# A model system to study contractions of the $(GAA)_n$ repeats in yeast *Saccharomyces cerevisiae*

A Senior Honors Thesis in Biology

By: Reed Shimberg

Tufts University Class of 2012

Advisor: Dr. Sergei Mirkin

Abstract:

DNA naturally contains many repetitive sequences some of which are dinucleotide, trinucleotide, pentanucleotide and dodecanucleotide repeats. Some repetitive sequences can form unusual structures in the DNA, which can interfere with replication, transcription, and repair processes and cause instability of the DNA. This instability can be detected as expansions or contractions in the length of the original repeats. Many human diseases are caused by the expansion of repetitive DNA. Expansion of a GAA repeat in the first intron of the frataxin (FXN) gene in humans is one such example. When the GAA run expands beyond a length threshold of approximately 70 repeats, *FXN* gene expression is affected, ultimately resulting in Freidreich's ataxia. While extensive research is being carried out on the molecular mechanisms of GAA repeat expansions, relatively few studies focus on their contractions. However, understanding contraction processes is important, as it might help revert the expanded *FXN* gene in diseased individuals and alleviate their symptoms. To understand mechanisms of contractions of GAA repeats, we were interested in identifying genes involved in the contraction process. Due to difficulties of running a genetic screen in human cells, we modeled GAA repeat contractions in a yeast system. This system was used to conduct a genetic screen to identify genes involved in the contraction process. In total, we identified 16 genes, some of which appear to affect the frequency of contractions.

Introduction:

Expandable DNA repeats can cause several genetic diseases in humans. They can be found in introns, exons, promoters and 5' and 3' untranslated regions of different genes and have been shown to be unstable through intergenerational transmission. This instability is often the cause of the disease, each of which is associated with the specific repetitive run. Trinucleotide repeats are naturally present in low to medium copy number in human genes, but their subsequent expansion can cause disease. The instability is thought to be grounded in the ability of these runs to potentially form complex secondary structures that can lead to expansions or contractions in the process of DNA repair, replication, or transcription (Mirkin, 2007). Expandable repeats associated with human diseases can be trinucleotide [$(CGG)_n \bullet (CCG)_n$,$(CAG)_n \bullet (CTG)_n$, $(GAA)_n \bullet (TTC)_n$ and $(GCX)_n \bullet (XGC)_n$ (where X represents any nucleotide)]; tetranucleotide $(CCTG)_n \bullet (CAGG)_n$) pentanucleotide $(ATTCT)_n \bullet (AGAAT)_n$, hexanucleotide $(GGGGCC)_n \bullet (CCCCGG)_n$, and dodecanucleotide $(C_4GC_4GCG)_n \bullet (CGCG_4CG_4)$ (Mirkin, 2007).

This paper concentrates on an expandable trinucleotide repeat, $(GAA)_n$. GAA repeats are naturally present in the first intron of the frataxin gene and exhibit a high degree of length polymorphism. Instability in GAA repeats is caused by the formation of H-DNA, sticky DNA, and slipped stranded structures that can occur in long GAA runs and cause expansions or contractions (Mirkin, 2007). The expansion of $(GAA)_n$ repeats leads to Friedreich's ataxia, the most common form of hereditary ataxia, affecting 1-2 out of 50,000 people (Cossee, 1997).

Friedreich's ataxia (FRDA) is an autosomal recessive disease that was discovered by Nicholaus Friedreich in 1863. It is characterized by progressive gait and limb ataxia, atrophy of the spinal cord, heart disease, diabetes as well as a large number of other clinical features (Marmolino, 2011). It is believed to directly affect Schwann cell viability and proliferation (Lu et al., 2009). Patients show severely reduced levels of a *FXN* gene-encoded mitochondrial protein called frataxin.

The *FXN* gene is located in humans on the long arm of chromosome 9 region 13. It is composed of 7 exons and is 6.5kb-long. Deletion of the *FXN* gene in mice is embryonically lethal (Cossee et al., 2000). The gene is expressed in all human cells with increased levels of frataxin protein seen in heart muscle and the spinal cord (Campuzano et al., 1996).

The frataxin protein in humans has a yeast homologue called *Yfh1* of which crystal structures are available (Dhe-Paganon et al., 2000). The protein is believed to be a mitochondrial matrix iron chaperone, which oxidizes and stores iron and promotes Fe-S cluster assembly. In *S. cerevisiae* the absence of *Yfh1* leads to general growth defects, reduced rate of respiration, loss of mtDNA and high sensitivity to oxidants (Babcock et al., 1997; Foury and Cazzalini, 1997; Koutnikova et al., 1997; Wilson and Roof, 1997).

Alleles of the frataxin gene are classified into categories based on the length of GAA repetitive tract and their subsequent ability to cause disease. Healthy individuals contain 7-38 GAA repeats. This length of the repetitive tract allows for sufficient transcription and translation, allowing adequate levels of frataxin protein to be produced. Between 38 and 65 repeats, the gene is considered to be in the "pre-mutation" stage. This means that there is an extremely high chance that an expansion event will occur, but the gene is still normally expressed. If the repeat exceeds 70 copies, frataxin gene expression goes down, and the diseased phenotype is displayed (Pollard et al., 2004). There are usually 500-1000 GAA repeats in most FRDA patients, but can be as high as 1700 repeats (Sharma et al., 2004). Alleles containing more than 70 GAA repeats are not inherited in a Mendelian manner; rather they follow a phenomenon termed genetic anticipation (Mirkin, 2007). This means that when it is passed down through generations, the phenotype appears stronger with an earlier age of onset in conjunction with the increased length of the repeat.

There are several reports suggesting that long GAA repeats impair transcription. It was also shown that an increase in the GAA repeat length from 30 to 100 units results in a three-fold decrease in the amount of transcript (Krasilnikova et al, 2006). The same inhibition of transcription was shown in HeLa cell extract and prokaryotic systems (Bidichandani et al, 1998). It was suggested in bacteria that RNA polymerase advancing within a long $(GAA)_n$ tract initiates the transient formation of the R-R-Y intermolecular DNA triplex. The non-transcribed $(GAA)_n$ strand falls back and forms a loop in the template and the RNA polymerase is stalled at the distal triplex-duplex junction (Grabczyk, 2000). The observation is that it is the secondary structure of the $(GAA)_n$ repeat rather than protein binding that is responsible for the transcription inhibition.

It is not fully known which mechanism is responsible for expansion or contraction of GAA repeats in individuals with diseased length repeats. Expansion or contraction can occur during replication, repair, or even transcription. We do not know what the causes of expansions are in humans, but studies of GAA repeat expansions in different models have shown a great potential of GAA repeats to cause replication fork stalling. Replication fork stalling was found to be orientation dependent and occurred when the repeat's homopurine strand served as the lagging strand template (Krasilnikova and Mirkin, 2004). It was also determined that knockouts of several genes encoding for stabilizers of replication fork and proteins involved in replication affected the rate of expansion. *TOF1* and *CSM3* were found to increase the rate of expansion while *RAD5, RAD 6* and *SGS1* were found to decrease the rate of expansion of GAA repeats (Shishkin et al., 2009). Several different models to study GAA repeats in human cells and in mice exist (Ditch, et al., 2009) (Al-Mahdawi, et al., 2010). However, yeast appears to be the most convenient system to study GAA repeat instability as they are easy to work with and have the most general information available. This is especially true because it is possible to model large-scale expansions in yeast (Shishkin et al., 2009).

Here, we have designed a yeast system to use for a genetic screen to identify genes involved in GAA repeat contractions. This system takes advantage of the *ADE2* and *SUP4* genes. Briefly, we used the *SUP4-o* gene encoding ochre suppressor tRNA, which was inactivated by the placing of GAA repeats in its natural intron, and an ochre nonsense mutation, *ADE2-1* as a reporter. Large-scale contractions of GAA repeats restore *SUP4-o* functioning, which can be accounted for by the appearance of Ade+ phenotype.

We then used a *LACZ/LEU2* minitransposon based insertion library to generate mutants that could be screened for the appearance of Ade+ phenotype. We have completed over 90% coverage of the yeast genome and uncovered 16 genes some of which increased the rate of contraction of the GAA repeats in the tRNA intron.

Material and Methods:

Plasmid:

All the plasmids used in this study were previously obtained by Robert Matera in our lab. The pRS316-WT*SUP4-o* plasmid (Fig. 1) carries the *SUP4-o* allele of the *SUP4* gene encoding Tyrosine tRNA. The natural intron of the *SUP4-o* was modified such that it has an artificial NruI site. The restriction site allowed us to insert various lengths of $(GAA)_n$ repeats into the intron of the *SUP4* gene (See Fig. 1). Variations of this plasmid were created having different GAA repeat lengths.

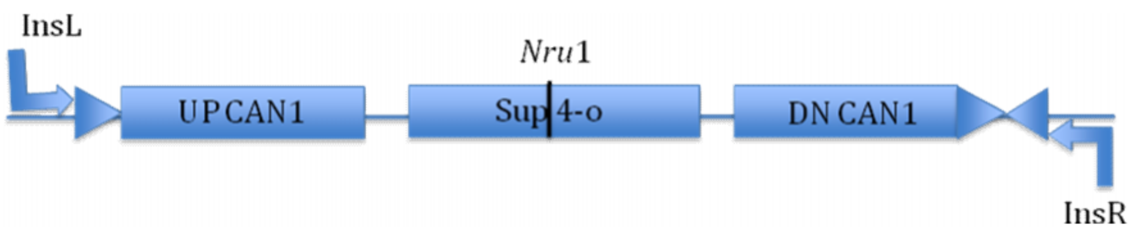


**Fig. 1** Map of the pRS316-WTSUP4-o plasmid. The plasmid contains *can1::SUP4-o* cassette where *SUP4* gene is placed inside the *CAN1* gene. *SUP4* natural intron was modified to contain an NruI artificial site where different number of GAA repeats can be placed inside. Using primers Ins-F and Ins-R, a PCR product containing either *can1::SUP4-o* cassette or *can1::SUP4-o(GAA)n* can be obtained.

Strains:

All the strains used in this study are listed in table 1. SMY600 carries the *URA3-Int-GAA300d* cassette integrated near *ARS306*. This cassette has a selectable *URA3* marker containing the intron from the *ACT1* gene (Shiskin et al 2009) and 300 GAA repeats inside the intron. CH1585 is a derivative of S288C strain (ATCC96098). SMY601-SMY606 strains carry *ade2-1* (UAA) nonsense mutation and a *trp1-289* (UAG) nonsense mutation. *SUP4-o* is an allele of the *SUP4* gene encoding tRNA-Tyr carrying a mutation in the anticodon loop, which makes this tRNA an ochre-suppressor. Strains SMY601 and SMY604 are derivatives of Δleu2-7B-YUNI300 and 9A-YPOM206 correspondingly where *SUP4-o* carrying a modified version of its natural intron was inserted in the *CAN1* locus. To integrate *can1::SUP4-o* cassette into yeast strains we used the PCR product from plasmid pRS316WTSUP4-o with primers Int-F and Int-R. To test the integration we used primers CsmIntCk-F and CsmIntCk-R. SMY602 and SMY605 carry *can1::SUP4-o* cassette with (GAA)$_{31}$ repeats added to the *SUP4-o* gene natural 22 bp intron after an NruI site was added. SMY603 and SMY606 had (GAA)$_{123}$ repeats added to the *SUP4-o* gene natural 22 bp intron.

**Table 1.** List of the name and genotypes of the strains used in the experiments

| Name: | Genotype: | Origin: |
|---|---|---|
| CH1585 | *MATα leu2Δ1 trpΔ 63 ura 3-52 his3-200* | (Smith, et al. 1971) |
| SMY600 | *MATα leu2Δ1 trpΔ 63 ura 3-52 his3-200 75423-75715 (ChrIII)::URA3-Int-GAA300d* | This study |
| Δleu2-7B-YUNI300 | *MATa CAN1 his7-2 Ieu2-3, 112 ura3-Δ trp1-289 ade2-1 lys2-B12* | (Pavlov, Newlon, and Kunkel, 2002) |
| SMY601 | *MATa can1::SUP4-o his7-2 Ieu2-3, 112 ura3-Δ trp1-289 ade2-1 lys2-B12* | This study |
| SMY602 | *MATa can1::SUP4-o::(GAA)32 his7-2 Ieu2-3, 112 ura3-Δ trp1-289 ade2-1 lys2-B12* | This study |
| SMY603 | *MATa can1::SUP4-o::(GAA)123 his7-2 Ieu2-3, 112 ura3-Δ trp1-289 ade2-1 lys2-B12* | This study |
| 9A-YPOM206 | *MATα CAN1 his7-2 Ieu2Δ-0 ura3-Δ trp1-289 ade2-1 lys2-ΔGG2899-2900* strain. | (Baranovskiy et al., 2008) |
| SMY604 | *MATα can1::SUP4-o his7-2 Ieu2Δ-0 ura3-Δ trp1-289 ade2-1 lys2-ΔGG2899-2900* strain. | This study |
| SMY605 | *MATα can1::SUP4-o::(GAA)32 his7-2 Ieu2Δ-0 ura3-Δ trp1-289 ade2-1 lys2-ΔGG2899-2900* strain. | This study |
| SMY606 | *MATα can1::SUP4-o::(GAA)123 his7-2 Ieu2Δ-0 ura3-Δ trp1-289 ade2-1 lys2-ΔGG2899-2900* strain. | This study |

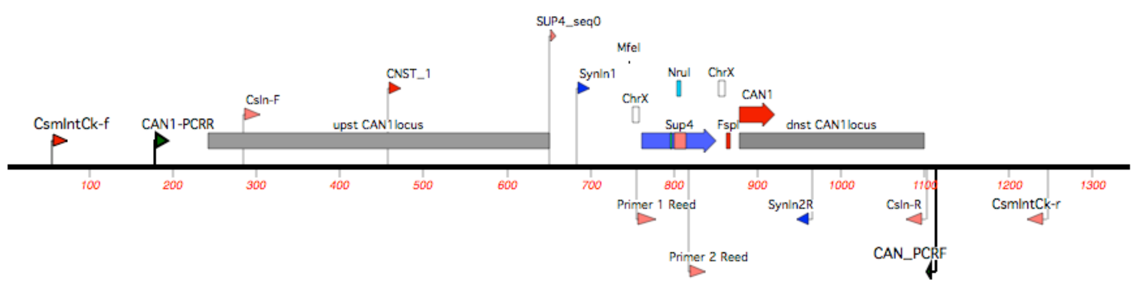


**Fig. 2.** Cassette with 0 GAA repeats showing the location of the primers and the genes.

## DNA Isolation:

DNA was isolated via total DNA prep: the procedure is a modified version from Methods in Yeast Genetics (Amberg et al., 2005). We used 290uL of 0.9M sorbitol, 0.1M EDTA pH7.5 solution and 10uL of 2.5mg/mL Lyticase solution (in 0.9M sorbitol, 0.1M EDTA) and suspended the colonies. After incubation they were spun down and the supernatant was discarded. The cells were then resuspended in 270uL of 50mM Tris pH7.5, 20mM EDTA solution and 20uL of 10% SDS, incubated at 65ºC for 20 minutes and 150uL of chilled 5M Potassium Acetate was added. They were incubated again at 4ºC for 30 minutes and then spun down. The supernatant was transferred to fresh tubes and 900uL of cold EtOH was added and it was incubated at -20ºC for 1 hour. Finally the tubes were spun down, the supernatant was discarded, and the pellet was washed with 70% EtOH. Then it was resuspended in 75uL of TE pH 8.0.

## Colony PCR:

Colonies were taken from Ade- or YPD plates and resuspended in 6uL of 2.5mg/mL Lyticase solution (in 0.9M Sorbitol, 0.1MEDTA) and incubated for 10 minutes at 37 degrees. The cells were spun down at 13,000rpm for 1 minute and then the supernatant was discarded and the pellet was resuspended in 20uL of water. It was boiled for 5 minutes, spun down and 1uL was used as template DNA for PCR.

**Table 2.** List of primers used with corresponding PCR program.

| Primer name: | Sequence | PCR Program |
|---|---|---|
| Ai-R | CTCGAGTCGAGACCTCAAATTCGATGA | 1 |
| Bn-F | CGATGTGCAGAACCTGAAGCTTGATCT | 1 |
| CsmInt Ck-F | GCAAACAGCAAAGGCCACAGAAC | 2 |

| | | |
|---|---|---|
| CsmInt Ck-R | CTAGGTTCGGGTGACGTGAAGATAA | 2 |
| Primer Reed-1 | TAAATACTCTCGGTAGCCAAGTTGGTT | 3 |
| Primer Reed-2 | ATCTTGAGATCGGGCGTTCGACTC | 3 |
| SynIn-1 | GCCGACATAAAGGAAAAGC | 3 |
| SynIn-2 | GAAATGGCGTGGAAATGTG | 3 |
| mTn | CGCCAGGGTTTTCCCAGTCACGAC | 5 |
| UV | CGAATCGTAACCGTTCGTACGAGAATCGCT | 5 |
| AB1 | GAAGGAGAGGACGCTGTCTGTCTGTCGAAGGTAAGGAACGG ACGAGAGAAGGGAGAG | 4 |
| AB2 | GACTCTCCCTTCTCGAATCGTAACCGTTCGTACGAGAATCGC TGTCCTCTCCTTC | 4 |

**Table 3.** List of PCR programs used.

| PCR Program: | Step 1: | Step 2: | Step 3: | Step 4 | Number of Cycles: |
|---|---|---|---|---|---|
| 1 | 94°C 15 seconds | 72°C 165 seconds | --------- | -------- | 31 |
| 2 | 98°C 20 seconds | 60°C 20 seconds | 72°C 150 seconds | -------- | 31 |
| 3 | 94°C 15 seconds | 56°C 20 seconds | 72°C 60 seconds | -------- | 30 |
| 4 | 22ºC 1 hour | 18ºC 8 hours | 16ºC 8 hours | -------- | 1 |
| 5 | 95ºC 15 seconds | 94 ºC seconds | 67 15 seconds | 72ºC 1 minute | 40 |

Repeat Length Determination:

Using Adobe Photoshop, the distance of every distinguishable ladder marker as well as each sample was determined from an arbitrary point on the gel using the measurement feature. A plot of the ladder marker length versus its measured distance was then constructed, and a logarithmic regression was calculated for the data. The equation of the regression was then used to determine the length of sample PCR products based on the measured data.

Library:

We used a *LACZ/LEU2* minitransposon library, a Tn3-derived minitransposon carrying the yeast selectable marker *LEU2* and a bacterial marker encoding resistance to ampicillin (Kumar and Snyder, 2000). The library was obtained from Michael Snyder's laboratory. The library was amplified in XL10-Gold strain. We amplified 8 pools (out of 15), and for each of these pools we got 50,000-200,000 transformants. The cells were collected and DNA was isolated and using Giga-prep DNA purification kit from QIAGEN. Next, we cut the library DNA with NotI and transformed linearized fragments into the yeast strain SMY606.

Gel Electrophoresis:

To separate DNA fragments obtained by PCR a 1.0%, 1.5%, or 2.0% agarose gel in TBE or TAE buffer at $5V/cm^2$ was run. The gel contained ethidium bromide at a final concentration of 5uL/100mL. The picture was taken under UV light 250nm.

Yeast Transformation:

The transformation protocol was a modified Gietz protocol. First, I inoculated a colony (approximately $10^7$ cells) from a plate into 2mL of YPD and grew overnight at 30°C. The next day the inoculum was diluted into fresh YPD media with a total volume of 25mL to get a concentration of $\sim 5x10^6$cells/mL. The cells were grown for 4 hours at

30°C until the cell titer was then ~$2x10^7$ /mL. The cells were spun down at 3000rpm for 2 minutes and the supernatant was discarded. The cells were gently resuspended in 25mL of water. After they were spun down again at 3000rpm for 2 minutes they were resuspended in 1mL of 0.1M LiAc and spun down at 3000rpm for 1 minute. The supernatant was removed again and 400uL of 0.1MLiAc was added and it was spun down again at 3000rpm for 1 minute. Next, 240uL of PEG (MW~3350, 50%w/v), 36uL of 1M LiAc, 10uL of ss-carrier DNA (10mg/mL), 5uL of plasmid DNA, and 69 uL of water were added in that order and briefly vortexed until a homogenous suspension was present. The transformation mixes were incubated at room temperature for 20 minutes and then heat shocked at 42°C for 30 minutes. After that, cells were spun down at 3000rpm for 1 minute, the supernatant was removed and cells were resuspended in 1mL of YPD and grown for 2-3 hours. After that they were plated on the necessary media.

As we did over 200 transformations, we began to modify the procedure to make it more efficient. In this way we created a master mix with 240uL of PEG, 36uL of 1MLiAc, 10uL of boiled carrier DNA, 10uL of insert, and 44uL of water. This could be multiplied to correspond to the number of transformations necessary. Once this was aliquotted into eppendorfs, the yeast were spun down at 3000rpm, washed with water, spun down again, and then resuspended in 1ml of 0.1M LiAc. They were spun down and resuspended in 200uL of 0.1M LiAc of which 20uL was added to the tubes with the aliquotted master mix. Then the same steps as before were taken with 20 minutes at room temperature followed by 25-30 minutes in a hot water bath at 42°C. Finally, it was spun down and 1mL of YPD was added and it was grown for 2-3 hours before being plated on media containing 1% geneticin.

Bacterial Transformation:

Chemically competent cells were thawed in an ice bath and 3uL of plasmid DNA was added to approximately 100uL of the thawed cells and mixed. The mixture was incubated on ice for 20 minutes and then heated to 42°C for 50 seconds. They were then immediately put back on ice for one minute. Afterwards, 400uL of pre-warmed LB broth was added and incubated at 37°C for an hour. The cells were then plated on media containing ampicillin.

MidiPrep:

The Qiagen MidiPrep kit was used to extract the plasmid from the bacteria.

Restriction Digest:

The pRS316-Sup4-o-GAA123 plasmid was digested with SpeI and KpnI. 40uL of plasmid, 5uL of each fast digest enzyme, 10uL of fast digest buffer, and 40uL of water were mixed and kept overnight at 37°C.

To prepare the DNA for vectorette PCR 1uL of fast digest AluI, 2uL of fast digest buffer, 10uL of sample, and 7uL of water were kept at 37°C overnight.

Gel Extraction:

The Fermentas Gene Jet Gel Extraction kit was used to purify the insert from the gel and prepare it for the transformation.

Steps for further analysis:

The DNA was digested with AluI and was purified using Qiagen PCR purification kit. To prepare the DNA for vectorette PCR, anchor bubbles had to be ligated to the ends of the DNA fragments. The following mix was used: 8uL of DNA, 0.2uL of AB1-AB2 (a stock oligo), 2uL of ligase buffer, 0.8uL of ligase (400,000 u/mL), and 9uL of water were added and PCR program 4 was run overnight (Table 2 and 3). A vectorette PCR (program 5) was run with primers mTn and UV (table 2 and 3). This amplified the sequence between the minitransposon and the added oligo. EXOSAP-IT was used to clean up the PCR product: 2uL of ExoSap was used in conjunction with 5uL of the PCR product.

Sequencing:

All DNA sequencing was done at the University of Chicago. The results were BLASTed on the Yeast Genome site to determine the identity of the gene where the minitransposon inserted.

Results:

In order to perform a screen to identify genes involved in GAA repeat contraction, we took advantage of a natural *SUP4* intron, which is only 22 bp-long. *SUP4* is a gene encoding tyrosine tRNA. Importantly, mutations in different tRNA genes were selected as nonsense suppressors. *SUP4* is one of these genes and the dominant *SUP4-o* allele behaves as an ochre suppressor. The G → T transversion in the 53$^{rd}$ position of the *SUP4-o* allele makes the anticodon loop of this tRNA completely complementary to the UAA stop codon. Another component of our new system is the *ade2-1* allele, which has a premature UAA (ochre) stop codon. Since *SUP4-o* is a gain of function mutation, the tRNA encoded by *SUP4-o* will suppress the *ade2-1* (UAA) nonsense mutation, making possible synthesis of the full-length Ade2. Normally tRNA genes have short introns, which cannot tolerate a significant change in length or sequence. Therefore, introducing a long GAA repeat in the *SUP4-o* intron should switch off the function of the gene. Importantly, the function will not be restored until the repeat significantly contracts. We introduced a unique *Nru*I restriction site into the *SUP4-o* intron and inserted GAA repeats of two different lengths in it: 32 GAA and 123 GAA repeats. The introduction of the NruI site alone into *SUP4-o* did not affect the suppression efficiency in our system. When GAA repeats were inserted, we observed no nonsense suppression (Matera, unpublished data) (See system scheme Fig. 3A and 3B).

*ADE2* encodes for phosphoribosylaminoimidazole carboxylase, which catalyzes the sixth step in the *de novo* biosynthesis of purine nucleotides (Jones, EW and Fink, GR, 1982). Mutations in the gene lead to adenine-auxotrophy and accumulation of purine precursors in the vacuole that cause the yeast color to be red (Zonneveld BJ and vander Zanden, AL, 1995). When the *ade2-1* mutation is manifested in the cells, they appear red due to the accumulation of the purine precursor. When the mutation is suppressed, the cells appear white. The frequency of GAA repeat contractions can be determined based on the color of the colonies or the growth on Ade- media (see Fig. 3A and 3B for details about the selection scheme).

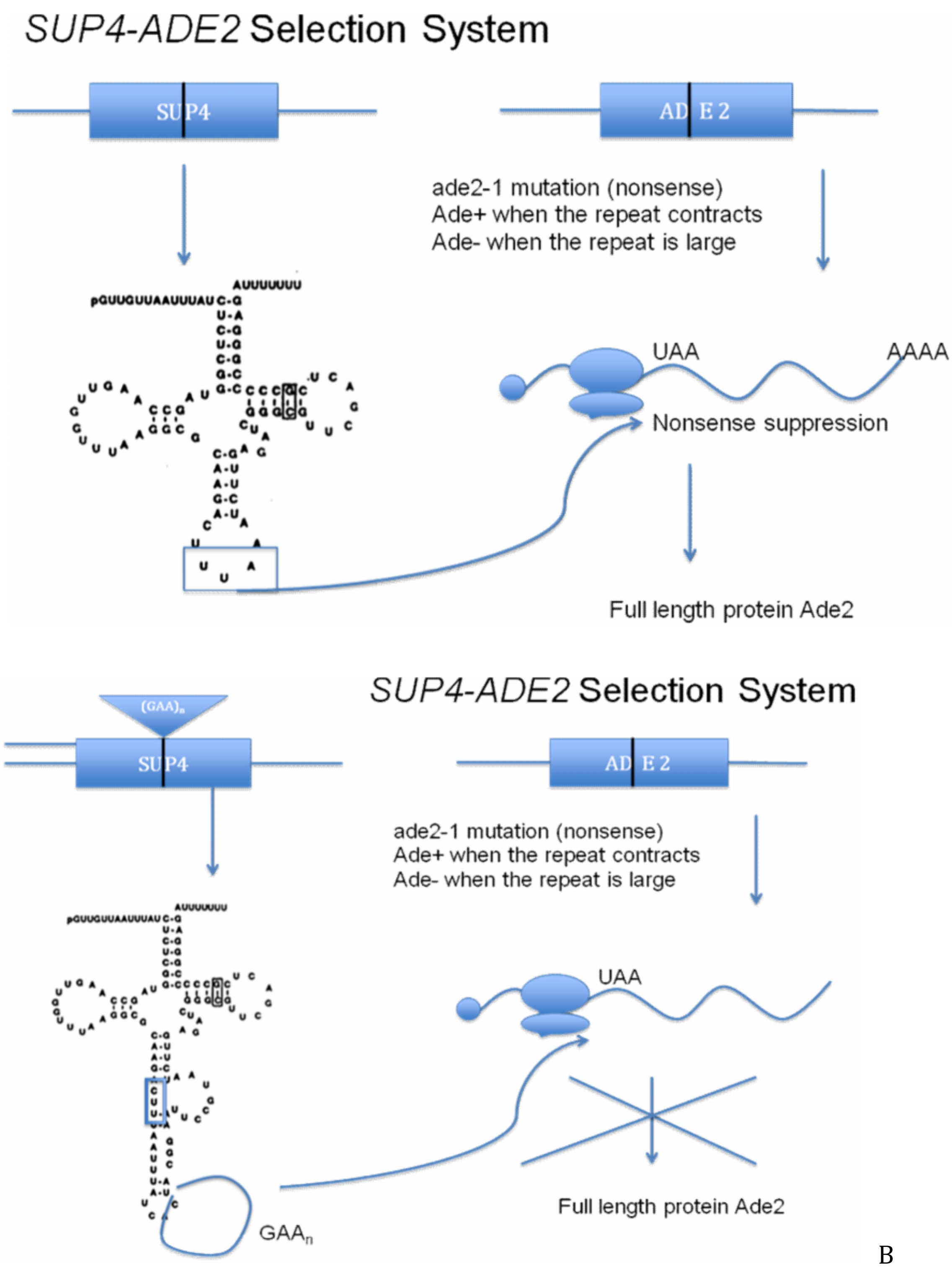


**Fig. 3A**: A representation of the system. The intron has been properly spliced out and UUA is the proper anticodon. When there is less than 10 GAA repeats in the intron of the *SUP4-o* nonsense suppression is manifested and a full length Ade2 protein is synthesized. 3B: Longer GAA tracts (more than 10 GAA repeats) affect tRNA maturation and splicing and this impairs nonsense suppression. The long repeat makes the incorrect anticodon in the proper location.

When 32 and 123 GAA repeats were placed into the intron of the *SUP4-o* tRNA gene, the intron became too large to be properly spliced out. This abolished suppression of the *ade2-1* mutation and hence cells were unable to grow on media lacking adenine. However, the GAA repeat can contract massively, such that only a few GAA repeats are left in the *SUP4-o* gene, which should be enough to restore suppression of the ade2-1 mutation, making the cells Ade+. Indeed when we plated SMY602 (32 GAA repeats) and SMY603 and SMY606 strains which contain 123 GAA repeats, we observed the appearance of white Ade+ colonies. The rate of these colonies was different depending on the strain and the number of GAA repeats (Table 4).

**Table 4.** Contraction frequency of GAA 32 and 123 repeats.

| Strain: | Number of initial Repeats | Rate of Contraction (CI) | Reference: |
|---|---|---|---|
| SMY602 | 32 | $7.8 \times 10^{-4}$ ($2.41 \times 10^{-4}$-$2.56 \times 10^{-3}$) | Robert Matera, unpublished data |
| SMY606 | 123 | $3.39 \times 10^{-6}$ ($2.97 \times 10^{-6}$-$3.87 \times 10^{-6}$) | This study |

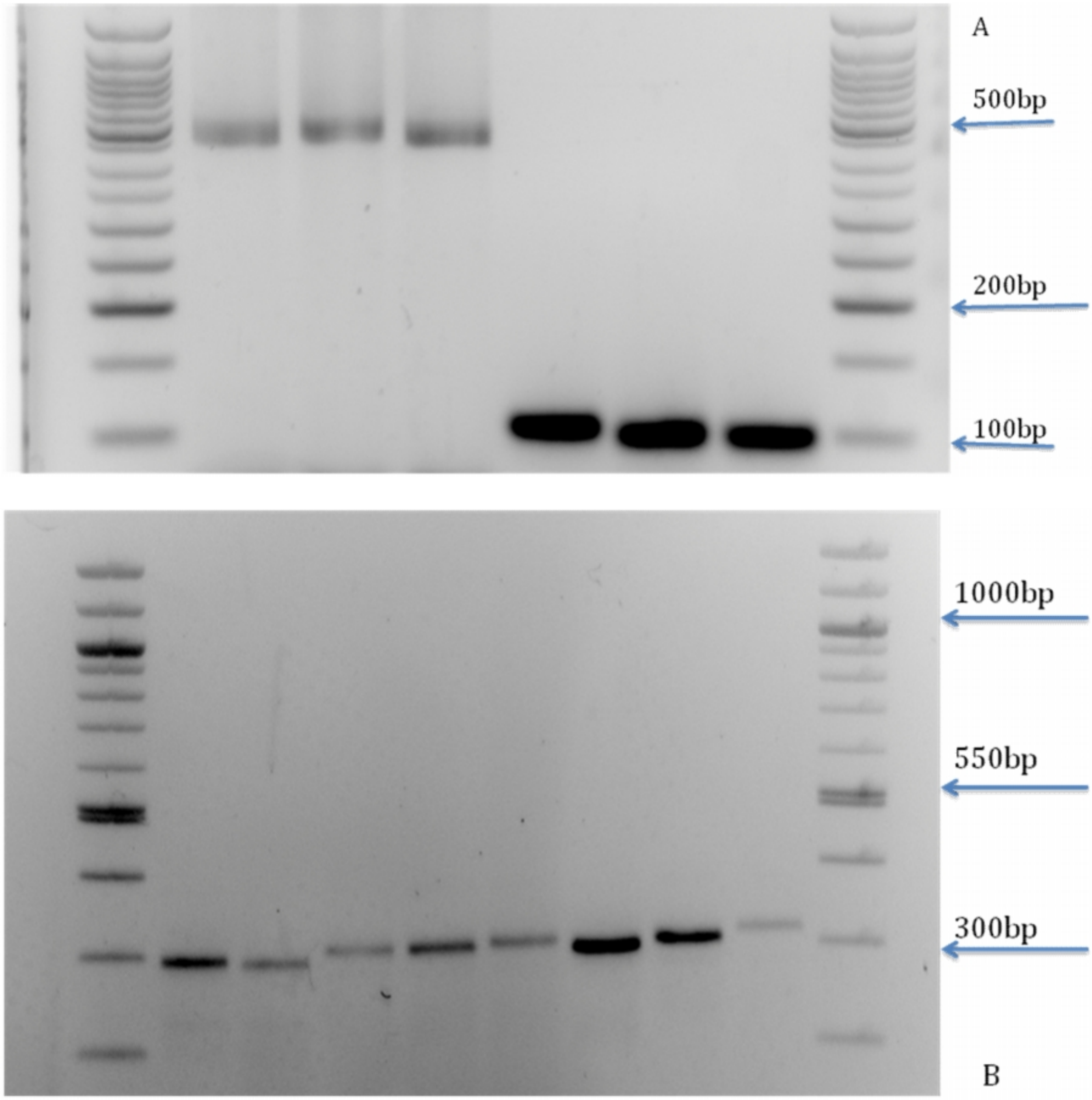


**Fig. 4.** 5A shows both the full length and contracted length of the GAA-123 repeat. The DNA was isolated from Ade+ clones and a PCR was run using primers PR-1 and PR-2. This is a 2.5% TBE agarose gel with a 50 bp ladder. The 3 lines on the left are the uncontracted length while the three on the right are the contracted length. These primers sat much closer to the repeat (Material and Methods). 5B shows only the contracted GAA repeat length from the same starting length as A. The DNA was isolated and a PCR was run using primers SynIn-1 and SynIn-2. This is a 2.5% TBE gel with 100bp ladder.

We then analyzed all the colonies, which grew on Ade- media by PCR using different primers in order to determine the length of the GAA repeat remaining in the *SUP4-o* intron. As mentioned previously, a significant contraction had to occur to allow these colonies to grow on Ade- media. Indeed, we found that all of the colonies carried a contracted version of the GAA repeat (See Fig. 4). We analyzed 51 colonies from SMY606 strain by PCR, with primers SynIn-1 and SynIn-2 (Material and Methods)

As seen in table 4 the contraction frequency for strain SMY606 is $3.39*10^{-6}$. This is two orders of magnitude lower than that of SMY602, which has a contraction frequency of $7.8*10^{-4}$. SMY606 had a longer GAA repeat (123) initially inserted into the intron and it contracts less frequently than the short GAA run, SMY602, with 32 GAA repeats initially placed into the intron. Such a difference in the contraction rate can most likely be explained by the requirement for multiple contraction steps for a longer repeat to reach the threshold length for intron splicing. The repeats contract to slightly varied lengths but within a strict limit of 5-10 GAA repeats remaining in the intron. Note that SMY606 is a derivate of 9A-YPOM206 strain and SMY602 is a derivate of Δleu2-7B-YUNI300 strain.

Based on the contraction rate of SMY606 we were able to begin the genetic screen. We used a *LACZ/LEU2* minitransposon insertion library (Fig. 6) (Kumar and Snyder, 2000). It is a transposon-based insertion library that allows us to rapidly generate a large set of mutations that can be screened for mutant phenotype and specific patterns of gene expressions. The library was obtained by mutagenizing a yeast genomic library with an mTn transposon in vivo in *E. coli* (Huisman et al., 1995). Mutant alleles are subsequently transferred into yeast via transformation and subsequent homologous recombination. The mTn transposon contains a *LEU2* reporter gene (for selection) and it contains *LACZ* reporter gene lacking its start codon and promoter allowing us to obtain fused genes with *LACZ* (Kumar and Snyder

2000).

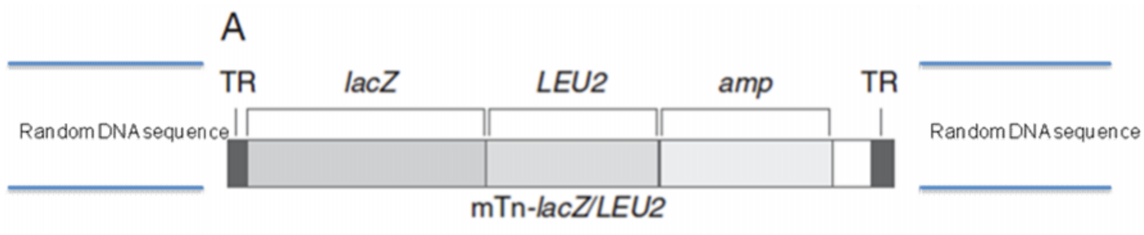


**Fig. 5.** *LACZ/LEU2* minitranspson Tn3-derived minitransposon capable of generating disruption alleles and reporter gene fusions. Proteins tagged by *LACZ/LEU2* minitransposon insertion may be immunolocalized using antibodies directed against Beta-gal. The minitransposon inserts into random yeast strains.

The library contains 15 independent pools of DNA, 8 of which were amplified in our lab (see material and methods) and 4 pools were used to transform the SMY606 strain. For the transformation, the library DNA was cut with *Not*I to get linear fragments containing mTn transposon cassette flanked by yeast DNA sequences. This represents approximately ≥90% coverage of the entire yeast genome.

We used Leu-, low adenine media to select transformants. We used 1% from a normal concentration (20mg/L) of adenine to allow all cells to undergo a few divisions. This would increase chances to select contraction events, which were detected based on a white or pink color. This differed from the uncontracted GAA repeats, which appeared red on the plate (Fig. 6).

The results for the number of clones that show a contraction in the GAA repeat length are presented in table 5.

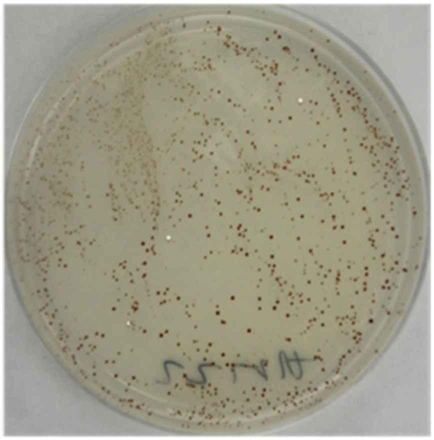

**Fig. 6**. A picture of a Leu- low adenine plate. The 3 off white colonies represent a contraction in the GAA repeat length in the *SUP4-o* intron.

Table 5. Number of clones from the first 4 pools analyzed.

| Pool Number | Number of Clones (pink or white) | Total amount of transformant analyzed |
|---|---|---|
| 21 | 36 | 26030 |
| 22 | 32 | 27075 |
| 23 | 60 | 24820 |
| 38 | 48 | 26100 |

Each clone was analyzed by PCR and gel electrophoresis to ensure that the repeats had actually contracted. However, it can be assumed that not all of the clones contracted as a direct result of the minitransposon interrupting a gene involved in repeat contraction. For this reason, each clone was further analyzed to see that it would contract again, thus attempting to ensure that the inactivated gene was the cause of the contracted repeat.

We constructed a plasmid with 123 $(GAA)_n$ repeats and the *KanMX* gene. To do this we cut out the *KanMX* gene out of the pFA6-KanMX4 plasmid. We ligated the insert into the pRS316 plasmid that already had the $(GAA)_{123}$ repeats within the *Sup4-o* gene to get the cassette for integration. The KanMX gene allows yeast to be grown on selective media with geneticin (G418). When we inserted the new plasmid into the mutagenized strain from which we got the original white colony, we could see if the repeats contract again. If it does contracts again we can assume that it is a direct result of the minitransposon inactivating a gene involved in the contraction process.

The yeast cells were plated on G418 selective media and the phenotype was analyzed. In the case where all of the yeast cells were white, we speculated that it was due to the minitransposon having inactivated a gene in the biosynthesis pathway of the red pigment or affecting its accumulation. This would make all cells appear white but they would still have the uncontracted repeat length. Because we suppose the minitransposon did not inactivate a gene that was directly linked to the contraction process, these clones were not analyzed any further.

If the yeast were all red, we believed that the initial contraction did not occur as a result of the inactivated gene but possibly by random chance or an unrelated mutation. Because the repeats did not contract again, these clones were thrown out and no further analysis was done.

However, if the plate showed a variety of red, white, and pink yeast colonies we assumed that the minitransposon had inactivated a gene in the contraction process and further analysis was necessary to determine which gene the minitransposon had interrupted.

**Table 6.** 150 clones were retransformed with the original repeat and the KanMX gene.

| Number of clones retransformed | Number of clones with all white phenotype | Number of clones with all red phenotype | Number of clones with red, white, and pink phenotype |
|---|---|---|---|
| 150 | 26 | 14 | 110 |

Of the 150 clones that were retransformed, 26 were all white and 14 were all red (Table 6). For the approximately 110 clones that showed a red/pink/white phenotype, the chromosomal DNA was isolated and digested with *Alu*I. *Alu*I is a blunt end restriction endonuclease that possesses a 4-base pair recognition sequence and cuts many times throughout the yeast genome. The DNA fragments were then ligated with DNA duplex adapters, which were generated by annealing two oligonucleotides (AB1 and AB2). This generated a structure with blunted double strand ends and bubble formed by imperfect basepairing. By ligating adapters to the digested chromosomal DNA we could PCR the sequences between the minitransposon and the oligonucleotide adapter (Fig. 7). The gene where the minitransposon inserted could then be determined.

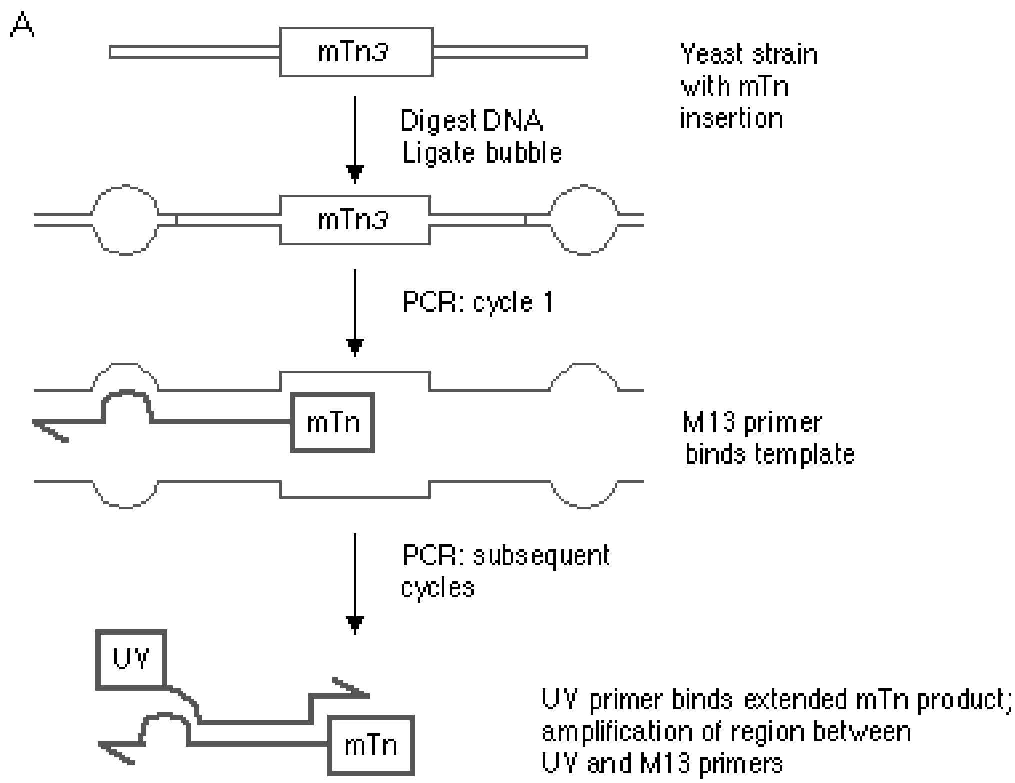


**Fig. 7**. Vectorette PCR. Shows how the ligating of the bubbles and the use of the two different primers allow the sequence of the gene with the mTn insert to be determined. Sequences of the ligated bubbles (AB1 and AB2) and the primers (mTn and UV) can be found in table 2.

The PCR was run out on a gel and any lane that had a band was sent for sequencing. When looking at the sequence with the mTn primer, we could determine if it was successful by looking for the sequence of the *LACZ*/*LEU2* minitransposon. If the sequence contained the junction with the *LACZ/LEU2* minitransposon adapter we would BLAST the part of the sequence following the adapter on the Yeast Genome website (http://www.yeastgenome.org/cgi-bin/blast-sgd.pl) to determine the point of insertion.

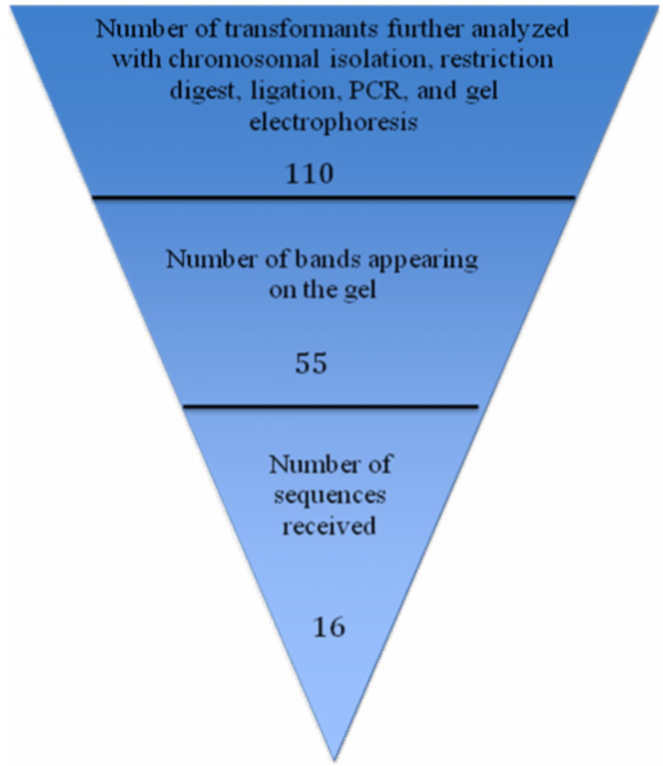


**Fig. 9.** Number of Transformants analyzed at each step.

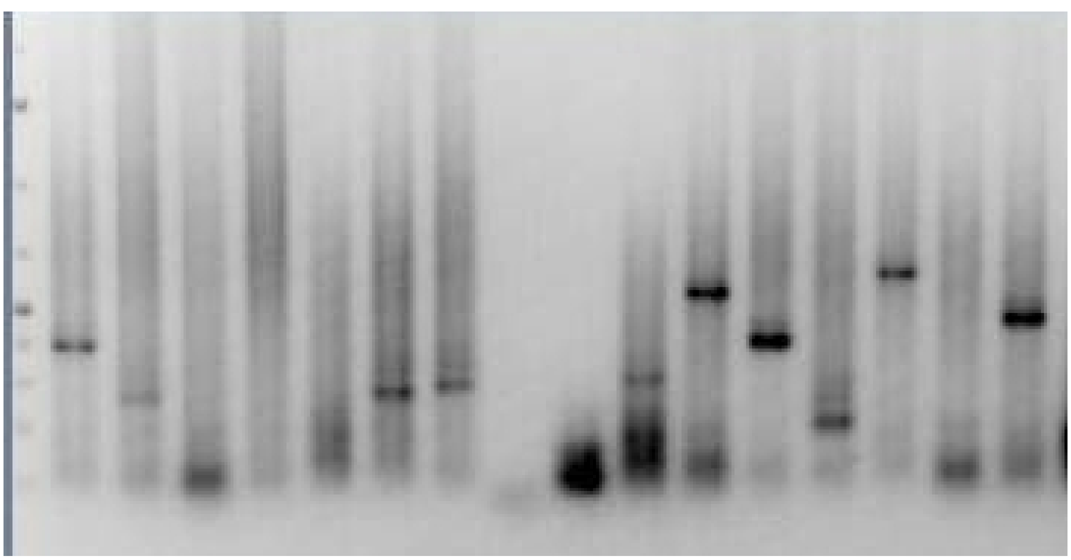

**Fig. 9.** 1% agarose 1X TAE gel with a 1kb+ ladder run at 90V. The black bands scattered across the lanes represent sample that would be sequenced due to the minitransposon inactivating a gene.

So far, there are 16 genes, one Ty element, and 30 random ribosomal DNA sequences that were identified after the results of the sequencing had been analyzed. The genes can be seen in Table 8. Each of these genes only appeared once in the screen.

**Table 8**. Each of these genes was identified when the samples were sent for sequencing.

| Genes: | BLAST hit location | Function |
| --- | --- | --- |
| YJR151W-A | In the gene | Protein-RNA complex assembly and mRNA processing |
| RGC1 | In the gene | Pleckstrin homology domain |
| TOF2 | In the gene | Condensin recruitment to the replication fork barrier site |
| MAP2 | In the gene | Methionine aminopeptidase |
| YND1 | In the gene | Apyrase (catalyzes hydrolysis of ATP) |
| RAD17 | In the gene | DNA damage repair |
| CRP1 | Insertion in the ARS sequence adjacent to the gene | Protein that binds to cruciform DNA structures |
| STH1 | Within 50 bp of the gene | ATPase component of the RSC chromatin remodeling complex |
| ENV7 | In the gene | Protein serine/threonine kinase activity |
| ACS1 | In the gene | Acetyl coenzyme-A synthetase |
| HAL5 | In the gene | Putative protein kinase |
| CYC3 | 200 bp away | Cytochrome c heme ligase |
| SWT1 | In the gene | RNA endoribonuclease |
| BSC2 | In the gene | Not fully known |
| VMR1 | In the gene | Vacuolar membrane protein |

Discussion:

GAA repeats are naturally present in the first intron of the human frataxin gene and can cause Freidreich's ataxia when expanded beyond the threshold length of 70 repeats. Long GAA repeats are a potent inhibitor of transcription. Therefore, humans who carry alleles with long GAA tracts exhibit a diseased phenotype. There is no known treatment for the disease. There can be two different directions in the research to find a treatment for these patients. One would be to find a way to prevent expansions of GAA repeats and the other one is to stimulate contractions. We hope that our study would be useful to find drugs, which would stimulate contractions. In addition, knowing which genes stimulate contractions would enhance our knowledge on instability of GAA repeats. No previous studies have been done on large-scale GAA repeat contractions. This study represents the first example of simulating a large scale GAA repeat contraction model suitable to perform a genetic screen.

A previous model used in our lab to study GAA repeat contractions revealed many difficulties including a high frequency of GAA repeat contractions. To avoid any difficulties associated with high contraction rate, we employed a novel system based on two genes, *SUP4* and *ADE2.* The *SUP4* gene encodes for tyrosine tRNA which is normally a latent nonsense suppressor due to wobble interaction in the third position of a codon. *SUP4-o* allele of this gene represents a potent ochre suppressor. We have chosen this tRNA, which has a natural intron to insert GAA repeats such that a long GAA tract would completely inhibit maturation and splicing of this tRNA. Using a reporter *ADE2* gene carrying a nonsense mutation (*ade2-1)* we are able to screen for contraction events. The contraction frequency of 123 GAA repeats was low enough to allow us to begin a genetic screen. To conduct this genetic screen we chose to use an mTn *LACZ/LEU2* library. Interestingly this library was previously used in another genetic screen to find genes involved in CAG repeat contractions. In this screen deletion of *MRC1, TOF1*, and *CSM1*, genes were found to increase contraction frequencies (Razidlo and Lahue, 2008).

This library was selected because it tags loci with an easily detectable DNA element. It inserts relatively randomly into the DNA so we can identify not only nonessential but possibly even essential genes involved in the process of GAA repeat contraction (Kumar and Snyder, 2000).

With each pool of the library DNA that we transformed into the SMY606 strain we found 30-60 isolates that arose due to a presumed contraction event. The full library should yield approximately 180,000 yeast transformants that should be screened for reporter activity in order to ensure 95% coverage of the genome (Kumar and Snyder, 2000). We screened approximately 100,000 transformants

The clones that show a contracted repeat were further analyzed to confirm that the minitransposon did indeed inactivate a gene that directly affected the contraction process. To do this we inserted a plasmid with 123 GAA repeats to see if the system would contract another time. We isolated 16 different genes, which we presume are involved in the instability of the repeat, the red pigment biosynthesis, or translation of our system.

We subdivided the 16 genes into 3 distinct groups based on the information in the yeast database and our extrapolation of their role in cells to understand the role they play in our system. It is important to note that this is a hypothetical division of categories due to the fact that not all functions of these genes are known or confirmed.

One group of genes seems to have affected the synthesis of the red pigment specifically affecting the mitochondria or vacuole. We have determined that *CYC3, ENV7, YND1, HAL5, and VMR1* are in this group. *CYC3* attaches heme to apo-cytochrome in the mitochondria. Affecting the mitochondria could be linked to the red pigment manifestation making some clones exhibit a white phenotype. *ENV7* is involved in vacuolar function, where the red pigment accumulates. Mutants of *ENV7* show defect in CPY processing. *YND1* is an apyrase with wide substrate specificity. It is able to hydrolyze nucleoside tri- and diphosphates. It is important in glycosylation. *HAL5* is a putative protein kinase that when disrupted increases cation and low pH sensitivity and impairs potassium uptake. *VMR1* is a vacuolar membrane protein. The disruption of any of these genes would not seem to have a direct affect on the stability of a long repeat and hence we have speculated that they played a small, irregular role in disrupting the synthesis or manifestation of the red pigment (Yeast Genome Database).

The next group of genes involves those, which affect translation and are significant due to the fact that our system involves nonsense suppression. We have speculated that *MAP2* and *BSC2* are showing up in our screen due to their role in translation and their possible effect on the nonsense suppression. *MAP2* catalyzes the cotranslational removal of N-terminal methionine from nascent polypeptides. *BSC2* is a protein of unknown function; its ORF exhibits genomic organization compatible with a translational read through-dependent mode of expression (Yeast Genome Database).

It is hard to determine which groupYJR151W-A fits in because it is not a fully characterized gene and was only recently determined to be a gene. However, based on studies that show protein interaction with YJR151W-A, it is believed to have a role in both protein-RNA complex assembly and mRNA processing. It interacts with at least 37 different other genes, including *TIF* and *TAF*, two universal transcription factors. Based on its interaction it could also be thought to affect transcription initiation. It also is believed to interact with some splicing factors, which could imply its role in affecting the tRNA splicing (Roy et al., 2009). This gene is truly unique and its function needs to be further determined before it can be properly placed in a category.

The final group of genes includes those that are likely involved in the instability of the repeat. We presumed that 7 of the genes that showed up in our screen played a role in the instability of the repeat. Due to the fact that they appear to be a somewhat scattered group of genes it would be difficult to propose a mechanism in which many or all of the genes were involved but I will speculate as to a possible role that each individual gene may have played in the contraction process.

*CRP1* is a protein that binds to cruciform DNA structures and has autoproteolytic activity. It works at three and four way DNA junctions that do not appear to be site specific. They are most commonly found at replication and recombination sites (Rass and Kemper 2002). GAA forms a hairpin, a weak slipped stranded structure that shows up as a three-way junction, and a recombination junction could interact with *CRP1*. It is possible that the *CRP1* can identify this structure and repair it.

*RGC1* is a pleckstrin homology domain containing protein that is involved in intracellular signalling. It plays a role in glycerol channel activation. *RGC1* mutants have cell wall stress and a build up of osmotic pressure that can ultimately lead to cell lysis. They also have a large build up of glycerol. It is important to note that humans do not have a homologue of this protein (Beese et al., 2009). I believe the influx of glycerol could have an affect on the transcription process, possibly upregulating transcription and increasing the chances of a contraction event occurring.

*STH1* is a gene with two distinct functions. On one hand it is an ATPase component of the RSC chromatin remodeling system, yet it also has helicase related function in transcriptional activation. It is part of the $G_2$/M checkpoint and affects the transcription of genes involved in this checkpoint (Laurent et al. 1992) (Du et al., 1998).

*TOF2* primarily binds to toposiomerase 1. Mutants of *TOF2*, however, do not affect topoisomerase function (Park and Sterglanz 1999). *TOF2* also is required for rDNA silencing and mitotic rDNA condensation. It clamps rDNA to a cohesin ring and restricts movement of rDNA sister chromatids. It has a direct link to chromosome segregation during meiosis I (Huang et al., 2006). Mutants of *TOF2* show delayed rDNA silencing with little overall effect on other cellular processes, however. Mutants also show impaired relocalization of condensin during anaphase. *TOF2* is also believed to stimulate Cdc14 phosphatase activity, which is required for mitotic exit. Disruption during mitosis or delay in rDNA silencing could affect the stability of the repeat and help induce contractions.

*SWT1* is an RNA endoribonuclease involved in prenuclear mRNP quality control via the turnover of aberrant, unprocessed pre-mRNAs. It is thought to play a role in transcription elongation and early steps in mRNA export. Deletion of *SWT1* leads to decreased transcription. It functions with the TREX complex, which connections transcription, mRNA packaging, and nuclear export, for efficient gene expression and is synthetically lethal when knocked out with the TREX complex (Rother et al., 2006).

*RAD17* is a protein involved in the DNA damage repair pathway (Parker et al., 1998). It sends arrests signals by associating with single stranded DNA in mitosis and meiosis (Lyndall et al., 1996). It forms a complex with *MEC3* and *DDC1* that shows similarities to the PCNA clamp. The role of this clamp is to search for DNA damage (Majka and Burgers 2003). Mutants of *RAD17* show delays in repair of double stranded breaks (Shinohara et al., 2003). Also, increased contractions of CAG repeats and increased chromosomal fragility was found in *RAD17* mutants.

*ACS1* is a gene that encodes for acetyl Coenzyme-A synthetase. It is a nuclear source of acetyl-coA for histone acetylation. It is important to growth under aerobic conditions which implies its working with the mitochondria (van der Berg et al., 2006).

It was determined that over 30 of the sequences were random ribosomal DNA. It seems like the minitransposon we used had an affinity for integrating in this region. It is unclear as to why this is possible. When these sequences were blasted against the entire genome it was clear that they came from ribosomal DNA. We also saw one instance of Ty come up in our screen. Ty is a transposable element that resembles a retrovirus and codes for an intermediate protein necessary in the transposition process.

Many of the transformations did not yield any results. I believe that future work on this project could find more genes simply by further analyzing the genes we have already determined show a change in contraction frequency. While identifying 16 genes should be considered a success, it begs the question of how many more can be isolated?

We expected to find genes involved in transcription, replication and repair when we complete our genetic screen. Formation of secondary structures, including H-DNA, sticky DNA, and slipped-stranded structures for GAA repeats, is common with repetitive DNA. These abnormal structures often disrupt the replication process and significantly damage the repetitive run. One such run is CAG repeats that form hairpin loops that are unstable during transcription, replication, and repair (Lenzmeier and Freudenreich, 2003).

Because this is the first major study on large scale GAA repeat contraction, we have no results to compare to ours. However, similar studies have been done involving GAA repeat expansions. In expansions, it has been shown that fork stabilizers such as *Tof1* and *Csm3* and DNA repair proteins such as *Rad5*, *Rad6,* and *Sgs1* affected the rate of expansion (Shishkin et al., 2003). None of these genes appeared in our screen, which could imply that the processes of contraction and expansion are very different.

There are many steps to be taken before this project will be over. First off, I think it is necessary to complete the analysis of the current strains I have determined to contract. It will be necessary to re-isolate their chromosomal DNA and the redo the proceeding steps. After analysis of all transformants from the first four pools is complete, another group of pools containing the minitransposon should be transformed into the strains. The transformants should be analyzed. Once a list of genes is completed, individual genes should be disrupted to determine the actual change in repeat contraction frequency that they cause. With the inclusion of further data, it is possible that a mechanism could be created. At this point our data suggests that a basic mechanism would involve transcription related instability for contractions.

**Acknowledgement:**

I cannot begin to adequately thank the members of the Mirkin lab for their invaluable help with my research. Primarily to Dr. Sergei Mirkin, for giving me this opportunity and believing that I could be a scientist. To Robert Matera A'10 who introduced me to this research and prepared me with the skills necessary to conduct research on my own. To Dr. Anna Aksenova, who was the best post-doc I could have asked for. She was always there to lead me in the right direction and her guidance has been exceedingly helpful. Dr. Jane Kim has been a resource for me since the day she arrived in lab and has been so kind to help answer all of my questions when I was confused. Finally, to Kartik Shah who was always there to help me with experimental technique. Thank you to you all.

## Literature Cited


Al-Mahdawi, S et al., GAA repeat expansion mutation mouse models of Freidreich ataxia exhibit oxidative stress leading to progressive neuronal and cardiac pathology. Genomics November 88(5):580-590. (2006).

Amberg, D., et al., Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual. Cold Springs Harbor library press, New York. 2005.

Babcock, M., de Silva, D., Oaks, R., et al., 1997. Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. Science 276, 1709-1712.

Baranovskiy et al. X-Ray structure of the complex regulatory subunits of human DNA polymerase δ. *Cell Cycle*. 2008 October; 7(19): 3026-3036).

Beese SE, *et al.* (2009) Identification of positive regulators of the yeast fps1 glycerol channel. *PLoS Genet* 5(11):e1000738

Bidichandani SI, Ashizawa T, Patel PI. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am J Hum Genet.* 1998; 62: 111–21.

Campuzano, V., Montermini, L., Molto, M.D., et al., 1996. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271, 1423-1427.

Cossee, M, Schmitt, M, Campuzano, V, Reutenauer, L, Moutou, C, Mandel JL, Koenig, M. Evolution of Friedreich's ataxia trinucleotide repeat expansion: founder effect and permutations *Proc. Natl. Acad. Sci. U.S.A.* 94: 7452-7457 (1997).

Cossee, M., Puccio, H., Gansmuller, A., et al., 2000. Inactivation of the Friedreichs ataxia mouse gene leads to early embryonic lethality without iron accumulation. Hum. Mol. Genet. 9, 1219-1226.

Dhe-Paganon, S., Shigeta, R., Chi, Y.I. et al., 2000. Crystal structure of human frataxin. J. Biol. Chem. 275, 30753-30756

Ditch S, Sammarco MC, Banerjee A, Grabczyk E (2009) Progresive GAA·TTC Repeat Expansion in Human Cell Lines. PLoS Genet 5(10): e10000704. Doi:10.1371/journal.pgen.10000704.

Du J, *et al.* (1998) Sth1p, a Saccharomyces cerevisiae Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins. *Genetics* 150(3):987-1005

Fourt, F., Cazzalini, O., 1997. Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria. FEBS Lett. 411, 373-377.

Gakh O, Park S, Liu G, Macomber L, Imlay JA, Ferreira GC, Isaya G. Mitochondrial iron detoxification is a primary function of frataxin that limits oxidative damage and preserves cell longevity. *Hum. Mol. Genet* 2005; 15:467-479.

Geil C, *et al.* (2008) A nucleolus-localized activator of Cdc14 phosphatase supports rDNA segregation in yeast mitosis. *Curr Biol* 18(13):1001-5

Grabczyk E, Usdin K. The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.* 2000; 28: 2815–22

Huang J, *et al.* (2006) Inhibition of homologous recombination by a cohesin-associated clamp complex recruited to the rDNA recombination enhancer.*Genes Dev* 20(20):2887-901

Huisman, T. et al. Isolation and characterization of chromosomal mTn*10*insertion mutations affecting K88 fimbriae production in *Escherichia coli*. Microbial Pathogenesis. 20:2. (1996).

Ji., H, et al., Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell* 73:1007-1018. 1993.

Jones EW and Fink GR (1982) "Regulation of amino acid and nucleotide biosynthesis in yeast." Pp.181-299 in The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression, edited by Strathern JN, Jones EW and Broach JR. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press

Koutnikova, H., Campuzano, V., Foury, F., et al., 1997. Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. Nat. Genet. 16, 345-351

Krasilnikova, M, Mirkin S, Replication Stalling at Friedreich's Ataxia $(GAA)_n$ Repeats in Vivo. Molecular and Cell Biology 24: 2286-2295. (2003)

Krasilnikova M, et al. Effects of Friedreich's ataxia $(GAA)_n{\cdot}(TTC)_n$ repeats on RNA synthesis and stability. Nucleic Acids Research 35.4: 1075-1084. (2006)

Kumar, A and Snyder M, Genome-Wide Transposon Mutagenesis in Yeast. Molecular Biology 13.3.1-13.3.15. (2000).

Lahue R and Razidlo D. Mrc1, Tof1 and Csm3 inhibit CAG ·CTG repeat instability by at least two mechanisms. DNA Repair 7 (2008) 633-640.

Laurent BC, *et al.* (1992) An essential Saccharomyces cerevisiae gene homologous to SNF2 encodes a helicase-related protein in a new family. *Mol Cell Biol* 12(4):1893-902

Lenzmeier, B. and Freudenreich, C. Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair Cytogenet Genome Res 100:7–24 (2003).

Lu, C., Schoenfeld, R., Shan, Y., et al., 2009. Frataxin deficiency induces Schwann cell inflammation and death. Biochim. Biophys. Acta 1792, 1052-1061

Lydall D, *et al.* (1996) A meiotic recombination checkpoint controlled by mitotic checkpoint genes. *Nature* 383(6603):840-3

Majka J and Burgers PM (2003) Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. *Proc Natl Acad Sci U S A* 100(5):2249-54

Mirkin, M. Sergei. Expandable DNA repeats and human disease. Nature 447, 932-940 (2007).

Park H and Sternglanz R (1999) Identification and characterization of the genes for two topoisomerase I-interacting proteins from Saccharomyces cerevisiae. *Yeast* 15(1):35-41

Parker AE, *et al.* (1998**)** A human homologue of the Schizosaccharomyces pombe rad1+ checkpoint gene encodes an exonuclease. *J Biol Chem*273(29):18332-9

Pavlov, Y, Newlon, C, Kenukel, T. Yeast origins establish a strand bias for replication mutagenesis. *Molecular Cell*. 2002, July. Vol. 10, 207-213.

Pelletier, R., Farrell, B., Miret, J., Lahue, R. Mechanistic features of CAG•CTG repeat contractions in cultured cells revealed by a novel genetic assay. Nucleic Acids Research 33 (2005): 5667-5676.

Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L, Monticelli A, Keats BJ, Bidichandani SI. Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res.* 2004;32:5962–5971

Rass U and Kemper B (2002) Crp1p, a new cruciform DNA-binding protein in the yeast Saccharomyces cerevisiae. *J Mol Biol* 323(4):685-700

Razidlo, D. and Lahue, R. Mrc1, Tof1 and Csm3 inhibit CAG CTG repeat instability by at least two mechanisms. *DNA Repair (Amst)*. 2008 April 2; 7(4): 633–640

Rother S, *et al.* (2006) Swt1, a novel yeast protein, functions in transcription. *J Biol Chem* 281(48):36518-25

Roy S, Martinez D, Platero H, Lane T, Werner-Washburne M (2009) Exploiting Amino Acid Composition for Predicting Protein-Protein Interactions. PLoS ONE4(11): e7813. doi:10.1371/journal.pone.0007813

Samadashwily, G.M., Raca, G., Mirkin, S.M. Trinucleotide repeats affect DNA replication *in vivo. Nature Genet.* 17, 298-304 (1997).

Sharma, R., De Biase, I., Gomez, M., et al., 2004. Friedreich ataxia in carriers of unstable borderline GAA triplet-repeat alleles. Ann. Nuerol. 56, 898-901.

Shinohara M, *et al.* (2003) The mitotic DNA damage checkpoint proteins Rad17 and Rad24 are required for repair of double-strand breaks during meiosis in yeast. *Genetics* 164(3):855-65

Shiskin, A, et al. Large-Scale Expansions of Friedreich's Ataxia GAA Repeats in Yeast. Molecular Cell 35, 82-92 (2009).

Smith, C, et al. The structure and coding specificitity of a Lysine transfer ribonucleic acid from the haploid yeast *Saccharomyces cerevisiae* S288C. *The Journal of Biological Chemistry*. 246: 24. 7817-7829 (1971).

van den Berg MA, *et al.* (1996) The two acetyl-coenzyme A synthetases of Saccharomyces cerevisiae differ with respect to kinetic properties and transcriptional regulation. *J Biol Chem* 271(46):28953-9

Wilson, R., Rood, D., 1997. Respiratory deficiency due to loss of mitochondrial DNA in yeast lacking the frataxin homologue. Nat. Genet. 16, 352-357.

Zonneveld BJ and van der Zanden AL (1995) The red ade mutants of Kluyveromyces lactis and their classification by complementation with cloned ADE1 or ADE2 genes from Saccharomyces cerevisiae. *Yeast* 11(9):823-7