Genetic analysis of length-dependent GAA/TTC and TTTCA/TGAAA repeat-induced GCR formation

Suzanne Malikova (S4566785)
Supervisor: Daniele Novarina

BSc Research Project Biomedical Sciences (WBBY902-10)

European Research Institute for Biology of Aging
University of Groningen
5/6/2023

Abstract

Currently, 13 short tandem repeat sequences are known to be involved in over 50 neurodegenerative human diseases. GAA repeat expansions in the *FXN* gene are the cause of Friedreich's ataxia, and Benign adult familial myoclonic epilepsy is associated with extending TTTCA repeats. In this pilot, we study these two repeats and their complementary sequences of several lengths, and how they affect repeat-induced fragility. By means of a GCR assay and high-throughput robotic replica-pinning, we can determine the frequency of Gross Chromosomal Rearrangement in repeat-containing yeast strains.

Here, we demonstrate a GAA/TTC repeat-induced length-dependent increase in GCR formation. Moreover, we show that GCR formation is orientation-dependent, as strains with complementary repeat sequences result in varying GCR frequencies. None of the strains reached the desired GCR frequency of 30-50%, but strains containing TTC_{160} and GAA_{95} can be used to identify genes that suppress GCRs. However, to identify genes that promote GAA/TTC repeat fragility, and genes that are involved in TTTCA/TGAAA repeat fragility in general, strains with longer repeat lengths should be tested to reach higher GCR frequencies to be useful in the genome-wide screen.

Introduction

Short tandem repeats (STRs) are nucleotide sequences that are repeated in the DNA. They are found in about 3% of the coding and non-coding regions of the human genome (Malik *et al*., 2021), and have a higher mutation rate compared to non-repetitive regions (Balzano *et al*., 2021). Expansion of repetitive DNA sequences can lead to Repeat Expansion Diseases (REDs). Currently, there are 13 different STRs identified that are associated with over 50 diseases in humans.

In Huntington's disease, CAG repeat expansions in the Huntingtin gene lead to a toxic gain of function at the protein level due to abnormally large polyglutamine tracts in the translated protein that cause aggregation (Jimenez-Sanchez *et al.*, 2017). GAA repeat expansions in the first intron of the *FXN* gene result in gene silencing, causing Friedreich's ataxia (FRDA) (Gottesfeld, 2019). Benign adult familial myoclonic epilepsy (BAFME) is associated with TTTCA repeat expansions in the *SAMD12* intron. However, Ishiura *et al*. (2018) suggest that similar TTTCA repeat patterns in other genes are also associated with BAFME pathogenesis.

It is evident that RED mechanisms differ substantially and can manifest themselves on various different levels, such as the DNA, RNA and protein level. What connects all the REDs is the inherent DNA instability of the repeat tracts. For instance, there are several DNA structures that these repeats are able to form that can interfere with replication, such as hairpins, DNA H-triplexes and G-quadruplexes (Khristich and Mirkin, 2020). Since there is still no cure for the majority of these diseases, it is essential to study the mechanisms behind DNA repeat instability.

In this pilot study, we aim to find the optimal lengths of GAA and TTTCA repeats, and their complementary sequences TTC and TGAAA, respectively, to be used in a future genome-wide screen for the identification of genes associated with repeat fragility. Here, we will use *Saccharomyces cerevisiae* as a model organism to study the formation of Gross Chromosomal Rearrangements (GCRs) as a result of these repeat sequences. Ideally, we will select strains with repeats that yield 30-50% GCR formation, and therefore we will test multiple sizes per sequence.

We will start by inserting the repeat sequences of varying lengths in yeast (CZY105) at the *PRB1* locus, upstream of the genetic markers *URA3* (essential for the production of uracil and sensitive to 5-fluoroorotic acid) and *CAN1* (sensitive to canavanine) on the left arm of chromosome V. The use of two counterselectable markers ensures that the selection is based on GCR formation, as mutations causing both genes to be inactivated is highly unlikely. To exclude GCR events before the assay, the strains are grown on media without uracil (-URA) first. Then, we can select for the loss of both genes induced by a GCR by transferring these cells onto selective media containing 5-fluoroorotic acid and canavanine. By arraying the strains in high-density format, and making use of the robotically replica-pinning approach described by Novarina *et al*. (2022), we can determine the frequency of GCR events for all repeat-containing strains.

Here, we show a length-dependent increase in GCR formation as a result of GAA/TTC repeats insertion, suggesting that a longer size increases the fragility of the repeat. In addition, we show that the GCR frequencies vary between complementary repeat sequences, confirming that repeat fragility is also dependent on orientation. Lastly, none of the tested strains yielded GCR frequencies in the preferred range of 30-50%, however, strains containing TTC_{160} and GAA_{95} can still be used in the genome-wide screen to identify genes that suppress GCR formation.

Methods

Plasmid extraction from E.coli

E. coli cells were grown on LB medium containing 100 µg/mL ampicillin, expressing one of the plasmids (Supplementary table S1). Incubation was done at 30°C to ensure stability of the repeats. A miniprep was conducted to extract the plasmids from *E. coli*, using the NucleoSpin® Plasmid kit from Macherey-Nagel.

Yeast transformation

Yeast (*Saccharomyces cerevisiae*) cells containing genetic markers for the GCR assay (CZY105) (Supplementary table S2) were streaked for single colonies, which were then used for transformation with the digested plasmids and PCR-amplified repeat fragments. Yeast cells that were successfully transformed were selected on YPD medium containing Hygromycin. Survivors were streaked for single colonies, from which several colonies for every length and repeat sequence were subsequently used to make temporary storage patches on YPD medium containing Hygromycin. Incubation was done at 30°C.

PCR

Genomic DNA was extracted from single colonies for PCR. $TTTCA_{42}$-HphMX and $TGAAA_{42}$-HphMX were amplified from pSO25 and pSO14 respectively, using primers pAG32_PRB1_fwd and pAG32_PRB1_rvs and Q5 polymerase, according to the New England BioLabs protocol (Supplementary table S3, S4). To verify that the repeat sequences were inserted at the *PRB1* locus, and to prepare samples to be sent for sequencing, primers HygB fwd and FR_PRB1_insertion_Rvs and Taq polymerase were used according to the New England BioLabs protocol (Supplementary table S3, S5).

To check the GAA repeat lengths, colony PCR was performed according to the New England BioLabs protocol, using Taq polymerase and primers expansion-FOR and expansion-REV (Supplementary table S3, S6). To check the TTC repeat lengths, colony PCR was performed according to the New England BioLabs protocol, using Taq polymerase and primers expansion-FOR and expansion-REV2 (Supplementary table S3, S7) that anneal near the repeat sequences in order to assess the repeat length more accurately. PCR clean-up was performed using the Nucleospin® Gel and PCR clean-up kit.

Digestion

Plasmids pDN41.1, pDN64.1, pDN59.1, pDN40.3, pDN63.1, and pDN58.1 were digested using SalI to linearize the repeat fragments. We digested 5 µg of plasmid with 100 units of SalI for 2 hours at 37 °C, followed by 20 minutes of heat activation at 65 °C.

Verification of repeat sequence and lengths

Integration of repeats at the *PRB1* locus was verified on a 1% agarose gel (5 µl ethidium bromide per 100 mL agarose solution). PCR products from GAA and TTC repeats were verified on a 2% agarose gel (5 µl ethidium bromide per 100 mL agarose solution).

Genomic DNA was extracted from storage patches for diagnostic PCR to verify that the repeats are integrated at the *PRB1* locus in the transformants, and that the repeats have the desired length. The PCR product was then analyzed by Sanger sequencing, using 2 µl of HygB fwd primer per sample, to determine the exact length and sequence of the repeats.

Patch-and-replica-plating

A patch-and-replica-plating experiment was performed to estimate the GCR frequencies of the strains containing the repeat sequences. First, to exclude GCRs before the start of the experiment, patches were made on SD-URA plates of the selected strains, as well as three positive control strains containing the following repeat sequences: CAG_{74}, CAG_{121}, and CAG_{161}. The patches were grown for 1 day at 30°C, and then replica-plated on 5-FOA/CAN plates to be grown for 4 days at 30°C. GCR frequencies were then estimated by counting the number of colonies per patch, as every colony stems from one GCR event.

GCR assay via robotic pinning

The strains containing the repeat sequences were streaked for single colonies on SD-URA. The selected colonies were manually arrayed in 96-format on SD-URA rectangular plates and grown for 1 day at 30°C. With the ROTOR HDA pinning robot, the colonies were arrayed from 96-format to 384-format, and from 384-format to 1536-format on SD-URA RotoR non-selective plates. In between pinning steps, the plates are incubated for 1 day at 30°C. The colonies were robotically replica-pinned onto new SD-URA RotoR plates and imaged.

The colonies were then robotically replica-pinned in 1536-format onto selective 5-FOA/CAN RotoR plates and incubated for 2 days at 30°C. These colonies were then again robotically replica-pinned in 1536-format onto selective 5-FOA/CAN RotoR plates and incubated for 1 day at 30°C. This selective plate was then also imaged. To calculate the GCR frequency for each repeat sequence and length, colonies on the non-selective and selective plates were counted manually. All imaging was done using a ScanMaker 9800XL scanner (MicroTek International, Inc.), scanning in 16-bit grayscale at 300-dpi resolution with the Transparent Media Adapter.

Results

Transformation of CZY105 yeast with repeat sequences of different lengths

The GAA and TTC sequences were digested with SalI to obtain linear repeat fragments of different sizes. Preliminary data suggested that slight variations in repeat lengths for TGAAA and TTTCA sequences, would already result in notable differences in GCR frequencies. Therefore, $TGAAA_{42}$-HphMX and $TTTCA_{42}$-HphMX were amplified using PCR in order to obtain different repeat lengths, due to expansions and contractions in the repeat tract. After transformation, yeast cells containing genetic markers for the GCR assay (Supplementary table S2), were grown on YPD+Hygromycin plates to select for successful transformants. Colony growth was observed for all sequences and repeat lengths, however, for GAA_{121} and GAA_{160} very little colony growth was observed.

To then verify that the repeat fragments were integrated into the correct location in the genome, the *PRB1* locus, a diagnostic PCR was performed on extracted genomic DNA. Integration and repeat lengths of all transformants were analyzed by Sanger sequencing. From this, we selected the following repeat lengths for the different sequences: $TGAAA_{40}$, $TGAAA_{41}$, $TGAAA_{43}$, $TTTCA_{41}$, $TTTCA_{42}$, $TTTCA_{43}$, GAA_{80}, GAA_{95}, and TTC_{81}. However, for the longer repeat sequences (roughly longer than 300 base pairs) the sequencing results became unreliable. Therefore PCR was performed on transformants containing GAA_{121}, GAA_{160}, TTC_{121}, and TTC_{160}, which were then validated on a 2% agarose gel.

From this, we selected the following repeat lengths for the different sequences: TTC_{121} and TTC_{160}. The GAA repeat sequences on the gel were shorter than required, therefore we only continued

with GAA_{95}, which was the result of the transformation with pDN58.1 and was verified by sequencing. To be able to continue the experiment with three different repeat lengths for GAA as well, we used a GAA_{160} which originated from a different background (DNY188), which was available at Chang's lab (ERIBA).

Verification of repeat sequences and lengths of selected strains to proceed with the GCR assay

To verify the repeat lengths of the different yeast strains before starting the GCR assay, PCR was performed. As a positive control, the plasmids were also amplified with PCR using the same primers. For the GAA and TTC repeat sequences, we validated the length of the PCR product by 2% agarose gel electrophoresis (Figure 1). We expected to see bands with the following sizes: 312 bp (GAA_{81}), 432 bp (GAA_{121}), 549 bp (GAA_{160}), 304 bp (TTC_{81}), 424 bp (TTC_{121}), 541 bp (TTC_{160}). Even though not all bands were at the expected locations in the gel based on the DNA ladder, the samples did match the size of the controls, suggesting that the number of repeats is the same. Due to unreliable sequencing data, we were unable to verify that the TGAAA and TTTCA repeat sequences in the different strains were still of the same size before starting the GCR assay.

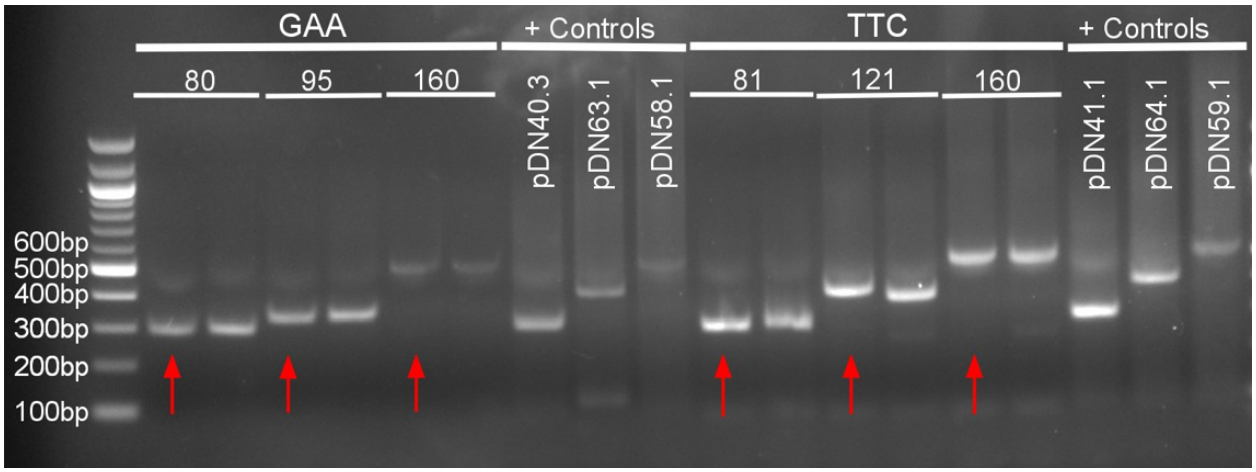

Figure 1: 2% agarose gel electrophoresis for verification of the repeat lengths before proceeding with the GCR assay. Repeat sequence lengths were of the expected size based on the bands of the controls on the gel. We proceeded the experiment with the first sample (red arrow) for every repeat length. A 100bp DNA ladder (NEB) was used.

Estimation of GCR frequency via patch-and-replica-plating

To obtain a rough estimation of the GCR frequency of each strain, we performed patch-and-replica-plating. From this we could observe whether there is a difference in colony growth between the patches (Figure 2), as a proxy for the GCR frequency. Here, each colony originates from a single GCR event, and therefore a higher number of colonies equals a higher GCR frequency. For the CAG repeat controls, an increased colony growth is observed for only CAG_{161} and not for the CAG_{121} and CAG_{74}. For the TTC repeats, an increase in colony growth is observed in TTC_{160} compared to TTC_{121} and TTC_{81} on the selective plate. For the other sequences, there is no clear difference in colony growth between the different repeat lengths.

TTC 81	TTTCA 41	TGAAA 40	GAA 80
TTC 121	TTTCA 42	TGAAA 41	GAA 95
TTC 160	TTTCA 43	TGAAA 43	GAA 160

CAG 161	CAG 121	CAG 74

Figure 2: Quick and dirty estimation of GCR frequency via patch-and-replica-plating. Patches were replica plated on CAN 5-FOA plates. On the left we see the CAN 5-FOA plate where we can observe increased colony growth for TTC_{160} and CAG_{161}. On the right we see a schematic overview of the patches on the plate.

Measuring the length-dependent GCR frequency via robotic pinning

To obtain a more precise GCR frequency, we measured the GCR frequency of all strains via robotic pinning. Colonies were first grown on nonselective SD-URA plates and arrayed from 96-format to 384-format and then to 1536-format (Figure 3A, 4A). All colonies were of equal size on the nonselective plates, except for GAA_{160}, which were visibly smaller. This could indicate that this strain grows slower.

Subsequently, the colonies were replica-pinned on selective CAN 5-FOA plates (Figure 3B). Surprisingly, the colonies on both selective plates (Figure 3B, 4B) were not all of the same size. The controls (CAG_{121}, CAG_{161}) indicate that a longer repeat length leads to a higher GCR frequency (Figure 3C, 3D, 4C, 4D). A similar pattern is observed for the TTC repeat lengths, where repeat length correlates with GCR frequency (Figure 3C). For the GAA repeat lengths, the difference in GCR frequency for GAA_{80} and GAA_{95} is rather large, but this increase in GCR frequency is not observed when the repeat length is increased to GAA_{160} (Figure 3D). The GCR frequency did not differ considerably between the different lengths for TTTCA (Figure 4C). For the TGAAA repeat lengths, we see a minor decrease in GCR frequency as the repeat length increases (Figure 4D).

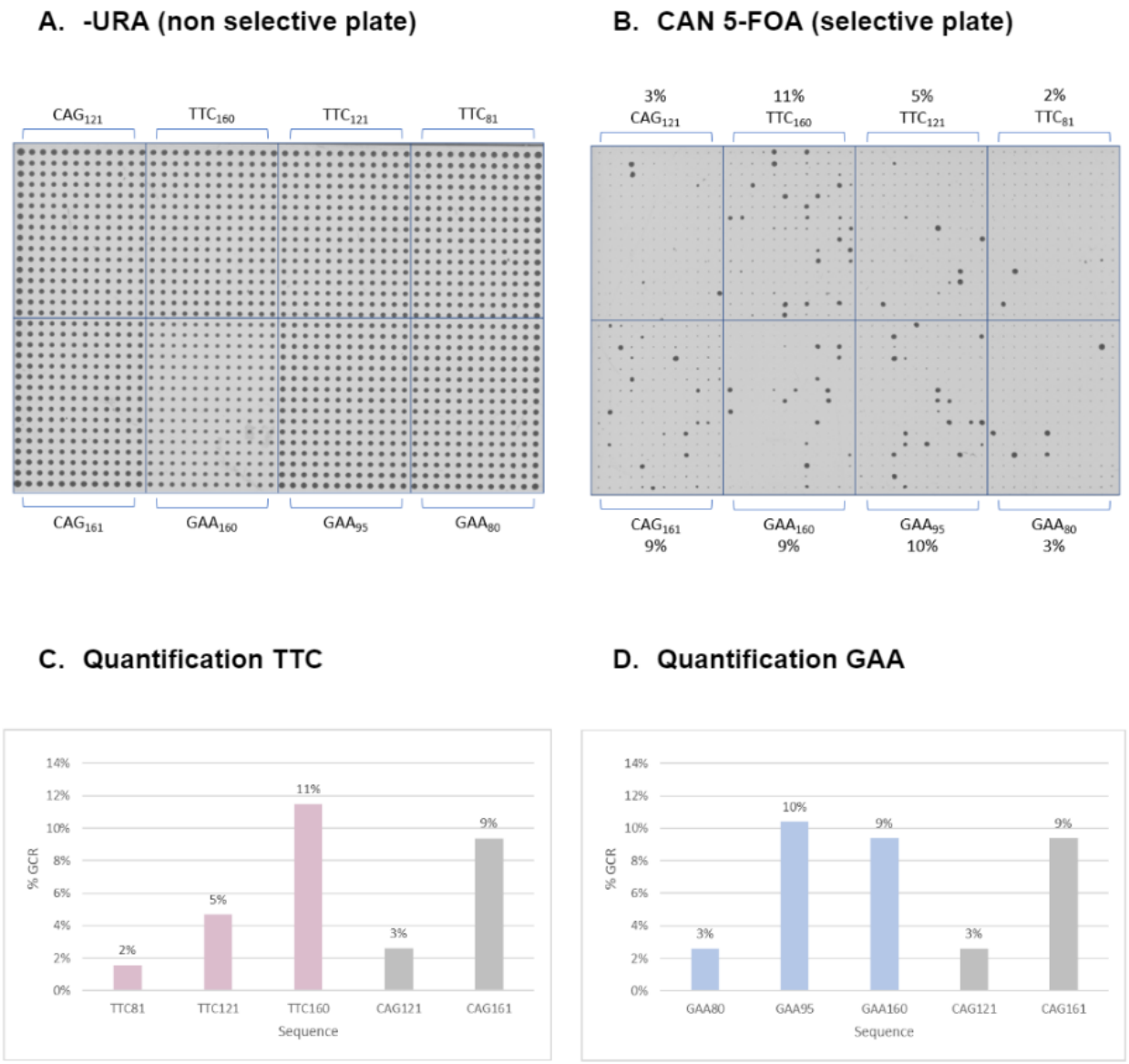

Figure 3: Result of GCR assay for different TTC and GAA repeat containing strains. **(A)** Colonies were grown on a non selective -URA plate to exclude GCRs before the start of the experiment. From the plate we can see that GAA_{160} colonies grew less than the other strains. **(B)** Colonies were grown on CAN 5-FOA plates to select colonies that have undergone a GCR event. Percentages of colonies that grew are displayed together with respective strains. Observed colonies are of different sizes. **(C)** Graphical representation of percentage of TTC repeat colonies that have undergone a GCR event. Here, a longer repeat sequence results in a higher GCR frequency. The CAG controls match this observation. **(D)** Graphical representation of the percentage of GAA colonies that have undergone a GCR event. Here, an increase in repeat length from 80 to 95 repeats already gives a high increase in GCR frequency. This increase is not visible at GAA_{160} repeats, which might be caused by its slower growth frequency.

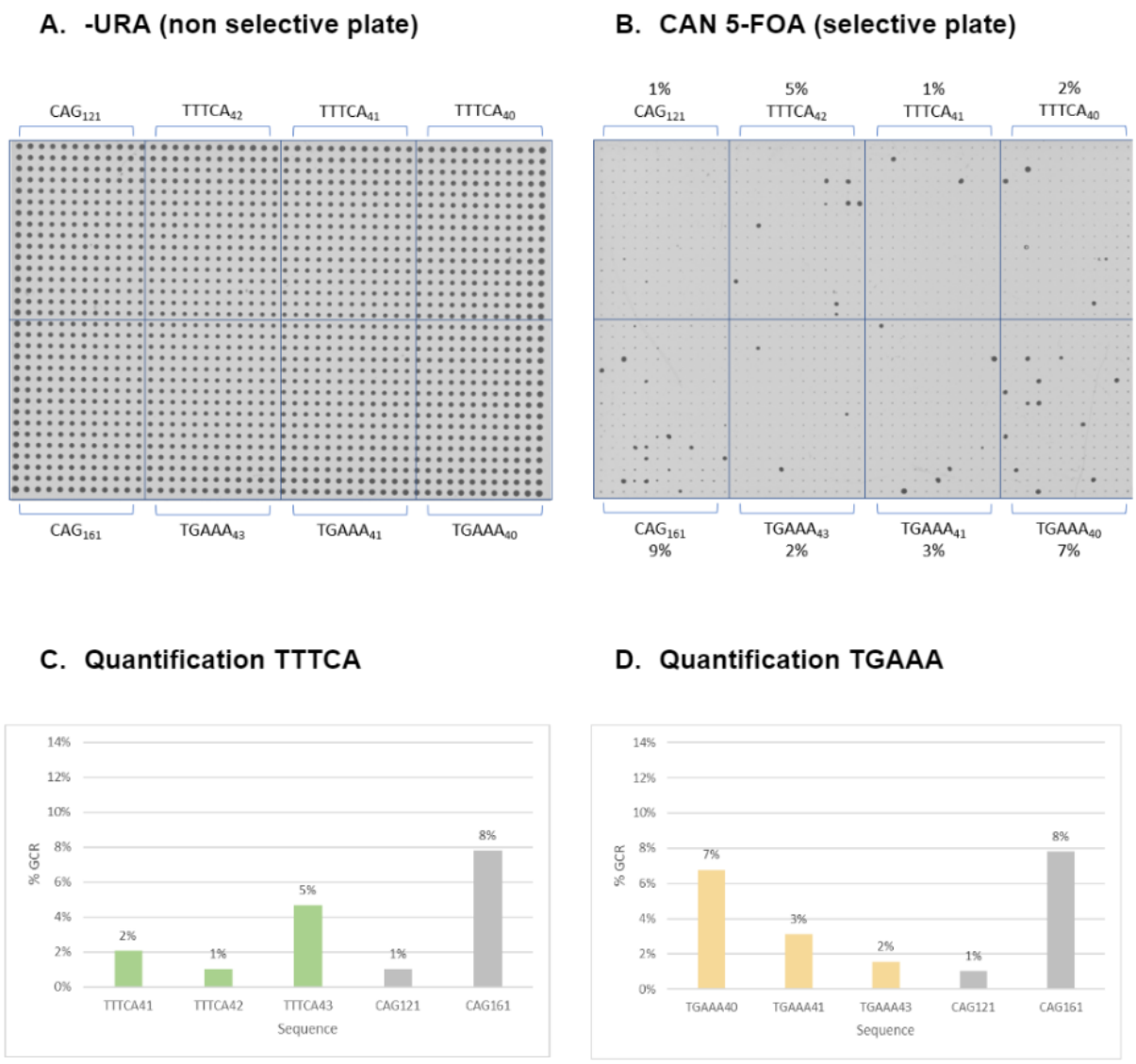

Figure 4: Result of GCR assay for different TTTCA and TGAAA repeat containing strains. **(A)** Colonies were grown on a non selective -URA plate to exclude GCRs before the start of the experiment. From the plate we can see that all colonies grew similarly. **(B)** Colonies were grown on CAN 5-FOA plates to select colonies that have undergone a GCR event. Percentages of colonies that grew are displayed together with respective strains. Observed colonies are of different sizes. **(C)** Graphical representation of percentage of TTTCA repeat colonies that have undergone a GCR event. Here, the GCR frequency does not seem to differ considerably between different repeat lengths. **(D)** Graphical representation of the percentage of TGAAA colonies that have undergone a GCR event. Here, an increase in repeat sequences seems to give a decrease in GCR frequency. However, because of the small differences, this could be due to random variability. For the CAG controls, the GCR frequency increases as the repeat length increases. This pattern is not observed for the TTTCA and TGAAA repeat containing strains.

Discussion

In this pilot study we aimed to find the optimal number of repeats in GAA, TTC, TTTCA and TGAAA repeat tracts that would result in GCR frequencies of about 30-50% by testing yeast strains with different repeat lengths in a GCR assay by robotic replica-pinning. This way, both decreases and increases in GCR frequencies can be detected in a future genome wide screen to identify genes involved in GCR formation and/or repeat fragility.

The patch-and-replica-plating experiment did not result in clear quantifiable differences in colony growth between the strains. Only the control strains containing CAG repeats and the strains containing TTC repeats showed a somewhat length-dependent increase in the number of colonies, but this is still not sufficient to estimate GCR frequencies. In the GCR assay by robotic pinning, the GCR frequency was determined by counting the number of colonies on the selective 5-FOA/CAN plates. Unexpectedly, the colonies were not uniform, but were of different sizes, contrary to previous findings in this lab that did not show this large of a variation in colony size. Possibly, the batches of 5-FOA/CAN media may have differed in concentrations, leading to differences in the degree of selectivity.

The GCR assay with the selected strains resulted in GCR frequencies between 1% and 11%, which is way below the ideal range. The TTC repeat containing strains showed an increase in GCR frequency as the repeat length increased again, which is in line with previous studies that show that repeat instability increases with the size of the repeat tract (Khristich and Mirkin, 2020). In the GCR assay, the TTC_{160} strain resulted in a GCR frequency of 11%, which could in principle still be used for the screen to clearly detect increases in GCR frequencies. This way, genes that suppress GCRs and/or repeat fragility can still be identified.

When comparing strains containing GAA_{80} to GAA_{95}, the GCR frequency increases more substantially than in the TTC repeat strains, indicating that the orientation of the repeat sequence results in different GCR frequencies. This is consistent with previous studies by Shishkin *et al*. (2009) and Pollard *et al.* (2004) that show differences in GAA repeat stability depending on whether the GAA_n tracts are located in the leading or the lagging strand.

Another difference is that the strain containing GAA_{160} has a slightly decreased frequency. A possible explanation for this could be that this GAA_{160} repeat is inserted in the DNY188 background instead of the CZY105 like the rest of the strains. Colonies from this GAA_{160} strain on the non selective robot plate were all equally and visibly smaller than the rest of the strains, suggesting that this strain grows slower than the others. The smaller colonies mean that there were less cells and therefore a lower chance of a GCR event occurring to begin with, which could explain why the GCR frequency is lower than expected. Preferably, this assay should be repeated with a GAA_{160} containing strain made in the CZY105 background, to examine whether the GCR frequency will increase. Then, like the TTC_{160} strain, the GAA_{95} strain could potentially also be used for the genetic screen to identify genes that suppress GCRs and/or repeat fragility.

Strains with TTTCA/TGAAA repeats did not show the same length-dependent increase in GCR frequency as the strains with GAA/TTC repeats did. Preliminary data suggested that slight differences in number of repeats for these particular sequences would result in large differences in GCR frequencies. However, in our GCR assay we only observe slight variations between 1-7% which is again way lower than the desired range. Moreover, we did not obtain reliable sequencing results from these strains, and therefore we were unable to verify that the strains contained the same repeat sequence and sizes before starting the GCR assay. In any case, it is essential to repeat this, but then to also test strains with larger

(differences in) repeat sizes, such as 50, 60 and 70 repeats, as this will likely increase GCR frequencies in a GCR assay.

To conclude, our findings show that GCR formation increases with the repeat size in yeast strains containing GAA/TTC repeats, indicating that repeat instability is length-dependent. Also in line with previous studies, our results indicate that the orientation of the repeat sequence influences repeat fragility, as complementary repeat sequences do not result in the same GCR frequencies. Even though further research is needed to obtain strains containing GAA, TTC, TTTCA and TGAAA repeat sequences that yield the desired GCR frequencies in a GCR assay, we provide two strains containing TTC_{160} and GAA_{95} sequences that can be used to identify genes involved in suppressing repeat fragility.

References

- Balzano, E., Pelliccia, F., & Giunta, S. (2021). Genome (in)stability at tandem repeats. Seminars in cell & developmental biology, 113, 97–112.

- Gottesfeld J. M. (2019). Molecular Mechanisms and Therapeutics for the GAA·TTC Expansion Disease Friedreich Ataxia. Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics, 16(4), 1032–1049.

- Ishiura, H., Doi, K., Mitsui, J., Yoshimura, J., Matsukawa, M. K., Fujiyama, A., Toyoshima, Y., Kakita, A., Takahashi, H., Suzuki, Y., Sugano, S., Qu, W., Ichikawa, K., Yurino, H., Higasa, K., Shibata, S., Mitsue, A., Tanaka, M., Ichikawa, Y., Takahashi, Y., … Tsuji, S. (2018). Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. Nature genetics, 50(4), 581–590.

- Jimenez-Sanchez, M., Licitra, F., Underwood, B. R., & Rubinsztein, D. C. (2017). Huntington's Disease: Mechanisms of Pathogenesis and Therapeutic Strategies. Cold Spring Harbor perspectives in medicine, 7(7), a024240.

- Khristich, A. N., & Mirkin, S. M. (2020). On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. The Journal of biological chemistry, 295(13), 4134–4170.

- Malik, I., Kelley, C. P., Wang, E. T., & Todd, P. K. (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. Nature reviews. Molecular cell biology, 22(9), 589–607.

- Novarina D., Rosas Bringas F.R., Rosas Bringas O.G., Chang M. High-throughput replica-pinning approach to screen for yeast genes controlling low-frequency events. STAR Protoc, 2022, 3(1):101082

- Pollard, L. M., Sharma, R., Go´mez, M., Shah, S., Delatycki, M. B., Pianese, L., Monticelli, A., Keats, B. J., and Bidichandani, S. I. (2004) Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res. 32, 5962–5971 CrossRef Medline

- Shishkin, A. A., Voineagu, I., Matera, R., Cherng, N., Chernet, B. T., Krasilnikova, M. M., Narayanan, V., Lobachev, K. S., and Mirkin, S. M. (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol. Cell 35, 82–92 CrossRef Medline 143

Supplementary information

Table S1. Plasmids used in this pilot study.

Plasmid	Repeat sequence + length	Selection marker
pS025	$TTTCA_{42}$	Ampicillin resistance
pS014	$TGAAA_{42}$	Ampicillin resistance
pDN41.1	TTC_{81}	Ampicillin resistance
pDN64.1	TTC_{121}	Ampicillin resistance
pDN59.1	TTC_{160}	Ampicillin resistance
pDN40.3	GAA_{81}	Ampicillin resistance
pDN63.1	GAA_{121}	Ampicillin resistance
pDN63.1	GAA_{160}	Ampicillin resistance

Table S2. Yeast strains used in this pilot study.

Strain name	Genotype
CZY105	*Matα hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + $TTTCA_{41}$	*Matα prb1::HPH:*$TTTCA_{41}$ *hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + $TTTCA_{42}$	*Matα prb1::HPH:*$TTTCA_{42}$ *hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + $TTTCA_{43}$	*Matα prb1::HPH:*$TTTCA_{43}$ *hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + $TGAAA_{40}$	*Matα prb1::HPH:*$TGAAA_{40}$ *hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + $TGAAA_{41}$	*Matα prb1::HPH:*$TGAAA_{41}$ *hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + $TGAAA_{43}$	*Matα prb1::HPH:*$TGAAA_{43}$ *hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*

CZY105 + TTC_{81}	*Matα prb1::HPH:TTC_{81} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + TTC_{121}	*Matα prb1::HPH:TTC_{121} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + TTC_{160}	*Matα prb1::HPH:TTC_{160} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + GAA_{80}	*Matα prb1::HPH:GAA_{80} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
CZY105 + GAA_{95}	*Matα prb1::HPH:GAA_{95} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
DNY188	*Matα prb1::HPH:GAA_{160} hxt13Δ::URA3 ARG4 CAN1 lys2Δ ADE2 leu2-3,112 his3-11,15 ura3-1 TRP1 RAD5*
DNY107	*Matα prb1::HPH:CAG_{74} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
DNY178	*Matα prb1::HPH:CAG_{121} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*
DNY160	*Matα prb1::HPH:CAG_{161} hxt13::URA3 mfa1::MFA1pr-HIS3 his3Δ1 ura3Δ0 lyp1Δ leu2Δ0 metl5Δ0*

Table S3. Primer sequences.

Primer name	**Sequence**
pAG32_PRB1_fwd	CAAACTTAAGAGTCCAATTAGCTTCATCGCCAATAAAAAAACAA ACTAAACCTAATTCTAACAgggttaattaaggcgcgccagatct
pAG32_PRB1_Rvs	TTGTAACCTCGAGACGCCTAAGGAAAGAAAAAGAAAAAAAAA AGCAGCTGAAATTTTTCTAAAcactatagggagaccggcagatccgl
HygB fwd	TCTTGACCAACTCTATCAGA
FR_PRB1_insertion_rvs	CGCAAATATGTAGTAATACGTGG
Expansion_FOR	ttctaaatgtcctgcaggcg
Expansion_REV	gtcctgcagttaatTAAGGTC
Expansion_REV2	tgctgtcgattcgaaggatc

Table S4. PCR program for amplification of TTTCA42-HphMX and TGAAA42-HphMX from pSO25 and pSO14, respectively.

Cycle step	Temperature	Time	Cycles
Initial denaturation	98°C	30 seconds	1
Denaturation	98°C	10 seconds	30
Annealing + extension	72°C	1 minute and 20 seconds	
Final extension	72°C	5 minutes	1
Hold	10°C	∞	

Table S5. PCR program for verification of sequence length and correct integration at the PRB1 locus.

Cycle step	Temperature	Time	Cycles
Initial denaturation	94°C	30 seconds	1
Denaturation	94°C	30 seconds	35
Annealing	45°C	30 seconds	
Extension	68°C	1 minute and 30 seconds	
Final extension	68°C	5 minutes	1
Hold	10°C	∞	

Table S6. Colony PCR program to check GAA repeats size.

Cycle step	Temperature	Time	Cycles
Initial denaturation	94°C	30 seconds	1
Denaturation	94°C	30 seconds	35
Annealing	51°C	30 seconds	
Extension	68°C	40 seconds	
Final extension	68°C	5 minutes	1
Hold	10°C	∞	

Table S7. Colony PCR program to check TTC repeats size.

Cycle step	Temperature	Time	Cycles
Initial denaturation	94°C	30 seconds	1
Denaturation	94°C	30 seconds	35
Annealing	47°C	30 seconds	
Extension	68°C	40 seconds	
Final extension	68°C	5 minutes	1
Hold	10°C	∞	