-D1, there are six contigs identical to the reference (lengths between 502 bp and 1,363 bp). The presence of the different sized L-A-D1 contigs in each *de-novo* assembly suggests that there are various L-A-D1s with different sizes in this strain. This might result from the deletion in the original sequence of L-A-D1, similar to X dsRNA in *S. cerevisiae*, which is derivative of L-A-L1 (Sommer and Wickner, 1982).

In terms of L-BC, its contigs were assembled in all *de-novo* assemblies run with medium-low sensitivity and all of the contigs were identical. In D3 and D6, which were run with 100% of the data, 2 contigs were assembled with a total number of 402 reads and identical with the consensus sequence made from mapping to L-BC. In other *de-novo* assemblies run with 25% of data, the contigs were broken into smaller contigs and some areas had no coverage. The number of reads in these contigs was between 4 and 28 reads. However, they are 99.8% to 100% identical with the contig mapped to L-BC. The areas without any coverage in mapping do not have coverage in the *de-novo* assemblies. There are no L-BC identical contigs in D7 and D8. This may result from the low number of reads in these assemblies (10% and 20% of the data).

Although the number of reads of L-A-1 and L-BC is very low, it appears that they specifically belong to their references, while changing the assembly settings did not change their sequences. The low number of reads might result from the low copy number of the dsRNAs in the cell.

The result of mapping to M28 using MCS1 setting showed that in D1, D2, D4, and D5 assemblies, the full sequence of M28 was assembled; whereas, in D3, D6, and D8, the M28 sequence is broken in the poly A area. The coding and non-coding sequences of all the M28 *de-novo* contigs are identical with the M28 consensus sequences of map-to-reference.

Overall, two types of L-A, L-A-D1 and L-A-L1, one type of L-BC, and one type of M dsRNA, M28, were found in this strain.

**CBS8441**

### Mapping to the references

The results of mapping CBS8441 reads to all references are summarised in Table 8-11. Aligning the consensus sequences of all mapping to the L-A references indicates that, apart from some differences in the sequence of both ends, the other part of the consensus sequences is identical. The identity of the consensus sequence to the known L-As is about 75%, which suggests a new type of L-A dsRNA in this strain. It is called L-A-C.

Table 8-11. The settings and results of mapping CBC8441 to all L-A dsRNA types.

| Reference | MMQ200 | Number of reads | % reference coverage | % identical sites with its reference |
|---|---|---|---|---|
| **L-A-L1** | - | 1,197,534 | 100% | 74.9% |
| | + | 69,062 | 100% | 75.1% |
| **L-A-2** | - | 1,142,193 | 100% | 75.0% |
| | + | 113,089 | 95.4% | 74.4% |
| **L-A-lus** | - | 1,156,705 | 100% | 74.8% |
| | + | 169,930 | 93.5% | 74.5% |
| **L-A-Q** | - | 1,199,137 | 100% | 76.5% |
| | + | 156,886 | 100% | 76.4% |
| **L-A-D2** | - | 1,204,637 | 98.7% | 75.1% |
| | + | 87,715 | 96.3% | 76.3% |
| **M1, M2, M28, M-lus** | ± | - | - | - |
| **L-BC** | ± | - | - | - |
| **T and W** | ± | - | - | - |

In terms of M dsRNA, no sequence of the known M dsRNA was found in this sample. It suggests that, because the host genome contamination is very low at 4.3% of the data, there are no M28 dsRNA reads in this sample.

## *De-novo* assemblies

Since the host genomes in this sample is very low and the number of the dsRNA reads is higher than the two other samples, just two *de-novos* with 25% of the data were run with (D1) and without (D2) the option *"do not merge variant with coverage over approximately 6"*. Mapping the consensus sequence of D1 and D2 contigs to L-A-C shows that contig 1 from D1 and contig 1 from D2, which are the longest contigs in both assemblies, are mapped to L-A-C. Apart from about 40 residues at the ends of the contigs, other parts of the sequences are identical to L-A-C. The alignments of the *de-novo* assembly of both contigs 1 show that, like mapping to the L-A references, the coverage of the reads is similar in most parts of the sequence, except for the ends. It suggests that, like DBVPG4650, the replication of the ends of the dsRNA was not really successful in the library preparation.

In terms of M dsRNA, contig 2 in D1 that is 3,166 bp long and composed of 39,947 reads shows the same structure as M dsRNA. There is an ORF with a length of 1,008 bp at 5', 1,284 bp poly A in the middle, and a non-coding part at 3' of the sequence. No similar sequences were found in blasting, which suggests it is a new type of M dsRNA and is called MC (Appendix; section 8.2.7).

To summarise, two types of dsRNA were found in this strain, L-A-C and MC.

### Pool1

Pool1 is composed of 17 strains from Silwood Park (T68.2, T68.3, T26.3, T21.4, T4b, T8.1, Q14.4, Q95.3, Q43.5, Q59.1, Q16.1, Q74.4, Y1, Y2.8, Y8.5, Y10, and Q62.5). Q62.5 was added as a control for the bioinformatics analyses. In addition to Q62.5, the gene part of M1 dsRNA and M28 dsRNA was sequenced in Q74.4 and T21.4 respectively by Chang et al. (Chang et al., 2015).

### Analysing the sequence of L-A-Q and MQ dsRNA of Q62.5 in Pool1

To discover the effect of pooling the strains on the dsRNA sequence, data analysis was started with MQ and L-A-Q in Q62.5. The sequences of these dsRNA were firstly studied by map-to-references. Then their sequences were analysed in a *de-novo* assembly that was run with a medium-low sensitivity setting called DML (*De-novo* with medium-low sensitivity). The results of the map-to-reference are in Table 8-12.

Table 8-12. Mapping Pool1 reads to MQ and L-A-Q.

| Reference | MMQ200 | Number of reads | % reference coverage | % identical sites with its reference |
|---|---|---|---|---|
| **MQ** | - | 325,978 | 100% | 98.9% |
| | + | 203,523 | 100% | 99.2% |
| **LQ** | - | 21,095,686 | 100% | 78% |
| | + | 19,543,643 | 100% | 78% |

Aligning the consensus sequences of the *de-novo* MQ contigs and mapping to the MQ with MQ indicates that, except for a few nucleotides before and after poly A and some residues at the ends, the rest of the sequences are identical. It suggests that, since the sequences of various types of M dsRNA are different, mixing the strains did not make a significant difference to the sequence of MQ. Only the sequences of both ends and beside poly A, which are similar in most of the M dsRNA, have been affected. However, this is not true about L-A-Q. In map-to-reference, about 75% of the reads were mapped to L-A-Q and the identity of its consensus sequence to the reference was 78%. In addition, no identical contigs were found in the *de-novo* contigs and the identity between the long contigs and L-A-Q was between 34% and 74%. Since the sequences of different types of dsRNA are similar (about 75% identity), it seems that the normal settings of mapping to a reference cannot categorise the reads of different types of L-A dsRNA in the assemblies. These suggest that the effect of mixing the strains in Pool1 and Pool2 on the sequence of L-A is like taking a metagenomics sample from nature. As a result, using metagenomics analysis might be helpful to categorise the reads of different types of L-A and find the real sequence of L-A dsRNA. To find the best setting for the metagenomics *de-novo* assembly, optimisation was started with the *de-novo* custom sensitivity (DCS) setting suggested by Geneious (Table 8-13). Then, different parameters were changed to improve the assembly, as shown in Table 8-14. The detail of optimisation is in the Appendix; section 8.2.5.

Table 8-13. *De-novo* custom sensitivity suggested by Geneious for metagenomics

□Don't merge variant with coverage over approximately 6

□Produce scaffold □Circulate contig with matching ends

■Allow gaps Maximum per read 10% Minimum gap size 2

□Minimum overlap

Maximum mismatch per read 20%

■Minimum overlap identity 98%

Table 8-14. *De-novo* custom sensitivity setting optimisation

| Name of the *de-novo* custom sensitivity | Maximum mismatch per read | Minimum overlap identity | Minimum overlap | Allow gaps | Using just paired end reads |
|---|---|---|---|---|---|
| **DCS1** | 20% | 98% | 200 bp | ON | OFF |
| **DCS2** | 2% | 98% | 200 bp | ON | OFF |
| **DCS3** | 20% | 98% | 200 bp | OFF | OFF |
| **DCS4** | 2% | 98% | 200 bp | OFF | OFF |
| **DCS5** | 2% | 98% | 200 bp | OFF | ON |
| **DCS6** | 1% | 99% | 200 bp | OFF | OFF |
| **DCS7** | 1% | 99% | 250 bp | OFF | OFF |
| **DCS8** | 1% | 99% | 300 bp | OFF | OFF |
| **DCS9** | 0 | 100 | 200 bp | OFF | OFF |
| **DCS10** | 5% | 95% | 200 bp | OFF | OFF |
| **DCS11** | 2% | 98% | 100 bp | OFF | OFF |
| **DCS12** | 2% | 98% | 50 bp | OFF | OFF |
| **DCS14** | 2% | 98% | OFF | OFF | OFF |
| **DCS15** | 2% | 98% | 200 bp | OFF | OFF |

### Analysing the M dsRNA in Pool1

In addition to the above *de-novo* assemblies, in order to increase the number of M dsRNA reads and remove interference of the L dsRNA, a *de-novo* assembly, DLS-L (*de-novo* with low sensitivity setting), was run with the unused reads from mapping Pool1 to LA dsRNA. The result was then compared to the other *de-novo* assemblies with the aim of selecting the best *de-novo* assemblies in which the MQ contigs were assembled.

Looking at the alignment of each MQ contig assembly and aligning the consensus sequences of all MQ contigs formed in the *de-novo* assemblies with the optimised settings, Table 8-14, showed that the best MQ contigs were assembled in DSC3, DCS4, DCS10, DCS11, DCS11-2, DML, and PLS-L. The MQ contigs from DCS3 had a longer identical sequence with MQ and each contig had more coverage when compared to others. All MQ contigs assembled in the selected *de-novo* assemblies were split in the poly A region. None of the contigs covered 44 nt from the beginning and 8 nt from the ends.

Since in some of the *de-novo* assemblies some parts of the contigs assembled better than others, to find the best sequence for each M dsRNA, all the M-identical contigs in the selected *de-novo* were compared. To separate the M dsRNA contigs in each *de-novo* assembly, the consensus sequences of the contigs of each *de-novo* were mapped to L-A-Q with the easy setting MCS1 (Table 8-22). Since the flexibility of this setting is high, all L-A dsRNA types were mapped to L-A-Q. By looking at the alignment, any contigs that were not aligned accurately with the reference or contained poly A were removed from the assembly and added to the unused consensus sequences. Then the unused consensus sequences were sorted by length and any sequence less than 500 bp, 100 reads and containing low coverage area, was removed. The number of consensus sequences remaining in each *de-novo* can be seen in (Table 8-15)

Table 8-15. Selected M dsRNA contigs in each *de-novo* assembly

| ***Name*** | ***Number of remaining sequences*** | ***Number of sequences containing poly A or poly T*** |
|---|---|---|
| ***Pool1 1% DCS3*** | 12 | 9 |
| ***Pool1 1% DCS4*** | 14 | 6 |
| ***Pool1 1% DCS10*** | 11 | 8 |
| ***Pool1 1% DCS11*** | 14 | 9 |
| ***Pool1 2% DCS11*** | 15 | 11 |
| ***Pool1 Ms*** | 9 | 5 |
| ***Pool1-L 1% LS*** | 11 | 8 |

A total of 86 consensus sequences were listed from the 7 *de-novo*. To categorize the M dsRNA candidate contigs, all M dsRNA candidate contigs from seven *de-novos* were mapped to the M dsRNA candidate of DCS3 with the easy custom sensitivity MCS1. The results indicated that all sequences with similar sequences to each contig were aligned to that contig. In this assembly, 100% of all contigs were covered with identical sequences, the minimum percentage of identical sites is 92%, and just one of the contigs did not map to any of the references (Table 8-16). The alignment of each mapping was studied separately to find the sequence of different M dsRNAs in Pool1. In each alignment, there was just one contig from DCS3 identical with the reference. This showed that all 12 contigs in DCS3 are different. The mapping alignment of each contig was studied separately to find the sequence of M dsRNA existing in the pool. After finding the sequence of the Ms, they were blasted to NCBL and nr. Since all of the *de-novo* assemblies were run with a maximum 2% of the data, to consider all variations, the Pool1 reads were mapped to the sequence of the locate Ms. The map-to-reference was run with medium-low sensitivity and with consideration of MMQ200. All the results are summarised in Table 8-16.

Table 8-16. The results of mapping all M dsRNA candidate contigs to DCS3 M contig

| Name of reference contig | Number of contigs aligned | Identity | Sequence Length | Poly A or Poly T position | ORF (bp) | Direction of ORF | Suggested position of the sequence on M | Name of the M | Number of reads mapped |
|---|---|---|---|---|---|---|---|---|---|
| **Contig6** | 8 | 97.9% | 1,603 | T at 5’ | 1002 | 3’ to 5’ | Reverse Seq. of Coding | M-P1G1 | 887,199 |
| **Contig 9.1** | 7 | 97.8% | 2,204 | T at 3’ | - | - | Seq. of Non-Coding | M-P1NC1 | 213,689 |
| **Contig 9.3** | 7 | 98.6% | 1,781 | A at 3’ | 678 | 5’ to 3’ | Seq. of Coding | M-P1G2 | 535,179 |
| **Contig 10.** | 7 | 97.1% | 2,126 | T at 5’ | - | - | Seq. of Non-Coding | M-P1NC2 | 467,503 |
| **Contig 11.1** | 10 | 96.6% | 1,415 | A at 3’ | 973 | 5’ to 3’ | Seq. of Coding | M-P1G3 | 603,066 |
| **Contig 11.2** | 1 | 90.3% | 507 | A at 5’ | - | - | - | - | - |
| **Contig 12** | 9 | 94.7% | 1,608 | T at 3’ | - | - | Seq. of Non-Coding | M-P1NC3 | 225,596 |
| **Contig 13** | 6 | 92.4% | 1,210 | A t 3’ | 470 | 5’ to 3’ | Seq. of Coding | M-P1SG | 317,407 |
| **Contig 14** | 6 | 99.8% | 1,523 | A t 3’ | 1045 | 5’ to 3’ | Seq. of Coding | M-P1G4 | 458,046 |
| **Contig 16** | 6 | 99.0% | 1,915 | T at 3’ | - | - | Seq. of Non-Coding | M-P1NC4 | 110,112 |
| **Contig 17** | 5 | 100.0% | 990 | A t 3’ | 780 | 5’ to 3’ | Seq. of Coding | M-P1G5 | 128,801 |

In total, six coding parts and four non-coding parts were found in the *de-novo* assembly. None of the sequences were reported before and they were not similar to the M dsRNA that was sequenced in the three single strains. Since there were two strains in Pool1, T21.4 and Q74.4, in which M28 and M1 were sequenced (Chang et al., 2015), we expected to find the sequence of these dsRNAs in the *de-novo* assemblies. However, nothing similar to these dsRNAs was found. As a result, the Pool1 reads were assembled to these M dsRNA sequences to determine if their reads exist in the pool. The result shows that the reads were assembled to none of the references. In addition to these M dsRNAs, the reads were mapped to other M dsRNAs: M2, M-lus, MQ, and MC. Apart from poly A sequences in the references, the reads were not assembled to the M dsRNA.

### Analysing L dsRNA in Pool1

In the first stage of analysing L dsRNA, the L-A-Q contigs assembled in the *de-novo* assemblies were compared to find the best *de-novo* assemblies for L-A dsRNA in Pool1. The result showed that the best L-A-Q sequences were assembled in DCS2, DCS4, DCS5, DCS6, DCS7, DCS11 and DCS11-2 assemblies. The whole sequence of L-A-Q only assembled in DCS11. In the next stage, all the L-A contigs in each *de-novo* were separated by mapping the consensus sequences of the contigs to L-A-Q by the easy setting, MCS1. All the mapped contigs were studied and the contigs with low coverage or a length of less than 500 bp were removed from the list. The selected L-A contigs of all *de-novo* were compiled in a list.

In order to separate each type of L-A, a custom mapping sensitivity was optimised, MCS3, which aligns only the same type of L-A dsRNA (Table 8-22). The list of L-A contigs was mapped to the best L-A contig assemblies in DCS11 using the MCS3 setting and the alignment of each contig was studied separately (Table 8-17). In each alignment, by looking at the assembly alignment of the contigs, the best contigs were selected. In the alignments where the reference contig did not have the whole sequence of the L and where the sequence of the other contigs overlapped with one other, the sequence of the L was determined based on the overlapped sequences. The sequence of each type of L-A was blasted in NCBI and nr. The results are summarised in Table 8-17 .

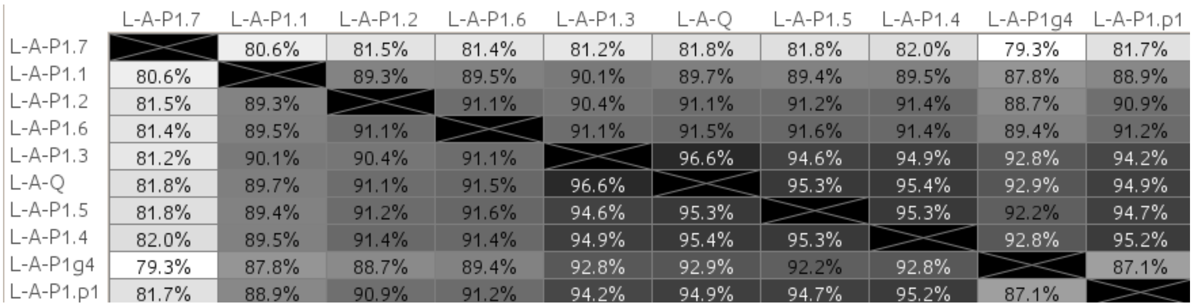
Table 8-17. The results of mapping all L dsRNA candidate contigs to DCS11 L dsRNA contigs

| Name of reference contig | Length of reference contig | Reference contig containing which part of L | Number of aligned contigs | Number of contigs that have a whole sequence | % Identical sites | Alignment length | Identity to the known L-A | Name of the L |
|---|---|---|---|---|---|---|---|---|
| Contig 3 1 | 4,452 | Whole Seq. | 12 | 3 | 99.7% | 5,529 | 74%-76% | L-A-P1.1 |
| Contig 1 2 | 4,569 | Whole Seq. | 49 | 3 | 95.1% | 5,505 | 75%-76% | L-A-P1.2 |
| **Contig 5 1** | 2,491 | ORF2 | 12 | 2 | 96.5% | 5,062 | 74%-76% | L-A-P1.3 |
| Contig 13 1 | 2,215 | ORF2 | 24 | 1 | 99.8% | 4,537 | 74%-76% | L-A-P1.4 |
| Contig 12 1 | 4,366 | Whole Seq. | 15 | 3 | 100.0% | 4,533 | 75%-76% | L-A-P1.5 |
| Contig 7 1 | 3,134 | ORF2 | 14 | 0 | 99.5% | 4,469 | 75% | L-A-P-p1 |
| Contig 16 1 | 2,242 | ORF1 | 10 | 2 | 99.8% | 4,465 | 75% | L-A-P1.6 |
| Contig 6 1 | 1,745 | Middle | 13 | 0 | 99.8% | 4,073 | 75%-76% | L-A-P1,7 |
| Contig 4 2 | 2,288 | ORF1 | 9 | 0 | 100.0% | 3,139 | 75% | L-A-P-g1 |
| Contig 11 1 | 2,910 | ORF1 | 10 | 0 | 98.9% | 3,094 | 76% | L-A-P-g2 |
| Contig 2 1 | 2,906 | ORF2 | 18 | 0 | 99.4% | 2,988 | 76% | L-A-P-p2 |
| Contig 18 1 | 2,100 | ORF1 | 4 | 0 | 100.0% | 2,493 | 75%-76% | L-A-P-g3 |
| Contig 9 1 | 2,444 | ORF1 | 6 | 0 | 98.4% | 2,446 | 75% | L-A-P-g4 |

In total, seven full sequences of L-A were found in Pool1: L-A-P1.1, L-A-P1.2, L-A-P1.3, L-A-P1.4, L-A-P1.5, L-A-P1.6 and L-A-P1.7. Moreover, four sequences of gag were assembled in the *de-novo* assemblies: L-A-P1g1, L-A-P1g2, L-A-P1g3 and L-A-P1g4; and two sequences of pol: L-A-P1p1 and L-A-P1p2. Aligning all the sequences showed that the sequences of L-A-P1g1, L-A-P1g2, L-A-P1g3 and L-A-P1p2 are identical with L-A-P1.3, L-A-P1.4, L-A-P1.6 and L-A-P1.7 respectively. Only the residues from beginning and the end of the sequences were different. Perhaps the software categorised the sequences separately based on the differences.

Aligning the L-A sequences with L-A-Q showed that the L-A dsRNA in the strains from Silwood Park is very close. Apart from L-A-P1.7 and L-A-P1.1, which have 79.3% - 81.7% and 80.6% - 90.1% identity with the other L-A dsRNA respectively, the identity between the rests of the L-A sequences is more than 90% (Table 8-18). The sequences of all the L-A dsRNAs are in the Appendix; section 8.2.8.

Table 8-18. The identity between different L-As in Pool1



| | L-A-P1.7 | L-A-P1.1 | L-A-P1.2 | L-A-P1.6 | L-A-P1.3 | L-A-Q | L-A-P1.5 | L-A-P1.4 | L-A-P1g4 | L-A-P1.p1 |
|---|---|---|---|---|---|---|---|---|---|---|
| L-A-P1.7 | | 80.6% | 81.5% | 81.4% | 81.2% | 81.8% | 81.8% | 82.0% | 79.3% | 81.7% |
| L-A-P1.1 | 80.6% | | 89.3% | 89.5% | 90.1% | 89.7% | 89.4% | 89.5% | 87.8% | 88.9% |
| L-A-P1.2 | 81.5% | 89.3% | | 91.1% | 90.4% | 91.1% | 91.2% | 91.4% | 88.7% | 90.9% |
| L-A-P1.6 | 81.4% | 89.5% | 91.1% | | 91.1% | 91.5% | 91.6% | 91.4% | 89.4% | 91.2% |
| L-A-P1.3 | 81.2% | 90.1% | 90.4% | 91.1% | | 96.6% | 94.6% | 94.9% | 92.8% | 94.2% |
| L-A-Q | 81.8% | 89.7% | 91.1% | 91.5% | 96.6% | | 95.3% | 95.4% | 92.9% | 94.9% |
| L-A-P1.5 | 81.8% | 89.4% | 91.2% | 91.6% | 94.6% | 95.3% | | 95.3% | 92.2% | 94.7% |
| L-A-P1.4 | 82.0% | 89.5% | 91.4% | 91.4% | 94.9% | 95.4% | 95.3% | | 92.8% | 95.2% |
| L-A-P1g4 | 79.3% | 87.8% | 88.7% | 89.4% | 92.8% | 92.9% | 92.2% | 92.8% | | 87.1% |
| L-A-P1.p1 | 81.7% | 88.9% | 90.9% | 91.2% | 94.2% | 94.9% | 94.7% | 95.2% | 87.1% | |

## Pool2

This sample was composed of seven strains: three strains from the Far East (N17, CBS8439, and CBS8444), and four strains from continental Europe (C02, C05, C07 and C10). Analysing the data of this sample was more complicated than Pool1. This complication arose from the genome contamination, which was considerably higher than in Pool1, and the insertion size, which was about half the size of the Pool1 insertion size. As a result, when analysing the data of Pool2, as with Q62.5 and DBVPG4650, the genome contamination was removed from the data and three *de-novo* assemblies, D1, D2 and D3, were run with DCS11, DCS12 and DCS14 settings (Table 8-14) with 5% of the data. In DCS12, the minimum overlap was 50 bp and in DCS14 this option was off to increase the flexibility of assembling the reads that were shorter than those of Pool1. Analysing the contig sequences assembled in these three assemblies showed that there was no full-length L or M dsRNA contig with high coverage of the reads along the full length of the contig. It suggested that, since the genomic contamination in this sample was high at 62%, and there was a possibility of similarity between the dsRNA and the genomic sequences, perhaps removing the contamination had eliminated some reads from L and M dsRNA. To test the hypothesis, a *de-novo* assembly, D4, was run with the DCS14 setting without removing the contamination. However, the result did not change and no high-coverage contig with full-length L and M dsRNA were formed. A comparison between the L dsRNA contigs in the four *de-novo* assemblies presented similarities with Q62.5; in all of the contigs the coverage at both ends was significantly higher than in the middle. In the middle of the contigs, in some areas the coverage was very low. As a consequence, the sequence was not reliable. To see the productivity of the library preparation, the Pool2 reads were mapped to three types of L-A dsRNA with a medium-low sensitivity setting. Like the L-A contigs in *de-novo* assemblies, the coverage of the ends, especially the 5' end, was significantly higher than that of the middle sequence. The coverage of the 5' peak was about 1,186 times greater than the mean coverage in the middle (Figure 8-1). These results suggested that, since just 5% of the data was used in the Pool2 *de-*

*novo* assemblies, most of the L-A reads were from the ends and there were not enough reads to assemble the sequence of the middle of the dsRNA.

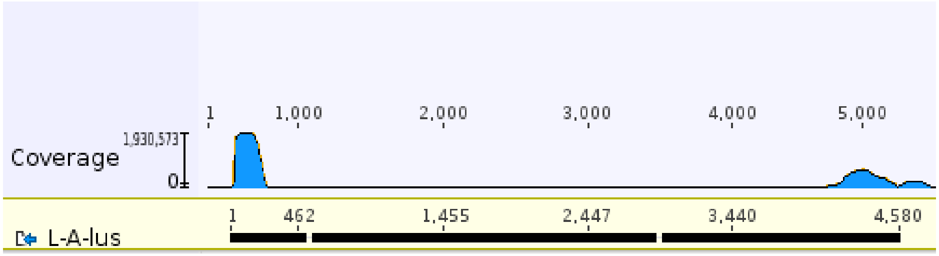


**Figure 8-1.** Mapping the Pool2 reads to L-A-lus. The coverage of both ends is significantly higher than the rest of the sequence.

## Analysing L-A dsRNA in Pool2

In order to have the same number of L-A reads all along the sequence, the Pool2 reads were firstly mapped to L-A-L1 with a medium-low sensitivity setting. Then, the 1,000 rows of reads assembled into the reference were selected and transferred into a folder. Next, the unused reads of the assembly were mapped to L-A-2 and, again, the same amount of reads in the assembly were selected and transferred into the same folder. These processes were repeated for L-A-lus, L-A-Q, L-A-D1 and L-A-C. In total, 51,145 reads were selected for the *de-novo* assembly of L-A. To increase the amount of data, the same procedure was repeated by selecting 2,000 rows. Overall 75,327 reads were selected in this group.

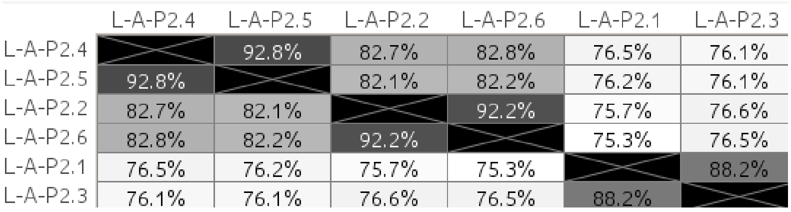
Three *de-novo* assemblies, DL1, DL2 and DL3, were run with the first group of reads with DCS11, DSC14 and DCS15 settings (Table 8-14); and two *de-novo* assemblies, DL4 and DL5, were run with the second group of L-A reads with DCS11 and DCS14. Analysis of the contigs in these assemblies showed that the assembly of the L-A contigs was considerably improved. Similar to Pool1, in each assembly the consensus sequences of the contigs were made and the consensus sequences of less than 100 reads and shorter than 500 reads were removed from the list. The contigs of the remaining sequences were studied to identify and remove the contigs that lacked a proper assembly. The selected consensus sequences of all assemblies were put in a folder and then mapped to the selected contigs from DL2. In DL2, there were eight contigs of between 2.1 kb to 4.6 kb in length. The consensus sequence of the contigs was blasted in NCBI and nr. All the contigs were blasted to the L-A dsRNA with a 74% - 76% identity. The alignment of the mapping of DL2 to each contig was studied separately. Given that the L-A reads in Pool2 were selected by mapping the known L-A dsRNAs, it is possible that some parts were missed that have a higher degree of diversity from these sequences. As a consequence, after finding the sequence of each type of L-A RNA, the Pool2 reads were mapped to the sequence with MCS2 setting (Table 8-22) and the assembly of the reads in the alignment was studied to be certain about the sequence. The consensus sequence of the mapping was compared to the *de-novo* contigs. The results of both the *de-novo* and map-to-reference are summarised in Table 8-19. Although neither contigs 6, 7 and 4, nor the contigs that were mapped to them, have the whole sequence of the L, the rest of the sequence was found using the overlapped sequence in their alignment. The sequences of the revealed Ls were identical with the consensus sequences of mapping the pool reads to their sequence with the MCS2 setting.

Table 8-19. The results of mapping all L dsRNA candidate contigs to D2 L contigs using MCS2 setting.

| Name of reference contig | Length of reference contig | Reference contig containing which part of L | Number of aligned contigs | Number of contigs in the alignment with whole sequence | % identical Sites | Name of the L | Number of the Pool2 reads mapped |
|---|---|---|---|---|---|---|---|
| **Contig 1** | 4,669 | Whole Seq | 8 | 1 | 99.9% | L-A-P2.1 | 398,174 |
| **Contig 3** | 4,592 | Whole Seq | 12 | 1 | 99.7% | L-A-P2.2 | 523,689 |
| **Contig 2** | 4,417 | Whole Seq | 6 | 4 | 96.5% | L-A-P2.3 | 47,052 |
| **Contig 6** | 2,043 | Gag ORF | 6 | 0 | 99.8% | L-A-P2.4 | 47,837 |
| **Contig 7** | 2,903 | Pol ORF | 4 | 0 | 100.0% | L-A-P2.5 | 84,702 |
| **Contig 4** | 2,523 | Gag ORF | 6 | 0 | 99.5% | L-A-P2.6 | 256,462 |

In total, the sequences of six different L-As were identified in Pool2: L-A-P2.1, L-A-P2.2, L-A-P2.3, L-A-P2.4, L-A-P2.5 and L-A-P2.6. Aligning the sequences showed that there are different L-A sequences with identity between 75.3% and 92.8%. Based on the identity, there are four groups of L dsRNA in Pool2: group one (L-A-P2.4 and L-A-P2.5); group two (L-A-P2.2 and L-A-P2.6); group three (L-A-P2.1); and group four (L-A-P2.3) (Table 8-20). In groups one and two the identity between the L dsRNA in each is 92%, and the groups are closer together (82% identity) than the other groups (76% identity). The identity between groups three (L-A-P2.1) and four (L-A-P2.3) is also higher (88.2%) than between the other groups.

Table 8-20**.** The identity between the six L dsRNA sequences in Pool2



| | L-A-P2.4 | L-A-P2.5 | L-A-P2.2 | L-A-P2.6 | L-A-P2.1 | L-A-P2.3 |
|---|---|---|---|---|---|---|
| L-A-P2.4 | | 92.8% | 82.7% | 82.8% | 76.5% | 76.1% |
| L-A-P2.5 | 92.8% | | 82.1% | 82.2% | 76.2% | 76.1% |
| L-A-P2.2 | 82.7% | 82.1% | | 92.2% | 75.7% | 76.6% |
| L-A-P2.6 | 82.8% | 82.2% | 92.2% | | 75.3% | 76.5% |
| L-A-P2.1 | 76.5% | 76.2% | 75.7% | 75.3% | | 88.2% |
| L-A-P2.3 | 76.1% | 76.1% | 76.6% | 76.5% | 88.2% | |

### Analysing M dsRNA in Pool2

As previously mentioned, since the sequence of the various types of M dsRNA is different, assembly of the coding and non-coding part of M dsRNA does not require custom sensitivity setting. As a result, in addition to the four *de-novo* assemblies with custom sensitivity, D1, D2, D3 and D4, two more *de-novo* assemblies, D5 and D6, were run using medium-low sensitivity with 10% of the data. The genome contamination was removed in D5. To find the M dsRNA contigs in each assembly, the L-A contigs were removed from the consensus sequence of each *de-novo* assembly by mapping to L-A with MCS1 setting (Table 8-22). Then the contigs with low coverage and length shorter than 500 bp were removed. The remaining contigs were blasted in NCBI and nr. Most of the contigs, especially the long contigs with high coverage, were blasted to the chromosomal DNA, particularly the rRNA genes in the genome. Even D4 and D5 assemblies, from which the genome contamination was removed, showed the same results. The remaining contigs did not have a high number of reads and had low coverage area in the sequence. It appears that, since the genome contamination is high at 62% of the data, by considering 5% to 10% of the data, there are not enough M dsRNA reads in the selected data to assemble. It is not possible to remove the genome contamination completely (mapping without considering MMQ200) because some of the dsRNA have similar sequences in the genome. Since most of the contigs' sequences belonged to rRNA sequences, to reduce the contamination and increase the number of reads, the Pool2 reads were mapped to the rRNA contigs. Then their unused reads were mapped to L-A dsRNA to remove L-A dsRNA reads. Following this, the 1,540,606 unused reads were used for *de-novo* assembly. Although 85% of the contaminant was removed, still most of the contigs were genome contamination.

Due to the complexity of the *de-novo* assembly of M dsRNA in Pool2, we tried to find the sequence of the M dsRNA by map-to-reference. The Pool2 reads were mapped to the M dsRNA reported in *S. cerevisiae* and *S. paradoxus* and the M sequences were found through this research. The results of the map-to-references are shown in Table 8-21.

Table 8-21. The results of mapping Pool2 to M dsRNA in *S. cerevisiae* and *S. paradoxus*

| **Reference** | **MQ200** | **Number of reads** | **The number of reads between the high coverage areas at the ends** | **Identity to the reference** |
|---|---|---|---|---|
| **MQ** | - | 192,628 | 19,331 | 85% |
| | + | 116,219 | 13,795 | 88.3% |
| **MC** | - | 103,309 | 11,713 | 91.9% |
| | + | 95,768 | 8,650 | 93.1% |
| **M-P1G2** | - | 3,618 | 3,618 | 80% |
| | + | 3,139 | 3,139 | 80.2% |
| ***S. cerevisiae* M28** | - | 1864 | 1864 | 92.2% |
| | + | 1640 | 1640 | 90.9% |

Of all M dsRNA, the Pool2 reads were assembled to only four types of M dsRNA: MQ, MC, M-p1g2 and M28. Analysing the assembly alignment of the M dsRNA with the full sequence of M dsRNA, MQ, MC and M28 indicated that most of the reads were assembled to the sequence of the ends, especially the 3' end. The number of the reads between these areas dramatically decreased between 1,640 and 19,331 reads (Table 8-21). That is why the sequence of the M dsRNA did not assemble in the *de-novo* assemblies when run with 5% of the data. Although poly A exists in all types of M dsRNA and in other samples, the number of the reads of poly A was high; the lowest coverage area in all of the M dsRNA in this sample comprised poly A sequences.

The identity between the sequences of M dsRNA of this sample and that of other samples is from 80% to 92.2%. The number of SNPs in the coding and non-coding part of MQ and MC, which have a higher number of reads when compared to other M dsRNAs, is significantly higher. This may be as a result of the polymorphism between different strains that have the same types of M dsRNA in the pool.

### 8.2.3 Optimising Map-to-references custom sensitivity settings

**Table 8-22. Map-to-references custom sensitivity settings**

| Setting name | Minimum Mapping Quality | Allow Gaps | | Minimum Overlap | Maximum Mismatch Per Read | Minimum Overlap Identity |
|---|---|---|---|---|---|---|
| | | Minimum Per Read | Maximum Gap size | | | |
| MCS1 | OFF | 60% | 1500 | OFF | 60% | OFF |
| MCS2 | 200 | OFF | | 200 | 2% | 98% |
| MCS3 | 200 | OFF | | 100 | 2% | 98% |
| MCS4 | 200 | OFF | | OFF | 2% | 98% |

### 8.2.4 Optimising *de-novo* custom sensitivity settings

Since the number of reads in Pool1 was high at 27,643,010 reads, and both the nuclear and mitochondrial contamination were very low (2% of the data), 1% of the data was used to test different sensitivities in order to save time.

After each *de-novo*, the consensus sequences of all contigs were made into a list. Then, the consensus sequences smaller than 500 bp and made of less than 50 reads were removed from the list. The contigs of the consensuses made of 50 to 100 reads were checked and any contig with low coverage areas was also removed. The remaining consensus sequences were mapped to MQ and L-A-Q with the easy custom sensitivity MCS1.

In the first *de-novo*, DCS1, no contigs were formed identically with the L-A-Q sequence. In addition, most the long contigs had a gap or a low coverage area in the middle. As a result, two unrelated sequences with high coverage were attached to each other using a gap or a low coverage area. By reducing the maximum mismatch per read to 2% in DCS2, some identical contigs with L-A-Q were formed. Although none of the contigs shared the whole sequence of L-A-Q, they overlapped with each other to cover the whole sequence of L-A-Q. There were still gaps and low coverage areas in the sequence of the contigs. In DCS3, turning off the "*allow gaps*" option in the software removed the gaps and low coverage areas from the middle of the contigs. Appling both changes together, i.e. reducing the maximum mismatch per read to 2% and turning off the allow-gaps in DCS4, in spite of improving the size and assembly of the contigs, could not merge the contigs which were identical to L-A-Q. Using just paired-end reads in DCS5 not only made the identical contigs of L-A-Q more fragmented, it also did not allow the contig identical to MQ to be assembled. Decreasing the maximum mismatch per read, increasing the minimum overlap identity, and increasing the minimum overlap in DCS6, DCS7, DCS8, and DCS9 reduced the size of the contigs and, as a consequence, increased the segmentation of identical contigs with L-A-Q and MQ. The coverage of the contigs was also reduced with the first two changes. These observations might suggest that, the higher the restrictions, the lower the size and quality of the contigs. In order to allow more flexibility to merge the fragments that have overlap, the maximum mismatch per read was increased to 5% and the minimum overlap identity was decreased to 95% in DCS10. The result showed that there were no identical large-size contigs mapped to L-A-Q. It appears that this setting is not restricted enough to separate the reads of different types of L dsRNA in the pool. Overall, results suggest that 2% maximum mismatch per read and 98% minimum overlap identity is the tipping point for categorising the reads. As a result, in DCS11, these two options were kept at 2% and 98% and the overlap identity was decreased to 100 bp to increase the flexibility of the assembly. In this assembly, one contig with 4,452 bp and 5,121 reads was identical with L-A-Q. The consensus sequence of this contig covered L-A-Q sequence from nucleotide 385 to 4,836. It is possible that decreasing the minimum overlap opened more space to attach more reads horizontally. Increasing the amount of data to 2% (DCS11-2 assembly) did not give the same result; the contigs identical with L-A-Q were split in three contigs which overlapped one another and covered the full sequence of the L-A-Q. Based on the Geneious team's opinion, the amount of data affects the result. On the one hand, contigs can merge with the addition of data. On the other hand, they may introduce more noise if some of the reads have sequencing errors or if a stringent condition is used that may be enough to split a contig apart.

### 8.2.5 Analysing the sequence of M dsRNAs and their preprotoxins

Aligning the sequence of all the coding sequences in Pool1 indicated that M-P1G3 and Mp1G4 are one type of M dsRNA with an identity of 89%. The other dsRNAs have different sequences. As a result, the name of the dsRNA changed to M-P1G3-1 and P1G3-2, respectively. Of all the M dsRNAs, the conserved sequence at the beginning of four of them did not assemble: MC, M-P1G1, M-P1G2 and M-P1G5. As a result, finding the start codon of these dsRNA was complicated.

In MC from CBS8441, the ORF starts from nucleotide 59 and the start codon is UUG. It has 16 nucleotides less than MC from Pool2 at 5' end. The AUG start codon of the MC from Pool2 is at residue 7. A comparison between the preprotoxin of both RNA also shows that the hydrophobic amino acids and the signal cleavage site, which exist at amino termini of all known preprotoxins, are missed in the preprotoxin of the MC from CBS8441. It seems that the first start codon in this dsRNA is missed during the sequencing. As a result, the preprotoxin of this dsRNA was translated from the beginning of the molecule. The start codon of M-P1G1 is at nucleotide 26. K-P1G1, similar to the other killer toxins, starts with hydrophobic amino acids. However, the signal cleavage site was not detected in the amino acids at the beginning of the preprotoxin. Therefore, three translation frames were studied from the beginning of the dsRNA. In frames 1 and 3, there were several stop codons in the sequence. Frame 2, which is the same frame as the predicted ORF, also had a stop codon before starting the ORF. It seems that the preprotoxin starts from the start codon of the ORF and the toxin has no signal cleavage site.

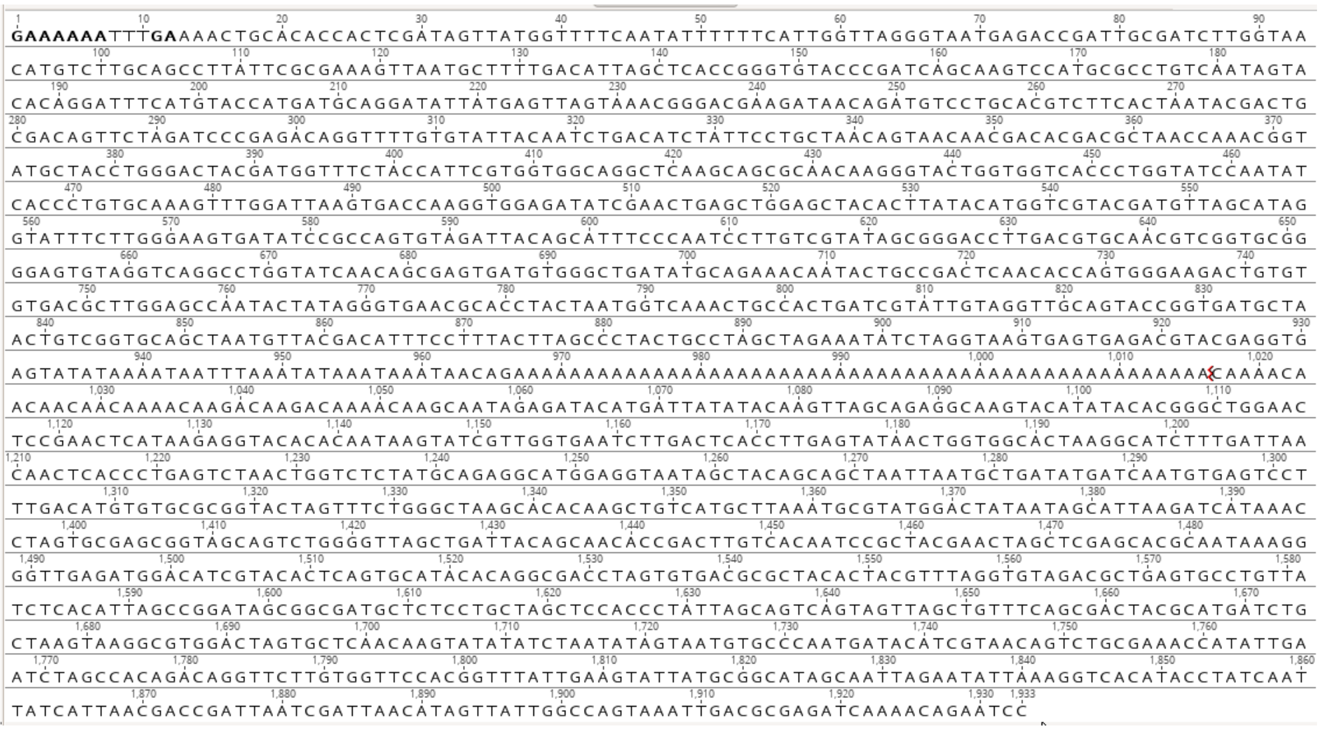
ORF of M-P1G2 starting from nucleotide 26. However, the start codon of this ORF is not AUG, too (it is UUG), and the length of the hydrophobic part of the predicted preprotoxin is very short. The first AUG in the sequence is at nucleotide 88 with frame 2, which seems unlikely to be the start codon of the preprotoxin. Frame 2 translations from beginning of the dsRNA, which is the same frame as the ORF, not only do not have a stop codon before the ORF but also have a stretch of hydrophobic amino acids at the beginning. It seems that, Similar to MC, the first start codon was in the sequences that were missing at the beginning of the dsRNA.

M-P1G5 has a length of 908 bp, which is less than the other M dsRNA. The AT-rich sequence is also shorter than that of other dsRNA. It seems that a longer part of the dsRNA is missing at the beginning of this M. Instead of starting from AUG, the ORFs starts with CUG and the hydrophobic amino acids and the cleavage sites are missing from the beginning of the protein. There is no stop codon at the six bp before the start codon. As a result translation of the preprotoxin was done from beginning of the dsRNA.

Aligning the M-P1G3-1 and M-P1G3-2 indicated that mutations in the first starting codon in M-P1G3-1 changed the AUG to AUA. Of the nucleotides aligned in the third position of the codon in M-P1G3-1, 94.3% are A. As a result, instead of the ORF starting from nucleotide 7, it starts from nucleotide 76. Therefore, the preprotoxin from M-P1G3-1 has 23 amino acids less than M-P1G3-2.

### 8.2.6 Sequences

#### MQ Q62.5



**GAAAAAA**TTT**GA**AAACTGCACACCACTCGATAGTTATGGTTTTCAATATTTTTTCATTGGTTAGGGTAATGAGACCGATTGCGATCTTGGTAA

CATGTCTTGCAGCCTTATTCGCGAAAGTTAATGCTTTTGACATTAGCTCACCGGGTGTACCCGATCAGCAAGTCCATGCGCCTGTCAATAGTA

CACAGGATTTCATGTACCATGATGCAGGATATTATGAGTTAGTAAACGGGACGAAGATAACAGATGTCCTGCACGTCTTCACTAATACGACTG

CGACAGTTCTAGATCCCGAGACAGGTTTTGTGTATTACAATCTGACATCTATTCCTGCTAACAGTAACAACGACACGACGCTAACCAAACGGT

ATGCTACCTGGGACTACGATGGTTTCTACCATTCGTGGTGGCAGGCTCAAGCAGCGCAACAAGGGTACTGGTGGTCACCCTGGTATCCAATAT

CACCCTGTGCAAAGTTTGGATTAAGTGACCAAGGTGGAGATATCGAACTGAGCTGGAGCTACACTTATACATGGTCGTACGATGTTAGCATAG

GTATTTCTTGGGAAGTGATATCCGCCAGTGTAGATTACAGCATTTCCCAATCCTTGTCGTATAGCGGGACCTTGACGTGCAACGTCGGTGCGG

GGAGTGTAGGTCAGGCCTGGTATCAACAGCGAGTGATGTGGGCTGATATGCAGAAACAATACTGCCGACTCAACACCAGTGGGAAGACTGTGT

GTGACGCTTGGAGCCAATACTATAGGGTGAACGCACCTACTAATGGTCAAACTGCCACTGATCGTATTGTAGGTTGCAGTACCGGTGATGCTA

ACTGTCGGTGCAGCTAATGTTACGACATTTCCTTTACTTAGCCCTACTGCCTAGCTAGAAATATCTAGGTAAGTGAGTGAGACGTACGAGGTG

AGTATATAAAATAATTTAAATATAAATAAATAACAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAACA

ACAACAACAAAACAAGACAAGACAAAACAAGCAATAGAGATACATGATTATATACAAGTTAGCAGAGGCAAGTACATATACACGGGCTGGAAC

TCCGAACTCATAAGAGGTACACACAATAAGTATCGTTGGTGAATCTTGACTCACCTTGAGTATAACTGGTGGCACTAAGGCATCTTTGATTAA

CAACTCACCCTGAGTCTAACTGGTCTCTATGCAGAGGCATGGAGGTAATAGCTACAGCAGCTAATTAATGCTGATATGATCAATGTGAGTCCT

TTGACATGTGTGCGCGGTACTAGTTTCTGGGCTAAGCACACAAGCTGTCATGCTTAAATGCGTATGGACTATAATAGCATTAAGATCATAAAC

CTAGTGCGAGCGGTAGCAGTCTGGGGTTAGCTGATTACAGCAACACCGACTTGTCACAATCCGCTACGAACTAGCTCGAGCACGCAATAAAGG

GGTTGAGATGGACATCGTACACTCAGTGCATACACAGGCGACCTAGTGTGACGCGCTACACTACGTTTAGGTGTAGACGCTGAGTGCCTGTTA

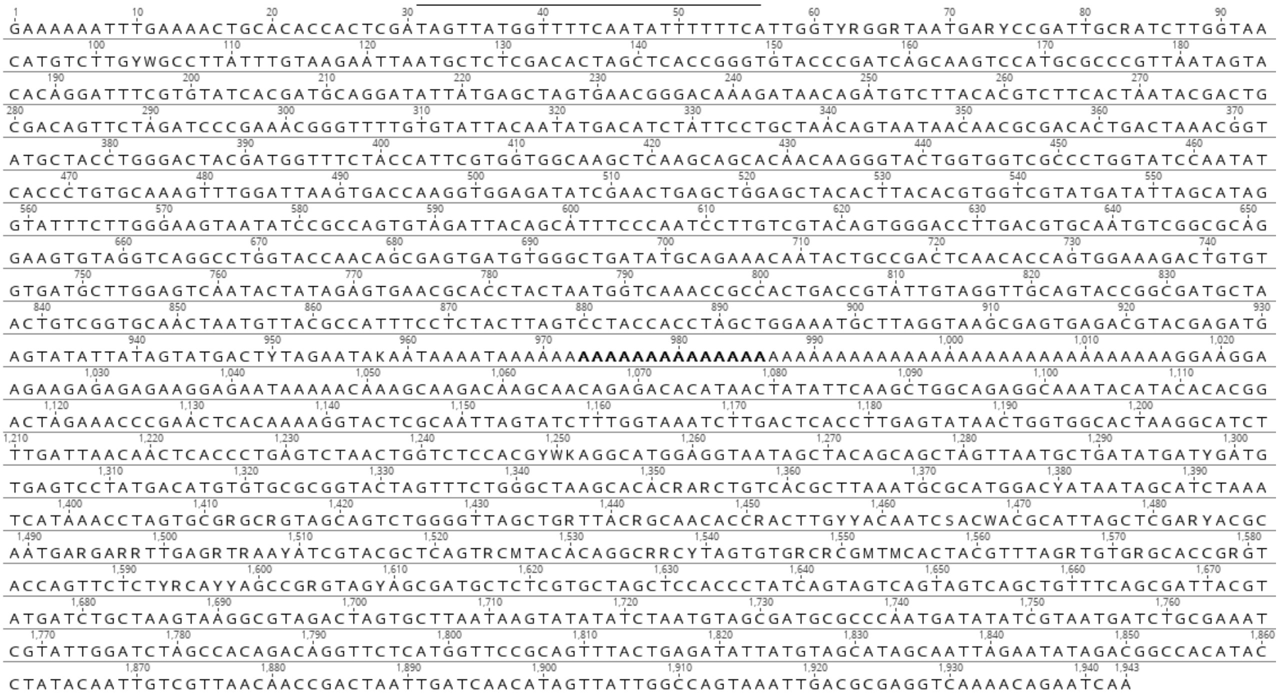
TCTCACATTAGCCGGATAGCGGCGATGCTCTCCTGCTAGCTCCACCCTATTAGCAGTCAGTAGTTAGCTGTTTCAGCGACTACGCATGATCTG

CTAAGTAAGGCGTGGACTAGTGCTCAACAAGTATATATCTAATATAGTAATGTGCCCAATGATACATCGTAACAGTCTGCGAAACCATATTGA

ATCTAGCCACAGACAGGTTCTTGTGGTTCCACGGTTTATTGAAGTATTATGCGGCATAGCAATTAGAATATTAAAGGTCACATACCTATCAAT

TATCATTAACGACCGATTAATCGATTAACATAGTTATTGGCCAGTAAATTGACGCGAGATCAAAACAGAATCC

#### MQ Pool2



GAAAAAATTTGAAAACTGCACACCACTCGATAGTTATGGTTTTCAATATTTTTTCATTGGTYRGGRTAATGARYCCGATTGCRATCTTGGTAA

CATGTCTTGYWGCCTTATTTGTAAGAATTAATGCTCTCGACACTAGCTCACCGGGTGTACCCGATCAGCAAGTCCATGCGCCCGTTAATAGTA

CACAGGATTTCGTGTATCACGATGCAGGATATTATGAGCTAGTGAACGGGACAAAGATAACAGATGTCTTACACGTCTTCACTAATACGACTG

CGACAGTTCTAGATCCCGAAACGGGTTTTGTGTATTACAATATGACATCTATTCCTGCTAACAGTAATAACAACGCGACACTGACTAAACGGT

ATGCTACCTGGGACTACGATGGTTTCTACCATTCGTGGTGGCAAGCTCAAGCAGCACAACAAGGGTACTGGTGGTCGCCCTGGTATCCAATAT

CACCCTGTGCAAAGTTTGGATTAAGTGACCAAGGTGGAGATATCGAACTGAGCTGGAGCTACACTTACACGTGGTCGTATGATATTAGCATAG

GTATTTCTTGGGAAGTAATATCCGCCAGTGTAGATTACAGCATTTCCCAATCCTTGTCGTACAGTGGGACCTTGACGTGCAATGTCGGCGCAG

GAAGTGTAGGTCAGGCCTGGTACCAACAGCGAGTGATGTGGGCTGATATGCAGAAACAATACTGCCGACTCAACACCAGTGGAAAGACTGTGT

GTGATGCTTGGAGTCAATACTATAGAGTGAACGCACCTACTAATGGTCAAACCGCCACTGACCGTATTGTAGGTTGCAGTACCGGCGATGCTA

ACTGTCGGTGCAACTAATGTTACGCCATTTCCTCTACTTAGTCCTACCACCTAGCTGGAAATGCTTAGGTAAGCGAGTGAGACGTACGAGATG

AGTATATTATAGTATGACTYTAGAATAKAATAAAATAAAAAA**AAAAAAAAAAAAAA**AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAGGA

AGAAGAGAGAGAAGGAGAATAAAAACAAAGCAAGACAAGCAACAGAGACACATAACTATATTCAAGCTGGCAGAGGCAAATACATACACACGG

ACTAGAAACCCGAACTCACAAAAGGTACTCGCAATTAGTATCTTTGGTAAATCTTGACTCACCTTGAGTATAACTGGTGGCACTAAGGCATCT

TTGATTAACAACTCACCCTGAGTCTAACTGGTCTCCACGYWKAGGCATGGAGGTAATAGCTACAGCAGCTAGTTAATGCTGATATGATYGATG

TGAGTCCTATGACATGTGTGCGCGGTACTAGTTTCTGGGCTAAGCACACRARCTGTCACGCTTAAATGCGCATGGACYATAATAGCATCTAAA

TCATAAACCTAGTGCGRGCRGTAGCAGTCTGGGGTTAGCTGRTTACRGCAACACCRACTTGYYACAATCSACWACGCATTAGCTCGARYACGC

AATGARGARRTTGAGRTRAAYATCGTACGCTCAGTRCMTACACAGGCRRCYTAGTGTGRCRCGMTMCACTACGTTTAGRTGTGRGCACCGRGT

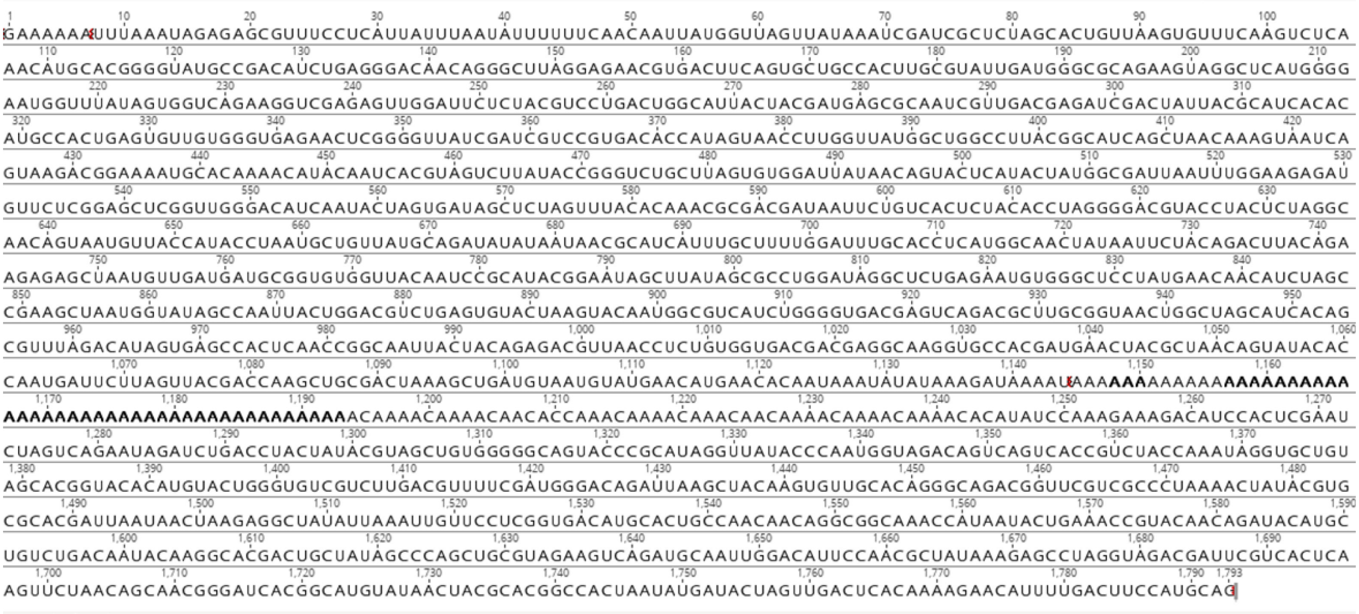
ACCAGTTCTCTYRCAYYAGCCGRGTAGYAGCGATGCTCTCGTGCTAGCTCCACCCTATCAGTAGTCAGTAGTCAGCTGTTTCAGCGATTACGT

ATGATCTGCTAAGTAAGGCGTAGACTAGTGCTTAATAAGTATATATCTAATGTAGCGATGCGCCCAATGATATATCGTAATGATCTGCGAAAT

CGTATTGGATCTAGCCACAGACAGGTTCTCATGGTTCCGCAGTTTACTGAGATATTATGTAGCATAGCAATTAGAATATAGACGGCCACATAC

CTATACAATTGTCGTTAACAACCGACTAATTGATCAACATAGTTATTGGCCAGTAAATTGACGCGAGGTCAAAACAGAATCAA

## M28 DBVPG4650



GAAAAAAUUUAAAUAGAGAGCGUUUCCUCAUUAUUUAAUAUUUUUUCAACAAUUAUGGUUAGUUAUAAAUCGAUCGCUCUAGCACUGUUAAGUGUUUCAAGUCUCA

AACAUGCACGGGGUAUGCCGACAUCUGAGGGACAACAGGGCUUAGGAGAACGUGACUUCAGUGCUGCCACUUGCGUAUUGAUGGGCGCAGAAGUAGGCUCAUGGGG

AAUGGUUUAUAGUGGUCAGAAGGUCGAGAGUUGGAUUCUCUACGUCCUGACUGGCAUUACUACGAUGAGCGCAAUCGUUGACGAGAUCGACUAUUACGCAUCACAC

AUGCCACUGAGUGUUGUGGGUGAGAACUCGGGGUUAUCGAUCGUCCGUGACACCAUAGUAACCUUGGUUAUGGCUGGCCUUACGGCAUCAGCUAACAAAGUAAUCA

GUAAGACGGAAAAUGCACAAAACAUACAAUCACGUAGUCUUAUACCGGGUCUGCUUAGUGUGGAUUAUAACAGUACUCAUACUAUGGCGAUUAAUUUGGAAGAGAU

GUUCUCGGAGCUCGGUUGGGACAUCAAUACUAGUGAUAGCUCUAGUUUACACAAACGCGACGAUAAUUCUGUCACUCUACACCUAGGGGACGUACCUACUCUAGGC

AACAGUAAUGUUACCAUACCUAAUGCUGUUAUGCAGAUAUAUAAUAACGCAUCAUUUGCUUUUGGAUUUGCACCUCAUGGCAACUAUAAUUCUACAGACUUACAGA

AGAGAGCUAAUGUUGAUGAUGCGGUGUGGUUACAAUCCGCAUACGGAAUAGCUUAUAGCGCCUGGAUAGGCUCUGAGAAUGUGGGCUCCUAUGAACAACAUCUAGC

CGAAGCUAAUGGUAUAGCCAAUUACUGGACGUCUGAGUGUACUAAGUACAAUGGCGUCAUCUGGGGUGACGAGUCAGACGCUUGCGGUAACUGGCUAGCAUCACAG

CGUUUAGACAUAGUGAGCCACUCAACCGGCAAUUACUACAGAGACGUUAACCUCUGUGGUGACGACGAGGCAAGGUGCCACGAUGAACUACGCUAACAGUAUACAC

CAAUGAUUCUUAGUUACGACCAAGCUGCGACUAAAGCUGAUGUAAUGUAUGAACAUGAACACAAUAAAUAUAUAAAGAUAAAAUAAA**AAA**AAAAAA**AAAAAAAAAA**

**AAAAAAAAAAAAAAAAAAAAAAAAAAA**ACAAAACAAAACAACACCAAACAAAACAAACAACAAAACAAAACAAAACACAUAUCCAAAGAAAGACAUCCACUCGAAU

CUAGUCAGAAUAGAUCUGACCUACUAUACGUAGCUGUGGGGGCAGUACCCGCAUAGGUUAUACCCAAUGGUAGACAGUCAGUCACCGUCUACCAAAUAGGUGCUGU

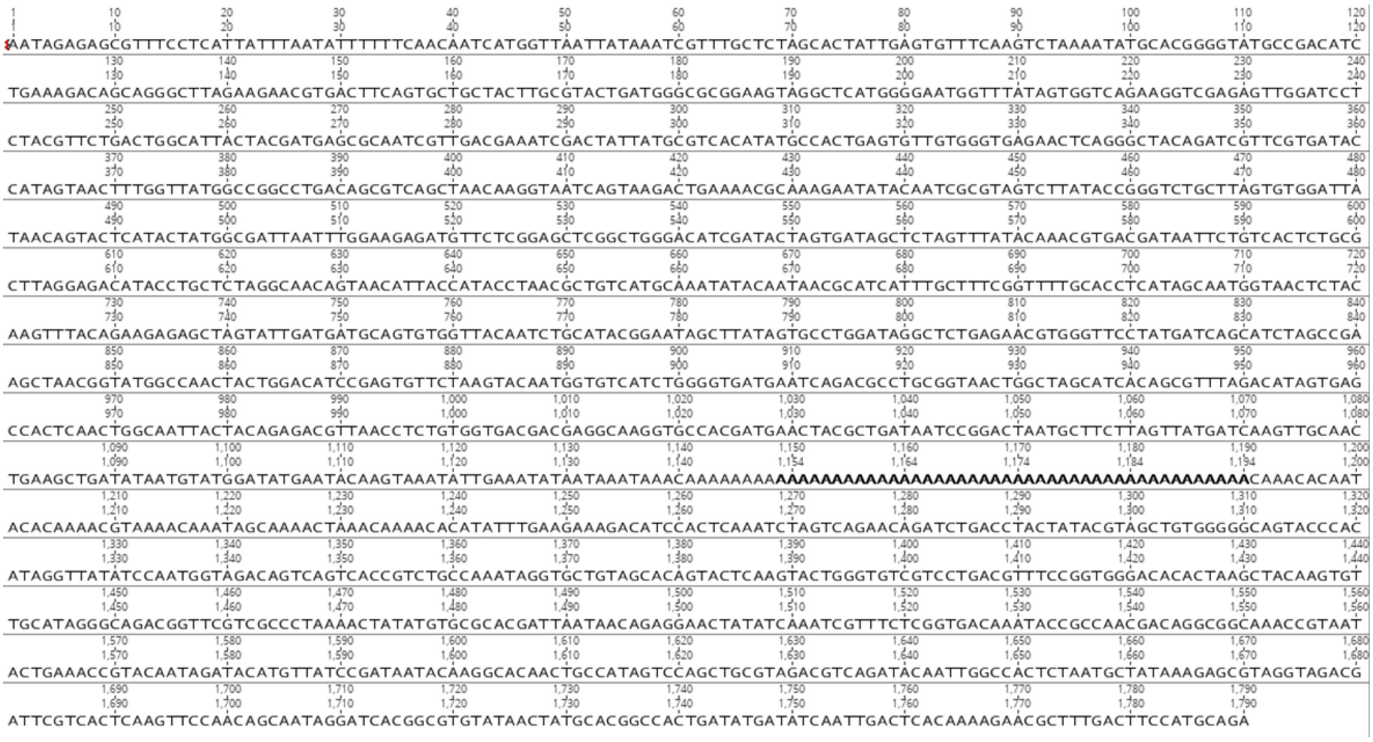
AGCACGGUACACAUGUACUGGGUGUCGUCUUGACGUUUCGAUGGGACAGAUUAAGCUACAAGUGUUGCACAGGGCAGACGGUUCGUCGCCCUAAAACUAUACGUG

CGCACGAUUAAUAACUAAGAGGCUAUAUUAAAUUGUUCCUCGGUGACAUGCACUGCCAACAACAGGCGGCAAACCAUAAUACUGAAACCGUACAACAGAUACAUGC

UGUCUGACAAUACAAGGCACGACUGCUAUAGCCCAGCUGCGUAGAAGUCAGAUGCAAUUGGACAUUCCAACGCUAUAAAGAGCCUAGGUAGACGAUUCGUCACUCA

AGUUCUAACAGCAACGGGAUCACGGCAUGUAUAACUACGCACGGCCACUAAUAUGAUACUAGUUGACUCACAAAAGAACAUUUUGACUUCCAUGCAG

## M28 Pool2



AATAGAGAGCGTTTCCTCATTATTTAATATTTTTTCAACAATCATGGTTAATTATAAATCGTTTGCTCTAGCACTATTGAGTGTTTCAAGTCTAAAATATGCACGGGGTATGCCGACATC

TGAAAGACAGCAGGGCTTAGAAGAACGTGACTTCAGTGCTGCTACTTGCGTACTGATGGGCGCGGAAGTAGGCTCATGGGGAATGGTTTATAGTGGTCAGAAGGTCGAGAGTTGGATCCT

CTACGTTCTGACTGGCATTACTACGATGAGCGCAATCGTTGACGAAATCGACTATTATGCGTCACATATGCCACTGAGTGTTGTGGGTGAGAACTCAGGGCTACAGATCGTTCGTGATAC

CATAGTAACTTTGGTTATGGCCGGCCTGACAGCGTCAGCTAACAAGGTAATCAGTAAGACTGAAAACGCAAAGAATATACAATCGCGTAGTCTTATACCGGGTCTGCTTAGTGTGGATTA

TAACAGTACTCATACTATGGCGATTAATTTGGAAGAGATGTTCTCGGAGCTCGGCTGGGACATCGATACTAGTGATAGCTCTAGTTTATACAAACGTGACGATAATTCTGTCACTCTGCG

CTTAGGAGACATACCTGCTCTAGGCAACAGTAACATTACCATACCTAACGCTGTCATGCAAATATACAATAACGCATCATTTGCTTTCGGTTTTGCACCTCATAGCAATGGTAACTCTAC

AAGTTTACAGAAGAGAGCTAGTATTGATGATGCAGTGTGGTTACAATCTGCATACGGAATAGCTTATAGTGCCTGGATAGGCTCTGAGAACGTGGGTTCCTATGATCAGCATCTAGCCGA

AGCTAACGGTATGGCCAACTACTGGACATCCGAGTGTTCTAAGTACAATGGTGTCATCTGGGGTGATGAATCAGACGCCTGCGGTAACTGGCTAGCATCACAGCGTTTAGACATAGTGAG

CCACTCAACTGGCAATTACTACAGAGACGTTAACCTCTGTGGTGACGACGAGGCAAGGTGCCACGATGAACTACGCTGATAATCCGGACTAATGCTTCTTAGTTATGATCAAGTTGCAAC

TGAAGCTGATATAATGTATGGATATGAATACAAGTAAATATTGAAATATAATAAATAAACAAAAAAAA**AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA**CAAACACAAT

ACACAAAACGTAAAACAAATAGCAAAACTAAACAAAACACATATTTGAAGAAAGACATCCACTCAAATCTAGTCAGAACAGATCTGACCTACTATACGTAGCTGTGGGGGCAGTACCCAC

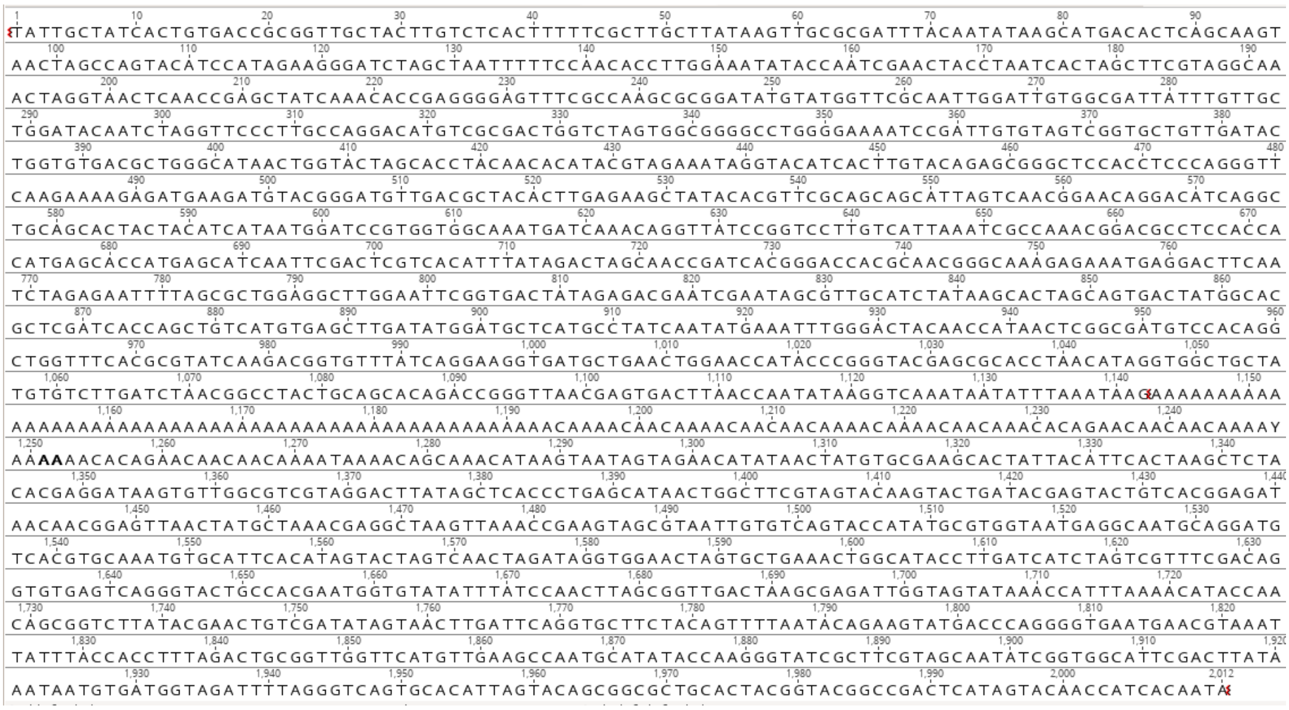
ATAGGTTATATCCAATGGTAGACAGTCAGTCACCGTCTGCCAAATAGGTGCTGTAGCACAGTACTCAAGTACTGGGTGTCGTCCTGACGTTTCCGGTGGGACACACTAAGCTACAAGTGT

TGCATAGGGCAGACGGTTCGTCGCCCTAAAACTATATGTGCGCACGATTAATAACAGAGGAACTATATCAAATCGTTTCTCGGTGACAAATACCGCCAACGACAGGCGGCAAACCGTAAT

ACTGAAACCGTACAATAGATACATGTTATCCGATAATACAAGGCACAACTGCCATAGTCCAGCTGCGTAGACGTCAGATACAATTGGCCACTCTAATGCTATAAAGAGCGTAGGTAGACG

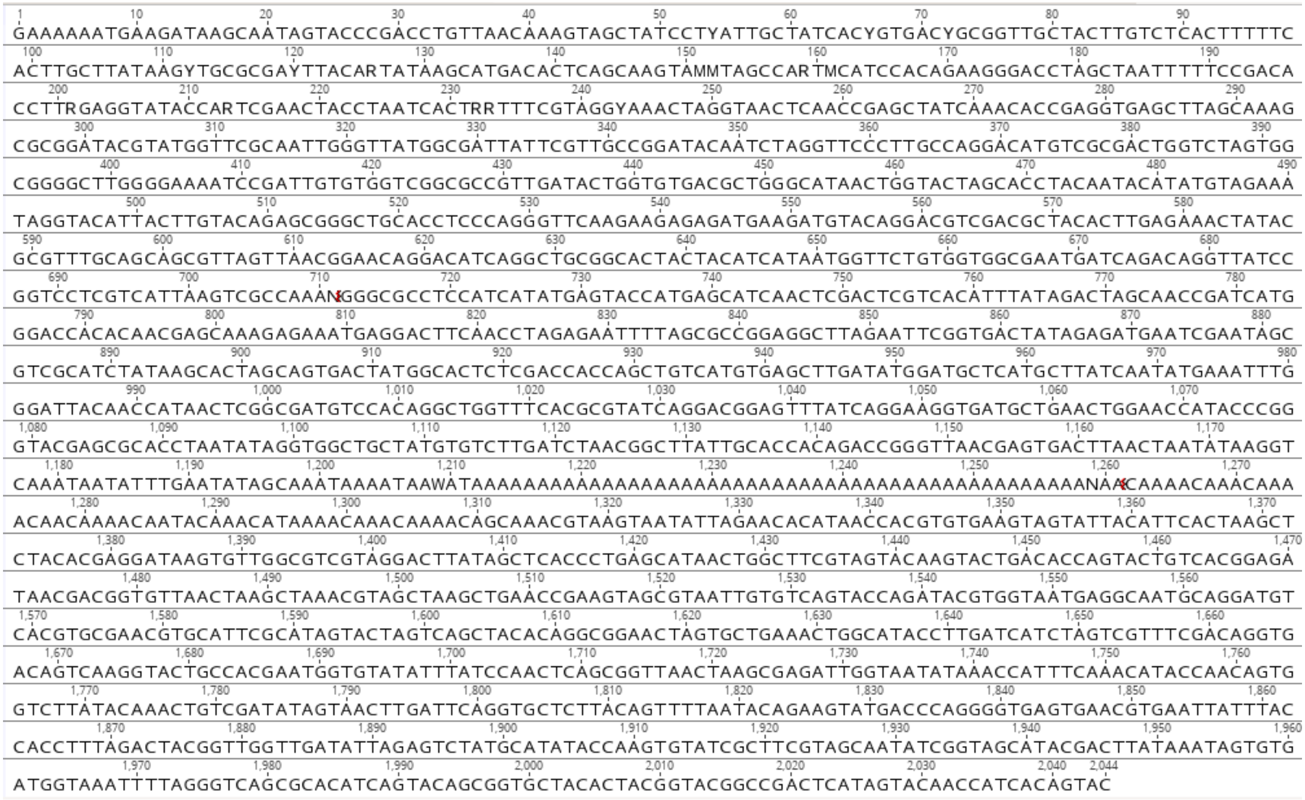
ATTCGTCACTCAAGTTCCAACAGCAATAGGATCACGGCGTGTATAACTATGCACGGCCACTGATATGATATCAATTGACTCACAAAAGAACGCTTTGACTTCCATGCAGA

## MC CBS8441



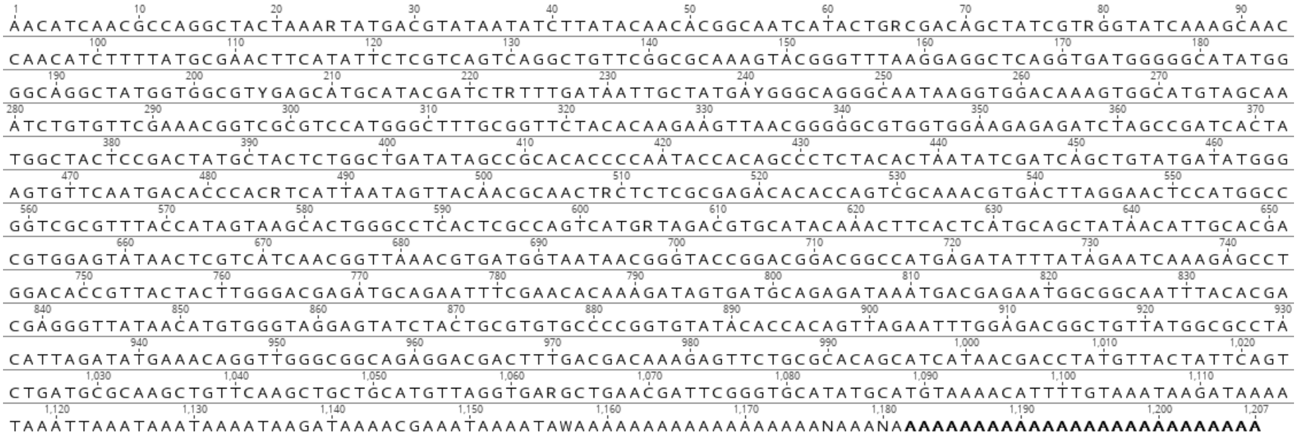
```
TATTGCTATCACTGTGACCGCGGTTGCTACTTGTCTCACTTTTTCGCTTGCTTATAAGTTGCGCGATTTACAATATAAGCATGACACTCAGCAAGT
AACTAGCCAGTACATCCATAGAAGGGATCTAGCTAATTTTTCCAACACCTTGGAAATATACCAATCGAACTACCTAATCACTAGCTTCGTAGGCAA
ACTAGGTAACTCAACCGAGCTATCAAACACCGAGGGGAGTTTCGCCAAGCGCGGATATGTATGGTTCGCAATTGGATTGTGGCGATTATTTGTTGC
TGGATACAATCTAGGTTCCCTTGCCAGGACATGTCGCGACTGGTCTAGTGGCGGGGCCTGGGGAAAATCCGATTGTGTAGTCGGTGCTGTTGATAC
TGGTGTGACGCTGGGCATAACTGGTACTAGCACCTACAACACATACGTAGAAATAGGTACATCACTTGTACAGAGCGGGCTCCACCTCCCAGGGTT
CAAGAAAAGAGATGAAGATGTACGGGATGTTGACGCTACACTTGAGAAGCTATACACGTTCGCAGCAGCATTAGTCAACGGAACAGGACATCAGGC
TGCAGCACTACTACATCATAATGGATCCGTGGTGGCAAATGATCAAACAGGTTATCCGGTCCTTGTCATTAAATCGCCAAACGGACGCCTCCACCA
CATGAGCACCATGAGCATCAATTCGACTCGTCACATTTATAGACTAGCAACCGATCACGGGACCACGCAACGGGCAAAGAGAAATGAGGACTTCAA
TCTAGAGAATTTTAGCGCTGGAGGCTTGGAATTCGGTGACTATAGAGACGAATCGAATAGCGTTGCATCTATAAGCACTAGCAGTGACTATGGCAC
GCTCGATCACCAGCTGTCATGTGAGCTTGATATGGATGCTCATGCCTATCAATATGAAATTTGGGACTACAACCATAACTCGGCGATGTCCACAGG
CTGGTTTCACGCGTATCAAGACGGTGTTTATCAGGAAGGTGATGCTGAACTGGAACCATACCCGGGTACGAGCGCACCTAACATAGGTGGCTGCTA
TGTGTCTTGATCTAACGGCCTACTGCAGCACAGACCGGGTTAACGAGTGACTTAACCAATATAAGGTCAAATAATATTTAAATAAGAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAACAACAAAACAACAACAAAACAAAACAACAAACACAGAACAACAACAAAAY
AAAAAACACAGAACAACAACAAAATAAAACAGCAAACATAAGTAATAGTAGAACATATAACTATGTGCGAAGCACTATTACATTCACTAAGCTCTA
CACGAGGATAAGTGTTGGCGTCGTAGGACTTATAGCTCACCCTGAGCATAACTGGCTTCGTAGTACAAGTACTGATACGAGTACTGTCACGGAGAT
AACAACGGAGTTAACTATGCTAAACGAGGCTAAGTTAAACCGAAGTAGCGTAATTGTGTCAGTACCATATGCGTGGTAATGAGGCAATGCAGGATG
TCACGTGCAAATGTGCATTCACATAGTACTAGTCAACTAGATAGGTGGAACTAGTGCTGAAACTGGCATACCTTGATCATCTAGTCGTTTCGACAG
GTGTGAGTCAGGGTACTGCCACGAATGGTGTATATTTATCCAACTTAGCGGTTGACTAAGCGAGATTGGTAGTATAAACCATTTAAAACATACCAA
CAGCGGTCTTATACGAACTGTCGATATAGTAACTTGATTCAGGTGCTTCTACAGTTTTAATACAGAAGTATGACCCAGGGGTGAATGAACGTAAAT
TATTTACCACCTTTAGACTGCGGTTGGTTCATGTTGAAGCCAATGCATATACCAAGGGTATCGCTTCGTAGCAATATCGGTGGCATTCGACTTATA
AATAATGTGATGGTAGATTTTAGGGTCAGTGCACATTAGTACAGCGGCGCTGCACTACGGTACGGCCGACTCATAGTACAACCATCACAATA
```

## MC Pool2



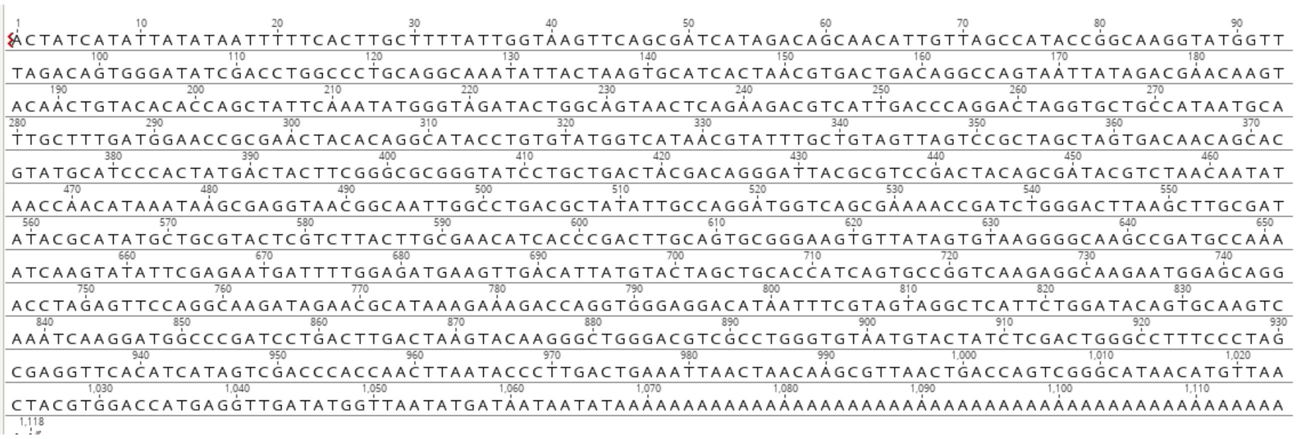
```
GAAAAAATGAAGATAAGCAATAGTACCCGACCTGTTAACAAAGTAGCTATCCTYATTGCTATCACYGTGACYGCGGTTGCTACTTGTCTCACTTTTTC
ACTTGCTTATAAGYTGCGCGAYTTACARTATAAGCATGACACTCAGCAAGTAMMTAGCCARTMCATCCACAGAAGGGACCTAGCTAATTTTTCCGACA
CCTTRGAGGTATACCARTCGAACTACCTAATCACTRRTTTCGTAGGYAAACTAGGTAACTCAACCGAGCTATCAAACACCGAGGTGAGCTTAGCAAAG
CGCGGATACGTATGGTTCGCAATTGGGTTATGGCGATTATTCGTTGCCGGATACAATCTAGGTTCCCTTGCCAGGACATGTCGCGACTGGTCTAGTGG
CGGGGCTTGGGGAAAATCCGATTGTGTGGTCGGCGCCGTTGATACTGGTGTGACGCTGGGCATAACTGGTACTAGCACCTACAATACATATGTAGAAA
TAGGTACATTACTTGTACAGAGCGGGCTGCACCTCCCAGGGTTCAAGAAGAGAGATGAAGATGTACAGGACGTCGACGCTACACTTGAGAAACTATAC
GCGTTTGCAGCAGCGTTAGTTAACGGAACAGGACATCAGGCTGCGGCACTACTACATCATAATGGTTCTGTGGTGGCGAATGATCAGACAGGTTATCC
GGTCCTCGTCATTAAGTCGCCAAANGGGCGCCTCCATCATATGAGTACCATGAGCATCAACTCGACTCGTCACATTTATAGACTAGCAACCGATCATG
GGACCACACAACGAGCAAAGAGAAATGAGGACTTCAACCTAGAGAATTTTAGCGCCGGAGGCTTAGAATTCGGTGACTATAGAGATGAATCGAATAGC
GTCGCATCTATAAGCACTAGCAGTGACTATGGCACTCTCGACCACCAGCTGTCATGTGAGCTTGATATGGATGCTCATGCTTATCAATATGAAATTTG
GGATTACAACCATAACTCGGCGATGTCCACAGGCTGGTTTCACGCGTATCAGGACGGAGTTTATCAGGAAGGTGATGCTGAACTGGAACCATACCCGG
GTACGAGCGCACCTAATATAGGTGGCTGCTATGTGTCTTGATCTAACGGCTTATTGCACCACAGACCGGGTTAACGAGTGACTTAACTAATATAAGGT
CAAATAATATTTGAATATAGCAAATAAAATAAWATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAACAAAACAAACAAA
ACAACAAAACAATACAAACATAAAACAAACAAAACAGCAAACGTAAGTAATATTAGAACACATAACCACGTGTGAAGTAGTATTACATTCACTAAGCT
CTACACGAGGATAAGTGTTGGCGTCGTAGGACTTATAGCTCACCCTGAGCATAACTGGCTTCGTAGTACAAGTACTGACACCAGTACTGTCACGGAGA
TAACGACGGTGTTAACTAAGCTAAACGTAGCTAAGCTGAACCGAAGTAGCGTAATTGTGTCAGTACCAGATACGTGGTAATGAGGCAATGCAGGATGT
CACGTGCGAACGTGCATTCGCATAGTACTAGTCAGCTACACAGGCGGAACTAGTGCTGAAACTGGCATACCTTGATCATCTAGTCGTTTCGACAGGTG
ACAGTCAAGGTACTGCCACGAATGGTGTATATTTATCCAACTCAGCGGTTAACTAAGCGAGATTGGTAATATAAACCATTTCAAACATACCAACAGTG
GTCTTATACAAACTGTCGATATAGTAACTTGATTCAGGTGCTCTTACAGTTTTAATACAGAAGTATGACCCAGGGGTGAGTGAACGTGAATTATTTAC
CACCTTTAGACTACGGTTGGTTGATATTAGAGTCTATGCATATACCAAGTGTATCGCTTCGTAGCAATATCGGTAGCATACGACTTATAAATAGTGTG
ATGGTAAATTTTAGGGTCAGCGCACATCAGTACAGCGGTGCTACACTACGGTACGGCCGACTCATAGTACAACCATCACAGTAC
```

**M-P1G1 Pool1**



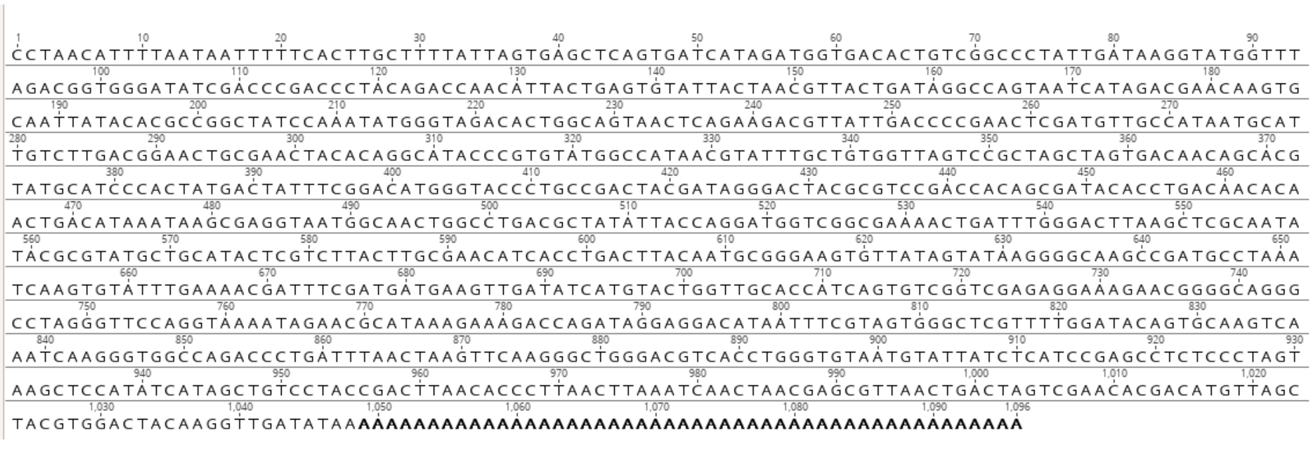
```
AACATCAACGCCAGGCTACTAAARTATGACGTATAATATCTTATACAACACGGCAATCATACTGRCGACAGCTATCGTRGGTATCAAAGCAAC
CAACATCTTTTATGCGAACTTCATATTCTCGTCAGTCAGGCTGTTCGGCGCAAAGTACGGGTTTAAGGAGGCTCAGGTGATGGGGGCATATGG
GGCAGGCTATGGTGGCGTYGAGCATGCATACGATCTRTTTGATAATTGCTATGAYGGGCAGGGCAATAAGGTGGACAAAGTGGCATGTAGCAA
ATCTGTGTTCGAAACGGTCGCGTCCATGGGCTTTGCGGTTCTACACAAGAAGTTAACGGGGGCGTGGTGGAAGAGAGATCTAGCCGATCACTA
TGGCTACTCCGACTATGCTACTCTGGCTGATATAGCCGCACACCCCAATACCACAGCCCTCTACACTAATATCGATCAGCTGTATGATATGGG
AGTGTTCAATGACACCCACRTCATTAATAGTTACAACGCAACTRCTCTCGCGAGACACACCAGTCGCAAACGTGACTTAGGAACTCCATGGCC
GGTCGCGTTTACCATAGTAAGCACTGGGCCTCACTCGCCAGTCATGRTAGACGTGCATACAAACTTCACTCATGCAGCTATAACATTGCACGA
CGTGGAGTATAACTCGTCATCAACGGTTAAACGTGATGGTAATAACGGGTACCGGACGGACGGCCATGAGATATTTATAGAATCAAAGAGCCT
GGACACCGTTACTACTTGGGACGAGATGCAGAATTTCGAACACAAAGATAGTGATGCAGAGATAAATGACGAGAATGGCGGCAATTTACACGA
CGAGGGTTATAACATGTGGGTAGGAGTATCTACTGCGTGTGCCCCGGTGTATACACCACAGTTAGAATTTGGAGACGGCTGTTATGGCGCCTA
CATTAGATATGAAACAGGTTGGGCGGCAGAGGACGACTTTGACGACAAAGAGTTCTGCGCACAGCATCATAACGACCTATGTTACTATTCAGT
CTGATGCGCAAGCTGTTCAAGCTGCTGCATGTTAGGTGARGCTGAACGATTCGGGTGCATATGCATGTAAAACATTTTGTAAATAAGATAAAA
TAAATTAAATAAATAAAATAAGATAAAACGAAATAAAATAWAAAAAAAAAAAAAAAAAAANAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

**M-P1G2 Pool1**



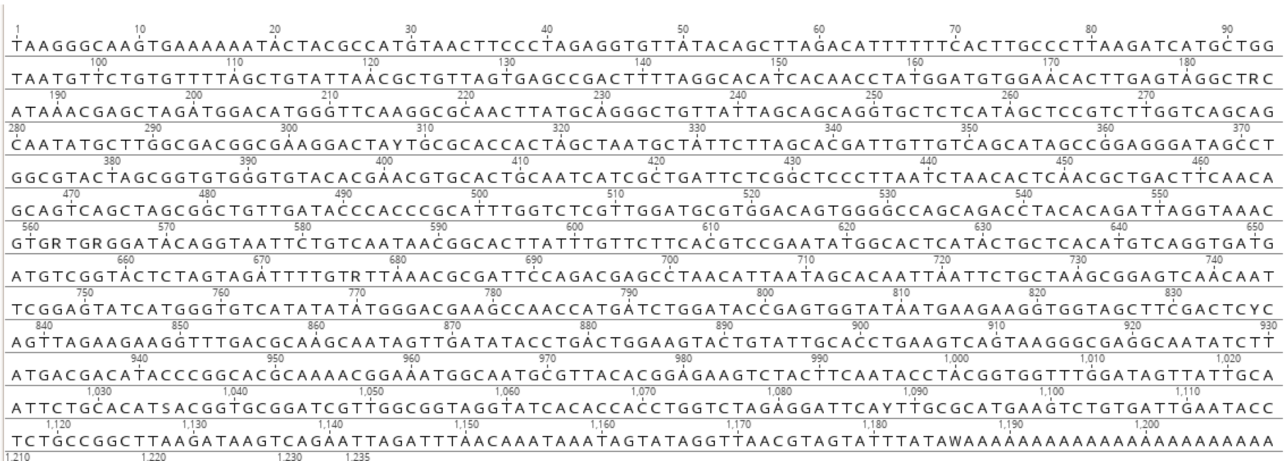
```
ACTATCATATTATATAATTTTTCACTTGCTTTTATTGGTAAGTTCAGCGATCATAGACAGCAACATTGTTAGCCATACCGGCAAGGTATGGTT
TAGACAGTGGGATATCGACCTGGCCCTGCAGGCAAATATTACTAAGTGCATCACTAACGTGACTGACAGGCCAGTAATTATAGACGAACAAGT
ACAACTGTACACACCAGCTATTCAAATATGGGTAGATACTGGCAGTAACTCAGAAGACGTCATTGACCCAGGACTAGGTGCTGCCATAATGCA
TTGCTTTGATGGAACCGCGAACTACACAGGCATACCTGTGTATGGTCATAACGTATTTGCTGTAGTTAGTCCGCTAGCTAGTGACAACAGCAC
GTATGCATCCCACTATGACTACTTCGGGCGCGGGTATCCTGCTGACTACGACAGGGATTACGCGTCCGACTACAGCGATACGTCTAACAATAT
AACCAACATAAATAAGCGAGGTAACGGCAATTGGCCTGACGCTATATTGCCAGGATGGTCAGCGAAAACCGATCTGGGACTTAAGCTTGCGAT
ATACGCATATGCTGCGTACTCGTCTTACTTGCGAACATCACCCGACTTGCAGTGCGGGAAGTGTTATAGTGTAAGGGGCAAGCCGATGCCAAA
ATCAAGTATATTCGAGAATGATTTTGGAGATGAAGTTGACATTATGTACTAGCTGCACCATCAGTGCCGGTCAAGAGGCAAGAATGGAGCAGG
ACCTAGAGTTCCAGGCAAGATAGAACGCATAAAGAAAGACCAGGTGGGAGGACATAATTTCGTAGTAGGCTCATTCTGGATACAGTGCAAGTC
AAATCAAGGATGGCCCGATCCTGACTTGACTAAGTACAAGGGCTGGGACGTCGCCTGGGTGTAATGTACTATCTCGACTGGGCCTTTCCCTAG
CGAGGTTCACATCATAGTCGACCCACCAACTTAATACCCTTGACTGAAATTAACTAACAAGCGTTAACTGACCAGTCGGGCATAACATGTTAA
CTACGTGGACCATGAGGTTGATATGGTTAATATGATAATAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

**M-P1G2 Pool2**



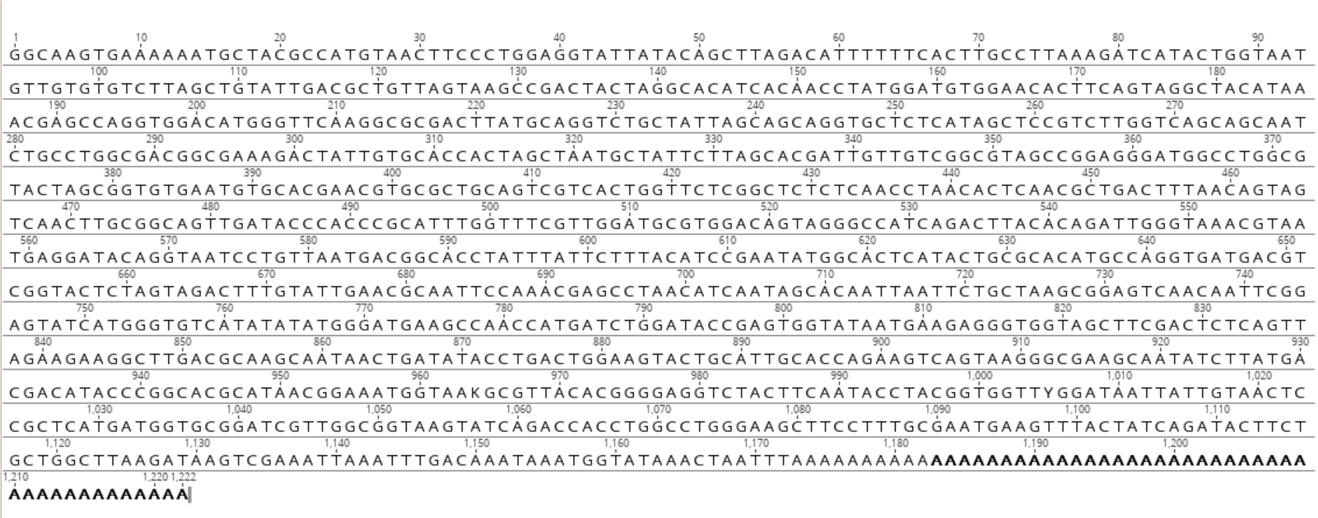
```
CCTAACATTTTAATAATTTTTCACTTGCTTTTATTAGTGAGCTCAGTGATCATAGATGGTGACACTGTCGGCCCTATTGATAAGGTATGGTTT
AGACGGTGGGATATCGACCCGACCCTACAGACCAACATTACTGAGTGTATTACTAACGTTACTGATAGGCCAGTAATCATAGACGAACAAGTG
CAATTATACACGCCGGCTATCCAAATATGGGTAGACACTGGCAGTAACTCAGAAGACGTTATTGACCCCGAACTCGATGTTGCCATAATGCAT
TGTCTTGACGGAACTGCGAACTACACAGGCATACCCGTGTATGGCCATAACGTATTTGCTGTGGTTAGTCCGCTAGCTAGTGACAACAGCACG
TATGCATCCCACTATGACTATTTCGGACATGGGTACCCTGCCGACTACGATAGGGACTACGCGTCCGACCACAGCGATACACCTGACAACACA
ACTGACATAAATAAGCGAGGTAATGGCAACTGGCCTGACGCTATATTACCAGGATGGTCGGCGAAAACTGATTTGGGACTTAAGCTCGCAATA
TACGCGTATGCTGCATACTCGTCTTACTTGCGAACATCACCTGACTTACAATGCGGGAAGTGTTATAGTATAAGGGGCAAGCCGATGCCTAAA
TCAAGTGTATTTGAAAACGATTTCGATGATGAAGTTGATATCATGTACTGGTTGCACCATCAGTGTCGGTCGAGAGGAAAGAACGGGGCAGGG
CCTAGGGTTCCAGGTAAAATAGAACGCATAAAGAAAGACCAGATAGGAGGACATAATTTCGTAGTGGGCTCGTTTTGGATACAGTGCAAGTCA
AATCAAGGGTGGCCAGACCCTGATTTAACTAAGTTCAAGGGCTGGGACGTCACCTGGGTGTAATGTATTATCTCATCCGAGCCTCTCCCTAGT
AAGCTCCATATCATAGCTGTCCTACCGACTTAACACCCTTAACTTAAATCAACTAACGAGCGTTAACTGACTAGTCGAACACGACATGTTAGC
TACGTGGACTACAAGGTTGATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

## M-P1G3-1 Pool1



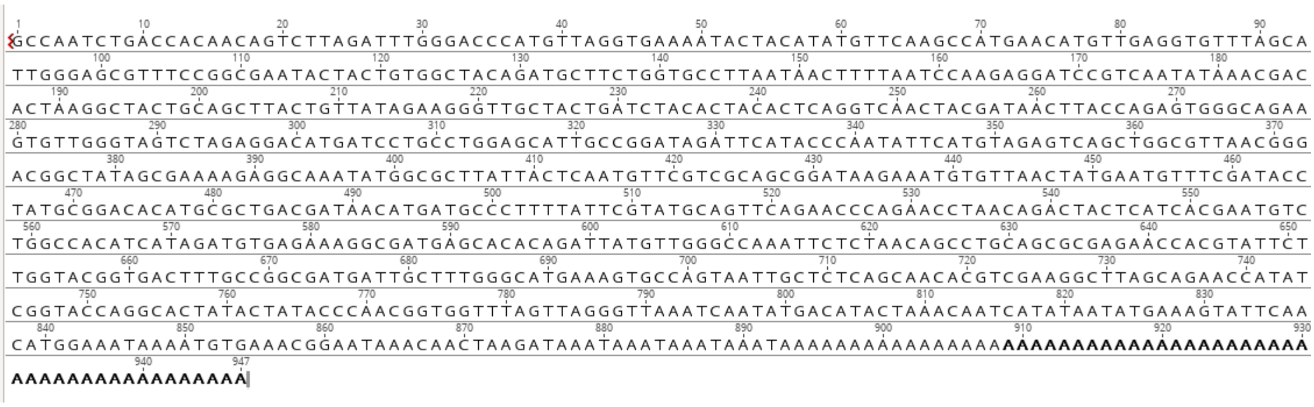
```
TAAGGGCAAGTGAAAAAATACTACGCCATGTAACTTCCCTAGAGGTGTTATACAGCTTAGACATTTTTTCACTTGCCCTTAAGATCATGCTGG
TAATGTTCTGTGTTTTAGCTGTATTAACGCTGTTAGTGAGCCGACTTTTAGGCACATCACAACCTATGGATGTGGAACACTTGAGTAGGCTRC
ATAAACGAGCTAGATGGACATGGGTTCAAGGCGCAACTTATGCAGGGCTGTTATTAGCAGCAGGTGCTCTCATAGCTCCGTCTTGGTCAGCAG
CAATATGCTTGGCGACGGCGAAGGACTAYTGCGCACCACTAGCTAATGCTATTCTTAGCACGATTGTTGTCAGCATAGCCGGAGGGATAGCCT
GGCGTACTAGCGGTGTGGGTGTACACGAACGTGCACTGCAATCATCGCTGATTCTCGGCTCCCTTAATCTAACACTCAACGCTGACTTCAACA
GCAGTCAGCTAGCGGCTGTTGATACCCACCCGCATTTGGTCTCGTTGGATGCGTGGACAGTGGGGCCAGCAGACCTACACAGATTAGGTAAAC
GTGRTGRGGATACAGGTAATTCTGTCAATAACGGCACTTATTTGTTCTTCACGTCCGAATATGGCACTCATACTGCTCACATGTCAGGTGATG
ATGTCGGTACTCTAGTAGATTTTGTRTTAAACGCGATTCCAGACGAGCCTAACATTAATAGCACAATTAATTCTGCTAAGCGGAGTCAACAAT
TCGGAGTATCATGGGTGTCATATATATGGGACGAAGCCAACCATGATCTGGATACCGAGTGGTATAATGAAGAAGGTGGTAGCTTCGACTCYC
AGTTAGAAGAAGGTTTGACGCAAGCAATAGTTGATATACCTGACTGGAAGTACTGTATTGCACCTGAAGTCAGTAAGGGCGAGGCAATATCTT
ATGACGACATACCCGGCACGCAAAACGGAAATGGCAATGCGTTACACGGAGAAGTCTACTTCAATACCTACGGTGGTTTGGATAGTTATTGCA
ATTCTGCACATSACGGTGCGGATCGTTGGCGGTAGGTATCACACCACCTGGTCTAGAGGATTCAYTTGCGCATGAAGTCTGTGATTGAATACC
TCTGCCGGCTTAAGATAAGTCAGAATTAGATTTAACAAATAAATAGTATAGGTTAACGTAGTATTTATAWAAAAAAAAAAAAAAAAAAAAAAA
```

## M-P1G3-2



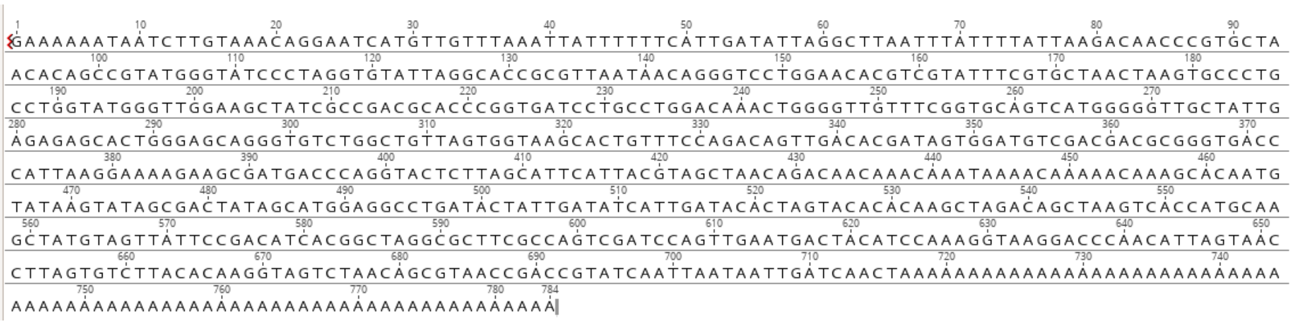
```
GGCAAGTGAAAAAATGCTACGCCATGTAACTTCCCTGGAGGTATTATACAGCTTAGACATTTTTTCACTTGCCTTAAAGATCATACTGGTAAT
GTTGTGTGTCTTAGCTGTATTGACGCTGTTAGTAAGCCGACTACTAGGCACATCACAACCTATGGATGTGGAACACTTCAGTAGGCTACATAA
ACGAGCCAGGTGGACATGGGTTCAAGGCGCGACTTATGCAGGTCTGCTATTAGCAGCAGGTGCTCTCATAGCTCCGTCTTGGTCAGCAGCAAT
CTGCCTGGCGACGGCGAAAGACTATTGTGCACCACTAGCTAATGCTATTCTTAGCACGATTGTTGTCGGCGTAGCCGGAGGGATGGCCTGGCG
TACTAGCGGTGTGAATGTGCACGAACGTGCGCTGCAGTCGTCACTGGTTCTCGGCTCTCTCAACCTAACACTCAACGCTGACTTTAACAGTAG
TCAACTTGCGGCAGTTGATACCCACCCGCATTTGGTTTCGTTGGATGCGTGGACAGTAGGGCCATCAGACTTACACAGATTGGGTAAACGTAA
TGAGGATACAGGTAATCCTGTTAATGACGGCACCTATTTATTCTTTACATCCGAATATGGCACTCATACTGCGCACATGCCAGGTGATGACGT
CGGTACTCTAGTAGACTTTGTATTGAACGCAATTCCAAACGAGCCTAACATCAATAGCACAATTAATTCTGCTAAGCGGAGTCAACAATTCGG
AGTATCATGGGTGTCATATATATGGGATGAAGCCAACCATGATCTGGATACCGAGTGGTATAATGAAGAGGGTGGTAGCTTCGACTCTCAGTT
AGAAGAAGGCTTGACGCAAGCAATAACTGATATACCTGACTGGAAGTACTGCATTGCACCAGAAGTCAGTAAGGGCGAAGCAATATCTTATGA
CGACATACCCGGCACGCATAACGGAAATGGTAAKGCGTTACACGGGGAGGTCTACTTCAATACCTACGGTGGTTYGGATAATTATTGTAACTC
CGCTCATGATGGTGCGGATCGTTGGCGGTAAGTATCAGACCACCTGGCCTGGGAAGCTTCCTTTGCGAATGAAGTTTACTATCAGATACTTCT
GCTGGCTTAAGATAAGTCGAAATTAAATTTGACAAATAAATGGTATAAACTAATTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAA
```

## M-P1G5



```
GCCAATCTGACCACAACAGTCTTAGATTTGGGACCCATGTTAGGTGAAAATACTACATATGTTCAAGCCATGAACATGTTGAGGTGTTTAGCA
TTGGGAGCGTTTCCGGCGAATACTACTGTGGCTACAGATGCTTCTGGTGCCTTAATAACTTTTAATCCAAGAGGATCCGTCAATATAAACGAC
ACTAAGGCTACTGCAGCTTACTGTTATAGAAGGGTTGCTACTGATCTACACTACACTCAGGTCAACTACGATAACTTACCAGAGTGGGCAGAA
GTGTTGGGTAGTCTAGAGGACATGATCCTGCCTGGAGCATTGCCGGATAGATTCATACCCAATATTCATGTAGAGTCAGCTGGCGTTAACGGG
ACGGCTATAGCGAAAAGAGGCAAATATGGCGCTTATTACTCAATGTTCGTCGCAGCGGATAAGAAATGTGTTAACTATGAATGTTTCGATACC
TATGCGGACACATGCGCTGACGATAACATGATGCCCTTTTATTCGTATGCAGTTCAGAACCCAGAACCTAACAGACTACTCATCACGAATGTC
TGGCCACATCATAGATGTGAGAAAGGCGATGAGCACACAGATTATGTTGGGCCAAATTCTCTAACAGCCTGCAGCGCGAGAACCACGTATTCT
TGGTACGGTGACTTTGCCGGCGATGATTGCTTTGGGCATGAAAGTGCCAGTAATTGCTCTCAGCAACACGTCGAAGGCTTAGCAGAACCATAT
CGGTACCAGGCACTATACTATACCCAACGGTGGTTTAGTTAGGGTTAAATCAATATGACATACTAAACAATCATATAATATGAAAGTATTCAA
CATGGAAATAAAATGTGAAACGGAATAAACAACTAAGATAAATAAATAAATAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAA
```

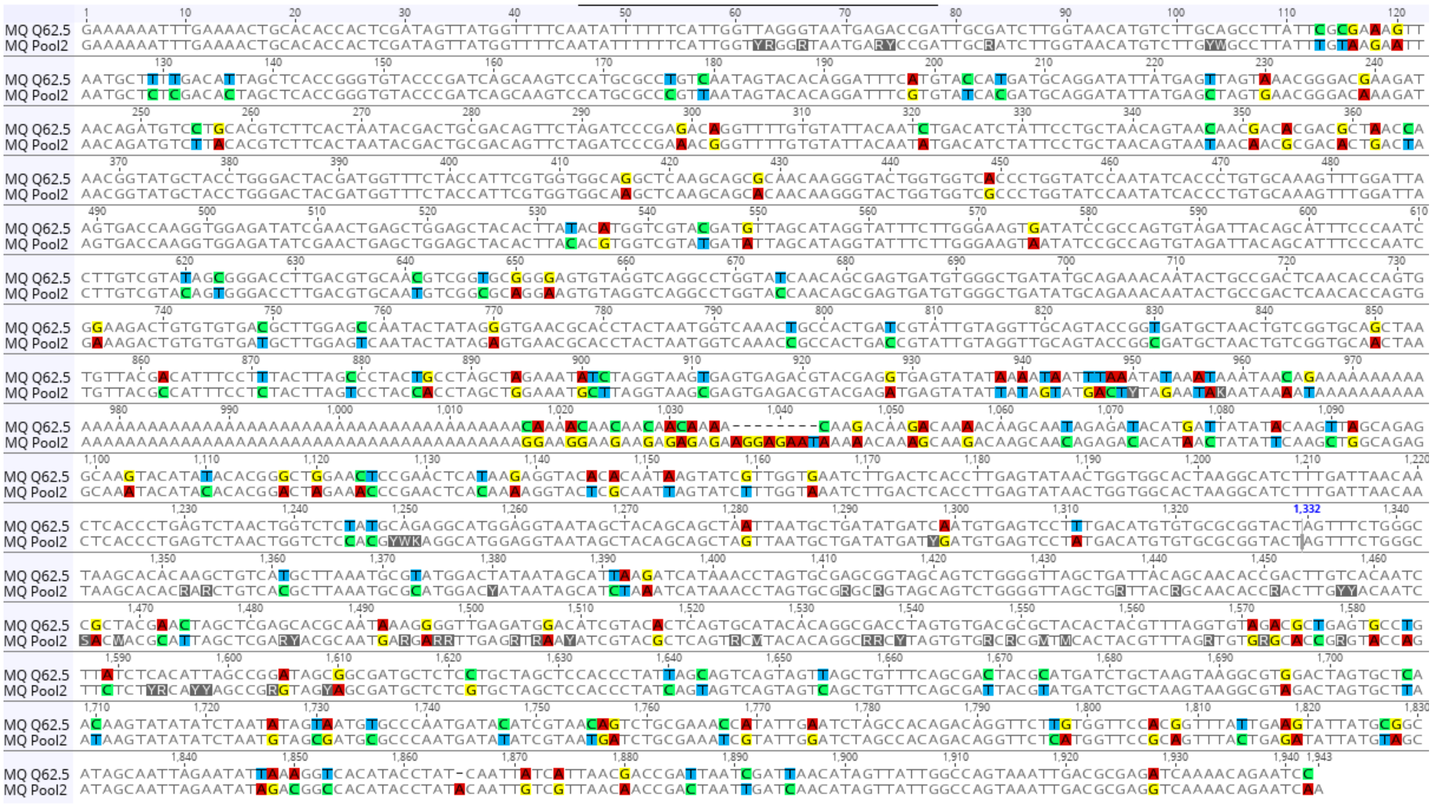
### M-P1SG



```
GAAAAAATAATCTTGTAAACAGGAATCATGTTGTTTAAATTATTTTTTCATTGATATTAGGCTTAATTTATTTTATTAAGACAACCCGTGCTA
ACACAGCCGTATGGGTATCCCTAGGTGTATTAGGCACCGCGTTAATAACAGGGTCCTGGAACACGTCGTATTTCGTGCTAACTAAGTGCCCTG
CCTGGTATGGGTTGGAAGCTATCGCCGACGCACCCGGTGATCCTGCCTGGACAAACTGGGGTTGTTTCGGTGCAGTCATGGGGGTTGCTATTG
AGAGAGCACTGGGAGCAGGGTGTCTGGCTGTTAGTGGTAAGCACTGTTTCCAGACAGTTGACACGATAGTGGATGTCGACGACGCGGGTGACC
CATTAAGGAAAAGAAGCGATGACCCAGGTACTCTTAGCATTCATTACGTAGCTAACAGACAACAAACAAATAAAACAAAAACAAAGCACAATG
TATAAGTATAGCGACTATAGCATGGAGGCCTGATACTATTGATATCATTGATACACTAGTACACACAAGCTAGACAGCTAAGTCACCATGCAA
GCTATGTAGTTATTCCGACATCACGGCTAGGCGCTTCGCCAGTCGATCCAGTTGAATGACTACATCCAAAGGTAAGGACCCAACATTAGTAAC
CTTAGTGTCTTACACAAGGTAGTCTAACAGCGTAACCGACCGTATCAATTAATAATTGATCAACTAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

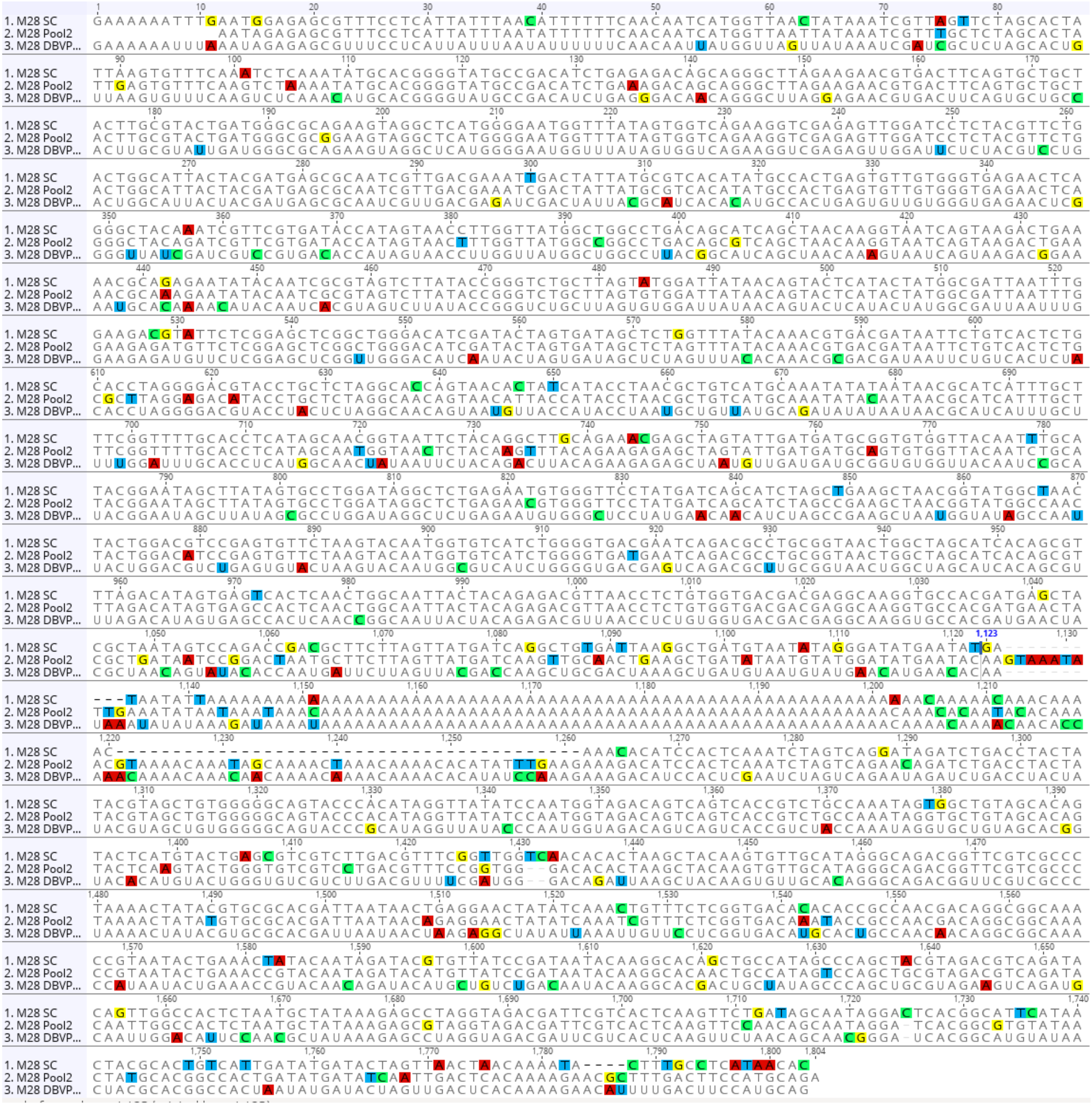
## 8.2.7 Alignments

### MQ alignment



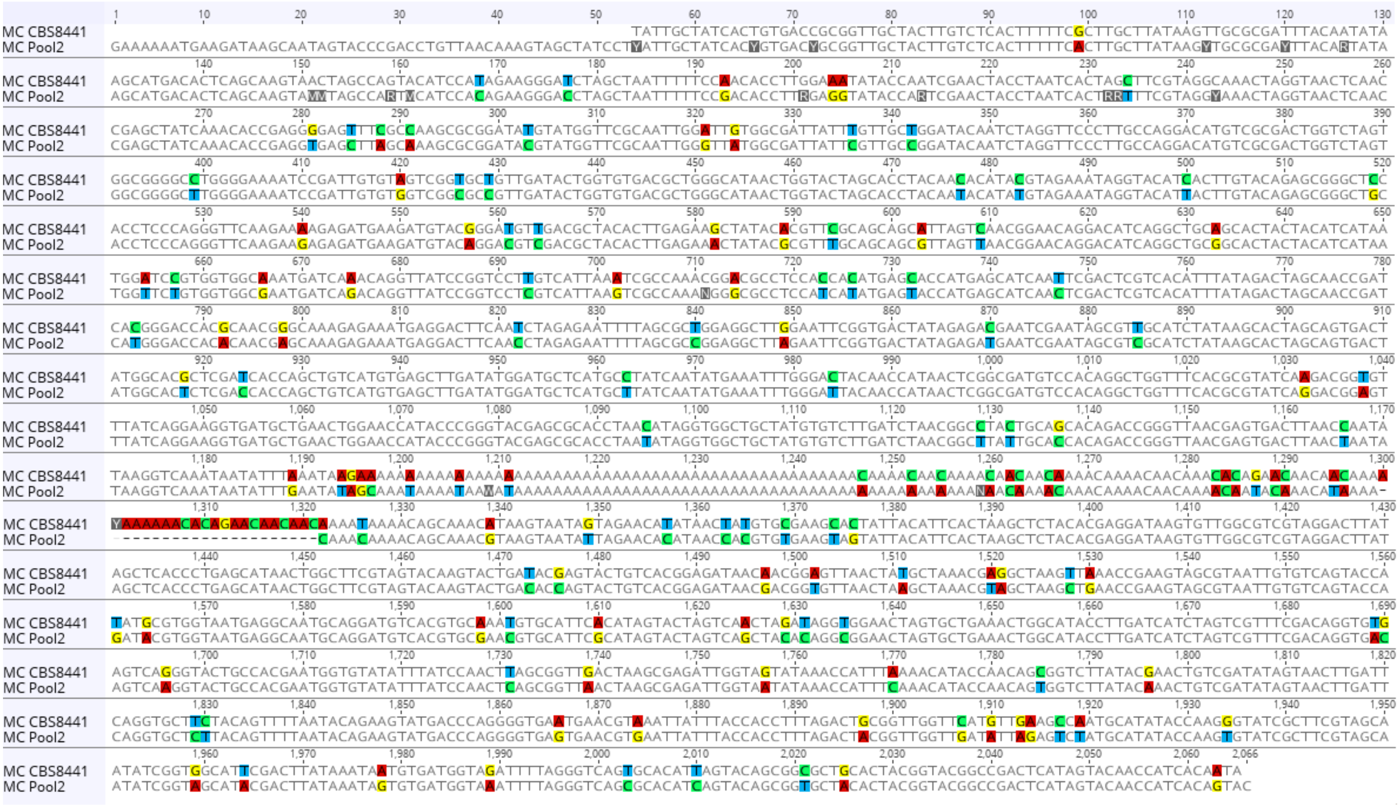
```
MQ Q62.5 GAAAAAATTTGAAAACTGCACACCACTCGATAGTTATGGTTTTCAATATTTTTTCATTGGTTAGGGTAATGAGACCGATTGCGATCTTGGTAACATGTCTTGCAGCCTTATTCGCGAAAGTT
MQ Pool2 GAAAAAATTTGAAAACTGCACACCACTCGATAGTTATGGTTTTCAATATTTTTTCATTGGTYRGGRTAATGARYCCGATTGCRATCTTGGTAACATGTCTTGYWGCCTTATTTGTAAGAATT

MQ Q62.5 AATGCTTTTGACATTAGCTCACCGGGTGTACCCGATCAGCAAGTCCATGCGCCTGTCAATAGTACACAGGATTTCATGTACCATGATGCAGGATATTATGAGTTAGTAAACGGGACGAAGAT
MQ Pool2 AATGCTCTCGACACTAGCTCACCGGGTGTACCCGATCAGCAAGTCCATGCGCCCGTTAATAGTACACAGGATTTCGTGTATCACGATGCAGGATATTATGAGCTAGTGAACGGGACAAAGAT

MQ Q62.5 AACAGATGTCCTGCACGTCTTCACTAATACGACTGCGACAGTTCTAGATCCCGAGACAGGTTTTGTGTATTACAATCTGACATCTATTCCTGCTAACAGTAACAACGACACGACGCTAACCA
MQ Pool2 AACAGATGTCTTACACGTCTTCACTAATACGACTGCGACAGTTCTAGATCCCGAAACGGGTTTTGTGTATTACAATATGACATCTATTCCTGCTAACAGTAATAACAACGCGACACTGACTA

MQ Q62.5 AACGGTATGCTACCTGGGACTACGATGGTTTCTACCATTCGTGGTGGCAGGCTCAAGCAGCGCAACAAGGGTACTGGTGGTCACCCTGGTATCCAATATCACCCTGTGCAAAGTTTGGATTA
MQ Pool2 AACGGTATGCTACCTGGGACTACGATGGTTTCTACCATTCGTGGTGGCAAGCTCAAGCAGCACAACAAGGGTACTGGTGGTCGCCCTGGTATCCAATATCACCCTGTGCAAAGTTTGGATTA

MQ Q62.5 AGTGACCAAGGTGGAGATATCGAACTGAGCTGGAGCTACACTTATACATGGTCGTACGATGTTAGCATAGGTATTTCTTGGGAAGTGATATCCGCCAGTGTAGATTACAGCATTTCCCAATC
MQ Pool2 AGTGACCAAGGTGGAGATATCGAACTGAGCTGGAGCTACACTTACACGTGGTCGTATGATATTAGCATAGGTATTTCTTGGGAAGTAATATCCGCCAGTGTAGATTACAGCATTTCCCAATC

MQ Q62.5 CTTGTCGTATAGCGGGACCTTGACGTGCAACGTCGGTGCGGGGAGTGTAGGTCAGGCCTGGTATCAACAGCGAGTGATGTGGGCTGATATGCAGAAACAATACTGCCGACTCAACACCAGTG
MQ Pool2 CTTGTCGTACAGTGGGACCTTGACGTGCAATGTCGGCGCAGGAAGTGTAGGTCAGGCCTGGTACCAACAGCGAGTGATGTGGGCTGATATGCAGAAACAATACTGCCGACTCAACACCAGTG

MQ Q62.5 GGAAGACTGTGTGTGACGCTTGGAGCCAATACTATAGGGTGAACGCACCTACTAATGGTCAAACTGCCACTGATCGTATTGTAGGTTGCAGTACCGGTGATGCTAACTGTCGGTGCAGCTAA
MQ Pool2 GAAAGACTGTGTGTGATGCTTGGAGTCAATACTATAGAGTGAACGCACCTACTAATGGTCAAACCGCCACTGACCGTATTGTAGGTTGCAGTACCGGCGATGCTAACTGTCGGTGCAACTAA

MQ Q62.5 TGTTACGACATTTCCTTTACTTAGCCCTACTGCCTAGCTAGAAATATCTAGGTAAGTGAGTGAGACGTACGAGGTGAGTATATAAAATAATTTAAATATAAATAAATAACAGAAAAAAAAAA
MQ Pool2 TGTTACGCCATTTCCTCTACTTAGTCCTACCACCTAGCTGGAAATGCTTAGGTAAGCGAGTGAGACGTACGAGATGAGTATATTATAGTATGACTYTAGAATAKAATAAAATAAAAAAAAAA

MQ Q62.5 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAACAACAACAACAAAA--------CAAGACAAGACAAAACAAGCAATAGAGATACATGATTATATACAAGTTAGCAGAG
MQ Pool2 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAGGAAGAAGAGAGAGAAGGAGAATAAAAAACAAAGCAAGACAAGCAACAGAGACACATAACTATATTCAAGCTGGCAGAG

MQ Q62.5 GCAAGTACATATACACGGGCTGGAACTCCGAACTCATAAGAGGTACACACAATAAGTATCGTTGGTGAATCTTGACTCACCTTGAGTATAACTGGTGGCACTAAGGCATCTTTGATTAACAA
MQ Pool2 GCAAATACATACACACGGACTAGAAACCCGAACTCACAAAAGGTACTCGCAATTAGTATCTTTGGTAAATCTTGACTCACCTTGAGTATAACTGGTGGCACTAAGGCATCTTTGATTAACAA

MQ Q62.5 CTCACCCTGAGTCTAACTGGTCTCTATGCAGAGGCATGGAGGTAATAGCTACAGCAGCTAATTAATGCTGATATGATCAATGTGAGTCCTTTGACATGTGTGCGCGGTACTAGTTTCTGGGC
MQ Pool2 CTCACCCTGAGTCTAACTGGTCTCCACGYWKAGGCATGGAGGTAATAGCTACAGCAGCTAGTTAATGCTGATATGATYGATGTGAGTCCTATGACATGTGTGCGCGGTACTAGTTTCTGGGC

MQ Q62.5 TAAGCACACAAGCTGTCATGCTTAAATGCGTATGGACTATAATAGCATTAAGATCATAAACCTAGTGCGAGCGGTAGCAGTCTGGGGTTAGCTGATTACAGCAACACCGACTTGTCACAATC
MQ Pool2 TAAGCACACRARCTGTCACGCTTAAATGCGCATGGACYATAATAGCATCTAAATCATAAACCTAGTGCGRGCRGTAGCAGTCTGGGGTTAGCTGRTTACRGCAACACCRACTTGYYACAATC

MQ Q62.5 CGCTACGAACTAGCTCGAGCACGCAATAAAGGGGTTGAGATGGACATCGTACACTCAGTGCATACACAGGCGACCTAGTGTGACGCGCTACACTACGTTTAGGTGTAGACGCTGAGTGCCTG
MQ Pool2 SACMACGCATTAGCTCGARYACGCAATGARGARRTTGAGRTRAAYATCGTACGCTCAGTRCMTACACAGGCRRCYTAGTGTGRCRCGMTMCACTACGTTTAGRTGTGRGCACCGRGTACCAG

MQ Q62.5 TTATCTCACATTAGCCGGATAGCGGCGATGCTCTCCTGCTAGCTCCACCCTATTAGCAGTCAGTAGTTAGCTGTTTCAGCGACTACGCATGATCTGCTAAGTAAGGCGTGGACTAGTGCTCA
MQ Pool2 TTCTCTYRCAYYAGCCGRGTAGYAGCGATGCTCTCGTGCTAGCTCCACCCTATCAGTAGTCAGTAGTCAGCTGTTTCAGCGATTACGTATGATCTGCTAAGTAAGGCGTAGACTAGTGCTTA

MQ Q62.5 ACAAGTATATATCTAATATAGTAATGTGCCCAATGATACATCGTAACAGTCTGCGAAACCATATTGAATCTAGCCACAGACAGGTTCTTGTGGTTCCACGGTTTATTGAAGTATTATGCGGC
MQ Pool2 ATAAGTATATATCTAATGTAGCGATGCGCCCAATGATATATCGTAATGATCTGCGAAATCGTATTGGATCTAGCCACAGACAGGTTCTCATGGTTCCGCAGTTTACTGAGATATTATGTAGC

MQ Q62.5 ATAGCAATTAGAATATTAAAGGTCACATACCTAT-CAATTATCATTAACGACCGATTAATCGATTAACATAGTTATTGGCCAGTAAATTGACGCGAGATCAAAACAGAATCC
MQ Pool2 ATAGCAATTAGAATATAGACGGCCACATACCTATACAATTGTCGTTAACAACCGACTAATTGATCAACATAGTTATTGGCCAGTAAATTGACGCGAGGTCAAAACAGAATCAA
```

## M28 alignment



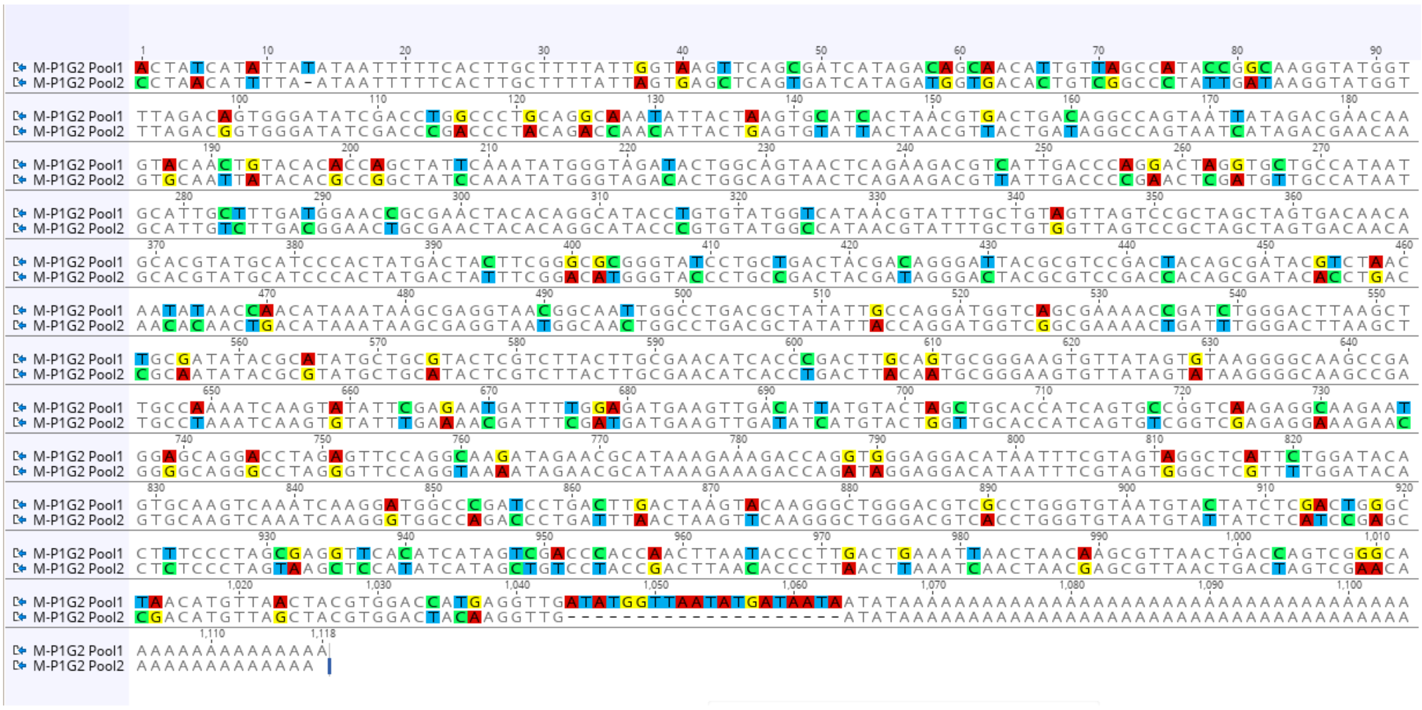
```
1. M28 SC      GAAAAAATTTGAATGGAGAGCGTTTCCTCATTATTTAACATTTTTTTCAACAATCATGGTTAACTATAAATCGTTAGTTCTAGCACTA
2. M28 Pool2              AATAGAGAGCGTTTCCTCATTATTTAATATTTTTTTCAACAATCATGGTTAATTATAAATCGTTTGCTCTAGCACTA
3. M28 DBVP...  GAAAAAAUUUAAAUAGAGAGCGUUUCCUCAUUAUUUAAUAUUUUUUCAACAAUUAUGGUUAGUUAUAAAUCGAUCGCUCUAGCACUG

1. M28 SC      TTAAGTGTTTCAAATCTCAAATATGCACGGGGTATGCCGACATCTGAGAGACAGCAGGGCTTAGAAGAACGTGACTTCAGTGCTGCT
2. M28 Pool2   TTGAGTGTTTCAAGTCTAAAATATGCACGGGGTATGCCGACATCTGAAAGACAGCAGGGCTTAGAAGAACGTGACTTCAGTGCTGCT
3. M28 DBVP... UUAAGUGUUUCAAGUCUCAAACAUGCACGGGGUAUGCCGACAUCUGAGGGACAACAGGGCUUAGGAGAACGUGACUUCAGUGCUGCC

1. M28 SC      ACTTGCGTACTGATGGGCGCAGAAGTAGGCTCATGGGGAATGGTTTATAGTGGTCAGAAGGTCGAGAGTTGGATCCTCTACGTTCTG
2. M28 Pool2   ACTTGCGTACTGATGGGCGCGGAAGTAGGCTCATGGGGAATGGTTTATAGTGGTCAGAAGGTCGAGAGTTGGATCCTCTACGTTCTG
3. M28 DBVP... ACUUGCGUAUUGAUGGGCGCAGAAGUAGGCUCAUGGGGAAUGGUUUAUAGUGGUCAGAAGGUCGAGAGUUGGAUUCUCUACGUCCUG

1. M28 SC      ACTGGCATTACTACGATGAGCGCAATCGTTGACGAAATTGACTATTATGCGTCACATATGCCACTGAGTGTTGTGGGTGAGAACTCA
2. M28 Pool2   ACTGGCATTACTACGATGAGCGCAATCGTTGACGAAATCGACTATTATGCGTCACATATGCCACTGAGTGTTGTGGGTGAGAACTCA
3. M28 DBVP... ACUGGCAUUACUACGAUGAGCGCAAUCGUUGACGAGAUCGACUAUUACGCAUCACACAUGCCACUGAGUGUUGUGGGUGAGAACUCG

1. M28 SC      GGGCTACAAATCGTTCGTGATACCATAGTAACCTTGGTTATGGCTGGCCTGACAGCATCAGCTAACAAGGTAATCAGTAAGACTGAA
2. M28 Pool2   GGGCTACAGATCGTTCGTGATACCATAGTAACTTTGGTTATGGCCGGCCTGACAGCGTCAGCTAACAAGGTAATCAGTAAGACTGAA
3. M28 DBVP... GGGUUAUCGAUCGUCCGUGACACCAUAGUAACCUUGGUUAUGGCUGGCCUUACGGCAUCAGCUAACAAAGUAAUCAGUAAGACGGAA

1. M28 SC      AACGCAGAGAATATACAATCGCGTAGTCTTATACCGGGTCTGCTTAGTATGGATTATAACAGTACTCATACTATGGCGATTAATTTG
2. M28 Pool2   AACGCAAAGAATATACAATCGCGTAGTCTTATACCGGGTCTGCTTAGTGTGGATTATAACAGTACTCATACTATGGCGATTAATTTG
3. M28 DBVP... AAUGCACAAAACAUACAAUCACGUAGUCUUAUACCGGGUCUGCUUAGUGUGGAUUAUAACAGUACUCAUACUAUGGCGAUUAAUUUG

1. M28 SC      GAAGACGTATTCTCGGAGCTCGGCTGGGACATCGATACTAGTGATAGCTCTGGTTTATACAAACGTGACGATAATTCTGTCACTCTG
2. M28 Pool2   GAAGAGATGTTCTCGGAGCTCGGCTGGGACATCGATACTAGTGATAGCTCTAGTTTATACAAACGTGACGATAATTCTGTCACTCTG
3. M28 DBVP... GAAGAGAUGUUCUCGGAGCUCGGUUGGGACAUCAAUACUAGUGAUAGCUCUAGUUUACACAAACGCGACGAUAAUUCUGUCACUCUA

1. M28 SC      CACCTAGGGGACGTACCTGCTCTAGGCACCAGTAACACTATCATACCTAACGCTGTCATGCAAATATATAATAACGCATCATTTGCT
2. M28 Pool2   CGCTTAGGAGACATACCTGCTCTAGGCAACAGTAACATTACCATACCTAACGCTGTCATGCAAATATACAATAACGCATCATTTGCT
3. M28 DBVP... CACCUAGGGGACGUACCUACUCUAGGCAACAGUAAUGUUACCAUACCUAAUGCUGUUAUGCAGAUAUAUAAUAACGCAUCAUUUGCU

1. M28 SC      TTCGGTTTTGCACCTCATAGCAACGGTAATTCTACAGGCTTGCAGAAACGAGCTAGTATTGATGATGCGGTGTGGTTACAATTTGCA
2. M28 Pool2   TTCGGTTTTGCACCTCATAGCAATGGTAACTCTACAAGTTTACAGAAGAGAGCTAGTATTGATGATGCAGTGTGGTTACAATCTGCA
3. M28 DBVP... UUUGGAUUUGCACCUCAUGGCAACUAUAAUUCUACAGACUUACAGAAGAGAGCUAAUGUUGAUGAUGCGGUGUGGUUACAAUCCGCA

1. M28 SC      TACGGAATAGCTTATAGTGCCTGGATAGGCTCTGAGAATGTGGGTTCCTATGATCAGCATCTAGCTGAAGCTAACGGTATGGCTAAC
2. M28 Pool2   TACGGAATAGCTTATAGTGCCTGGATAGGCTCTGAGAACGTGGGTTCCTATGATCAGCATCTAGCCGAAGCTAACGGTATGGCCAAC
3. M28 DBVP... UACGGAAUAGCUUAUAGCGCCUGGAUAGGCUCUGAGAAUGUGGGCUCCUAUGAACAACAUCUAGCCGAAGCUAAUGGUAUAGCCAAU

1. M28 SC      TACTGGACGTCCGAGTGTTCTAAGTACAATGGTGTCATCTGGGGTGACGAATCAGACGCCTGCGGTAACTGGCTAGCATCACAGCGT
2. M28 Pool2   TACTGGACATCCGAGTGTTCTAAGTACAATGGTGTCATCTGGGGTGATGAATCAGACGCCTGCGGTAACTGGCTAGCATCACAGCGT
3. M28 DBVP... UACUGGACGUCUGAGUGUACUAAGUACAAUGGCGUCAUCUGGGGUGACGAGUCAGACGCUUGCGGUAACUGGCUAGCAUCACAGCGU

1. M28 SC      TTAGACATAGTGAGTCACTCAACTGGCAATTACTACAGAGACGTTAACCTCTGTGGTGACGACGAGGCAAGGTGCCACGATGAGCTA
2. M28 Pool2   TTAGACATAGTGAGCCACTCAACTGGCAATTACTACAGAGACGTTAACCTCTGTGGTGACGACGAGGCAAGGTGCCACGATGAACTA
3. M28 DBVP... UUAGACAUAGUGAGCCACUCAACCGGCAAUUACUACAGAGACGUUAACCUCUGUGGUGACGACGAGGCAAGGUGCCACGAUGAACUA

                                                                                         1,123
1. M28 SC      CGCTAATAGTCCAGACCGACGCTTCTTAGTTATGATCAGGCTGTGATTAAGGCTGATGTAATATAGGGATATGAATATGA-------
2. M28 Pool2   CGCTGATAATCCGGACTAATGCTTCTTAGTTATGATCAAGTTGCAACTGAAGCTGATATAATGTATGGATATGAATACAAGTAAATA
3. M28 DBVP... CGCUAACAGUAUACACCAAUGAUUCUUAGUUACGACCAAGCUGCGACUAAAGCUGAUGUAAUGUAUGAACAUGAACACAA-------

1. M28 SC      ---TAATATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAACAAACAAA
2. M28 Pool2   TTGAAATATAATAAATAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAACACAATACACAAA
3. M28 DBVP... UAAAUAUAUAAAGAUAAAAUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAACAAAACAACACC

1. M28 SC      AC--------------------------------------AAACACATCCACTCAAATCTAGTCAGGATAGATCTGACCTACTA
2. M28 Pool2   ACGTAAAACAAATAGCAAAACTAAACAAAACACATATTTGAAGAAAGACATCCACTCAAATCTAGTCAGAACAGATCTGACCTACTA
3. M28 DBVP... AAACAAAACAAACAACAAAACAAAACAAAACACAUAUCCAAAGAAAGACAUCCACUCGAAUCUAGUCAGAAUAGAUCUGACCUACUA

1. M28 SC      TACGTAGCTGTGGGGGCAGTACCCACATAGGTTATATCCAATGGTAGACAGTCAGTCACCGTCTGCCAAATAGTGGCTGTAGCACAG
2. M28 Pool2   TACGTAGCTGTGGGGGCAGTACCCACATAGGTTATATCCAATGGTAGACAGTCAGTCACCGTCTGCCAAATAGGTGCTGTAGCACAG
3. M28 DBVP... UACGUAGCUGUGGGGGCAGUACCCGCAUAGGUUAUACCCAAUGGUAGACAGUCAGUCACCGUCUACCAAAUAGGUGCUGUAGCACGG

1. M28 SC      TACTCATGTACTGAGCGTCGTCTTGACGTTTCGGTTGGTCAACACACTAAGCTACAAGTGTTGCATAGGGCAGACGGTTCGTCGCCC
2. M28 Pool2   TACTCAAGTACTGGGTGTCGTCCTGACGTTTCCGGTGG--GACACACTAAGCTACAAGTGTTGCATAGGGCAGACGGTTCGTCGCCC
3. M28 DBVP... UACACAUGUACUGGGUGUCGUCUUGACGUUUUCGAUGG--GACAGAUUAAGCUACAAGUGUUGCACAGGGCAGACGGUUCGUCGCCC

1. M28 SC      TAAAACTATACGTGCGCACGATTAATAACTGAGGAACTATATCAAACTGTTTCTCGGTGACACACACCGCCAACGACAGGCGGCAAA
2. M28 Pool2   TAAAACTATATGTGCGCACGATTAATAACAGAGGAACTATATCAAATCGTTTCTCGGTGACAAATACCGCCAACGACAGGCGGCAAA
3. M28 DBVP... UAAAACUAUACGUGCGCACGAUUAAUAACUAAGAGGCUAUAUUAAAUUGUUCCUCGGUGACAUGCACUGCCAACAACAGGCGGCAAA

1. M28 SC      CCGTAATACTGAAACTATACAATAGATACGTGTTATCCGATAATACAAGGCACAGCTGCCATAGCCCAGCTACGTAGACGTCAGATA
2. M28 Pool2   CCGTAATACTGAAACCGTACAATAGATACATGTTATCCGATAATACAAGGCACAACTGCCATAGTCCAGCTGCGTAGACGTCAGATA
3. M28 DBVP... CCAUAAUACUGAAACCGUACAACAGAUACAUGCUGUCUGACAAUACAAGGCACGACUGCUAUAGCCCAGCUGCGUAGAAGUCAGAUG

1. M28 SC      CAGTTGGCCACTCTAATGCTATAAAGAGCCTAGGTAGACGATTCGTCACTCAAGTTCTGATAGCAATAGGACTCACGGCATTCATAA
2. M28 Pool2   CAATTGGCCACTCTAATGCTATAAAGAGCGTAGGTAGACGATTCGTCACTCAAGTTCCAACAGCAATAGGA-TCACGGCGTGTATAA
3. M28 DBVP... CAAUUGGACAUUCCAACGCUAUAAAGAGCCUAGGUAGACGAUUCGUCACUCAAGUUCUAACAGCAACGGGA-UCACGGCAUGUAUAA

                                                                                  1,804
1. M28 SC      CTACGCACTGTCATTGATATGATACTAGTTAACTAACAAAATA----CTTTGCCTCATAACAC
2. M28 Pool2   CTATGCACGGCCACTGATATGATATCAATTGACTCACAAAAGAACGCTTTGACTTCCATGCAGA
3. M28 DBVP... CUACGCACGGCCACUAAUAUGAUACUAGUUGACUCACAAAAGAACAUUUUGACUUCCAUGCAG
```

## MC alignment



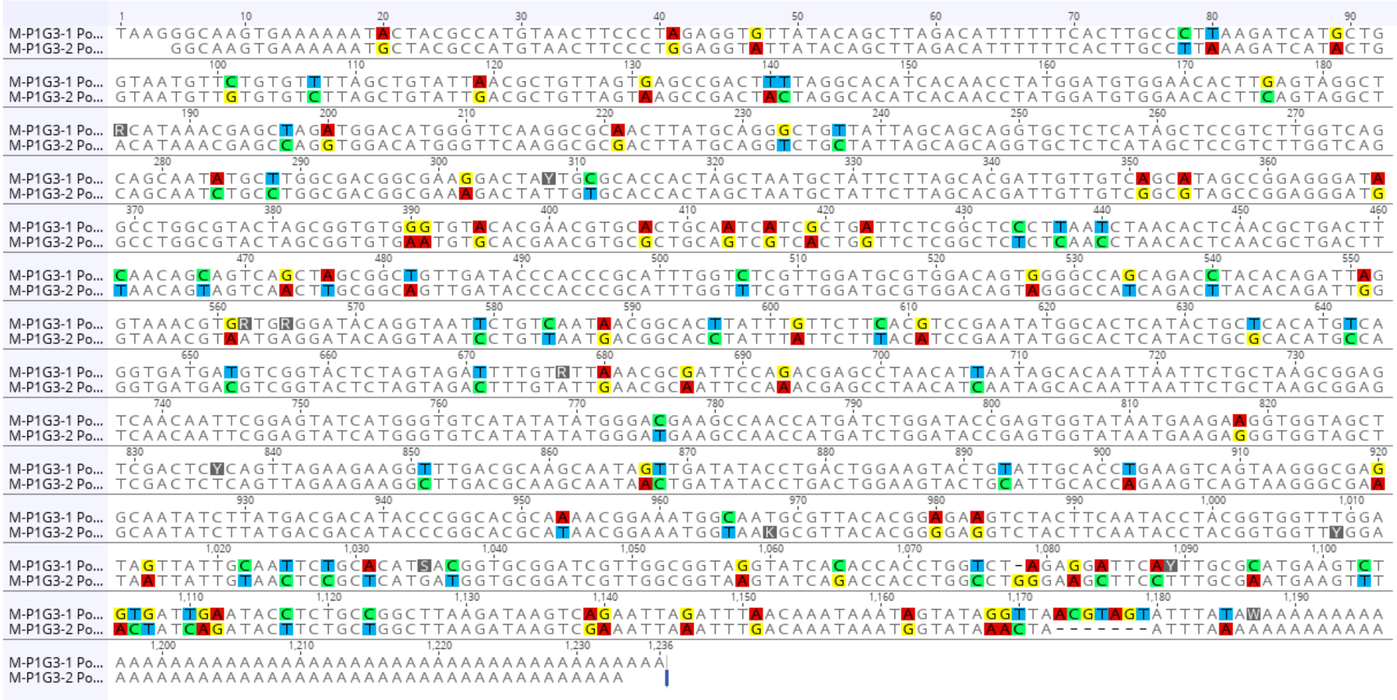
```
MC CBS8441                                                      TATTGCTATCACTGTGACCGCGGTTGCTACTTGTCTCACTTTTTCGCTTGCTTATAAGTTGCGCGATTTACAATATA
MC Pool2    GAAAAAATGAAGATAAGCAATAGTACCCGACCTGTTAACAAAGTAGCTATCCTYATTGCTATCACYGTGACYGCGGTTGCTACTTGTCTCACTTTTTCACTTGCTTATAAGYTGCGCGAYTTACARTATA

MC CBS8441  AGCATGACACTCAGCAAGTAACTAGCCAGTACATCCATAGAAGGGATCTAGCTAATTTTTCCAACACCTTGGAAATATACCAATCGAACTACCTAATCACTAGCTTCGTAGGCAAACTAGGTAACTCAAC
MC Pool2    AGCATGACACTCAGCAAGTAMMTAGCCARTMCATCCACAGAAGGGACCTAGCTAATTTTTCCGACACCTTRGAGGTATACCARTCGAACTACCTAATCACTRRTTTCGTAGGYAAACTAGGTAACTCAAC

MC CBS8441  CGAGCTATCAAACACCGAGGGGAGTTTCGCCAAGCGCGGATATGTATGGTTCGCAATTGGATTGTGGCGATTATTTGTTGCTGGATACAATCTAGGTTCCCTTGCCAGGACATGTCGCGACTGGTCTAGT
MC Pool2    CGAGCTATCAAACACCGAGGTGAGCTTAGCAAAGCGCGGATACGTATGGTTCGCAATTGGGTTATGGCGATTATTCGTTGCCGGATACAATCTAGGTTCCCTTGCCAGGACATGTCGCGACTGGTCTAGT

MC CBS8441  GGCGGGGCCTGGGGAAAATCCGATTGTGTAGTCGGTGCTGTTGATACTGGTGTGACGCTGGGCATAACTGGTACTAGCACCTACAACACATACGTAGAAATAGGTACATCACTTGTACAGAGCGGGCTCC
MC Pool2    GGCGGGGCTTGGGGAAAATCCGATTGTGTGGTCGGCGCCGTTGATACTGGTGTGACGCTGGGCATAACTGGTACTAGCACCTACAATACATATGTAGAAATAGGTACATTACTTGTACAGAGCGGGCTGC

MC CBS8441  ACCTCCCAGGGTTCAAGAAAAGAGATGAAGATGTACGGGATGTTGACGCTACACTTGAGAAGCTATACACGTTCGCAGCAGCATTAGTCAACGGAACAGGACATCAGGCTGCAGCACTACTACATCATAA
MC Pool2    ACCTCCCAGGGTTCAAGAAGAGAGATGAAGATGTACAGGACGTCGACGCTACACTTGAGAAACTATACGCGTTTGCAGCAGCGTTAGTTAACGGAACAGGACATCAGGCTGCGGCACTACTACATCATAA

MC CBS8441  TGGATCCGTGGTGGCAAATGATCAAACAGGTTATCCGGTCCTTGTCATTAAATCGCCAAACGGACGCCTCCACCACATGAGCACCATGAGCATCAATTCGACTCGTCACATTTATAGACTAGCAACCGAT
MC Pool2    TGGTTCTGTGGTGGCGAATGATCAGACAGGTTATCCGGTCCTCGTCATTAAGTCGCCAAANGGGCGCCTCCATCATATGAGTACCATGAGCATCAACTCGACTCGTCACATTTATAGACTAGCAACCGAT

MC CBS8441  CACGGGACCACGCAACGGGCAAAGAGAAATGAGGACTTCAATCTAGAGAATTTTAGCGCTGGAGGCTTGGAATTCGGTGACTATAGAGACGAATCGAATAGCGTTGCATCTATAAGCACTAGCAGTGACT
MC Pool2    CATGGGACCACACAACGAGCAAAGAGAAATGAGGACTTCAACCTAGAGAATTTTAGCGCCGGAGGCTTAGAATTCGGTGACTATAGAGATGAATCGAATAGCGTCGCATCTATAAGCACTAGCAGTGACT

MC CBS8441  ATGGCACGCTCGATCACCAGCTGTCATGTGAGCTTGATATGGATGCTCATGCCTATCAATATGAAATTTGGGACTACAACCATAACTCGGCGATGTCCACAGGCTGGTTTCACGCGTATCAAGACGGTGT
MC Pool2    ATGGCACTCTCGACCACCAGCTGTCATGTGAGCTTGATATGGATGCTCATGCTTATCAATATGAAATTTGGGATTACAACCATAACTCGGCGATGTCCACAGGCTGGTTTCACGCGTATCAGGACGGAGT

MC CBS8441  TTATCAGGAAGGTGATGCTGAACTGGAACCATACCCGGGTACGAGCGCACCTAACATAGGTGGCTGCTATGTGTCTTGATCTAACGGCCTACTGCAGCACAGACCGGGTTAACGAGTGACTTAACCAATA
MC Pool2    TTATCAGGAAGGTGATGCTGAACTGGAACCATACCCGGGTACGAGCGCACCTAATATAGGTGGCTGCTATGTGTCTTGATCTAACGGCTTATTGCACCACAGACCGGGTTAACGAGTGACTTAACTAATA

MC CBS8441  TAAGGTCAAATAATATTTAAATAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAACAACAAAACAACAAAACAAAACAACAAAACACAGAACAACAACAAAA
MC Pool2    TAAGGTCAAATAATATTTGAATATAGCAAATAAAATAAMATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAACAAAACAAACAAAACAACAAAACAATACAAACATAAAA-

MC CBS8441  YAAAAAACACAGAACAACAACAAAATAAAACAGCAAACATAAGTAATAGTAGAACATATAACTATGTGCGAAGCACTATTACATTCACTAAGCTCTACACGAGGATAAGTGTTGGCGTCGTAGGACTTAT
MC Pool2    ---------------------CAAACAAAACAGCAAACGTAAGTAATATTAGAACACATAACCACGTGTGAAGTAGTATTACATTCACTAAGCTCTACACGAGGATAAGTGTTGGCGTCGTAGGACTTAT

MC CBS8441  AGCTCACCCTGAGCATAACTGGCTTCGTAGTACAAGTACTGATACGAGTACTGTCACGGAGATAACAACGGAGTTAACTATGCTAAACGAGGCTAAGTTAAACCGAAGTAGCGTAATTGTGTCAGTACCA
MC Pool2    AGCTCACCCTGAGCATAACTGGCTTCGTAGTACAAGTACTGACACCAGTACTGTCACGGAGATAACGACGGTGTTAACTAAGCTAAACGTAGCTAAGCTGAACCGAAGTAGCGTAATTGTGTCAGTACCA

MC CBS8441  TATGCGTGGTAATGAGGCAATGCAGGATGTCACGTGCAAATGTGCATTCACATAGTACTAGTCAACTAGATAGGTGGAACTAGTGCTGAAACTGGCATACCTTGATCATCTAGTCGTTTCGACAGGTGTG
MC Pool2    GATACGTGGTAATGAGGCAATGCAGGATGTCACGTGCGAACGTGCATTCGCATAGTACTAGTCAGCTACACAGGCGGAACTAGTGCTGAAACTGGCATACCTTGATCATCTAGTCGTTTCGACAGGTGAC

MC CBS8441  AGTCAGGGTACTGCCACGAATGGTGTATATTTATCCAACTTAGCGGTTGACTAAGCGAGATTGGTAGTATAAACCATTTAAAACATACCAACAGCGGTCTTATACGAACTGTCGATATAGTAACTTGATT
MC Pool2    AGTCAAGGTACTGCCACGAATGGTGTATATTTATCCAACTCAGCGGTTAACTAAGCGAGATTGGTAATATAAACCATTTCAAACATACCAACAGTGGTCTTATACAAACTGTCGATATAGTAACTTGATT

MC CBS8441  CAGGTGCTTCTACAGTTTTAATACAGAAGTATGACCCAGGGGTGAATGAACGTAAATTATTTACCACCTTTAGACTGCGGTTGGTTCATGTTGAAGCCAATGCATATACCAAGGGTATCGCTTCGTAGCA
MC Pool2    CAGGTGCTCTTACAGTTTTAATACAGAAGTATGACCCAGGGGTGAGTGAACGTGAATTATTTACCACCTTTAGACTACGGTTGGTTGATATTAGAGTCTATGCATATACCAAGTGTATCGCTTCGTAGCA

MC CBS8441  ATATCGGTGGCATTCGACTTATAAATAATGTGATGGTAGATTTTAGGGTCAGTGCACATTAGTACAGCGGCGCTGCACTACGGTACGGCCGACTCATAGTACAACCATCACAATA
MC Pool2    ATATCGGTAGCATACGACTTATAAATAGTGTGATGGTAAATTTTAGGGTCAGCGCACATCAGTACAGCGGTGCTACACTACGGTACGGCCGACTCATAGTACAACCATCACAGTAC
```

## M-P1G2 alignment



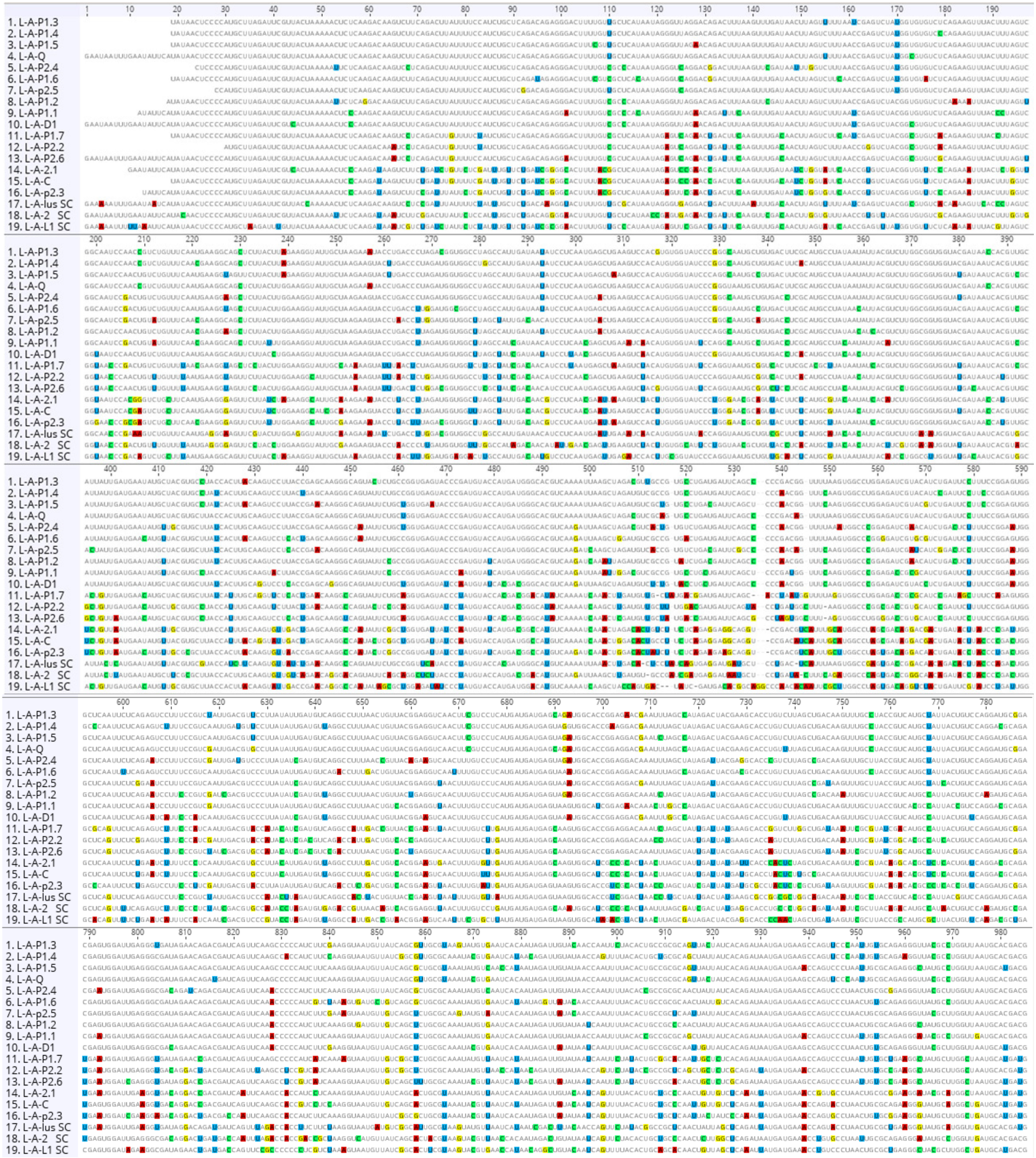
```
M-P1G2 Pool1  ACTATCATATTATATAATTTTTCACTTGCTTTTATTGGTAAGTTCAGCGATCATAGACAGCAACATTGTTAGCCATACCGGCAAGGTATGGT
M-P1G2 Pool2  CCTAACATTTTA-ATAATTTTTCACTTGCTTTTATTAGTGAGCTCAGTGATCATAGATGGTGACACTGTCGGCCCTATTGATAAGGTATGGT

M-P1G2 Pool1  TTAGACAGTGGGATATCGACCTGGCCCTGCAGGCAAATATTACTAAGTGCATCACTAACGTGACTGACAGGCCAGTAATTATAGACGAACAA
M-P1G2 Pool2  TTAGACGGTGGGATATCGACCCGACCCTACAGACCAACATTACTGAGTGTATTACTAACGTTACTGATAGGCCAGTAATCATAGACGAACAA

M-P1G2 Pool1  GTACAACTGTACACACCAGCTATTCAAATATGGGTAGATACTGGCAGTAACTCAGAAGACGTCATTGACCCAGGACTAGGTGCTGCCATAAT
M-P1G2 Pool2  GTGCAATTATACACGCCGGCTATCCAAATATGGGTAGACACTGGCAGTAACTCAGAAGACGTTATTGACCCCGAACTCGATGTTGCCATAAT

M-P1G2 Pool1  GCATTGCTTTGATGGAACCGCGAACTACACAGGCATACCTGTGTATGGTCATAACGTATTTGCTGTAGTTAGTCCGCTAGCTAGTGACAACA
M-P1G2 Pool2  GCATTGTCTTGACGGAACTGCGAACTACACAGGCATACCCGTGTATGGCCATAACGTATTTGCTGTGGTTAGTCCGCTAGCTAGTGACAACA

M-P1G2 Pool1  GCACGTATGCATCCCACTATGACTACTTCGGGCGCGGGTATCCTGCTGACTACGACAGGGATTACGCGTCCGACTACAGCGATACGTCTAAC
M-P1G2 Pool2  GCACGTATGCATCCCACTATGACTATTTCGGACATGGGTACCCTGCCGACTACGATAGGGACTACGCGTCCGACCACAGCGATACACCTGAC

M-P1G2 Pool1  AATATAACCAACATAAATAAGCGAGGTAACGGCAATTGGCCTGACGCTATATTGCCAGGATGGTCAGCGAAAACCGATCTGGGACTTAAGCT
M-P1G2 Pool2  AACACAACTGACATAAATAAGCGAGGTAATGGCAACTGGCCTGACGCTATATTACCAGGATGGTCGGCGAAAACTGATTTGGGACTTAAGCT

M-P1G2 Pool1  TGCGATATACGCATATGCTGCGTACTCGTCTTACTTGCGAACATCACCCGACTTGCAGTGCGGGAAGTGTTATAGTGTAAGGGGCAAGCCGA
M-P1G2 Pool2  CGCAATATACGCGTATGCTGCATACTCGTCTTACTTGCGAACATCACCTGACTTACAATGCGGGAAGTGTTATAGTATAAGGGGCAAGCCGA

M-P1G2 Pool1  TGCCAAAATCAAGTATATTCGAGAATGATTTTGGAGATGAAGTTGACATTATGTACTAGCTGCACCATCAGTGCCGGTCAAGAGGCAAGAAT
M-P1G2 Pool2  TGCCTAAATCAAGTGTATTTGAAAACGATTTCGATGATGAAGTTGATATCATGTACTGGTTGCACCATCAGTGTCGGTCGAGAGGAAAGAAC

M-P1G2 Pool1  GGAGCAGGACCTAGAGTTCCAGGCAAGATAGAACGCATAAAGAAAGACCAGGTGGGAGGACATAATTTCGTAGTAGGCTCATTCTGGATACA
M-P1G2 Pool2  GGGGCAGGGCCTAGGGTTCCAGGTAAAATAGAACGCATAAAGAAAGACCAGATAGGAGGACATAATTTCGTAGTGGGCTCGTTTTGGATACA

M-P1G2 Pool1  GTGCAAGTCAAATCAAGGATGGCCCGATCCTGACTTGACTAAGTACAAGGGCTGGGACGTCGCCTGGGTGTAATGTACTATCTCGACTGGGC
M-P1G2 Pool2  GTGCAAGTCAAATCAAGGGTGGCCAGACCCTGATTTAACTAAGTTCAAGGGCTGGGACGTCACCTGGGTGTAATGTATTATCTCATCCGAGC

M-P1G2 Pool1  CTTTCCCTAGCGAGGTTCACATCATAGTCGACCCACCAACTTAATACCCTTGACTGAAATTAACTAACAAGCGTTAACTGACCAGTCGGGCA
M-P1G2 Pool2  CTCTCCCTAGTAAGCTCCATATCATAGCTGTCCTACCGACTTAACACCCTTAACTTAAATCAACTAACGAGCGTTAACTGACTAGTCGAACA

M-P1G2 Pool1  TAACATGTTAACTACGTGGACCATGAGGTTGATATGGTTAATATGATAATAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
M-P1G2 Pool2  CGACATGTTAGCTACGTGGACTACAAGGTTG--------------------ATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

M-P1G2 Pool1  AAAAAAAAAAAAAA
M-P1G2 Pool2  AAAAAAAAAAAAA
```

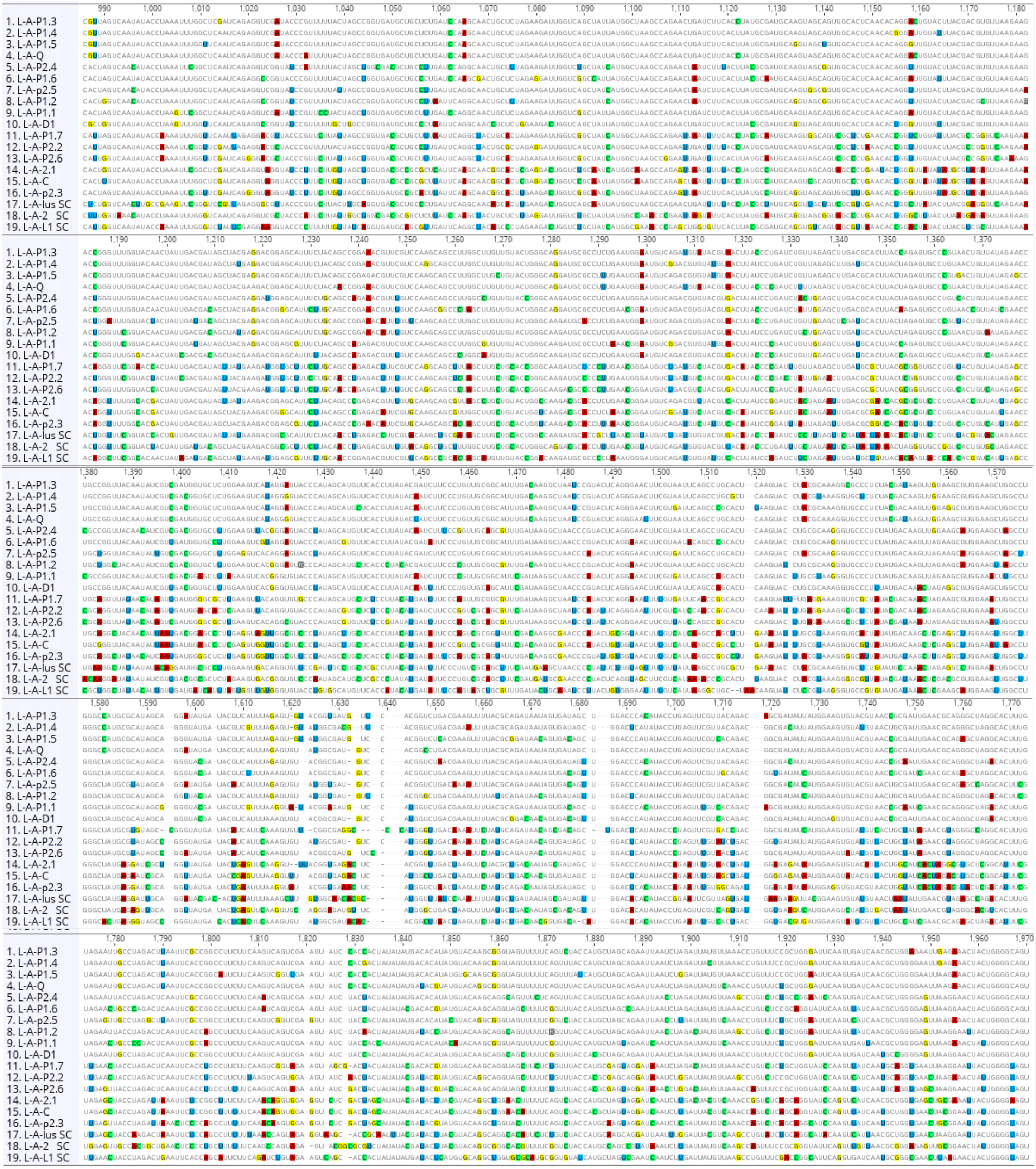
## M-P1G3-1 and M-P1G3-2 alignment

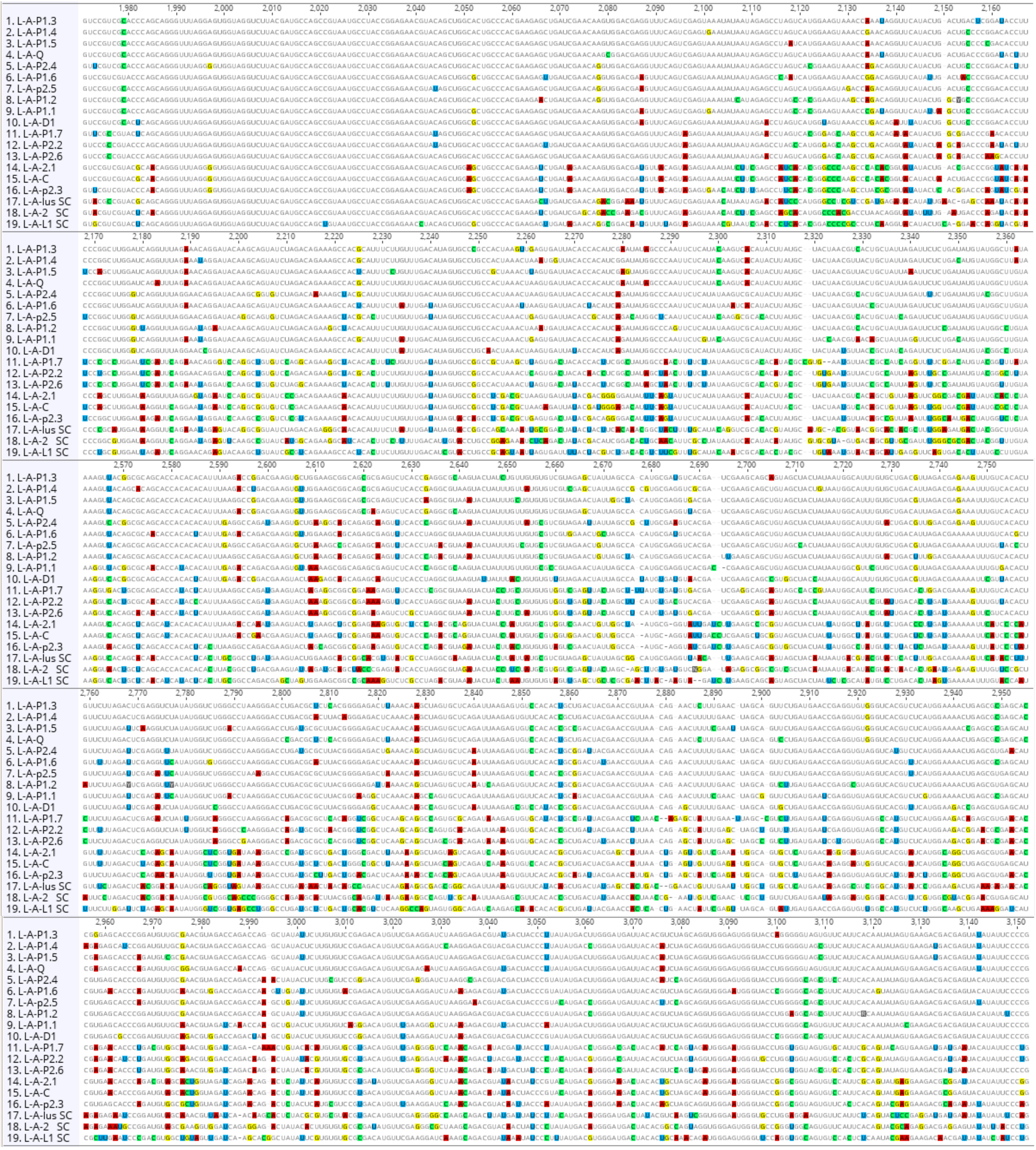


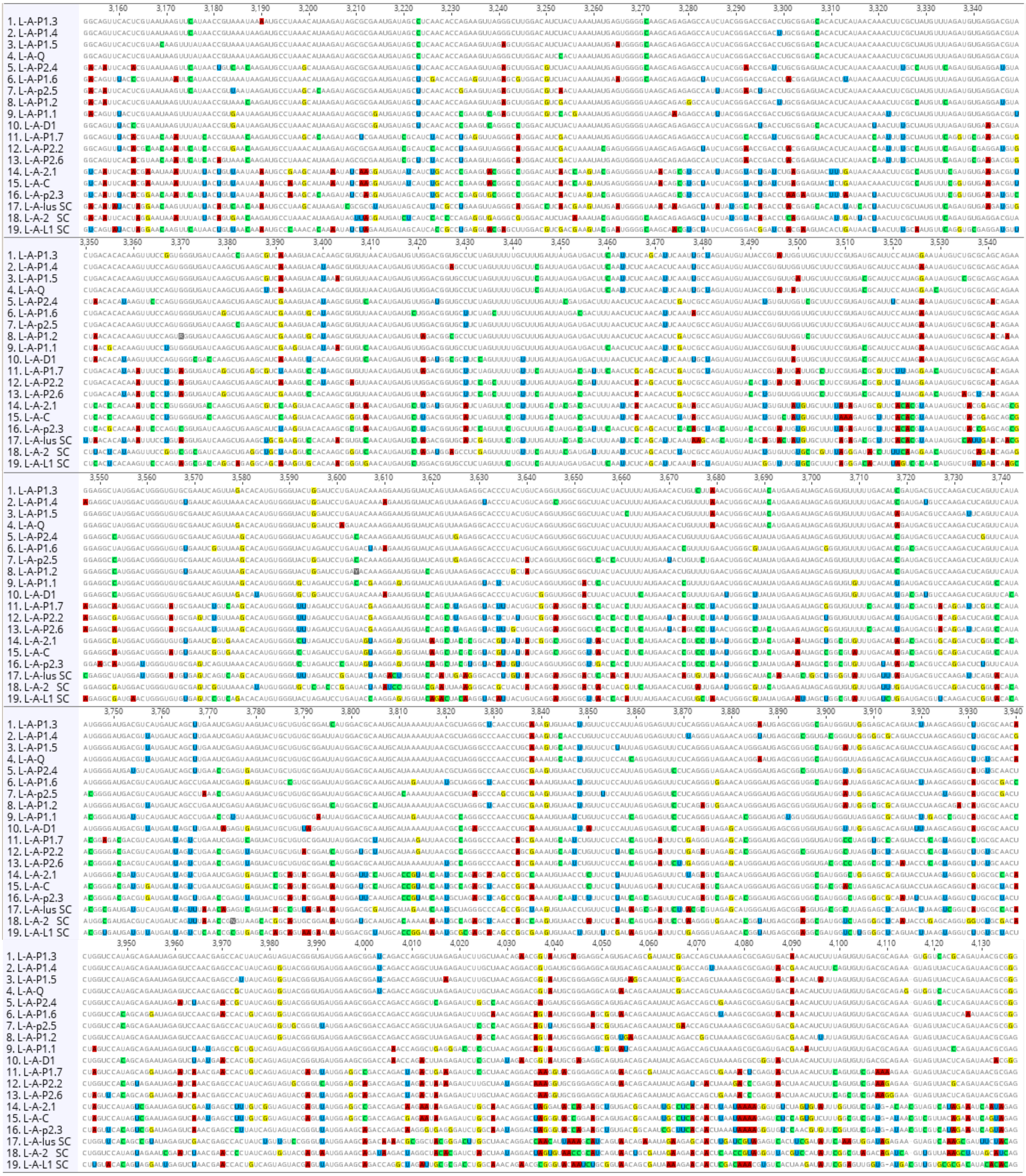
```
               1        10        20        30        40        50        60        70        80        90
M-P1G3-1 Po... TAAGGGCAAGTGAAAAAATACTACGCCATGTAACTTCCCTAGAGGTGTTATACAGCTTAGACATTTTTTCACTTGCCCTTAAGATCATGCTG
M-P1G3-2 Po...     GGCAAGTGAAAAAATGCTACGCCATGTAACTTCCCTGGAGGTATTATACAGCTTAGACATTTTTTCACTTGCCTTAAAGATCATACTG
                        100       110       120       130       140       150       160       170       180
M-P1G3-1 Po... GTAATGTTCTGTGTTTTAGCTGTATTAACGCTGTTAGTGAGCCGACTTTTAGGCACATCACAACCTATGGATGTGGAACACTTGAGTAGGCT
M-P1G3-2 Po... GTAATGTTGTGTGTCTTAGCTGTATTGACGCTGTTAGTAAGCCGACTACTAGGCACATCACAACCTATGGATGTGGAACACTTCAGTAGGCT
                      190       200       210       220       230       240       250       260       270
M-P1G3-1 Po... RCATAAACGAGCTAGATGGACATGGGTTCAAGGCGCAACTTATGCAGGGCTGTTATTAGCAGCAGGTGCTCTCATAGCTCCGTCTTGGTCAG
M-P1G3-2 Po... ACATAAACGAGCCAGGTGGACATGGGTTCAAGGCGCGACTTATGCAGGTCTGCTATTAGCAGCAGGTGCTCTCATAGCTCCGTCTTGGTCAG
                  280       290       300       310       320       330       340       350       360
M-P1G3-1 Po... CAGCAATATGCTTGGCGACGGCGAAGGACTAYTGCGCACCACTAGCTAATGCTATTCTTAGCACGATTGTTGTCAGCATAGCCGGAGGGATA
M-P1G3-2 Po... CAGCAATCTGCCTGGCGACGGCGAAAGACTATTGTGCACCACTAGCTAATGCTATTCTTAGCACGATTGTTGTCGGCGTAGCCGGAGGGATG
                370       380       390       400       410       420       430       440       450       460
M-P1G3-1 Po... GCCTGGCGTACTAGCGGTGTGGGTGTACACGAACGTGCACTGCAATCATCGCTGATTCTCGGCTCCCTTAATCTAACACTCAACGCTGACTT
M-P1G3-2 Po... GCCTGGCGTACTAGCGGTGTGAATGTGCACGAACGTGCGCTGCAGTCGTCACTGGTTCTCGGCTCTCTCAACCTAACACTCAACGCTGACTT
                         470       480       490       500       510       520       530       540       550
M-P1G3-1 Po... CAACAGCAGTCAGCTAGCGGCTGTTGATACCCACCCGCATTTGGTCTCGTTGGATGCGTGGACAGTGGGCCAGCAGACCTACACAGATTAG
M-P1G3-2 Po... TAACAGTAGTCAACTTGCGGCAGTTGATACCCACCCGCATTTGGTTTCGTTGGATGCGTGGACAGTAGGGCCATCAGACTTACACAGATTGG
                       560       570       580       590       600       610       620       630       640
M-P1G3-1 Po... GTAAACGTGRTGRGGATACAGGTAATTCTGTCAATAACGGCACTTATTTGTTCTTCACGTCCGAATATGGCACTCATACTGCTCACATGTCA
M-P1G3-2 Po... GTAAACGTAATGAGGATACAGGTAATCCTGTTAATGACGGCACCTATTTATTCTTTACATCCGAATATGGCACTCATACTGCGCACATGCCA
                    650       660       670       680       690       700       710       720       730
M-P1G3-1 Po... GGTGATGATGTCGGTACTCTAGTAGATTTTGTRTTAAACGCGATTCCAGACGAGCCTAACATTAATAGCACAATTAATTCTGCTAAGCGGAG
M-P1G3-2 Po... GGTGATGACGTCGGTACTCTAGTAGACTTTGTATTGAACGCAATTCCAAACGAGCCTAACATCAATAGCACAATTAATTCTGCTAAGCGGAG
                   740       750       760       770       780       790       800       810       820
M-P1G3-1 Po... TCAACAATTCGGAGTATCATGGGTGTCATATATATGGGACGAAGCCAACCATGATCTGGATACCGAGTGGTATAATGAAGAAGGTGGTAGCT
M-P1G3-2 Po... TCAACAATTCGGAGTATCATGGGTGTCATATATATGGGATGAAGCCAACCATGATCTGGATACCGAGTGGTATAATGAAGAGGGTGGTAGCT
                 830       840       850       860       870       880       890       900       910       920
M-P1G3-1 Po... TCGACTCYCAGTTAGAAGAAGGTTTGACGCAAGCAATAGTTGATATACCTGACTGGAAGTACTGTATTGCACCTGAAGTCAGTAAGGGCGAG
M-P1G3-2 Po... TCGACTCTCAGTTAGAAGAAGGCTTGACGCAAGCAATAACTGATATACCTGACTGGAAGTACTGCATTGCACCAGAAGTCAGTAAGGGCGAA
                         930       940       950       960       970       980       990      1,000     1,010
M-P1G3-1 Po... GCAATATCTTATGACGACATACCCGGCACGCAAAACGGAAATGGCAATGCGTTACACGGAGAAGTCTACTTCAATACCTACGGTGGTTTGGA
M-P1G3-2 Po... GCAATATCTTATGACGACATACCCGGCACGCATAACGGAAATGGTAAKGCGTTACACGGGGAGGTCTACTTCAATACCTACGGTGGTTYGGA
                      1,020     1,030     1,040     1,050     1,060     1,070     1,080     1,090     1,100
M-P1G3-1 Po... TAGTTATTGCAATTCTGCACATSACGGTGCGGATCGTTGGCGGTAGGTATCACACCACCTGGTCT-AGAGGATTCAYTTGCGCATGAAGTCT
M-P1G3-2 Po... TAATTATTGTAACTCCGCTCATGATGGTGCGGATCGTTGGCGGTAAGTATCAGACCACCTGGCCTGGGAAGCTTCCTTTGCGAATGAAGTTT
                    1,110     1,120     1,130     1,140     1,150     1,160     1,170     1,180     1,190
M-P1G3-1 Po... GTGATTGAATACCTCTGCCGGCTTAAGATAAGTCAGAATTAGATTTAACAAATAAATAGTATAGGTTAACGTAGTATTTATTAWAAAAAAAAA
M-P1G3-2 Po... ACTATCAGATACTTCTGCTGGCTTAAGATAAGTCGAAATTAAATTTGACAAATAAATGGTATAAACTA-------ATTTAAAAAAAAAAAA
                  1,200     1,210     1,220     1,230     1,236
M-P1G3-1 Po... AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
M-P1G3-2 Po... AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

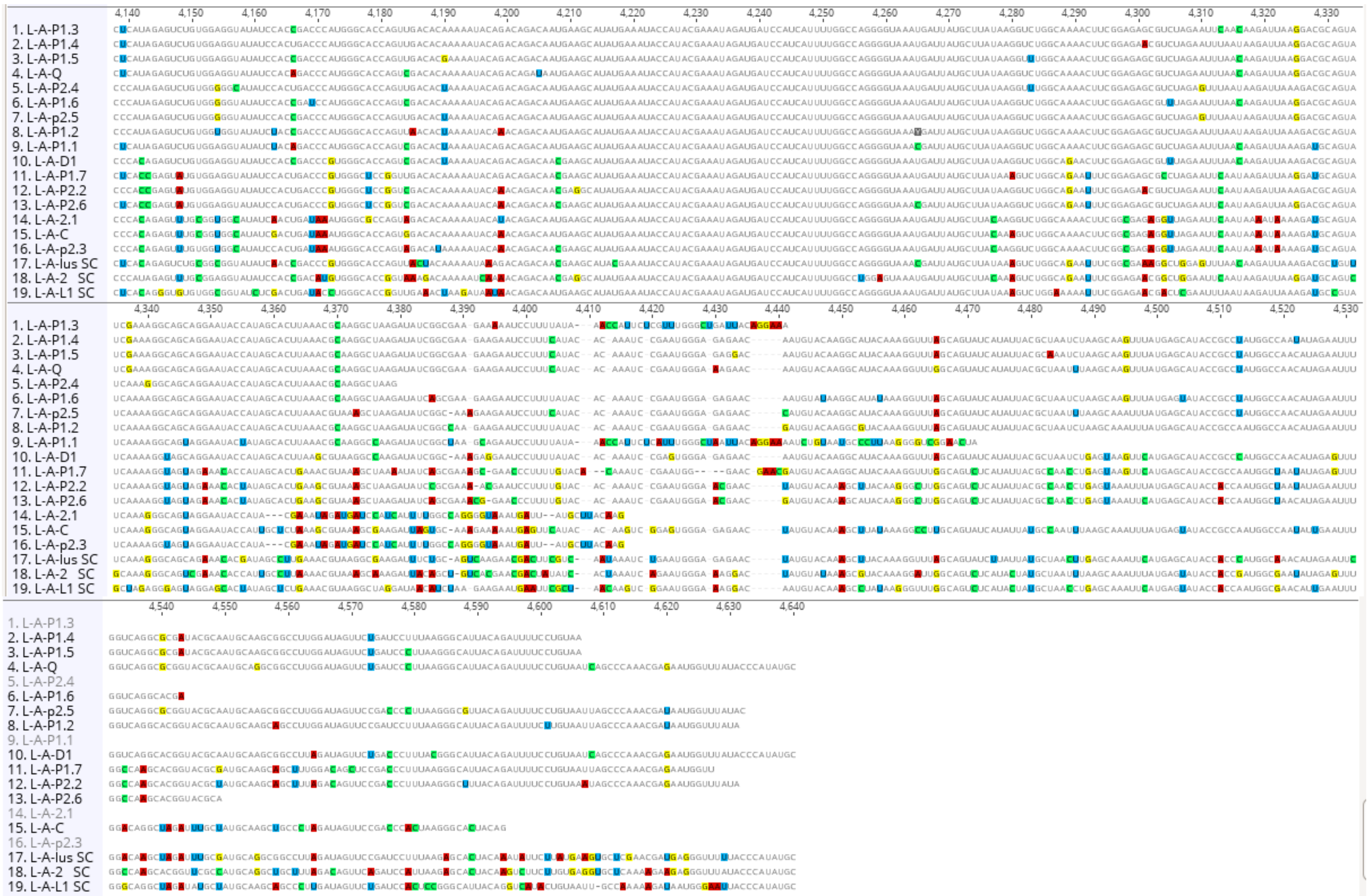
**Nucleotide alignment of all L-A.** The first 16 sequences are from *S. paradoxus* and the three last sequences belong to *S. cerevisiae.* SC is abbreviation of *S. cerevisiae.*



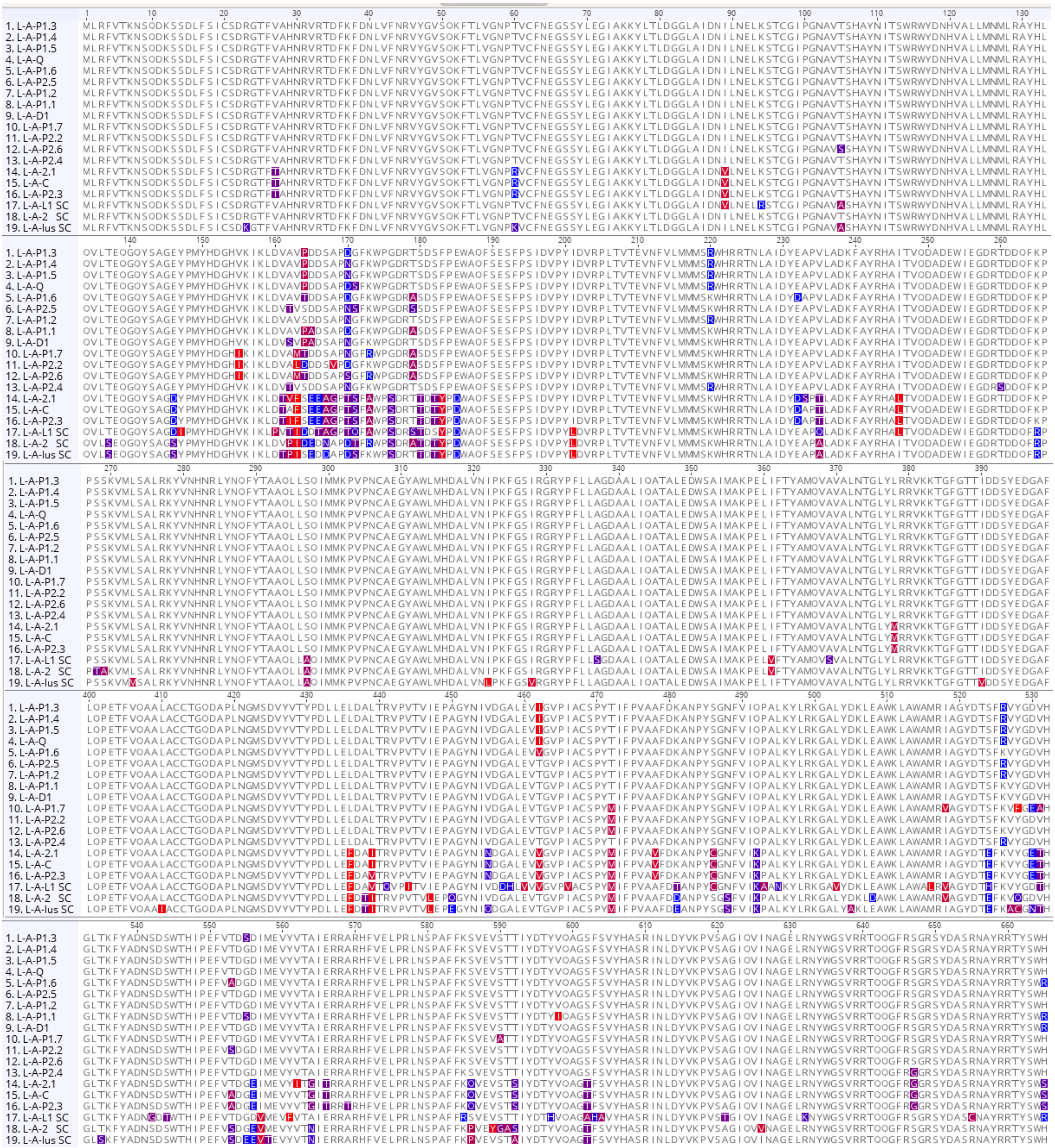

990
1,000
1,010
1,020
1,030
1,040
1,050
1,060
1,070
1,080
1,090
1,100
1,110
1,120
1,130
1,140
1,150
1,160
1,170
1,180
1,190
1,200
1,210
1,220
1,230
1,240
1,250
1,260
1,270
1,280
1,290
1,300
1,310
1,320
1,330
1,340
1,350
1,360
1,370
1,380
1,390
1,400
1,410
1,420
1,430
1,440
1,450
1,460
1,470
1,480
1,490
1,500
1,510
1,520
1,530
1,540
1,550
1,560
1,570
1,580
1,590
1,600
1,610
1,620
1,630
1,640
1,650
1,660
1,670
1,680
1,690
1,700
1,710
1,720
1,730
1,740
1,750
1,760
1,770
1,780
1,790
1,800
1,810
1,820
1,830
1,840
1,850
1,860
1,870
1,880
1,890
1,900
1,910
1,920
1,930
1,940
1,950
1,960
1,970
1. L-A-P1.3
2. L-A-P1.4
3. L-A-P1.5
4. L-A-Q
5. L-A-P2.4
6. L-A-P1.6
7. L-A-p2.5
8. L-A-P1.2
9. L-A-P1.1
10. L-A-D1
11. L-A-P1.7
12. L-A-P2.2
13. L-A-P2.6
14. L-A-2.1
15. L-A-C
16. L-A-p2.3
17. L-A-lus SC
18. L-A-2 SC
19. L-A-L1 SC


1. L-A-P1.3
2. L-A-P1.4
3. L-A-P1.5
4. L-A-Q
5. L-A-P2.4
6. L-A-P1.6
7. L-A-p2.5
8. L-A-P1.2
9. L-A-P1.1
10. L-A-D1
11. L-A-P1.7
12. L-A-P2.2
13. L-A-P2.6
14. L-A-2.1
15. L-A-C
16. L-A-p2.3
17. L-A-lus SC
18. L-A-2 SC
19. L-A-L1 SC
1,980
1,990
2,000
2,010
2,020
2,030
2,040
2,050
2,060
2,070
2,080
2,090
2,100
2,110
2,120
2,130
2,140
2,150
2,160
2,170
2,180
2,190
2,200
2,210
2,220
2,230
2,240
2,250
2,260
2,270
2,280
2,290
2,300
2,310
2,320
2,330
2,340
2,350
2,360
2,570
2,580
2,590
2,600
2,610
2,620
2,630
2,640
2,650
2,660
2,670
2,680
2,690
2,700
2,710
2,720
2,730
2,740
2,750
2,760
2,770
2,780
2,790
2,800
2,810
2,820
2,830
2,840
2,850
2,860
2,870
2,880
2,890
2,900
2,910
2,920
2,930
2,940
2,950
2,960
2,970
2,980
2,990
3,000
3,010
3,020
3,030
3,040
3,050
3,060
3,070
3,080
3,090
3,100
3,110
3,120
3,130
3,140
3,150

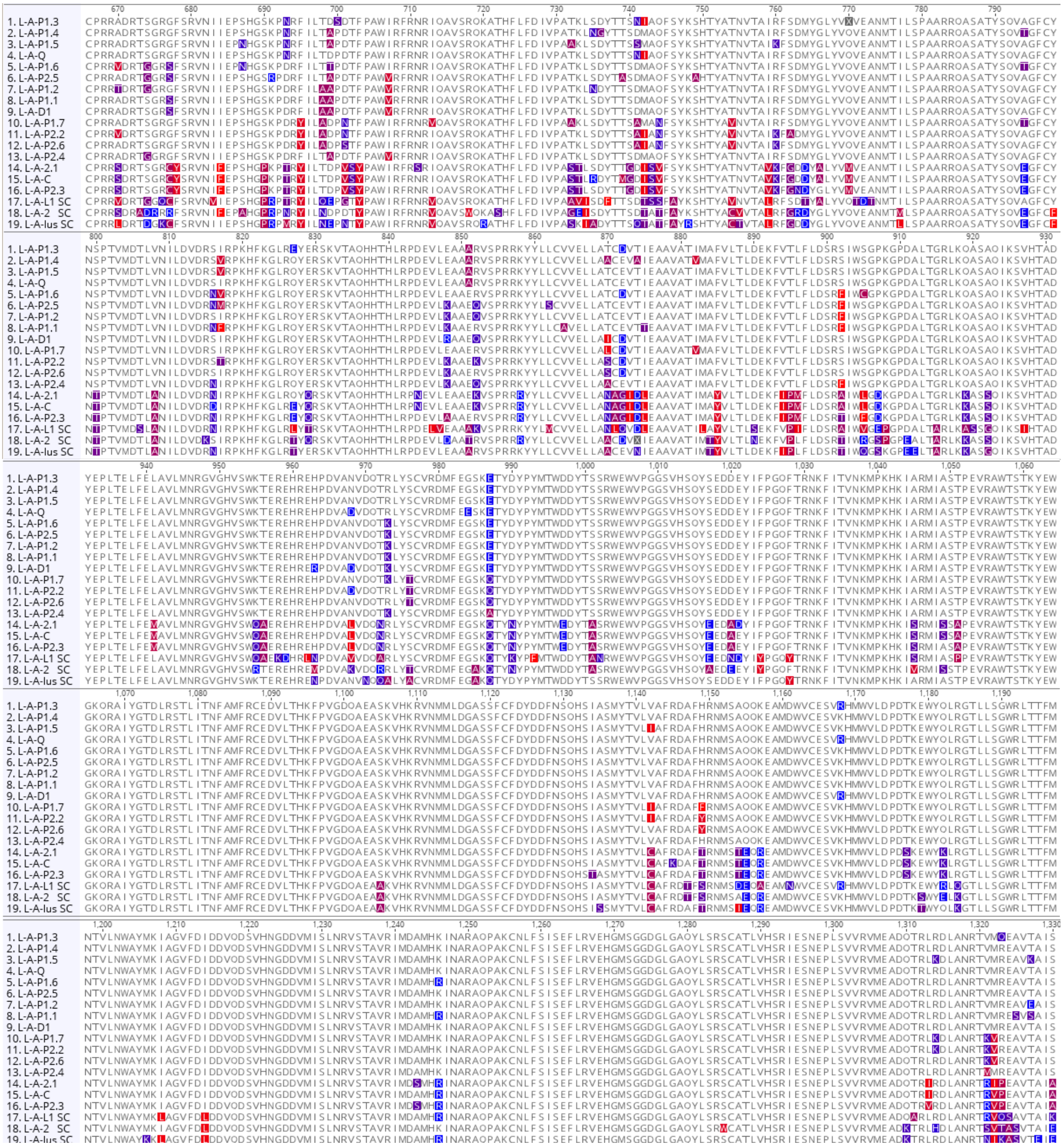

3,160
3,170
3,180
3,190
3,200
3,210
3,220
3,230
3,240
3,250
3,260
3,270
3,280
3,290
3,300
3,310
3,320
3,330
3,340
3,350
3,360
3,370
3,380
3,390
3,400
3,410
3,420
3,430
3,440
3,450
3,460
3,470
3,480
3,490
3,500
3,510
3,520
3,530
3,540
3,550
3,560
3,570
3,580
3,590
3,600
3,610
3,620
3,630
3,640
3,650
3,660
3,670
3,680
3,690
3,700
3,710
3,720
3,730
3,740
3,750
3,760
3,770
3,780
3,790
3,800
3,810
3,820
3,830
3,840
3,850
3,860
3,870
3,880
3,890
3,900
3,910
3,920
3,930
3,940
3,950
3,960
3,970
3,980
3,990
4,000
4,010
4,020
4,030
4,040
4,050
4,060
4,070
4,080
4,090
4,100
4,110
4,120
4,130
1. L-A-P1.3
2. L-A-P1.4
3. L-A-P1.5
4. L-A-Q
5. L-A-P2.4
6. L-A-P1.6
7. L-A-p2.5
8. L-A-P1.2
9. L-A-P1.1
10. L-A-D1
11. L-A-P1.7
12. L-A-P2.2
13. L-A-P2.6
14. L-A-2.1
15. L-A-C
16. L-A-p2.3
17. L-A-lus SC
18. L-A-2 SC
19. L-A-L1 SC


4,140
4,150
4,160
4,170
4,180
4,190
4,200
4,210
4,220
4,230
4,240
4,250
4,260
4,270
4,280
4,290
4,300
4,310
4,320
4,330
4,340
4,350
4,360
4,370
4,380
4,390
4,400
4,410
4,420
4,430
4,440
4,450
4,460
4,470
4,480
4,490
4,500
4,510
4,520
4,530
4,540
4,550
4,560
4,570
4,580
4,590
4,600
4,610
4,620
4,630
4,640
1. L-A-P1.3
2. L-A-P1.4
3. L-A-P1.5
4. L-A-Q
5. L-A-P2.4
6. L-A-P1.6
7. L-A-p2.5
8. L-A-P1.2
9. L-A-P1.1
10. L-A-D1
11. L-A-P1.7
12. L-A-P2.2
13. L-A-P2.6
14. L-A-2.1
15. L-A-C
16. L-A-p2.3
17. L-A-lus SC
18. L-A-2 SC
19. L-A-L1 SC

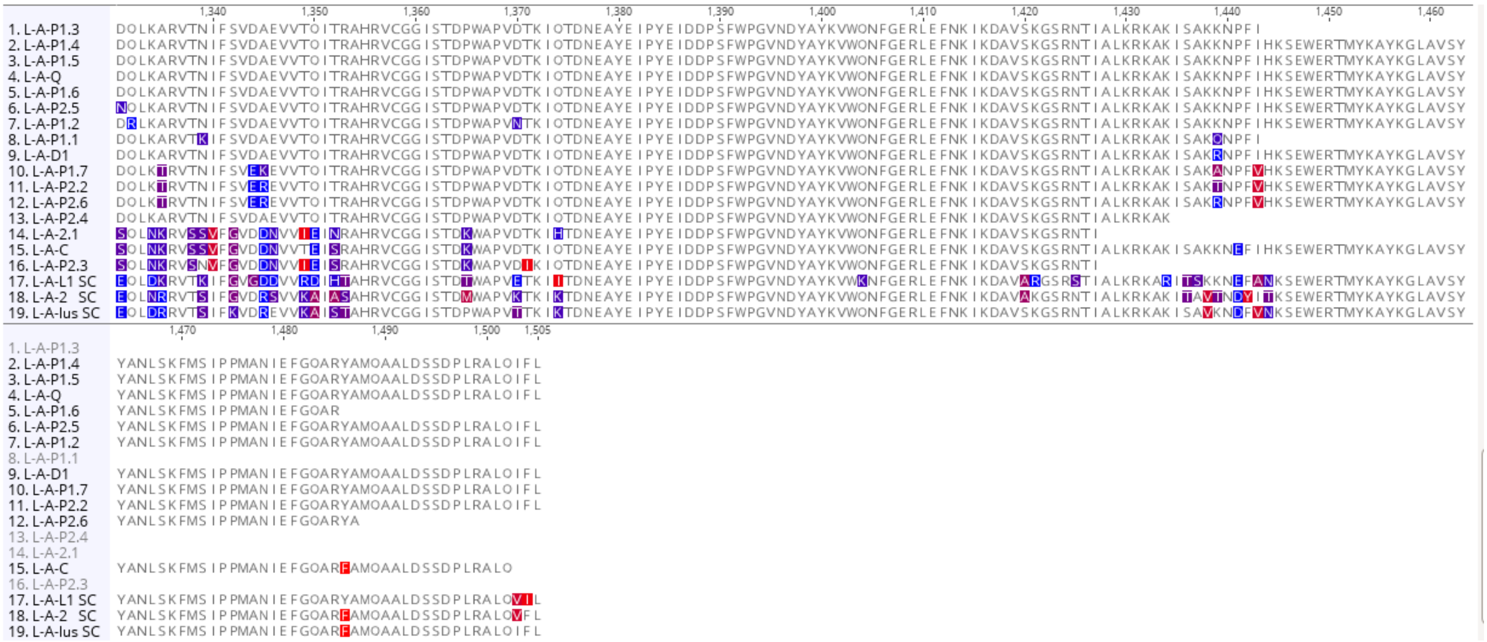
**Gag and Pol fusion protein sequence alignment of all L-A dsRNA in *S. cerevisiae* and *S. paradoxus.***
The first 16 sequences are from *S. paradoxus* and the three last sequences belong to *S. cerevisiae.* SC is abbreviation of *S. cerevisiae.*



```
                1       10        20        30        40        50        60        70        80        90       100       110       120       130
1. L-A-P1.3     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
2. L-A-P1.4     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
3. L-A-P1.5     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
4. L-A-Q        MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
5. L-A-P1.6     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
6. L-A-P2.5     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
7. L-A-P1.2     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
8. L-A-P1.1     MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
9. L-A-D1       MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
10. L-A-P1.7    MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
11. L-A-P2.2    MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
12. L-A-P2.6    MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVSSHAYNITSWRWYDNHVALLMNMLRAYHL
13. L-A-P2.4    MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
14. L-A-2.1     MLRFVTKNSQDKSSDLFSICSDRGTFTAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPRVCFNEGSSYLEGIAKKYLTLDGGLAIDNVLNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
15. L-A-C       MLRFVTKNSQDKSSDLFSICSDRGTFTAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPRVCFNEGSSYLEGIAKKYLTLDGGLAIDNVLNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
16. L-A-P2.3    MLRFVTKNSQDKSSDLFSICSDRGTFTAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPRVCFNEGSSYLEGIAKKYLTLDGGLAIDNVLNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
17. L-A-L1 SC   MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNVLNELRSTCGIPGNAVASHAYNITSWRWYDNHVALLMNMLRAYHL
18. L-A-2  SC   MLRFVTKNSQDKSSDLFSICSDRGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPTVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVTSHAYNITSWRWYDNHVALLMNMLRAYHL
19. L-A-lus SC  MLRFVTKNSQDKSSDLFSICSDKGTFVAHNRVRTDFKFDNLVFNRVYGVSQKFTLVGNPKVCFNEGSSYLEGIAKKYLTLDGGLAIDNILNELKSTCGIPGNAVASHAYNITSWRWYDNHVALLMNMLRAYHL

                    140       150       160       170       180       190       200       210       220       230       240       250       260
1. L-A-P1.3     QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVPDDSAPDGFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSRWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
2. L-A-P1.4     QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVPDDSAPNGFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSRWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
3. L-A-P1.5     QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVPDDSAPNGFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSRWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
4. L-A-Q        QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVPDDSAPDSFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSRWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
5. L-A-P1.6     QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVTDDSAPDGFKWPGDRASDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYDAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
6. L-A-P2.5     QVLTEQGQYSAGEYPMYHDGHVKIKLDVTVSDDSAPNSFKWPGDRSSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
7. L-A-P1.2     QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVSDDSAPNGFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSRWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
8. L-A-P1.1     QVLTEQGQYSAGEYPMYHDGHVKIKLDVAVPADSAPDGFKWPGDRASDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
9. L-A-D1       QVLTEQGQYSAGEYPMYHDGHVKIKLDVSVPADSAPNGFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
10. L-A-P1.7    QVLTEQGQYSAGEYPMYHDGHIKIKLDVAMTDDSAPNGFRWPGDRASDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
11. L-A-P2.2    QVLTEQGQYSAGEYPMYHDGHIKIKLDVALDDDSVPDGFKWPGDRASDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
12. L-A-P2.6    QVLTEQGQYSAGEYPMYHDGHIKIKLDVAMTDDSAPSGFRWPGDRASDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRTDDQFKP
13. L-A-P2.4    QVLTEQGQYSAGEYPMYHDGHVKIKLDVTVSDDSAPNGFKWPGDRTSDSFPEWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSRWHRRTNLAIDYEAPVLADKFAYRHAITVQDADEWIEGDRSDDQFKP
14. L-A-2.1     QVLTEQGQYSAGDYPMYHDGHVKIKLDTVFSEEAGPTSFAWPSDRTTDTYPDWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYDSPTLADKFAYRHALTVQDADEWIEGDRTDDQFKP
15. L-A-C       QVLTEQGQYSAGDYPMYHDGHVKIKLDTAFSEEAGPTSFAWPSDRTTDTYPDWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYDAPTLADKFAYRHALTVQDADEWIEGDRTDDQFKP
16. L-A-P2.3    QVLTEQGQYSAGDYPMYHDGHVKIKLDTIFSEEAGPTSFAWPSDRTTDTYPDWAQFSESFPSIDVPYIDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYDAPTLADKFAYRHALTVQDADEWIEGDRTDDQFKP
17. L-A-L1 SC   QVLTEQGQYSAGDIPMYHDGHVKIKLPVTIDDTAGPTQFAWPSDRSTDSYPDWAQFSESFPSIDVPYLDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPQLADKFAYRHALTVQDADEWIEGDRTDDQFRP
18. L-A-2  SC   QVLSEQGQYSAGSYPMYHDGHVKIKLDVPIDEDNAPDTFRWPSDRATDTYPDWAQFSESFPSIDVPYLDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPALADKFAYRHAITVQDADEWIEGDRTDDQFRP
19. L-A-lus SC  QVLSEQGQYSAGSYPMYHDGHVKIKLDTPISEDDAPDSFKWPSDRTTDTYPDWAQFSESFPSIDVPYLDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPALADKFAYRHAITVQDADEWIEGDRTDDQFRP

                 270       280       290       300       310       320       330       340       350       360       370       380       390
1. L-A-P1.3     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
2. L-A-P1.4     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
3. L-A-P1.5     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
4. L-A-Q        PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
5. L-A-P1.6     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
6. L-A-P2.5     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
7. L-A-P1.2     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
8. L-A-P1.1     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
9. L-A-D1       PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
10. L-A-P1.7    PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
11. L-A-P2.2    PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
12. L-A-P2.6    PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
13. L-A-P2.4    PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
14. L-A-2.1     PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYMRRVKKTGFGTTIDDSYEDGAF
15. L-A-C       PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYMRRVKKTGFGTTIDDSYEDGAF
16. L-A-P2.3    PSSKVMLSALRKYVNHNRLYNQFYTAAQLLSQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYMRRVKKTGFGTTIDDSYEDGAF
17. L-A-L1 SC   PSSKVMLSALRKYVNHNRLYNQFYTAAQLLAQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLSGDAALIQATALEDWSAIMAKPELVFTYAMQVSVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
18. L-A-2  SC   PTAKVMLSALRKYVNHNRLYNQFYTAAQLLAQIMMKPVPNCAEGYAWLMHDALVNIPKFGSIRGRYPFLLAGDAALIQATALEDWSAIMAKPELVFTYAMQVAVALNTGLYLRRVKKTGFGTTIDDSYEDGAF
19. L-A-lus SC  PSSKVMMSALRKYVNHNRLYNQFYTAAQLLAQIMMKPVPNCAEGYAWLMHDALVNLPKFGSVRGRYPFLLAGDAALIQATALEDWSAIMAKPELIFTYAMQVAVALNTGLYLRRVKKTGFGTTVDDSYEDGAF

                400     410       420       430       440       450       460       470       480       490       500       510       520       530
1. L-A-P1.3     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVIGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
2. L-A-P1.4     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVIGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
3. L-A-P1.5     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVIGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
4. L-A-Q        LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVIGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
5. L-A-P1.6     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVVGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFKVYGDVH
6. L-A-P2.5     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
7. L-A-P1.2     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
8. L-A-P1.1     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFKVYGDVH
9. L-A-D1       LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFKVYGDVH
10. L-A-P1.7    LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYMIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRVAGYDTSFKVFGEAH
11. L-A-P2.2    LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYMIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFKVYGDVH
12. L-A-P2.6    LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYMIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFKVYGDVH
13. L-A-P2.4    LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLELDALTRVPVTVIEPAGYNIVDGALEVTGVPIACSPYTIFPVAAFDKANPYSGNFVIQPALKYLRKGALYDKLEAWKLAWAMRIAGYDTSFRVYGDVH
14. L-A-2.1     LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLEFDAITRVPVTVIEPAGYNINDGALEVVGVPIACSPYMIFPVAVFDKANPYCGNFVIKPALKYLRKGALYDKLEAWKLAWAMRIAGYDTEFKVYGETH
15. L-A-C       LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLEFDAITRVPVTVIEPAGYNINDGALEVVGVPIACSPYMIFPVAVFDKANPYCGNFVIKPALKYLRKGALYDKLEAWKLAWAMRIAGYDTEFKVYGETH
16. L-A-P2.3    LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLEFDAVTRVPVTVIEPAGYNINDGALEVVGVPIACSPYMIFPVAVFDKANPYCGNFVIKPALKYLRKGALYDKLEAWKLAWAMRIAGYDTEFKVYGETH
17. L-A-L1 SC   LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLEFDAVTQVPITVIEPAGYNIVDDHLVVVGVPVACSPYMIFPVAAFDTANPYCGNFVIKAANKYLRKGAVYDKLEAWKLAWALRVAGYDTHFKVYGDTH
18. L-A-2  SC   LQPETFVQAALACCTGQDAPLNGMSDVYYVTYPDLLEFDTITRVPVTVLEPQGYNIVDGALEVTGVPIACSPYMIFPVAAFDDANPYSGSFVIKPALKYLRKGALYDKLDAWKLAWAMRVAGYDTEFKVQGDVH
19. L-A-lus SC  LQPETFVQAAIACCTGQDAPLNGMSDVYYVTYPDLLEFDTITRVPVTVLEPEGYNIQDGALEVTGVPIACSPYMIFPVAAFDEANPYSGSFVIKPALKYLRKGALYAKLEAWKLAWAMRIAGYDTEFKACGNTH

                    540       550       560       570       580       590       600       610       620       630       640       650       660
1. L-A-P1.3     GLTKFYADNSDSWTHIPEFVTDSDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
2. L-A-P1.4     GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
3. L-A-P1.5     GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
4. L-A-Q        GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
5. L-A-P1.6     GLTKFYADNSDSWTHIPEFVADGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWR
6. L-A-P2.5     GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
7. L-A-P1.2     GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
8. L-A-P1.1     GLTKFYADNSDSWTHIPEFVTDSDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYIQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWR
9. L-A-D1       GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWR
10. L-A-P1.7    GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVATTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
11. L-A-P2.2    GLTKFYADNSDSWTHIPEFVSDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
12. L-A-P2.6    GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
13. L-A-P2.4    GLTKFYADNSDSWTHIPEFVTDGDIMEVYVTAIERRARHFVELPRLNSPAFFKSVEVSTTIYDTYVQAGSFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRGGRSYDASRNAYRRTYSWH
14. L-A-2.1     GLTKFYADNSDSWTHIPEFVTDGEIMEVYITGITRRARHFVELPRLNSPAFFKQVEVSTSIYDTYVQAGTFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRGGRSYDASRNAYRRTYSWS
15. L-A-C       GLTKFYADNSDSWTHIPEFVADGEIMEVYVTGITRRARHFVELPRLNSPAFFKQVEVSTSIYDTYVQAGTFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRGGRSYDASRNAYRRTYSWS
16. L-A-P2.3    GLTKFYADNSDSWTHIPEFVADGEIMEVYVTGITRRTRHFVELPRLNSPAFFKQVEVSTSIYDTYVQAGTFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRGGRSYDASRNAYRRTYSWS
17. L-A-L1 SC   GLTKFYADNGDTWTHIPEFVTDGDVMEVFVTAIERRARHFVELPRLNSPAFFRSVEVSTTIYDTHVQAGAHAVYHASRINLDYVKPVSTGIQVINAGELKNYWGSVRRTQQGFRSGRSYDASCNAYRRTYSWR
18. L-A-2  SC   GLTKFYADNSDSWTHIPEFVSDGEVMEVYVTNIERRARHFVELPRLNSPAFFKPVEYGASIYDTYVQAGTFSVYHASRINLDYVKPVSAGIQVVNAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
19. L-A-lus SC  GLSKFYADNSDSWTHIPEFVSDEEVTEVYVTNIERRARHFVELPRLNSPAFFKPVEVSTAIYDTYVQAGTFSVYHASRINLDYVKPVSAGIQVINAGELRNYWGSVRRTQQGFRSGRSYDASRNAYRRTYSWH
```



```
                 670       680       690       700       710       720       730       740       750       760       770       780       790
1. L-A-P1.3      CPRRADRTSGRGFSRVNIIEPSHGSKPNRFILTDSDTFPAWIRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSNIAOFSYKSHTYATNVTAIRFSDMYGLYVXVEANMTILSPAARROASATYSOVAGFCY
2. L-A-P1.4      CPRRADRTSGRGFSRVNIIEPSHGSKPNRFILTAPDTFPAWIRFRNRIOAVSROKATHFLFDIVPATKLNGYTTSDMAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVTGFCY
3. L-A-P1.5      CPRRADRTSGRGFSRVNIIEPNHGSKPNRFILTAPDTFPAWIRFRNRIOAVSROKATHFLFDIVPAAKLSDYTTSSMAOFSYKSHTYATNVTAIKFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
4. L-A-Q         CPRRADRTSGRGFSRVNIIEPSHGSKPNRFILTDPDTFPAWIRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSNIAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
5. L-A-P1.6      CPRRVDRTGGRSFSRVNIIEPNHGSKPDRFILTTPDTFPAWIRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSDMAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVTGFCY
6. L-A-P2.5      CPRRADRTGGRSFSRVNIIEPSHGSRPDRFILTAPDTFPAWVRFRNRIOAVSROKATHFLFDIVPATKLSDYTASDMAOFSYKAHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
7. L-A-P1.2      CPRRTDRTGGRGFSRVNIIEPSHGSKPDRFILAAPDTFPAWVRFRNRIOAVSROKATHFLFDIVPATKLNDYTTSDMAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
8. L-A-P1.1      CPRRADRTSGRSFSRVNIIEPSHGSKPDRFILAAPDTFPAWVRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSDMAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
9. L-A-D1        CPRRADRTSGRSFSRVNIIEPSHGSKPDRFILAAPDTFPAWVRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSDMAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
10. L-A-P1.7     CPRRADRTSGRGFSRVNIIEPSHGSKPDRYILADPNTFPAWIRFRNRVOAVSROKATHFLFDIVPAAKLSDYTTSAMANFSYKSHTYAVNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVTGFCY
11. L-A-P2.2     CPRRVDRTSGRGFSRVNIIEPSHGSKPDRYILADPNTFPAWIRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSAIANFSYKSHTYAVNVTAIKFADMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
12. L-A-P2.6     CPRRADRTSGRGFSRVNIIEPSHGSKPDRYILADPSTFPAWIRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSAIANFSYKSHTYAVNVTAIKFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
13. L-A-P2.4     CPRRADRTGGRGFSRVNIIEPSHGSKPDRFILTAPDTFPAWVRFRNRIOAVSROKATHFLFDIVPATKLSDYTTSDMAOFSYKSHTYATNVTAIRFSDMYGLYVOVEANMTILSPAARROASATYSOVAGFCY
14. L-A-2.1      CPRRSDRTSGRCYSRVNIFEPSHGPKPTRYILTDPVSYPAWIRFRSRIOAVSROKATHFLFDIVPASTLSDYTTGDISVFSYKSHTYATNVTAVKFGDDYALYVMVEANMTILSPAARROASATYSOVEGFCY
15. L-A-C        CPRRSDRTSGRCYSRVNIFEPSHGPKPTRYILTDPVSYPAWIRFRNRIOAVSROKATHFLFDIVPASTLRDYTMGDISVFSYKSHTYATNVTAVKFGDDYALYVMVEANMTILSPAARROASATYSOVEGFCY
16. L-A-P2.3     CPRRSDRTSGRCYSRVNIFEPSHGPKPTRYILTDPVSYPAWIRFRNRIOAVSROKATHFLFDIVPASTLSDYTTGDISVFSYKSHTYATNVTAVKFGNDYGLYVMVEANMTILSPAARROASATYSOVEGFCY
17. L-A-L1 SC    CPRRVDRTGGOCFSRVNVIEPSHGPRPTRYILOEPGTYPAWIRFRNRVOAVSROKATHFLFDIVPAAVISDFTTSDTSSFAYKSHTYAVNVTALRFSDTYALYVOTDTNMTILSPAARROASATYSOVAGFCY
18. L-A-2  SC    CPRRSDRADRRRFSRVNIFEPAHGPRPNRYILNDPDTYPAWIRFRNRVOAVSMOKASHFLFDIVPAGEISDYTTSDTATFAYKSHTYACVVTALRFGRDYGLYVOVEANMTMLSPAARROASATYSOVEGFCF
19. L-A-lus SC   CPRRLDRTDGKCFSRVNIIEPSHGPRPMRYILNEPNTYPAWIRFRNRVOAVSRORATHFLFDIVPASKIADYTTSOTATFAYRSHTYACTVTALRFGDDYGLYVOVEANMTLLSPAARROASATYSOVEGFCF

                 800       810       820       830       840       850       860       870       880       890       900       910       920       930
1. L-A-P1.3      NSPTVMDTLVNILDVDRSIRPKHFKGLREYERSKVTAOHHTHLRPDEVLEAAARVSPRRKYYLLCVVELLATCDVTIEAAVATIMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
2. L-A-P1.4      NSPTVMDTLVNILDVDRSVRPKHFKGLROYERSKVTAOHHTHLRPDEVLEAAARVSPRRKYYLLCVVELLAACEVAIEAAVATVMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
3. L-A-P1.5      NSPTVMDTLVNILDVDRSVRPKHFKGLROYERSKVTAOHHTHLRPDEVLEAAARVSPRRKYYLLCVVELLATCEVTIEAAVATIMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
4. L-A-Q         NSPTVMDTLVNILDVDRSIRPKHFKGLROYERSKVTAOHHTHLRPDEVLEAAARVSPRRKYYLLCVVELLATCEVTIEAAVATIMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
5. L-A-P1.6      NSPTVMDTLVNILDVDRNVRPKHFKGLROYERSKVTAOHHTHLRPDEVLEAAERVSPRRKYYLLCVVELLATCDVTIEAAVATIMAFVLTLDEKFVTLFLDSRFIWCGPKGPDALTGRLKOASAOIKSVHTAD
6. L-A-P2.5      NSPTVMDTLVNILDVDRNMRPKHFKGLROYERSKVTAOHHTHLRPDEVLKAAEOVSPRRKYYLSCVVELLATCEVTIEAAVATIMAFVLTLDEKFVTLFLDSRFIWSGPKGPDALTGRLKOASAOIKSVHTAD
7. L-A-P1.2      NSPTVMDTLVNILDVDRSIRPKHFKGLROYERSKVTAOHHTHLRPDEVLKAAEOVSPRRKYYLLCVVELLATCEVTIEAAVATIMAFVLTLDEKFVTLFLDSRFIWSGPKGPDALTGRLKOASAOIKSVHTAD
8. L-A-P1.1      NSPTVMDTLVNILDVDRNFRPKHFKGLROYERSKVTAOHHTHLRPDEVLKAAERVSPRRKYYLLCAVELLATCEVTTEAAVATIMAFVLTLDEKFVTLFLDSRFIWSGPKGPDALTGRLKOASAOIKSVHTAD
9. L-A-D1        NSPTVMDTLVNILDVDRSIRPKHFKGLROYERSKVTAOHHTHLRPDEVLRAAEOVSPRRKYYLLCVVELLAICDVTIEAAVATIMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
10. L-A-P1.7     NSPTVMDTLVNILDVDRSIRPKHFKGLROYERSKVTAOHHTHLRPDEVLEAAERVSPRRKYYLLCVVELLALCDVTIEAAVATVMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
11. L-A-P2.2     NSPTVMDTLVNILDVDRSTRPKHFKGLROYERSKVTAOHHTHLRPDEVLKAAEKVSPRRKYYLLCVVELLASCDVTIEAAVATIMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
12. L-A-P2.6     NSPTVMDTLVNILDVDRSIRPKHFKGLROYERSKVTAOHHTHLRPDEVLKAAERVSPRRKYYLLCVVELLASCDVTIEAAVATIMAFVLTLDEKFVTLFLDSRSIWSGPKGPDALTGRLKOASAOIKSVHTAD
13. L-A-P2.4     NSPTVMDTLVNILDVDRNIRPKHFKGLROYERSKVTAOHHTHLRPDEVLKAAEOVSPRRKYYLLCVVELLAACEVTIEAAVATIMAFVLTLDEKFVTLFLDSRFIWSGPKGPDALTGRLKOASAOIKSVHTAD
14. L-A-2.1      NTPTVMDTLANILDVDRNIRPKHFKGLROYORSKVTAOHHTHLRPNEVLEAAEKVSPRRRYYLLCVVELLANAGIDLEAAVATIMAYVLTLDEKFIPMFLDSRAIWLGDKGPDALTGRLKKASSOIKSVHTAD
15. L-A-C        NTPTVMDTLANILDVDRDIRPKHFKGLREYORSKVTAOHHTHLRPNEVLEAAEKVSPRRRYYLLCVVELLANAGIDLEAAVATIMAYVLTLDEKFIPMFLDSRAIWLGDKGPDALTGRLKKASSOIKSVHTAD
16. L-A-P2.3     NTPTVMDTLANILDVDRNIRPKHFKGLREYORSKVTAOHHTHLRPDEVLAAAERVSPRRRYYLLCVVELLANAGIDLEAAVATIMAYVLTLDEKFIPMFLDSRTIWFGDKGPDALTGRLKKASSOIKSVHTAD
17. L-A-L1 SC    NTPTVMDSLANILDVDRNIRPKHFKGLRLYTRSKVTAOHHTHLRPDELVEAAAKVSPRRKYYLMCVVELLANLOVDLEAAVATILAYVLTLSEKFVPIFLDSRAIWVGEPGPDALTARLKASGOIKSIHTAD
18. L-A-2  SC    NTPTVMDTLANILDVDKSIRPKHFKGLRTYORSKVTAOHHTHLRPDEVLDAATRVSPRRRYYLLCVVELLAACDVXIEAAVATIMTYVLTLNEKFVPLFLDSRTIWRGSPGPEALTARLKKASSOIKSVHTAD
19. L-A-lus SC   NTPTVMDTLANILDVDRNIRPKHFKGLRTYERSKVTAOHHTHLRPDEVLEAAARVSPRRKYYLLCVVELLAACEVNIEAAVATIMTYVLTLDEKFIPLFLDSRTIWOGSKGPEELTARLKKASGOIKSVHTAD

                 940       950       960       970       980       990       1,000     1,010     1,020     1,030     1,040     1,050     1,060
1. L-A-P1.3      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTRLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
2. L-A-P1.4      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTRLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
3. L-A-P1.5      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTRLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
4. L-A-Q         YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVADVDOTRLYSCVRDMFEESKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
5. L-A-P1.6      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTKLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
6. L-A-P2.5      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTKLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
7. L-A-P1.2      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTKLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
8. L-A-P1.1      YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTKLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
9. L-A-D1        YEPLTELFELAVLMNRGVGHVSWKTEREHRERPDVADVDOTKLYSCVRDMFEGSKETYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
10. L-A-P1.7     YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTKLYTCVRDMFEGSKOTYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
11. L-A-P2.2     YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVADVDOTRLYTCVRDMFEGSKOTYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
12. L-A-P2.6     YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTRLYTCVRDMFEGSKOTYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
13. L-A-P2.4     YEPLTELFELAVLMNRGVGHVSWKTEREHREHPDVANVDOTKLYSCVRDMFEGSKATYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOFTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW
14. L-A-2.1      YEPLTELFEMAVLMNRGVGHVSWOAEREHREHPDVALVDONRLYSCVRDMFEGSKOTYNYPYMTWEDYTASRWEWVPGGSVHSOYEEDADYIFPGOFTRNKFITVNKMPKHKISRMISSAPEVRAWTSTKYEW
15. L-A-C        YEPLTELFEMAVLMNRGVGHVSWOAEREHREHPDVALVDONRLYSCVRDMFEGSKOTYNYPYMTWEDYTASRWEWVPGGSVHSOYEEDAEYIFPGOFTRNKFITVNKMPKHKISRMISSAPEVRAWTSTKYEW
16. L-A-P2.3     YEPLTELFEMAVLMNRGVGHVSWOAEREHREHPDVALVDONRLYSCVRDMFEGSKOTYNYPYMTWEDYTASRWEWVPGGSVHSOYEEDAEYIFPGOFTRNKFITVNKMPKHKISRMIASAPEVRAWTSTKYEW
17. L-A-L1 SC    YEPLTELFELAVLMNRGVGHVSWOAEKDHRLNPDVAVVDOARLYSCVRDMFEGSKOTYKYPFMTWDDYTANRWEWVPGGSVHSOYEEDNDYIYPGOYTRNKFITVNKMPKHKISRMIASPPEVRAWTSTKYEW
18. L-A-2  SC    YEPLTELFELAVLMNRGVGHVSWRTEREHREMPDVAKVDORRLYTCVRDMFEGAKOTYNYPYMTWDDYTASRWEWVPGGSVHSOYAEDEEYIYPGOFTRNKFITVNKMPKHKIVRMISSTPEVRAWTSTKYEW
19. L-A-lus SC   YEPLTELFELAVLMNRGVGHVSWKTEREHRENPDVANVNOOALYACVRDMFEGAKOTYDYPYMTWDDYTSSRWEWVPGGSVHSOYSEDDEYIFPGOYTRNKFITVNKMPKHKIARMIASTPEVRAWTSTKYEW

                 1,070     1,080     1,090     1,100     1,110     1,120     1,130     1,140     1,150     1,160     1,170     1,180     1,190
1. L-A-P1.3      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVRHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
2. L-A-P1.4      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
3. L-A-P1.5      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLIAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
4. L-A-Q         GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVRHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
5. L-A-P1.6      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
6. L-A-P2.5      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
7. L-A-P1.2      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
8. L-A-P1.1      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
9. L-A-D1        GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVRHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
10. L-A-P1.7     GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLIAFRDAFFRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
11. L-A-P2.2     GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLIAFRDAFYRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
12. L-A-P2.6     GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFYRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
13. L-A-P2.4     GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLVAFRDAFHRNMSAOOKEAMDWVCESVKHMWVLDPDTKEWYOLRGTLLSGWRLTTFM
14. L-A-2.1      GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLCAFRDAFTRNMSTEOREAMDWVCESVKHMWVLDPDSKEWYKLRGTLLSGWRLTTFM
15. L-A-C        GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLCAFKDAFTRNMSTEOREAMDWVCESVKHMWVLDPDSKEWYKLRGTLLSGWRLTTFM
16. L-A-P2.3     GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEASKVHKRVNMMLDGASSFCFDYDDFNSOHSTASMYTVLCAFRDAFTRNMSTEOREAMDWVCESVKHMWVLDPDSKEWYKLRGTLLSGWRLTTFM
17. L-A-L1 SC    GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEAAKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLCAFRDTFSRNMSDEOAEAMNWVCESVRHMWVLDPDTKEWYRLOGTLLSGWRLTTFM
18. L-A-2  SC    GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEAAKVHKRVNMMLDGASSFCFDYDDFNSOHSIASMYTVLCAFRDTFSRNMSAEOREAMDWVCESVKHMWVLDPDTKSWYELKGTLLSGWRLTTFM
19. L-A-lus SC   GKORAIYGTDLRSTLITNFAMFRCEDVLTHKFPVGDOAEAAKVHKRVNMMLDGASSFCFDYDDFNSOHSISSMYTVLCAFRDAFTRNMSIEOREAMDWVCESVKHMWVLDPDTKTWYOLKGTLLSGWRLTTFM

                 1,200     1,210     1,220     1,230     1,240     1,250     1,260     1,270     1,280     1,290     1,300     1,310     1,320     1,330
1. L-A-P1.3      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMOEAVTAIS
2. L-A-P1.4      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMREAVTAIS
3. L-A-P1.5      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLKDLANRTVMREAVKAIS
4. L-A-Q         NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMREAVTAIS
5. L-A-P1.6      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMREAVTAIS
6. L-A-P2.5      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMREAVTAIS
7. L-A-P1.2      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMREAVEAIS
8. L-A-P1.1      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMRESVSAIS
9. L-A-D1        NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTVMREAVTAIS
10. L-A-P1.7     NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLKDLANRTKVREAVTAIS
11. L-A-P2.2     NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLKDLANRTKVREAVTAIS
12. L-A-P2.6     NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTKVREAVTAIS
13. L-A-P2.4     NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRLRDLANRTMMREAVTAIS
14. L-A-2.1      NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDSMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRIRDLANRTRIPEAVTAIA
15. L-A-C        NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDAMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRIRDLANRTRVPEAVTAIA
16. L-A-P2.3     NTVLNWAYMKIAGVFDIDDVODSVHNGDDVMISLNRVSTAVRIMDSMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOTRVRDLANRTRVPEAVTAIA
17. L-A-L1 SC    NTVLNWAYMKLAGVFDLDDVODSVHNGDDVMISLNRVSTAVRIMDAMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADOARLRDLANRTRVOSAVTAIK
18. L-A-2  SC    NTVLNWAYMKIAGVFDLDDVODSVHNGDDVMISLNRVSTAVRIMDAMHKINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRMCATLVHSRIESNEPLSVVRVMEADKTRLHDLANRTSVTASVTAIE
19. L-A-lus SC   NTVLNWAYKKLAGVFDLDDVODSVHNGDDVMISLNRVSTAVRIMDAMHRINARAOPAKCNLFSISEFLRVEHGMSGGDGLGAOYLSRSCATLVHSRIESNEPLSVVRVMEADKTRLRDLANRTNIKASVTEIE
```



```
                 1,340     1,350     1,360     1,370     1,380     1,390     1,400     1,410     1,420     1,430     1,440     1,450     1,460
1. L-A-P1.3      DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFI
2. L-A-P1.4      DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFIHKSEWERTMYKAYKGLAVSY
3. L-A-P1.5      DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFIHKSEWERTMYKAYKGLAVSY
4. L-A-Q         DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFIHKSEWERTMYKAYKGLAVSY
5. L-A-P1.6      DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFIHKSEWERTMYKAYKGLAVSY
6. L-A-P2.5      NOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFIHKSEWERTMYKAYKGLAVSY
7. L-A-P1.2      DRLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVNTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNPFIHKSEWERTMYKAYKGLAVSY
8. L-A-P1.1      DOLKARVTKIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKONPFI
9. L-A-D1        DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKRNPFIHKSEWERTMYKAYKGLAVSY
10. L-A-P1.7     DOLKTRVTNIFSVEKEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKANPFVHKSEWERTMYKAYKGLAVSY
11. L-A-P2.2     DOLKTRVTNIFSVEREVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKTNPFVHKSEWERTMYKAYKGLAVSY
12. L-A-P2.6     DOLKTRVTNIFSVEREVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKRNPFVHKSEWERTMYKAYKGLAVSY
13. L-A-P2.4     DOLKARVTNIFSVDAEVVTOITRAHRVCGGISTDPWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAK
14. L-A-2.1      SOLNKRVSSVFGVDDNVVIEINRAHRVCGGISTDKWAPVDTKIHTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTI
15. L-A-C        SOLNKRVSSVFGVDDNVVTEISRAHRVCGGISTDKWAPVDTKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAKKNEFIHKSEWERTMYKAYKGLAVSY
16. L-A-P2.3     SOLNKRVSNVFGVDDNVVIEISRAHRVCGGISTDKWAPVDIKIOTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTI
17. L-A-L1 SC    EOLDKRVTKIFGVGDDVVRDIHTAHRVCGGISTDTWAPVETKIITDNEAYEIPYEIDDPSFWPGVNDYAYKVWKNFGERLEFNKIKDAVARGSRSTIALKRKARITSKKNEFANKSEWERTMYKAYKGLAVSY
18. L-A-2  SC    EOLNRRVTSIFGVDRSVVKAIASAHRVCGGISTDMWAPVKTKIKTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVAKGSRNTIALKRKAKITAVTNDYITKSEWERTMYKAYKGLAVSY
19. L-A-lus SC   EOLDRRVTSIFKVDREVVKAISTAHRVCGGISTDPWAPVTTKIKTDNEAYEIPYEIDDPSFWPGVNDYAYKVWONFGERLEFNKIKDAVSKGSRNTIALKRKAKISAVKNDFVNKSEWERTMYKAYKGLAVSY

                      1,470     1,480     1,490     1,500 1,505
1. L-A-P1.3
2. L-A-P1.4      YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
3. L-A-P1.5      YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
4. L-A-Q         YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
5. L-A-P1.6      YANLSKFMSIPPMANIEFGOAR
6. L-A-P2.5      YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
7. L-A-P1.2      YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
8. L-A-P1.1
9. L-A-D1        YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
10. L-A-P1.7     YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
11. L-A-P2.2     YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOIFL
12. L-A-P2.6     YANLSKFMSIPPMANIEFGOARYA
13. L-A-P2.4
14. L-A-2.1
15. L-A-C        YANLSKFMSIPPMANIEFGOARFAMOAALDSSDPLRALO
16. L-A-P2.3
17. L-A-L1 SC    YANLSKFMSIPPMANIEFGOARYAMOAALDSSDPLRALOVIL
18. L-A-2  SC    YANLSKFMSIPPMANIEFGOARFAMOAALDSSDPLRALOVFL
19. L-A-lus SC   YANLSKFMSIPPMANIEFGOARFAMOAALDSSDPLRALOIFL
```