Studying DNA Polymerase Slippage at Simple Sequence Repeats

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Rajdip Sarkar

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

November, 2022

Supervisor: Prof. Saikrishnan Kayarat

Department of Biology, IISER Pune

Certificate

This is to certify that this dissertation entitled *Studying DNA Polymerase Slippage at Simple Sequence Repeats* towards the partial fulfilment of the BS-MS dual degree programme at Indian Institute of Science Education and Research, Pune represents study/work carried out by Rajdip Sarkar at Indian Institute of Science Education and Research under the supervision of Prof. Saikrishnan Kayarat, Professor, Department of Biology, during the academic year 2021-2022.

Prof. Saikrishnan Kayarat

Committee:

Prof. Saikrishnan Kayarat

Dept. of Biology

IISER Pune

Dr. Sunish Radhakrishnan

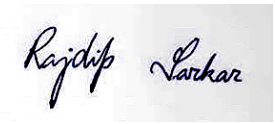
Dept. of Biology

IISER Pune

This thesis is dedicated to my family

Declaration

I hereby declare that the matter embodied in the report entitled *Studying DNA Polymerase Slippage at Simple Sequence Repeats* are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Saikrishnan Kayarat and the same has not been submitted elsewhere for any other degree.



Rajdip Sarkar

Date: 31/10/2022

Contents

Abstract

Short tandem repeats (STR) or simple sequence repeats (SSR), or microsatellites are 1-6 nucleotide or sometimes 1-10 nucleotide long repetitive motifs of DNA. These regions of DNA are known to be prone to de novo mutations, occurring during replication.

Encountering SSR during replication, DNA polymerases slip or stall, hypothesised to be because of formation of secondary structures by the single stranded DNA. During replication, at SSR, polymerases transiently detaches, and with it, the newly formed strand, from the parent strand at the SSR and anneals again and polymerase starts replication further. During this process, often the SSR lead to mismatch during re-annealing. As a result, there occurs expansion or contraction in the DNA strand. During stalling of DNA polymerase, random point mutations occur. This leads to InDel mutations causing frameshift.

In *Mycoplasma bovis*, in type III restriction modification system's methyl-transferase *mod* gene there is an SSR of AG-repeats. With this gene *mod*, the gene *res* is present in the same open-reading frame (ORF) which has variable expression patterns owing to the presence of AG-repeats in the preceding gene *mod*.

We studied patterns of slippage at AG-repeats, AT-repeats and G-repeats inserting them in SSR-part of *mod* (call *mod'*) by *in vitro* experiments. We found the threshold for slippage remains very close and it depends on the number of nucleotides in the entire stretch for Pfu DNA polymerase. For AG-repeats it is $(AG)_6$, for other constructs as well it has been 12 nucleotide-long stretch of SSR.

Acknowledgement

I would like to thank IISER Pune and Prof. Saikrishnan Kayarat for providing me with the opportunity to work in his excellent lab and Dr. Sunish Radhakrishnan for being the expert member of my TAC. I thank all the members of SK Lab and G3 Lab. I thank Dr. Vishal and Akash for motivating me whenever required. I appreciate Rushik Bhatti for mentoring me.

Chapter 1: Introduction

One of the most important and most widespread genome markers is (Ellegen, 2004). Short tandem repeats (STR) or simple sequence repeats (SSR) or microsatellites are 1-6 (or sometimes considered to be 1-10) nucleotide-long motifs of repeating itself. They are often 10-60 repetition of nucleotide motifs. With accumulation of genomic data, it has been understood that SSR are ubiquitous across all prokaryotes and eukaryotes. It has been observed that they are often 1-10 nucleotide-long, but sometimes these motifs of nucleotide-sequences are of much longer. It has also been observed that there exists repetitive motifs within a larger motifs. Sometimes they are also cryptic because of imperfection in the repetitive motifs.

The omnipresent nature of the the SSR has drawn quite a lot of attention. Often designated as junk DNA they do hold significance. Though they are found ubiquitously in all genomes, their origin and role and significance in evolution are still not fully understood. There are motifs of SSR, which are found in genomes of distantly related species. For example, $(GATA)_n$ and $(GACA)_n$ are present in taxa such as *Mus*, *Drosophila*, *Elaphe*. Probability of finding $(CA)_n$ in entire human genome twice is very much rare. Not only this, many such repetitive stretches of di-, tri- or mono-nucleotide sequences are present. In type III restriction systems of *Mycoplasma bovis*, the gene *mod* has the gene *res* in the same ORF. In gene *mod* there is an STR of $(AG)_n$, to variable expression of *res*. Such phase-varions help in regulation of expression of many genes by epigenetic means like methylation. Bacteria like *Haemophilus influenzae, Streptococcus pneumoniae* have genes that lead to pathogenesis, which remain under regulation by presence of SSR and pathogenesis is not always found of same severity. The independent possibility of evolution of same repeat-motifs in different taxa suggests presence of some common mechanisms. One of the possible accepted hypothesis for mechanism for origin of long stretches of SSR is thought to be slipped strand mispairing (SSM) of the DNA double-helix at the SSR (Levinson, Gutman, 1987). We are going to discuss about that.

1.1 Slipped Strand Mispairing:

Slipped strain mispairing (SSM) describes one possible mechanism for evolution and propensity of *de novo* mutagenesis at the SSR. During replication on encountering SSR, DNA polymerase slips or stalls, and newly synthesised DNA strand detaches transiently

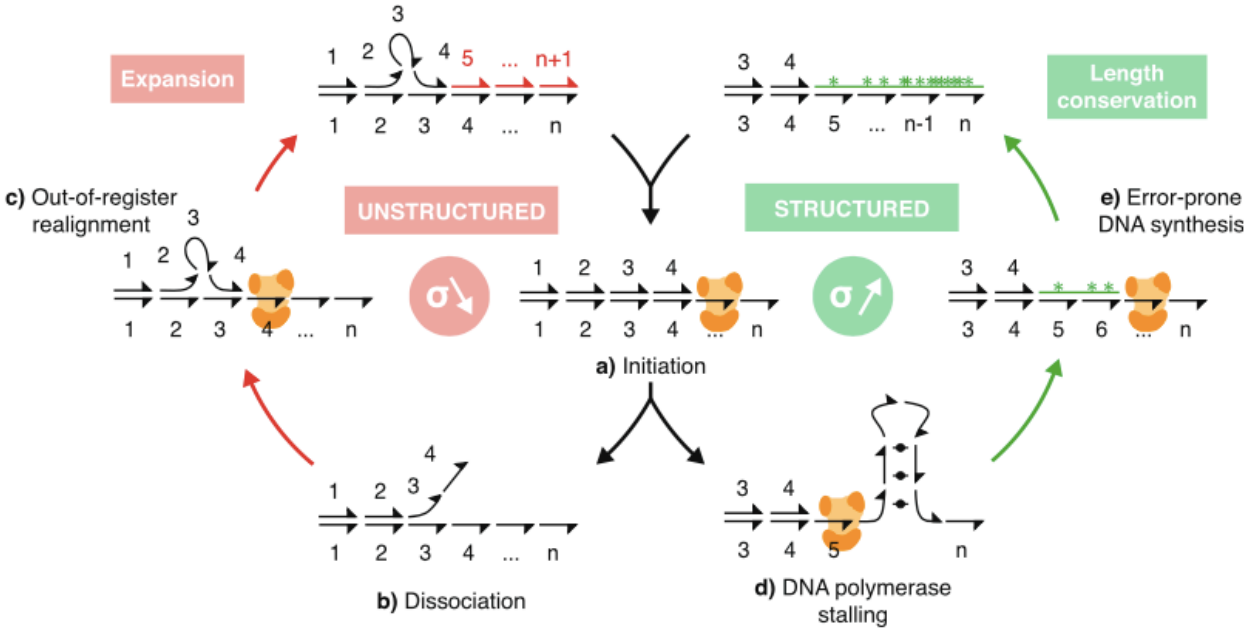


Fig 1.1: Stalling or slippage leading to de novo mutations (Murat et al, 2020)

Secondary structures lead to slippage or stalling at the SSR.
A. On encountering the SSR, DNA polymerases slip or stall. As a result of that the nascent DNA strand detaches from the template strand. Next the nascent strand anneals again. Because of having repetitive nucleotides often out-of-register alignment takes place at the SSR. And the polymerase binds at the 3'-end of the nascent strand which results in shortening or expansion of the repeat region.
B. When the polymerases do not slip on encountering secondary structures, then changes in length do not occur, but, at the point of stalling, inappropriate incorporation of nucleotides occur, which lead to de novo point mutations.

and re-anneals. During annealing, at the repeats, often mismatch occurs at the SSR, leading to out-of-register alignment. The polymerase binds again and starts replication leading to change in number of nucleotide repeats. As a result of that expansions or contractions of length of the SSR-stretches occur strand occurs. After several rounds of contraction and expansions, propensities to expand and to contract come to equilibrium, leading to apparent stability of the lengths. During stalling of DNA polymerase at the SSR, error-prone synthesis takes place.

Some discussions suggested that because of formation of different secondary structures by different nucleotide stretches of the single-stranded DNA at the SSR during replication slippage or stalling occurs (Murat et al, 2020, Levinson, Gutman, 1989). This concept is also supported by the observations that for different polymerases with induction of strand-displacement activity that is the ability of the DNA polymerases to displace downstream DNA and to polymerase only one strand at a time (Viguera et al, 2001).

But in reality, various repair mechanisms like mismatch repair (MMR), base excision repair (BER), exonuclease activity of the DNA polymerases come into play during *in vivo* replication of DNA. These repair processes do help in genomic stability. But with increase in length tendency to change increase.

1.2 Implications:

Simple sequence repeats often lead to changes in the lengths of DNA. Many neurodegenerative disorders like Huntington disease, various axiomatic etc., some cancers are found in patents with high microsatellite instability (MSI-H) or in patents with mismatch-repair deficiency (dMMR). Many polygenic disease cease to appear in populations because of slippage. Pathogenicity of some pathogens can vary which can help in more or less elicitation of immune-response in host. This variation often comes out of expression of particular proteins which are expressed under stress. These are contingency genes which are expressed when under selection-pressure. These genes are contingency genes. In genomic studies these repeat regions are used extensively as markers. Repair-mechanisms generally fix the mutations arising at the SSR, so their mutability are not often obvious. Often because of distortion in the motifs, SSR remain cryptic.

1.3 Techniques used: Restriction-Free cloning:

Traditional cloning methods use restriction enzymes to cut plasmid DNA at specific sites and insertion of the genes of interest using enzyme DNA ligase. But this method has its own limitations because of dependence of restriction-sites. On the other hand restriction-free cloning can be done to insert any gene to any circular plasmid irrespective of presence of any restriction-sites. In RF cloning, a forward primer is taken with a stretch of the nucleotide-sequence to be inserted in a plasmid vector with a nucleotide-stretch next to the insert similar to that of in the plasmid. Similarly a reverse primer is taken. With high-fidelity PCR reactions, the primers are made to anneal to the vector target (presence of complementarity) and are amplified giving a linear PCR product containing the vector nucleotide-sequence and insert sequence in between. These are called the ‘megaprimers’, the reaction is the primary PCR. The megaprimers are again amplified with the vectors giving circular plasmids with the insert sequence. This is the secondary PCR. This PCR product is subjected to digestion with a restriction enzyme, which chops the vectors (sourced from cells, have methylated restriction sites.) and the circular DNA with the inserted nucleotide-sequence (secondary PCR products) remains intact. Transforming some competent cells with these plasmids, we can get cells with plasmids with the insert.

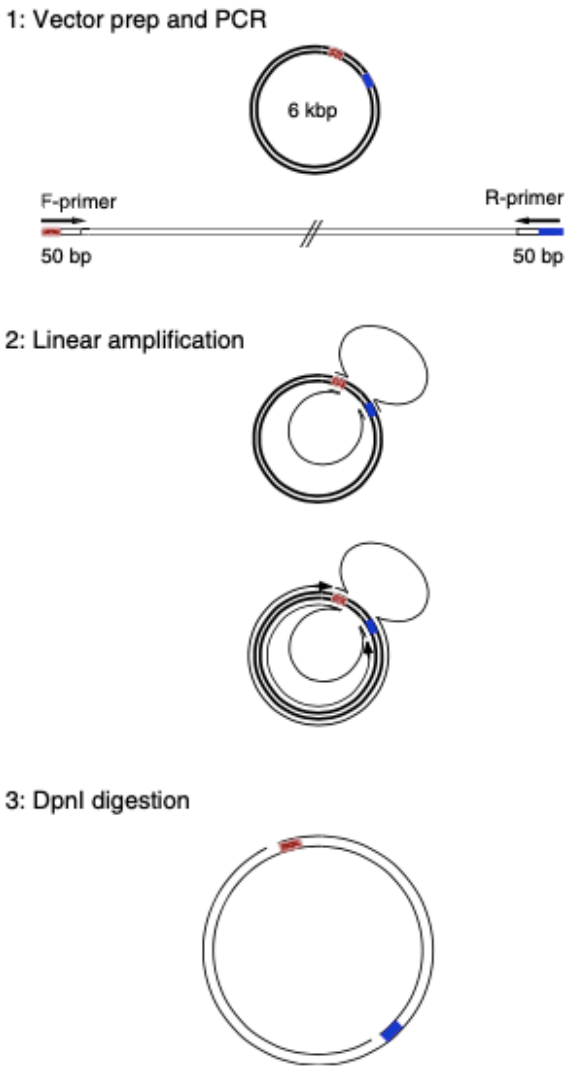


Fig. 1. Schematic representation of restriction-free (RF) cloning. 1. A circular vector isolated from a *dam*$^+$ strain has unique priming sites (depicted in red and blue) and is combined with a PCR product encoding the gene of interest flanked by priming sites complementary to those in the vector. 2. The PCR product acts as a primer pair in a linear amplification reaction. Once annealed to the vector, *PfuTurbo*™ extends and incorporates the gene into a nicked, circular DNA molecule. 3. The parental vector is digested with *DpnI* and the double nicked, circular, double-stranded DNA is transformed into a suitable host cell. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig 1.2 Schematic diagram of restriction-free cloning

Here a schematic diagram for restriction-free cloning has been shown using a 6 kb dam+ vector. Dpn1 cuts DNA at the methylated GATC sites

In this project to prepare clones with SSR of interest, plasmid pHis17 was used as a vector, which has ampicillin-resistance gene in it, along with T7 promoter, T7 terminator sequences. Between T7 promoter and T7 terminator, the genes with the SSR of interest were inserted. Restriction enzymes Dpn1 (when the vector was purified from dam+ cells) or SauUSI (when the vector plasmids were purified from dcm+ cells). Schematic has been shown in Fig 1.2 from Fusinita van den Ent, Jan Löwe, Journal of Biochemical and

Biophysical Methods,Volume 67, Issue 1, 2006. Here the schematic has been shown for a 6 kb vector. Details are provided in Chapter 2: Materials and Methods.

1.4 Expression strains of *Escherichia coli* BL21(DE3):

To identify probable positive clones from the LB-agar plates after cloning we have used EGFP, about which we shall be discussing in Chapter 2: Materials and Methods.

To get basal level expression of EGFP, we have used expression strain of *E. coli* BL21(DE3).

BL21 (DE3) has DE3 prophage having T7 RNA polymerase under lacUV5 promoter. The leaky expression of T7 RNA polymerase helps in basal level expression of EGFP in the cell. This property was used in cloning extensively in the project.

Plasmid construct used: pHIS17

Plasmid pHis17 was used as vectors where the SSR-stretches had been inserted. It is a high-copy number plasmid i.e., a plasmid, number of which in the cell is 500-700. In a cell. Plasmid pHis17 has T7 promoter and T7 terminator in it with Ampicillin-R in it.

When T7 RNA polymerase is expressed, it binds to the T7 promoter and it expresses the genes in the ORF. Schematic has been shown in Fig 1.3.

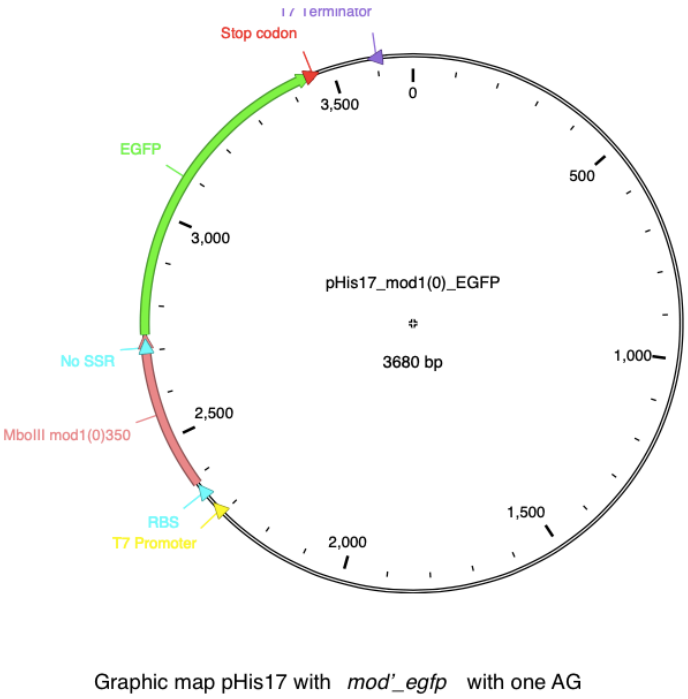


Fig 1.3: pHis17 plasmid with the inserted part of gene *mod* with single AG-repeat with gene *egfp* downstream to it.

Here pHis17 has been shown with the T7 promoter upstream to the inserted gene, T7 terminator downstream to it. This promoter region is activated by the T7 RNA polymerase expressed by the expression of the T7 prophage present in the *E. coli* strain BL21(DE3)

1.5 SSR-motifs studied:

In this thesis, we are presenting studies involving (AG)-repeats, (AT)-repeats, and (G)-repeats. In gene *mod* of type-III restriction-modification system of *Mycoplasma bovis* there is a stretch of AG-repeats. And next to this, there is gene *res*

responsible for cleaving of the non-methylated DNA in the same ORF. Because of the repeats in this SSR of AG-repeats we observe replication slippages which often lead to phase variation of expression of gene *res*.

We have inserted different SSR-motifs replacing SSR regions of a part of gene *mod* (we call this *mod'*) one AG and two AG.

AT-repeats were chosen as another di-nucleotide sequence and G-repeats, to study G-4 structures.

1.5 Secondary structures involved with the motifs of interest:

The AG-repeats are sequences with high G-content. High G-contents are known to form secondary structure G-quadruplexes or G-quartets G-4 complexes (Fig 1.4). The complementary strand here is $(CT)_n$. Cytosines are known to form secondary structures known as i-motifs (Fig.1.5). The repeat-region in question is $(AG/CT)_n$ here, role of which we shall be studying in slippage. To study formation of G-4 complexes from only G-stretches we inserted G-nucleotide sequence which for our study is $(G/C)_n$ stretches.

Apart from this we studied slippage at another di-nucleotide sequence (AT)-repeats, inserting the same at the place of AG-repeats. AT-repeats for hairpin loops or cruciform DNA structures (Fig. 1.6) on one strand and the complementary strand also has AT-repeats. This kind of structures are formed when the nucleotide-sequence are formed of

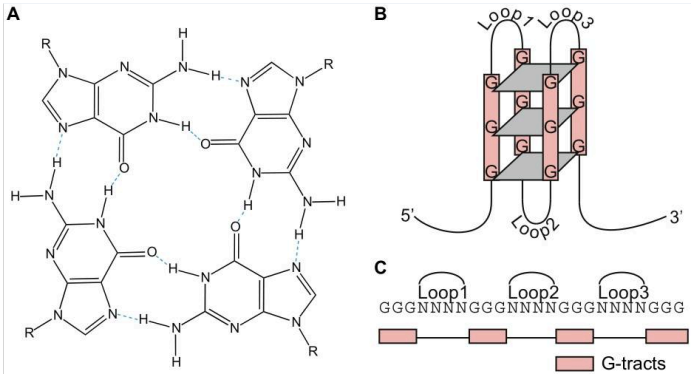


Fig 1.4: Formation of G-4 structures

Nucleotide-sequences rich in guanine are known to form secondary structures known as G-quartets, forming different loops

alternate complimentary bases.

1.6 DNA Polymerases:

We have compared patterns of appearance of slippage bands for some of the SSR-motifs using different polymerases: Taq polymerase, Pfu polymerase, New England Biology®

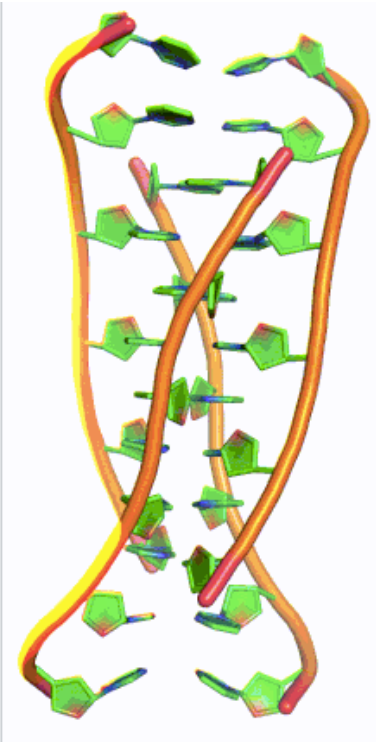


Fig 1.5: Formation of i-motif

Nucleotide-sequences rich in cytosine are known to form secondary structures known as i-motifs, forming different loops

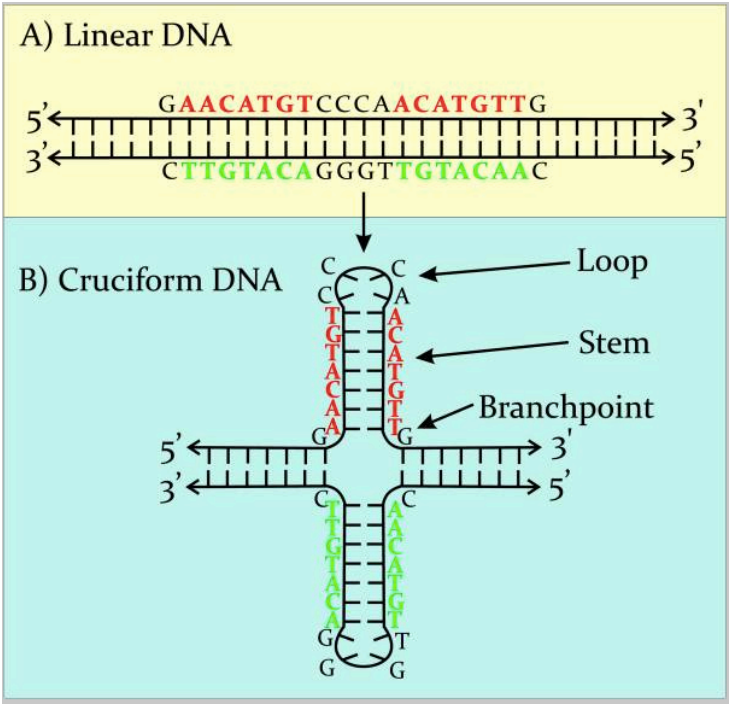


Fig 1.6: Formation of cruciform or hairpin-structure

Formation of cruciform DNA secondary structures by complimentary bases

Q5 HiFi DNA polymerase. Different polymerases have different strand-displacement activity, different fidelity. Fidelity means accuracy of a polymerase to insert correct nucleotides. With higher rate of addition of nucleotides, generally fidelity decreases. Strand-displacement activity means property of a polymerase to move along a DNA strand. With increased strand displacement activity propensity to slip decreases.

1.7 Aim of the project:

Our aim was to look at different patterns of *in vitro* slippage bands appearing for different numbers of repeats of the nucleotide-motifs, and that too for different SSR-sequences. Apart from that, we tried to find threshold for slippage in each cases. We also looked at whether the overall lengths of the SSR-stretches would be affecting the slippage pattern or not, or this phenomenon has something to do with the number of repeats in the SSR.

We have found that using Pfu DNA polymerase, slippage for AT-repeats starts when number of repeats is somewhere between six or eight. For G-repeats it has been seen for eleven G-repeats, and for six AG-repeats slippage was prominent. For five AG-repeats slippage bands were not visible.

Chapter 2: Materials & Methods

We used plasmid pHis17 which has a T7 promoter and downstream to it T7 terminator regions in it with ampicillin-resistance gene.

We inserted a part of gene *mod* with different numbers of repeats of AG-, AT- and G-repeats along with gene *egfp* in same ORF. With this rationale, we are to call the clones of interest for n number of repeats of nucleotide-motif x as pHis17_poly(nx)_egfp (or may refer to them as poly-nx or nx.).

We wanted to use fluorescence of EGFP to look for positive clones. SSR of interest would affect the expressions of EGFP (*refer to annexure*).

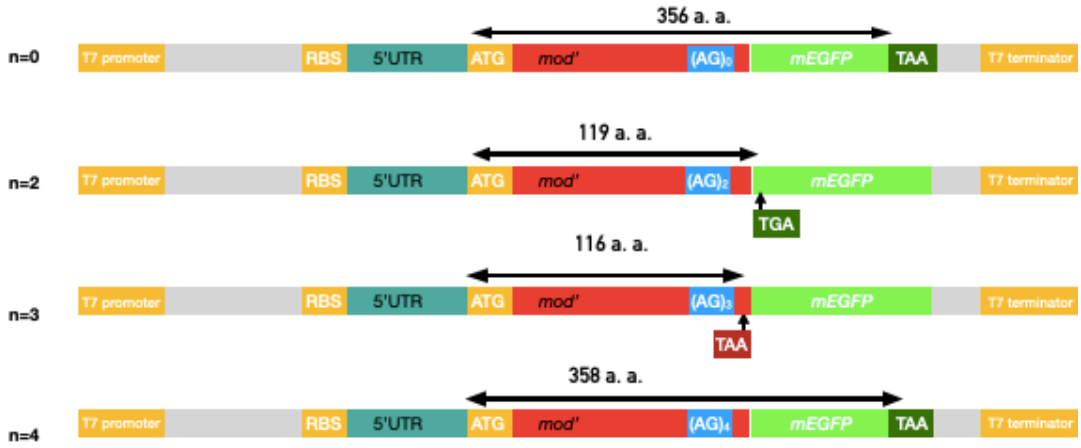


Fig 2.1 Expression of EGFP varies because of SSR causing stop codons to be in ORF

There are stop codons which is not always in frame, but because of the SSR, stop codons do not allow expression of EGFP

Table 2-1: Pattern of EGFP Expression

#Repeats	0	2	3	4	5	6	7	…
EGFP for $(AG)_n$ and $(AT)_n$	On	Off	Off	On	Off	Off	On	…
EGFP for $(G)_n$	Off	On	Off	Off	On	Off	Off	…

2.1 Obtaining clones:

Parts of genes *mod* with the AG-repeat stretches of one AG and two AG (we shall call it *mod'*) were inserted into plasmid pHis17 with gene *egfp* in the same ORF next to the gene *mod'*. In *Mycoplasma bovis,* with the gene *mod* there is gene *res* in the same open reading frame (ORF). At first *mod'* with zero AG-repeats i.e., single AG and two AG were inserted into the plasmid pHis17. The clones were inserted into *E. coli* strains of DH5α or NEB Turbo Chemical competent cells (NTC) doing heat-shock transformation at 37° C or 42° C.

Now, technique of RF-cloning was used to create constructs of *mod*' in pHis17 inserted with different numbers of AT-repeats, AG-repeats and G-repeats, choosing the above mentioned constructs pHis17_mod(0)_EGFP and pHis17_mod(2)_EGFP. Preliminary discussions can be found in *Chapter 1: Introduction.*

2.1.1 CHOOSING VECTOR FOR RF CLONING:

The main part comes to the choice of vectors for successful cloning. During RF cloning, after plating the transformed bacteria we wanted to choose correct colonies for inoculations. Our desired plasmids were pHis17 with *mod*' (mentioned earlier) and EGFP in same ORF with the SSR of particular numbers of repeats. As we have discussed earlier about expression pattern of gene *res* in type III RM system, similar to this, because of presence of the repeat-region or the SSR in *mod*', we are supposed to see variable expression pattern of the gene *egfp*. During overlap-extension PCR or secondary PCR (2° PCR), we would be getting our desired plasmids, mixed with the vectors. After digestion with restriction-enzymes (Dpn1 or SauUSI) of sufficient volume of PCR-products we electroplated the plasmids into *E. coli* cells of strain BL21(DE3). This strain of *E. coli* is an expression-strain, so we would see basal-level expression of the genes present in the plasmid. If for number of repeats, say x, for motif, say y, we get constructs with *egfp* on, we choose $(AG)_2$-clone as vector, and if number of repeats, say a, for motif, say b, we get *egfp* off, we choose $(AG)_0$-clone as the vector.

The pHis17 plasmids with one AG- and two AG-repeats were used as vectors since they do not slip since in $(AG)_0$ there is no repeat and in $(AG)_2$, the stretch is short. The primary PCR reactions were set up according to the tables Table 2-2 and Table 2-3.

The primary PCR products were checked on 1% gel before proceeding further. These are the megaprimers. They were purified using commercially available PCR purification kit

Table 2-2: Reaction Mix for Primary PCR

Component	Control (μl)	Test (3-4 reactions)(μl)
Deionised water	41.5	40.5
Forward primer (20 μM)	1	1
Reverse primer (EGFP_R) (20 μM)	1	1
Vector (100 ng/ μl)	-	1
dNTP (2.5 mM)	1	1
Pfu buffer (10X)	5	5
Pfu polymerase	0.5	0.5

Table 2-3: Thermocycles for Primary PCR

	X30				
95° C	95° C	55° C	72° C	72° C	4° C
2’	30”	45”	2’	3’	Hold

QIAquick® PCR Purification Kit (50). Then the yield of the PCR products were checked. Example of an agarose gel-image is below in Fig 2.2.

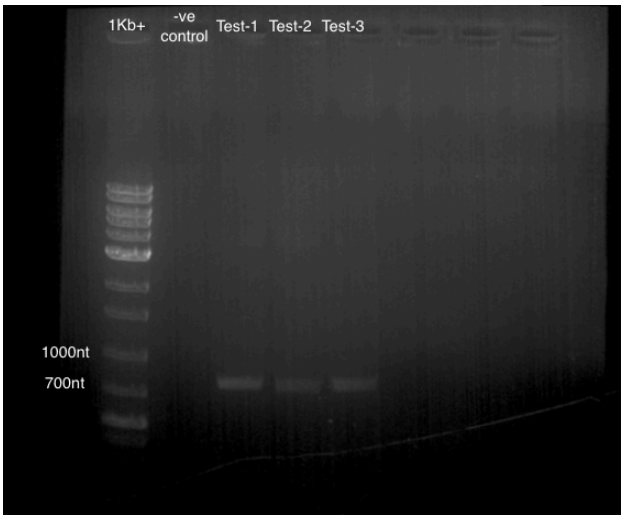


Fig 2.2 1% agarose gel image with poly-2G megaprimers

Checking 1° PCR products during attempt to make poly-2G clones. The lanes left to right have DNA marker, -ve control, and three test reaction samples from poly-2G reactions

After purification of the megaprimers, concentrations were checked. Lengths of the megaprimers were approximately 770 nt. Lengths of ORF in the plasmid

pHis17_mod(0)_EGFP and pHis17_mod(2)_EGFP were approximately 3700 nt (3680 nt). The megaprimers or the inserts for the secondary PCR reactions. Vector and insert were taken in 1:100 molar ratio for the OE PCR or the secondary PCR. The calculations were done according to the following:

If C ng/ μl is the contraction of the purified megaprimers,

Then for vector: insert molar ratio of 1:100, 1 ng of vector corresponds to

(770/3700)*100 ng of insert.

For setting up the PCR reactions, we used template or vector concentrations of 100 ng/ μl.

So for 100 ng i.e., 1 μl of vector, we need (770/3700)*100*100 ng of insert.

So for setting up the secondary reaction, we needed ((770/3700)*100*100/C) μl of insert.

Following the above calculation, we set up the secondary PCR reactions according the the tables Table 2-4, Table 2-5:

Table 2-4: Secondary PCR

Component	Control (μl)	Test(μl)
Deionised water	38.5	38.5-((770/3700)*100*100/C)
Insert	-	((770/3700)*100*100/C)
Vector (100 ng/ μl)	1	1
dNTP (2.5 mM each)	5	5
Pfu buffer (10X)	5	5
Pfu polymerase	0.5	0.5

Table 2-5: Thermocycles for Secondary PCR

	X30				
95° C	95° C	60° C	72° C	72° C	4° C
5’	2’	40”	5’	7’	Hold

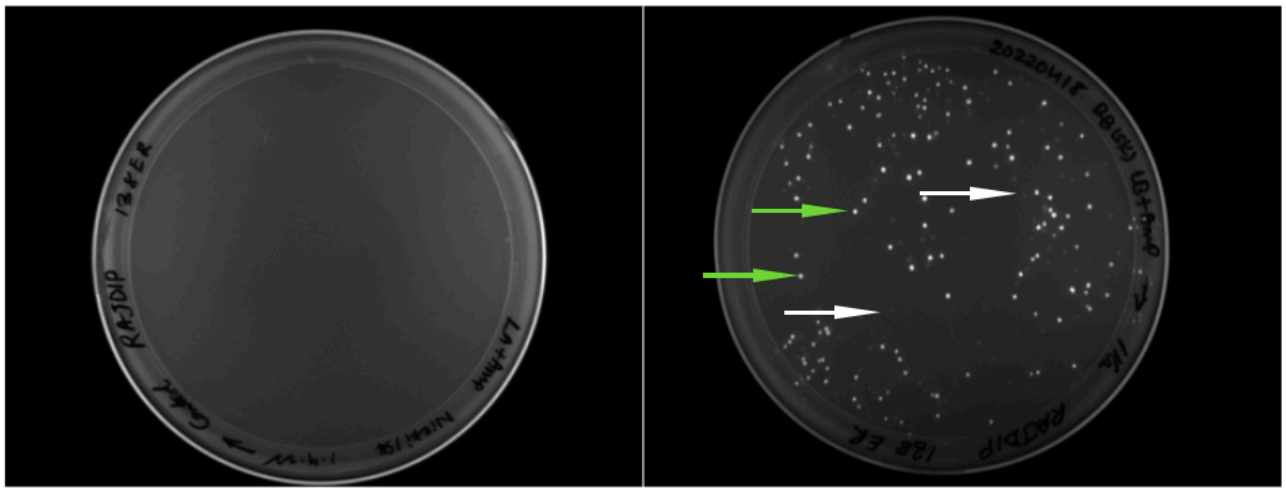
Following the secondary PCR, sufficient amount of the secondary PCR products were taken for restriction-digestion which were to be used for electroporation into expression

strains of *E. coli* BL21(DE3) electro-competent cells. For electroporation 2 mm BioRad single-use electrode cuvettes were used. Using electroporation pulse instrument BioRad Gene Pulsar XCell™ applying 2.5 kV we carried out electroporations.For digestion, we used Dpn1 or SauUSI. The principles of working of these two enzymes are given in the table Table 2-6.

Table 2-6: Restriction-digestion by SauUSI and Dpn1

Enzyme	Plasmid purified from *Escherischia coli* strain	Target recognises
Dpn1	NEB Turbo Chemical Comp cells	5'-GA|TC-3'
SauUSI	DH5α	5'-SC^{5m}NGS-3'

After electroporation, the cells were plated on two LB-ampicilin plates, one for test plasmids, one for control plasmids. There would be no colonies on the control, and many colonies on the test. Images of plates for cloning of Poly-11G are shown bellow in Fig 2.3.



Control plate for Poly-11G clones: No colonies **Test plate for Poly-11G clones: Green colonies to be picked up**

Fig 2.3 Plates after electroporation after secondary PCR

Plates during cloning of poly-11G are shown. Two bright colonies on Test shown marked with green arrows, two white colonies are shown with white arrows

Now based on the status of EGFP ON/OFF of the desired clone, the fluorescent or non-fluorescent colonies were chosen. 10 ml LB with ampicillin in test-tubes were inoculated with picking the chosen colonies. After the test-tube culture became turbid, 5 ml culture was purified using commercially available plasmid-purification kit QIAprep® Miniprep Kit and the clones were sent for sequencing. If the clones turned out to be positive, we

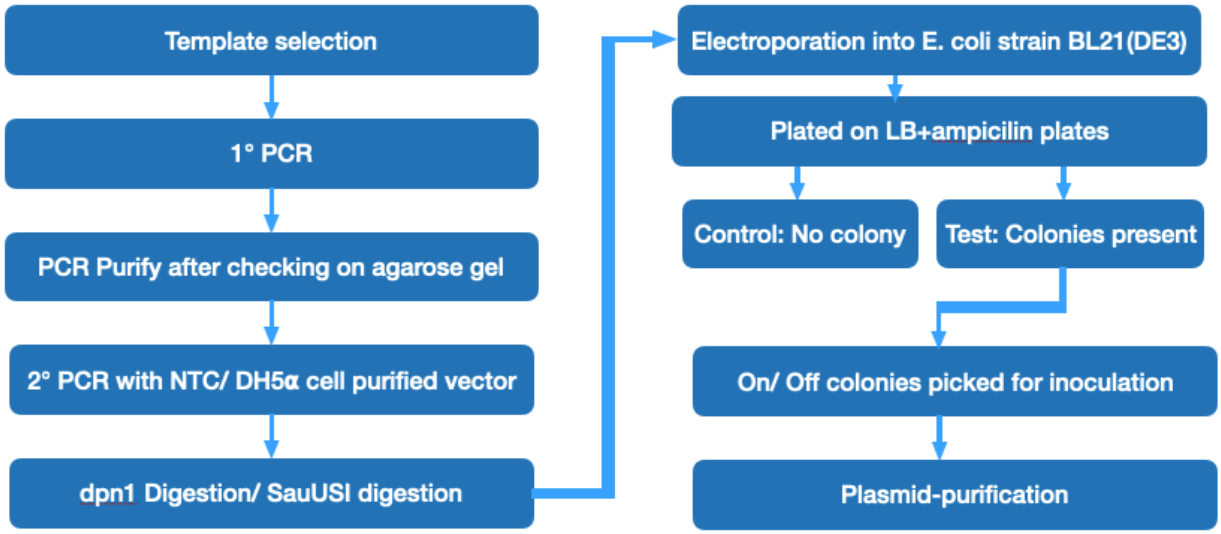


Fig 2.4: Flowchart showing the cloning procedure

carried forward to study the slippage bands after PCR reactions.

2.1.3 CHECKING SEQUENCING RESULTS:

At first from the Sanger sequencing data, reliable regions were chosen and then sequences were checked using software Jalview 2.11.2. 4. During Sanger sequencing also we might encounter slippage, giving false sequencing results. For this reason we looked at the QV for all nucleotides and considered QV> 20.

2.2 Studying slippage bands:

For studying slippage bands we had to separate DNA bands with 1-2 nucleotide difference. After setting PCRs, we proceeded with denaturing urea PAGE to observe the bands. For urea PAGE, we would not be able to

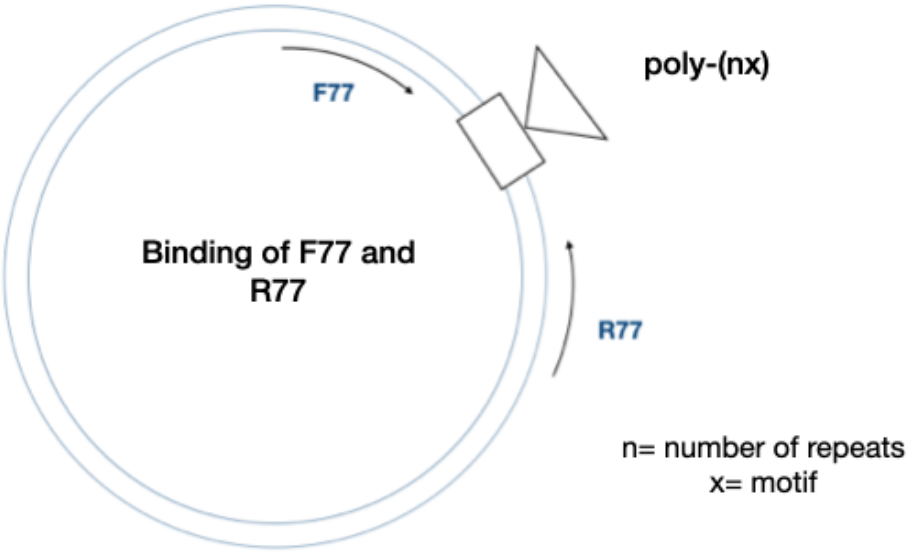


Fig 2.5: Schematic showing binding positions of the primers F77 and R77

F77 binds to the reverse strand and R77 to the forward strand

make the whole plasmid of length 3.7 kb enter the gel. For this reason, one forward and one revers primers were chosen (named them as F77, R77 respectively), such that length of the stretch after binding with the 1AG gives PCR products of length 77 nucleotide (Ref to *Annexure*).

Chapter 3: Results & Discussions

We are going to discuss observations of the cloning process and then we shall be proceeding to the observations of results of experiments to observe slippage.

3.1 Restriction-free cloning:

3.1.1 PRIMARY PCR:

We started with RF-cloning to get different constructs of pHis17_poly(nx)_egfp (n is number of repeats, x is the nucleotide motif of interest).

To gain EGFP on constructs we started with vector pHis17_poly(2AG)_egfp and for clones with EGFP off, pHis17_poly(1AG)_egfp plasmids were chosen as vectors. We set four or five of the primary PCR reactions where one is no polymerase negative control, for each construct.

We loaded the samples from the reaction mix after PCR on 1% agarose gel. Where bands of expected length appeared, we selected the corresponding PCR products and mixed them, and purified using commercially available PCR purification kit QIAquick® PCR Purification Kit (50). Then the yield of the PCR products were checked. This was the megaprimer for a particular construct. 1% agarose gel with megaprimers for poly-2G clones is shown in Fig. 3.1

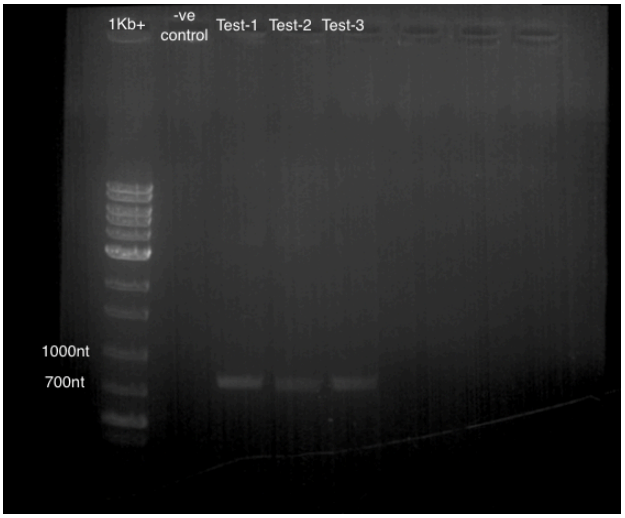


Fig3.1 Checking 1° PCR products during attempt to make poly-2G clones.

Lanes from left to right are loaded with DNA Ladder 1Kb Plus, negative control and three tests of megaprimers of poly-2G

Many times cloning of more than one constructs have been done or the bands on the gel did not run at the same levels, then the bands running at the expected levels where chosen. One such agarose gel image (Fig 3.2) with primary PCR products for Poly-12A and Poly -12G is as below. In this case, Test-1, Test-3, Test-4 for Poly-12G and Test-1, Test-2, Test-4 for the Poly-12A constructs were selected for purification of megaprimers.

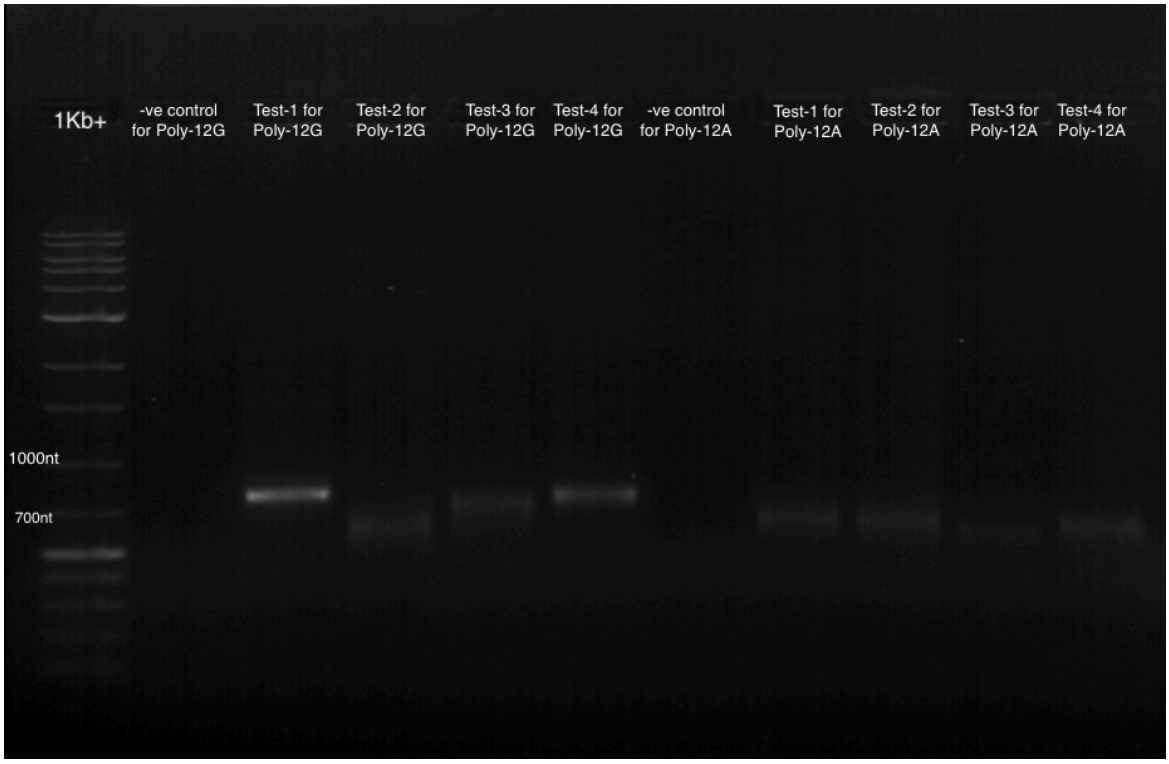


Fig. 3.2: Checking primary PCR products for poly-12A and poly-12G constructs on 1% agarose gel

From left to right: Lane-1has 1Kb+ DNA Ladder, no polymerase negative control for poly-12A, lane-2 to lane-5 have four tests for poly-A primary PCR, lane-6 has no polymerase negative control for poly-12G, lane-2 to lane-5 have four tests for poly-G primary PCR

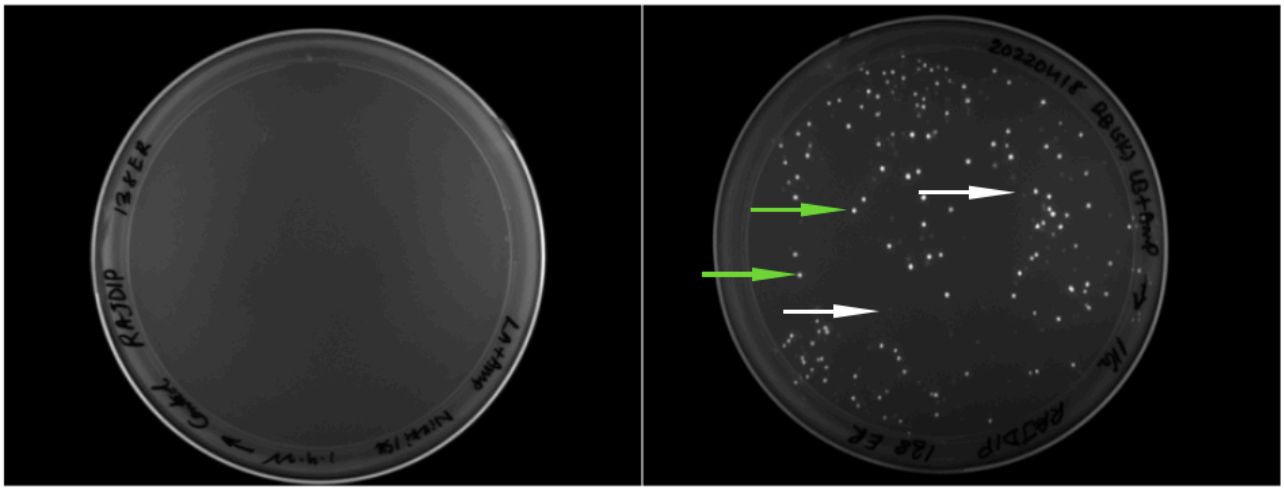
3.1.2 SECONDARY PCR:

After we obtained the megaprimers, their yields were checked, and we proceeded for the secondary PCR or the overlap extension PCR, taking the vector and insert in 1:100 molar ratio. Calculations were done as explained in *Chapter 2: Materials & Methods*. So for setting up the secondary reaction, we needed ((770/3700)*100*100/C) μl of insert.

Along with the test, 'no insert' negative control PCR reaction was set.

After completion of the secondary PCR, based on the source of the vector i.e., if they were DH5α purified or NTC purified, 10 μl of plasmid sample was prepared for restriction

digestion from the 2° PCR products, with the restriction enzymes Dpn1 or SauUSI (following the protocol as mentioned in *Chapter 2: Materials and Methods*). Then the PCR products digested with restriction-enzyme were electroplated into *E. coli* strain BL21(DE3) electrocompetent cells (following protocol mentioned in *Chapter 2: Materials and Methods*). The plasmid pHis17 has ampicillin-resistance gene in it. After electroporation the cells were plated on two LB-amp plates- Control and Test plates. After 12 hour of plating Control did not have any colonies, but the Test got colonies. Using UV radiation, images of the plates were taken. While Control had no colonies, Test had two kinds of colonies, brightly glowing green colonies and white colonies (Fig 3.3).



Control plate for Poly-11G clones: No colonies

Test plate for Poly-11G clones: Green colonies to be picked up

Fig 3.3 Plates after electroporation after secondary PCR

Plates during cloning of poly-11G are shown. Two bright colonies on Test shown marked with green arrows, two white colonies are shown with white arrows

Now depending upon the desired plasmid's EGFP ON/ OFF (in frame or out of frame), the green or white colonies were chosen and were inoculated into 10 ml of LB + ampicillin broth for overnight growth of culture. After the culture became turbid, we proceeded to plasmid purification, which was carried out using commercially available plasmid purification kits (QIAprep® Spin Miniprep Kit (50)) After purification we screened the purified plasmids setting up PCR reactions as per the standardised protocol of PCR with the primers F77 and R77. If the main band (the darkest band) appeared to be of the expected lengths, the plasmids were sent for Sanger sequencing.

3.2 Setting up PCR with primers F77 and R77 with Pfu DNA polymerase:

We set up PCR reactions with F77 and R77 primers for poly-(1AG) and poly-(2AG) to standardise the protocols for the slippage experiments. These two constructs were chosen because they have the lowest tendency to show polymerase slippage. Fig 3.4 shows the image of a 12% urea PAGE gel. It shows no slippage band. The protocol fixed for F77 and R77 is given in table 3-1 with the thermocycle conditions given in table 3-2.

Once the PCR protocols were fixed, we proceeded with PCR with the primers F77 and R77 for different constructs. We aimed for finding the threshold for slippage. Apart from that, we looked for presence of specific patterns of the slippage bands.

Setting up PCR with primers F77 and R77 with Pfu DNA polymerase of AG-clones:

Table 3-1

Component	Control (μl)	Test(μl)
Deionised water	41	40.5
Forward primer (20 μM)	1	1
Reverse primer (20 μM)	1	1
Vector (100 ng/ μl)	1	1
dNTP (2.5 mM each)	1	1
Pfu buffer (10X)	5	5
Pfu polymerase	-	0.5

Table 3-2

Initial denaturation	X20			Final Extension	Hold
95° C	95° C	53° C	72° C	72° C	4° C
1'	30"	20"	30"	1'	∞

Observations:

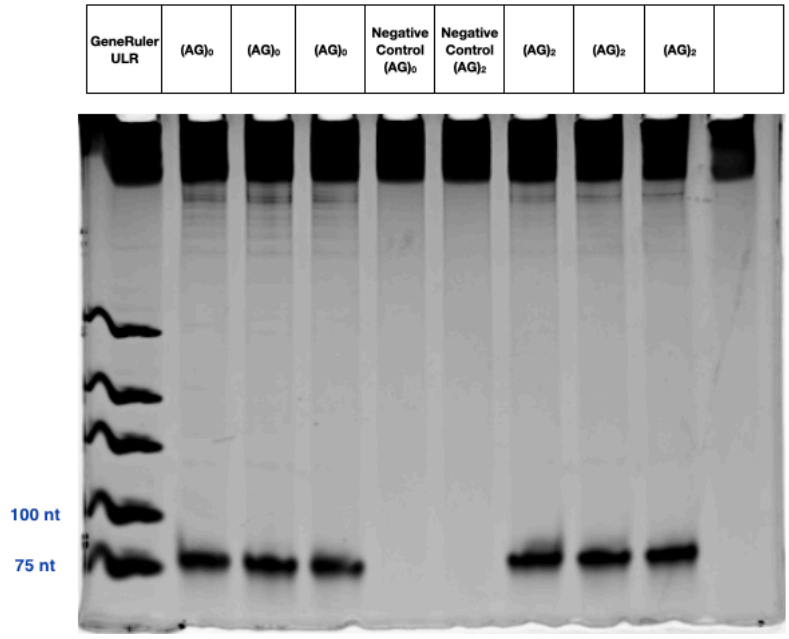


Fig 3.4: Checking the PCR amplified products for 1-AG and 2-AG with 12 % urea PAGE gel

From left to right: DNA marker, tests of poly-(1AG), negative control of poly-(1AG), negative control of poly-(2AG), tests of poly-(2AG). We loaded 1 μl of samples of PCR products amplified with primers F77 and R77 with 6 μl of 2X formamide dye. 1.5 μl of Thermo Scientific Ultra Low Range DNA Ladder with 6 μl of 2X formamide was loaded as ladder. The PCR amplified products of pHis17_Poly(1AG)_EGFP and pHis17_Poly(1AG)_EGFP for F77 and R77 have been shown when no extra bands were present. The polyacrylamide gel is a 12% urea denaturing gel with 7 M urea.

We set up PCR of different clones with the primers F77 and R77. We have found different have found some patterns in the appearance of the stutter bands and near similarity of the threshold for length of the SSR where slippage starts. We have found that for nucleotide-motifs of length around 12 nucleotide leads to slippage using Pfu polymerase. Fig. 3.5, Fig 3.6 are showing gels loaded with samples of different clones. In Fig 3.6, the gel has no DNA ladder since we were checking different patterns. For poly-(6AG,) in Fig 3.5 there is one slippage band appearing, which is very faint. In another set triplicates of PCR reactions with the primers F77 and R77 of the clone poly-(6AG) and poly-(7AG) on a different gel the slippage band has been prominent.The gel-image for these triplicates is shown in Fig 3.7. We could not get positive clone of poly-(9AG) and poly-(10AG). Poly-(7AG) gives more than one slippage band. Poly-(8AG), the next higher slippage bands are more prominent. But we can see there is drastic jump in patterns of the slippage bands

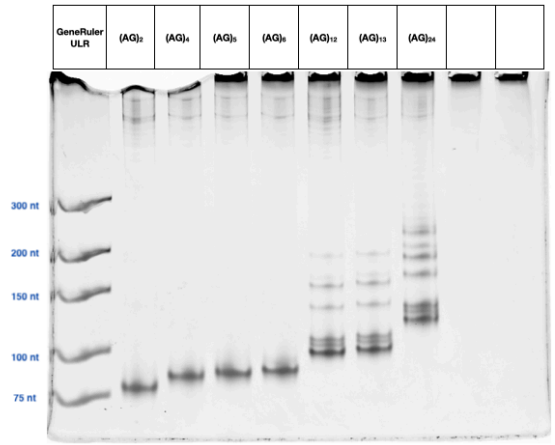


Fig 3.5: Observation of PCR amplified products of AG-clones with 14% urea denaturing PAGE with F77, R77

From left tor right: GeneRuler ULR, PCR products of poly-(2AG), poly-(4AG), poly-(5AG), poly-(6AG), poly-(12AG), poly-(13AG), poly-(24AG). Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V

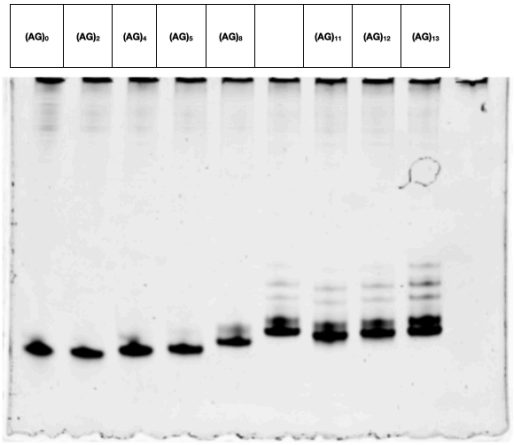


Fig 3.6: Observation of PCR amplified products of AG-clones with 14% urea denaturing PAGE with F77, R77

From left tor right: PCR products of poly-(1AG), poly-(2AG), poly-(4AG), poly-(5AG), poly-(8AG), a clone with poly-(10AG and AT inserts) not mentioned on the image, poly-(11AG), poly-(12AG), poly-(13AG). Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V, ladder not added as here different products were compared.

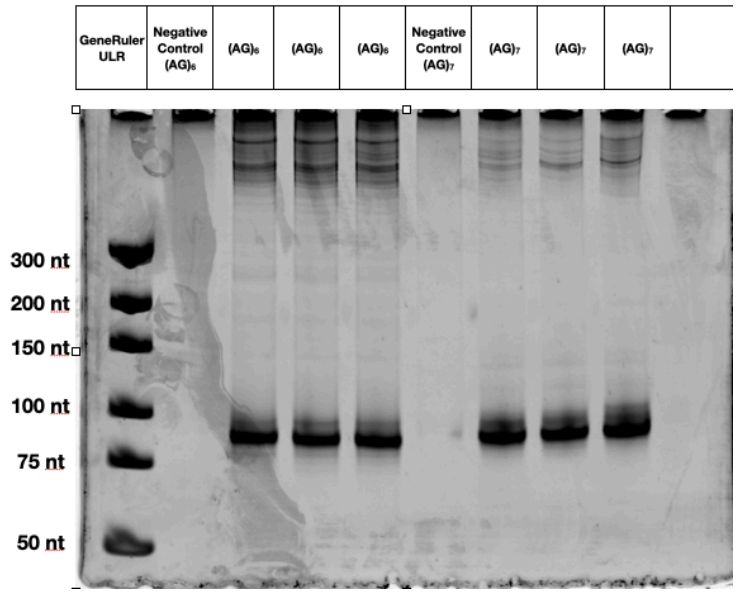


Fig 3.7: Observation of PCR amplified triplicates of poly-(6AG) and poly-(7AG) with 14% urea denaturing PAGE with primers F77 and R77

From left tor right: DNA ladder, PCR products of negative control of poly-(6AG), triplicate reactions of poly-(6AG), negative control of poly-(7AG), triplicate reactions of poly-(7AG). Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V, ladder not added as here different products were compared

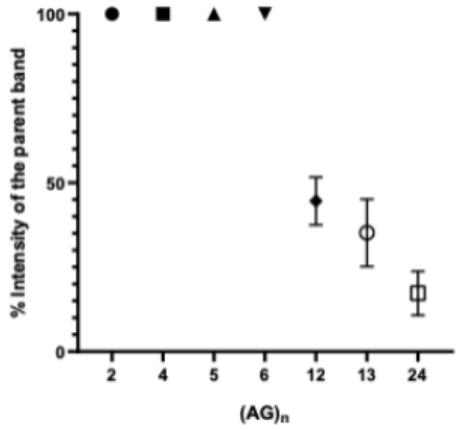


Fig 3.8 Gels of PCR-products of AG-clones were analysed using ImageQuant TL

Analysing the gels of the AG-clones we have found that with increase in number of repeats tendency to slip increases

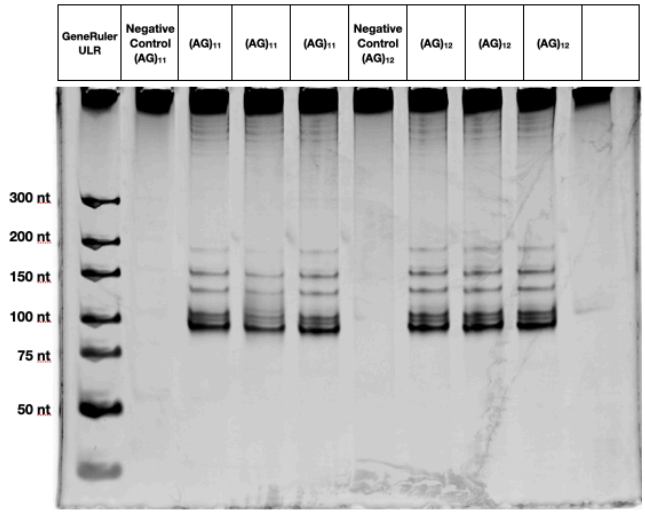


Fig 3.9: 14% urea PAGE gel with Pfu PCR-reactions with F77 and R77 for poly-(11AG) and poly-(12AG)

Lane-4 with PCR-products of poly-(11AG) has a different banding pattern

for the poly-(11AG)

The PCR reaction products shown in the gel of Fig 3.5 were carried out thrice and were checked on three different gels of 14% urea PAGE gel and were analysed using software ImageQuantTL. We calculated the total volume of the bands analysing the gel for each lane and then checked what percentage or fraction of the main band (the most intense band) remained to check propensity to slip. We have found that with increase in number

of repeats retention of PCR product in the main band decreased in each lane, i.e., slippage occurred more.

One peculiar observation we found for of a set of triplicate PCR-reactions with F77 and R77 for poly-(11AG) and poly-(12AG).For one of the lanes of poly-(11AG), the pattern of slippage bands were completely different, where in the same gel patterns of the other two PCR-reaction products of poly-(11G) appeared similar to those of PCR-products with F77 and R77 of poly-(12AG). The gel image is shown in Fig 3.9. Here lane-4 has PCR-products of poly-(11AG) with three immediate slippage expansion bands but in all other tests (those of poly-(11AG) and poly-(12AG)) have two immediate slippage bands.

Similarly, we checked slippage with Pfu polymerase for AT-repeats. We could prepare

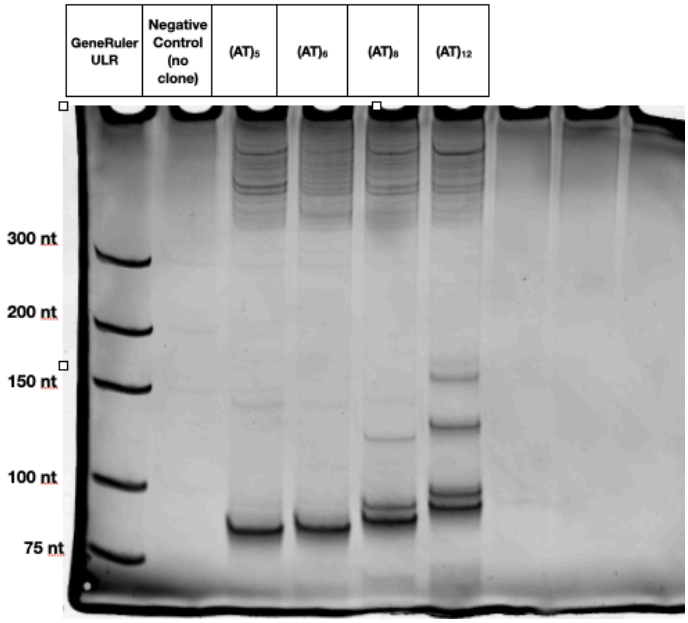


Fig 3.10: Observation of PCR amplified products of AT-clones with 14% urea denaturing PAGE with F77, R77

From left tor right: DNA ladder, PCR products of negative control without any template, poly-(5AT), poly-(6AT), poly-(8AT), poly-(12AT) for reactions with F77 and R77. Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V, ladder not added as here different products were compared

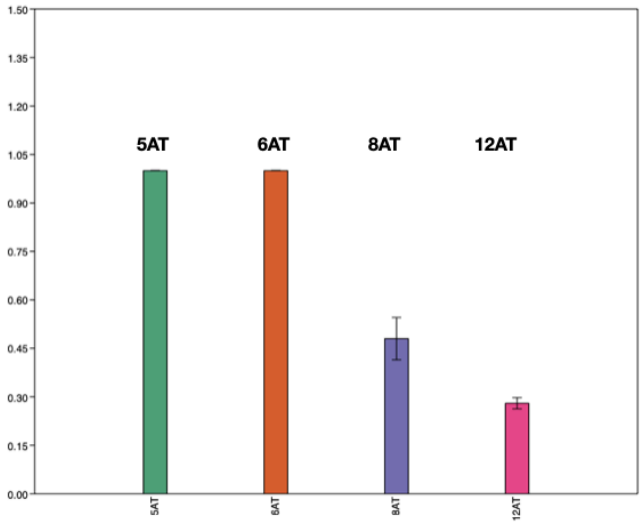


Fig 3.11: Retention of PCR amplifies DNA in main band

With increase in number of AT-repeats, fractions of PCR-products in the main bands have been found to have decreased.

clones for poly-(5AT), poly-(6AT), poly-(8AT), poly-(12AT). For poly-(12AT) clone shown here, there are deletions of three nucleotides between the AT-repeats and the binding site

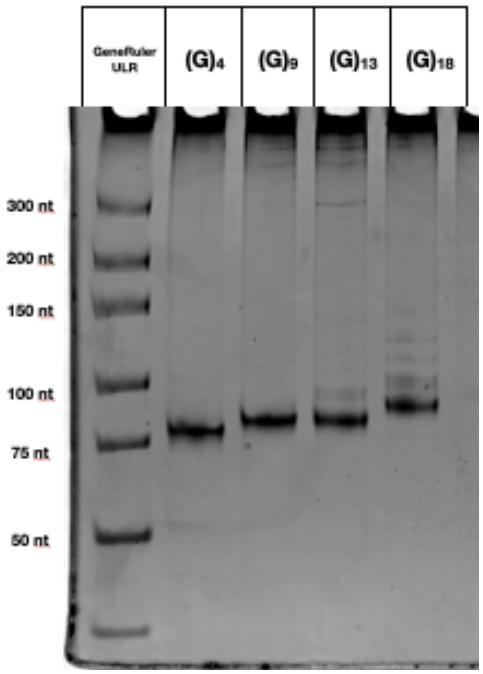


Fig 3.12: 14% urea PAGE to observe slippage for G-clones

In this image it is seen that for G-clones slippage occurs for poly-(13G), not for poly-(9G)

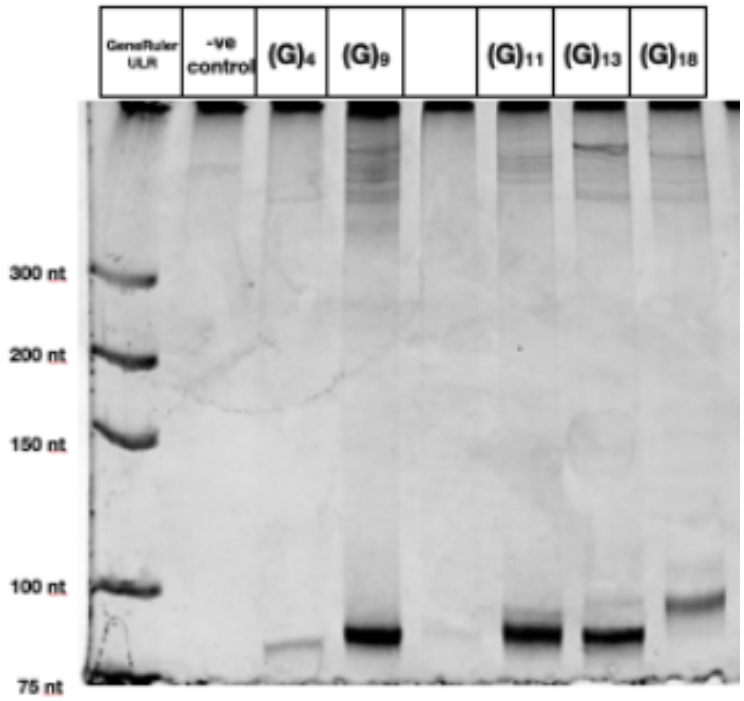


Fig 3.13: 20% urea PAGE to observe slippage for G-clones

The gel was run for 7 hrs, to separate the bands and because of heating smeared bands appeared. For poly-(11G) slippage is visible here

of F77, the forward primer (Refer to Annexure). This should not affect the propensity to slippage, but the bands shall appear two-nucleotide shorter. One image of 14% urea PAGE gel is shown in Fig 3.10. There were no slippage bands present for products of PCR with F77 and R77 for poly-(5AT) and poly-(6AT). For poly-(8AT) and poly-(12AT), we have found appearance of slippage bands. The threshold for slippage remains at 7AT-repeats or 8AT-repeats. For 8AT-repeats the first slippage band was quite intense comparatively to that of 6AG-repeats and one more band is present next to the first slippage band. There is a high-possibility for 7-AT to be the threshold for slippage. This was replicated thrice independently. Similar to the analysis of retention of the DNA products in the main bands for the AG-clone, we calculated the fraction for the AT-repeats using Image-QuantTL.

For G-clones, we checked slippage of Pfu polymerase only once since they were hard to differentiate because of being mono-nucleotide motifs. The gel images are in Fig 3.12, and Fig 3.13.

3.3 Setting up PCR with primers F77 and R77 with other DNA polymerases for some clones:

We have also compared patterns of slippage bands appearingfor different DNA polymerases namely, Q5 HiFi DNA Polymerase, Taq DNA Polymerase for some of the clones. We set up reactions following a standard protocol keeping the conditions as much as similar to those for the PCR reactions with Pfu Polymerase and to decrease non-specific binding of the primers.

The reactions were set up according to the protocol mentioned in tables Table 3-3, Table 3-4, Table 3-5 and Table 3-6.

With Q5 HiFi DNA Polymerase we set up PCR reactions for the 24-AG clones and we have found that the intensities for the immediate two slippage bands next to the main band for the Q5 reaction product were much more than those for the Pfu products of reaction with Q5 polymerase, the intensities were less and for higher bands they might have been too low to be visible clearly. Fig 3.14 show the gel image for these bands.

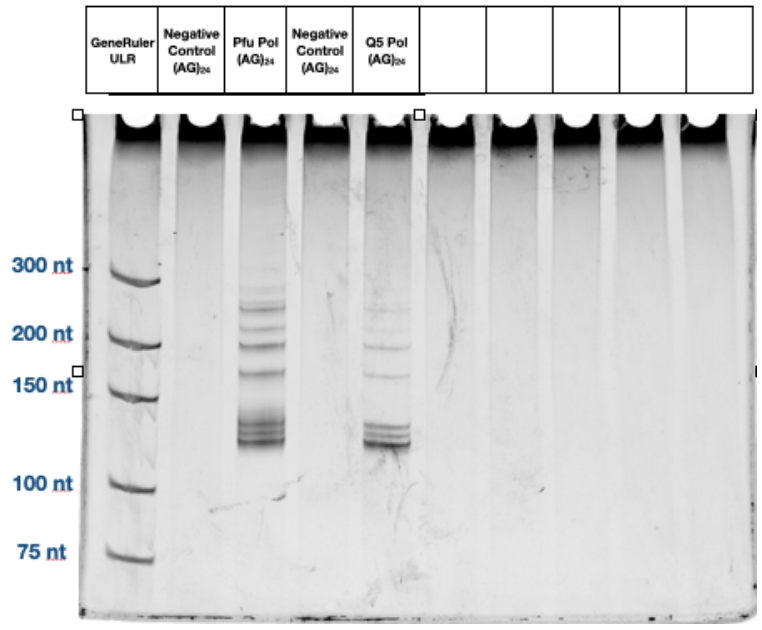


Fig 3.14: Observation of PCR amplified products with Pfu polymerase and Q5 HiFi Polymerase of poly-(24AG) with 14% urea denaturing PAGE

From left tor right: GeneRuler ULR, PCR products of no polymerase negative control of poly-(24AG), test with Pfu polymerase PCR of 24AG-clone, negative control for PCR of 24-AG clone for Q5, test with Pfu polymerase PCR of 24AG-clone. Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V

Table 3-3: Reaction Table for PCR with NEB® Q5 DNA polymerase

Component	Control (μl)	Test(μl)
Deionised water	36	35.7
Forward primer (F77) (20 μM)	1	1
Reverse primer (R77) (20 μM)	1	1
Vector (100 ng/ μl)	1	1
dNTP (2.5 mM each)	1	1
Q5 buffer (5X)	10	10
DNA polymerase	-	0.3

Table 3-4: Thermocycle for Q5 Polymerase with primer F77, R77

Initial denaturation	X20			Final Extension	Hold
95° C	95° C	53° C	72° C	72° C	4° C
30”	10”	20”	3”	3”	∞

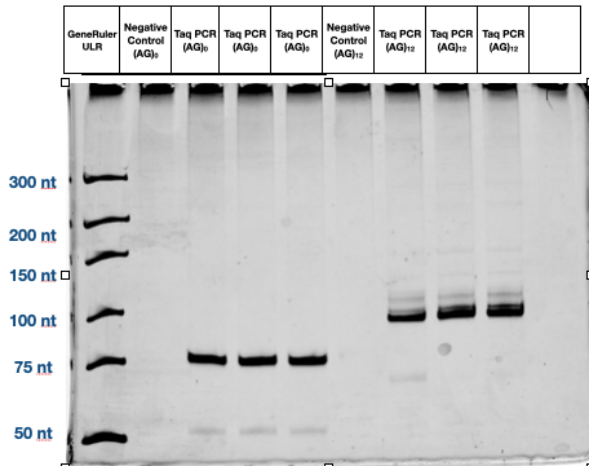


Fig 3.15: Observation of PCR amplified products by Taq polymerase and poly-(1AG) and poly-(12AG) with 14% urea denaturing PAGE

From left tor right: DNA ladder, Taq PCR products of negative control of poly-(1AG), triplicate reactions of poly-(1AG), negative control of poly-(12AG), triplicate reactions of poly-(12AG). Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V, ladder not added as here different products were compared

Table 3-5: Reaction Table for PCR with Taq Polymerase

Component	Control (μl)	Test(μl)
Deionised water	41	40.5
Forward primer (20 μM)	1	1
Reverse primer (20 μM)	1	1
Vector (100 ng/ μl)	1	1
dNTP (2.5 mM each)	1	1
Taq buffer (10X)	5	5
DNA polymerase	-	0.5

Table 3-6: Thermocycle for Taq Polymerase with primer F77, R77

Initial denaturation	X20			Final Extension	Hold
95° C	95° C	53° C	68° C	68° C	4° C
1’	30”	20”	15”	1’	∞

For Taq polymerase, we have found for poly-(12AG) the PCR products have more intense main band. Patterns of slippage band are also different for this case. The second slippage band looks to be of longer length than for the case of PCR-reaction with Pfu DNA polymerase. For poly-(1AG), no slippage band is appearing for Taq polymerase as expected. Fig 3.15 shows the urea PAGE gel image.

3.4 Comparing PCR slippage with primers F77 and R77 with Pfu DNA polymerase with different constructs of SSR:

We checked slippage of Pfu polymerase at SSR-motifs of poly-(12AT) and poly-(12AG) to compare the patterns. The gel image is shown in Fig 3.16. It is observed that though the immediate slippage bands were showing similar patterns for the constructs poly-(12AG) and poly-(12AT), extra longer slippage bands were appearing in the lanes of PCR-products of construct poly-(12AG).

3.5 Conclusion:

Some important observations are:

We have seen that for AG-repeats, from 6AG-repeats, slippage starts. 7AG-repeats and 8AG-repeats give more than one bands and there is a significant difference in the slippage-band patterns for 11AG and higher number of repeats. For AT-repeats slippage most probably starts from 7AT, since for 8AT we saw more than one slippage-band, and for 6AT, there was not a single slippage-band visible. For G-repeats 11G was leading to slippage of Pfu polymerase. The entire observation might indicate a close relation between propensity to slip and total length (number of nucleotides) of the SSR. Also,

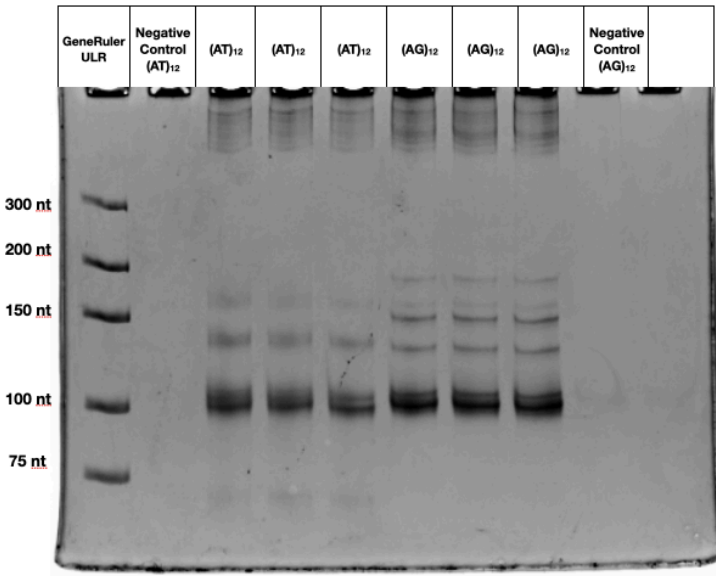


Fig 3.16: Observation of PCR amplified products by Pfu polymerase of poly-(12AT) and poly-(12AG) with 14% urea denaturing PAGE

From left tor right: DNA ladder, negative control of poly-(12AT), triplicate reactions of poly-(12AT), triplicate reactions of poly-(12AG), negative control of poly-(12AG). Loaded in each lane 1 μl of sample with 6 μl of 2X formamide dye, the last lane was loaded with DNA-loading dye containing xylene cyanol. The gel was run at 100 V preceded by pre-run for 1 hr at 150 V, ladder not added as here different products were compared

previously it has been observed that slippage threshold for CA-repeats is from 6CA-repeats, which is similar to our results of threshold of 6AG-repeats for slippage. We have also seen that for Taq DNA polymerase which has lower fidelity (that is probability of incorporating wrong nucleotide is more) than Pfu pol, it leads to lower slippage, suggesting more possibility of stalling. The immediate next slippage band for Taq polymerase for 12AG-clone was also further away from the main band comparing to main band of the 12AG-clone's Pfu PCR. Interestingly for the conditions used for the PCR reactions, for G-repeats, AG-repeats, or AT-repeats we have not seen any contraction. PCR with Q5 Polymerase showed more intense bands with less intense immediate next bands. Understanding propensity to slip, we may be able to guess how mutation rates can be.

Annexure

Primers used were:

F8AT	CTTAAAAATCTTATAGTGATCG**ATATATATATAT ATAT**TCACAGTCAGAGACTG
F5AT	CTTAAAAATCTTATAGTGATCG**ATATATATAT**TC ACAGTCAGAGACTG
F14G	CTTAAAAATCTTATAGTGATCG**GGGGGGGGG GGGGG**TCACAGTCAGAGACTG
F12AT	CTTAAAAATCTTATAGTGATCG**ATATATATATAT ATATATATATAT**TCACAGTCAGAGACTG
F6AT	CTTAAAAATCTTATAGTGATCG**ATATATATATAT** TCACAGTCAGAGACTG
F7AT	CTTAAAAATCTTATAGTGATCG**ATATATATATAT AT**TCACAGTCAGAGACTG
F8G	CTTAAAAATCTTATAGTGATCG**GGGGGGGG**TC ACAGTCAGAGACTG
F11G	CTTAAAAATCTTATAGTGATCG**GGGGGGGGG GG**TCACAGTCAGAGACTG
F4AT	CTTAAAAATCTTATAGTGATCG**ATATATAT**TCAC AGTCAGAGACTG
F6AT	CTTAAAAATCTTATAGTGATCG**ATATATATATAT** TCACAGTCAGAGACTG
F7AT	CTTAAAAATCTTATAGTGATCG**ATATATATATAT AT**TCACAGTCAGAGACTG
3AG	CTTAAAAATCTTATAGTGATCGA**AGAGAG**TCA CAGTCAGAGACTG
4AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAG**TC ACAGTCAGAGACTG
5AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG** TCACAGTCAGAGACTG
6AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AG**TCACAGTCAGAGACTG
7AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAG**TCACAGTCAGAGACTG
8AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAGAG**TCACAGTCAGAGACTG
9AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAGAGAG**TCACAGTCAGAGACTG
10AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAGAGAGAG**TCACAGTCAGAGACTG
11AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAGAGAGAGAG**TCACAGTCAGAGACTG
12AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAGAGAGAGAGAG**TCACAGTCAGAGACTG
13AG	CTTAAAAATCTTATAGTGATCG**AGAGAGAGAG AGAGAGAGAGAGAGAG**TCACAGTCAGAGAC

F9G	CTTAAAAATCTTATAGTGATCG**GGGGGGGGG**T CACAGTCAGAGACTG
F10G	CTTAAAAATCTTATAGTGATCG**GGGGGGGGGG G**TCACAGTCAGAGACTG
F19G	CTTAAAAATCTTATAGTGATCG**GGGGGGGGG GGGGGGGGGG**TCACAGTCAGAGACTG
F4G	CTTAAAAATCTTATAGTGATCG**GGGG**TCACAG TCAGAGACTG
F4G	CTTAAAAATCTTATAGTGATCG**GGGG**TCACAG TCAGAGACTG
F77	GCGAAAACTATGATGCTCTTAAAAAT
R77	CTTGCTCACCTCGTAATTAACAG

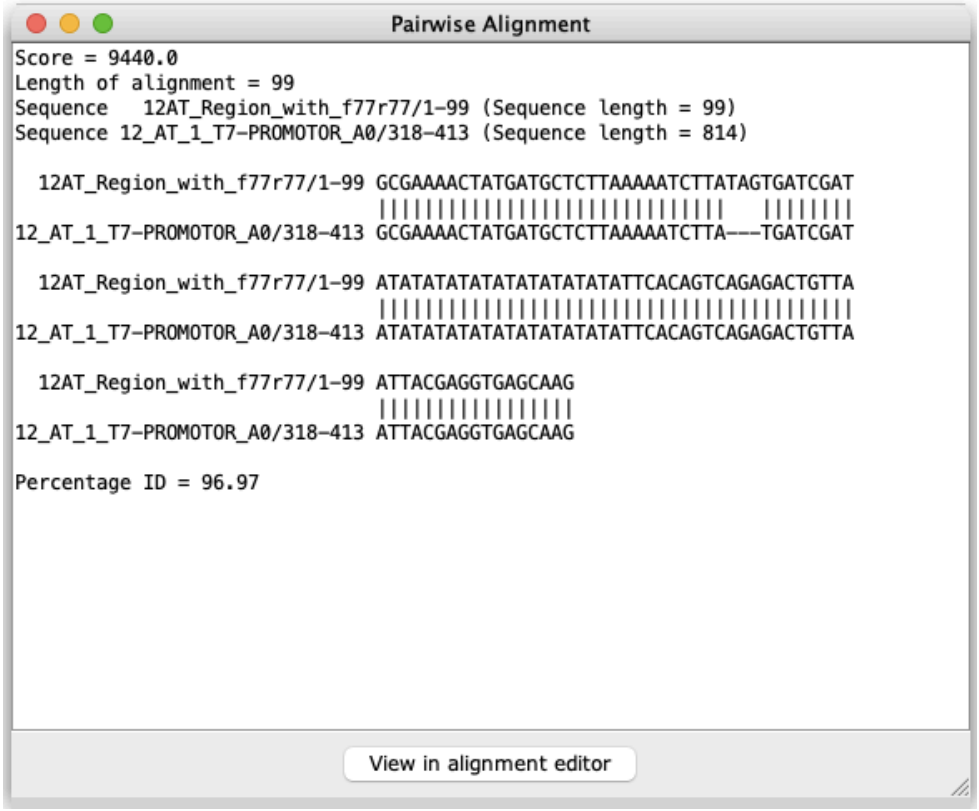
Sequences

Sequence of pHis17_poly(1AG)_EGFP:

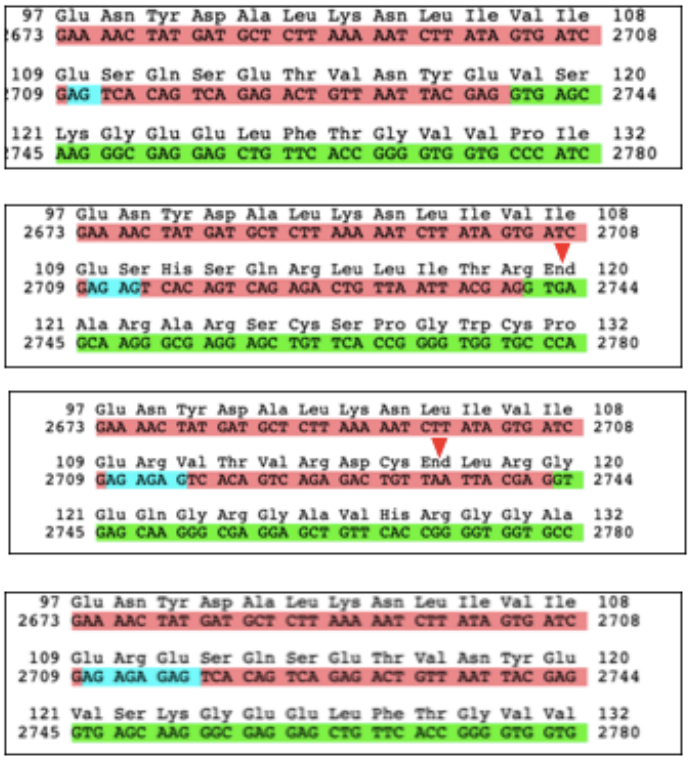
TGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCG
CGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA
TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGC
TGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGA
TGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACT
TGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGG
CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCG
GAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTA
CTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCT
GATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGG
GGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATAT
ACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCC
ACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTAC
CAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGT
AGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAG
TCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGA
GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG
CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGA
GCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTT
GCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGA
GCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGGTGCACTCTCAGT
ACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGC
TGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTC
ATCACCGAAACGCGCGAGGCAGGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTA
ACTTTAAGAAGGAGATATACATATGAATACTATAAAACAGGATTATATAGAAAAAGCTAATGCCTTAAGCCTTAGCAATGAACTAAATCAAGATCAAAAAGATC
TTATTCTTTCAATCATTGATAAATTTGAGGACAACGACCCAGCTTTACATAATGTTTACCAACTTTTAATCAAAAGAGTAAAACTAGGTTTTGTATTTGATATT
GCTCCTTCAGTTAATGCTTCTGAAATAGCTTTATTCAAAAAGGATGAAAAATTATCGTTTAATAATGACAATAACAAGCCTACTAATACACTAATAATTG**GCGA**
AAACTATGATGCTCTTAAAAATCTTATAGTGATCG**AG**TCACAGTCAGAGA**CTGTTAATTACGAG**GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG
GTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTAC
CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGG
GGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
AGCACCCAGTCCAAGTTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
GGACGAGCTGTACAAGTAATGGATCCCATCATCATCATCATCATTAAAAGCTTCACCACCACCACCACCACTAAGAATTCCATCATCATCATCATCATTAA
CGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTAGGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGT
CTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATAATTCT

Colours indicate:
Yellow : T7 promoter
Cyan : Ribosome binding site
Blue : *mod*'
Orange : Repeats of (AG), (AT) or (G) were inserted here, in *mod* this place has AG repeats (here 1-AG)
Green : EGFP in the same ORF
Red : Stop codon for (AG)
In bold, binding sites of F77, R77

Deletion of three nucleotides in pHis17_poly(12AT)_EGFP:



Stop codons in frame for some constructs:



References

1. Bond SR, Naus CC. RF-Cloning.org: an online tool for the design of restriction-free cloning projects. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W209-13. doi: 10.1093/nar/gks396. Epub 2012 May 8. PMID: 22570410; PMCID: PMC3394257
2. Fox KL, Srikhanta YN, Jennings MP. Phase variable type III restriction-modification systems of host-adapted bacterial pathogens. Mol Microbiol. 2007 Sep;65(6):1375-9. doi: 10.1111/j.1365-2958.2007.05873.x. Epub 2007 Aug 21. PMID: 17714447
3. Phillips ZN, Husna AU, Jennings MP, Seib KL, Atack JM. Phasevarions of bacterial pathogens - phase-variable epigenetic regulators evolving from restriction-modification systems. Microbiology (Reading). 2019 Sep;165(9):917-928. doi: 10.1099/mic.0.000805. Epub 2019 Apr 17. PMID: 30994440
4. Fusinita van den Ent, Jan Löwe,. RF cloning: A restriction-free method for inserting target genes into plasmids, Journal of Biochemical and Biophysical Methods,Volume 67, Issue 1, 2006, Pages 67-74, ISSN 0165-022X, https://doi.org/10.1016/j.jbbm.2005.12.008
5. Bayliss CD, Field D, Moxon ER. The simple sequence contingency loci of Haemophilus influenzae and Neisseria meningitidis. J Clin Invest. 2001 Mar;107(6):657-62. doi: 10.1172/JCI12557. PMID: 11254662; PMCID: PMC208953
6. Viguera, E., Canceill, D., & Ehrlich, S. D. (2001). In vitro replication slippage by DNA polymerases from thermophilic organisms. *Journal of molecular biology*, *312*(2), 323-333.
7. Canceill, D., Viguera, E., & Ehrlich, S. D. (1999). Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency. *Journal of Biological Chemistry*, *274*(39), 27481-27490.
8. Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*, *4*(3), 203-221.
9. Li, Y. C., Korol, A. B., Fahima, T., & Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution*, *21*(6), 991-1007.
10. Gemayel, R., Vinces, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual review of genetics*, *44*, 445-477.
11. Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews genetics*, *5*(6), 435-445.

12. Shinde, D., Lai, Y., Sun, F., & Arnheim, N. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis:(CA/GT) n and (A/T) n microsatellites. *Nucleic acids research*, *31*(3), 974-980.
13. Petruska, J., Hartenstine, M. J., & Goodman, M. F. (1998). Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *Journal of Biological Chemistry*, *273*(9), 5204-5210.
14. Schlotterer, C., & Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic acids research*, *20*(2), 211-211.
15. Shioi, S., Shimamoto, A., Nakagami, Y., Qin, L., Shimokawa, M., & Oda, S. (2021). Precision length determination and in silico simulation in PCR of microsatellite repeat sequences. *Electrophoresis*, *42*(12-13), 1323-1332.
16. Murat, P., Guilbaud, G., & Sale, J. E. (2020). DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome biology*, *21*(1), 1-26.
17. Rao, D. N., Dryden, D. T., & Bheemanaik, S. (2014). Type III restriction-modification enzymes: a historical perspective. *Nucleic acids research*, *42*(1), 45-55.
18. Abou Assi H, Garavís M, González C, Damha MJ. i-Motif DNA: structural features and significance to cell biology. Nucleic Acids Res. 2018 Sep 19;46(16):8038-8056. doi: 10.1093/nar/gky735. PMID: 30124962; PMCID: PMC6144788.
19. Capra JA, Paeschke K, Singh M, Zakian VA. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. PLoS Comput Biol. 2010 Jul 22;6(7):e1000861. doi: 10.1371/journal.pcbi.1000861. PMID: 20676380; PMCID: PMC2908698.
20. Brázda V, Laister RC, Jagelská EB, Arrowsmith C. Cruciform structures are a common DNA feature important for regulating biological processes. BMC Mol Biol. 2011 Aug 5;12:33. doi: 10.1186/1471-2199-12-33. PMID: 21816114; PMCID: PMC3176155.