DESAI, PARTH H., M.S. Selective Enrichment of Rare Mutations as a New Biotechnology to Study DNA Mismatch Repair Processes in Bacteria. (2020) Directed by Dr. Eric A. Josephs. 115 pp.

DNA mismatch repair (MMR) is the key process which ensures the incorporation of correct nucleotides during DNA replication by recognizing and removing of incorrectly paired nucleotides from DNA. DNA replication can introduce a mismatched nucleotide at a rate of 10^{-5} to 10^{-6} nts per replication cycle. If this mismatch is not corrected, then it becomes a permanent mutation after the next round of replication. Understanding the MMR mechanism can yield important insights into many aspects of human health, like the emergence of cancer and drug resistance in bacteria. To overcome experimental challenges with studying this process in living cells, we have developed a new method to enrich the rare genomic mutations by genotypic selection in a way that allows us to study the mismatch repair process in *Escherichia coli*, a model organism for MMR. We have shown the maximum 705,000-fold enrichment of DNA with a mutation even after a 10^{-6} times dilution by DNA with the wild-type sequence *in vitro*. After further optimization, we could then use this technique to directly measure MMR activity occurring in living *E. coli (in vivo)*. We expect this technique will open up new opportunities and research directions to study MMR-like processes in *E. coli* as well as different organisms, including Actinobacteria.

SELECTIVE ENRICHMENT OF RARE MUTATIONS AS A NEW

BIOTECHNOLOGY TO STUDY DNA MISMATCH REPAIR

PROCESSES IN BACTERIA

by

Parth H. Desai

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2020

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis written by Parth H. Desai has been approved by the following

committee of the Faculty of The Graduate School at The University of North Carolina at

Greensboro.

Committee Chair   _____

Committee Members   _____

                                                           _____

_____
Date of Acceptance by Committee

04/03/2020_____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Introduction to the Thesis:

For my master's thesis, I have been developing new biotechnologies to study how cells avoid genetic mutations during replication. This mutation avoidance is a Mismatch Repair (MMR) process which is a key guardian of genetic information. Proteins associated with MMR recognize mismatched nucleotides on the DNA, then remove and restore the original genetic information before the incorrect nucleotide can be passed on to the next generation. Here, we have developed a molecular technique to isolate rare mutations from a population in a way that allows us to directly quantify MMR activity as it occurs in living cells (*in vivo*). This technique makes use of a 'genotypic' screen, so we can isolate the rare mutations from the DNA itself, rather than 'phenotypic' screens that require those mutations to change the phenotype of the organism. In our technique, we continuously degrade unmutated or wild-type DNA using a thermostable restriction enzyme while using a polymerase chain reaction (PCR) to enrich the DNA with the mutation. First, we validated this approach to enrich the rare mutations on the plasmid DNA containing mutations of the recognition site of the thermostable restriction enzyme and found it to be highly selective during this assay.

We found this approach produced a maximum of a 705000-fold enrichment of the mutated to wild-type DNA. Then, we applied this approach to a technique to study MMR in living *Escherichia coli*, a model organism to study MMR by using a gene-editing to introduce mismatched nucleotides directly into its genome. Previous approaches to this technique used a phenotypic screen to identify if the mismatch had been repaired, but for the first time, we could use genotypic selection. Here, the data and the analysis of the experiment and a detailed description of the protocol is described and discussed. The results suggest that this approach could be a powerful new tool to address several questions about how MMR is coordinated in living cells that would otherwise be nearly impossible to study.

1.2 DNA Replication and DNA Damage Repair In *Escherichia coli*:

1.2.1 <u>DNA Replication In *E.coli*</u>:

DNA stores the genetic information in living organisms and serves as a template for replication and transcription. The *E. coli* genome contains over 4.6M basepairs (bps) [11]. Replication of chromosomal DNA is a well-orchestrated process requires the assembly and action of molecular machines known as replisome that coordinate all the necessary enzymatic activities for DNA replication. DNA replication in *E.coli* initiates at the origin of replication [12] and proceeds bidirectionally resulting in two "replication forks" that travel in opposite directions from the origin. In 1963, John Cairns [13] reported evidence of replicating the chromosomal DNA with two replication forks, the active DNA synthesis site.

At each of these sites, parental DNA must be unwound into the two template strands that are copied simultaneously with DNA polymerase by extending preexisting primers from the 3'-OH end. This process is highly efficient and well maintained. At the replication fork, the two nascent DNA strands, known as 'leading' and 'lagging,' are synthesized differently. The leading strand is synthesized continuously, while the lagging strand is synthesized discontinuously as short Okazaki fragments [14]. Both strands are carried out by the replisome even if they differ in their processing nature.

DNA replication starts when DnaA recognizes and binds to specific sequences within the OriC region and unwinds that region. Then, DnaA interacts with DnaB helicase in loading the DnaB-DnaC complex on each strand. DnaG is a primase which produces primers which will be extended during replication. DnaG function dissociates DnaC-DnaB complex and then DnaB helicase unwinds parental DNA strand by interactions with cellular replication enzymes, DNA polymerase III holoenzyme and primase as primers have formed that help in extension and duplication of the chromosome [15]. The replisome duplicates DNA with a very low mutation rate of 10^{-5} to 10^{-6}- bp misincorporation of nucleotides per genomic replication [16]. Erroneous or unrepaired mismatches become permanent mutations if they are not repaired.

1.2.2 Repair of Misincorporated Nucleotides During Replication:

DNA Mismatch Repair (MMR) pathway is involved in the repair of mis-incorporated nucleotides due to errors in replication or incorrect recombination between non-identical DNA, preventing the permanent incorporation of these mismatches into the genome. MMR is the key guardian that ensures the recognition and removal of the

3

mismatched nucleotides. Loss or mutation of genes involved in this repair system leads to

dramatic increases of spontaneous mutations that can cause harmful effects to the cell

[17]. This system is well studied and characterized in the *E.coli* and consists of different

mutator (Mut) proteins that coordinate to accurately repair the mismatch.

1.2.3 Proteins of The MMR Pathway:

The MMR pathway proofreads newly replicated DNA and is coupled to the

replication. In *E.coli*, MMR starts when a MutS homodimer binds to a mismatch. MutS

recognizes all types of DNA mismatches, except for dC-dC mismatches [18] or

nucleotide insertion/deletions (indels) larger than 3 nucleotides [19]. MMR is initiated by

the binding of MutS to a mismatch: MutS binds to different mismatches from highest

binding to weakly binding according to: dG-dT, dA-dC, dA-dA, dG-dG > dT-dT, dT-dC,

dA-dG >> dC-dC [19],[20],[21], where dC-dC [7] is weakly recognized and poorly

corrected by MMR system. After binding to mismatch, MutS recruits the MutL protein

and they slide along the DNA strand together [18], [22] and recruit the MutH protein.

Activated MutH has endonuclease activity and it nicks the newly synthesized DNA

strand can be distinctly differentiated from methylated template strand at d(GATC) sites

to initiate the repair process because the newly synthesized DNA is still hemi-methylated

at those sites. MutL loads helicase UvrD [23] at the site of nick which uses this single-

strand break as the entry point to unwind the strand and generated single-stranded DNA

is bound by single-stranded binding protein and protected from the nuclease attack [24].

Now, depending on the site generated relative to the mismatch, if it is at 3'-side

then ExoI and ExoX are going to be activated, or if it is at 5'-side thenExovII or RecJ

will be activated and shred the strand up to mismatch. The resulting single-stranded gap

undergoes DNA synthesis and ligation by DNA polymerase III holoenzyme, SSB and

DNA ligase [25]. The efficient process of the digestion and resynthesis of the DNA

strand between nick and the erroneous nucleotides, which can be separated by hundreds

of base pairs, is called Long Patch Repair (LPR) [26].

1.2.3.1 Mismatch Recognition ATPase MutS:

The protein MutS directly interacts with the DNA strand and binds to mismatched

nucleotides. It is a 95kDa protein with 835 amino acids and in solution exists as a blend

of dimers and tetramers. The N-terminus and the C-terminus of MutS have distinct roles

in the functioning of the protein. The N-terminus has a DNA binding motif through

which it directly interacts with the mismatch on the DNA strand and has a MutL

interaction domain. The conserved Phe-39 residue of MutS enters the DNA helix at

mismatched nucleotides. The C-terminus is required for the oligomerization and thus

specifies the stoichiometry of MutS during the repair. The protein also has an ATPase

domain as binding and hydrolysis of ATP is necessary for its function. ADP-bound MutS

specifically recognizes a mismatch and binds to DNA. MutS possesses a highly

conserved walker type-A nucleotide-binding motif which is involved in ATP binding and

hydrolysis. In the MutS N-2 and N-3' nucleotide-binding motifs are unique that

recognizes the four nucleotides on the DNA. N-2 is involved in ATP hydrolysis.

This MutS-mismatch binding reduces the affinity of MutS to DNA and triggers

ADP-ATP exchange [27]. This ATP binding to the MutS functions as a checkpoint and

allow MutS to convert to a long-lived 'sliding clamp' conformation that diffuses quickly

along the DNA molecule. In the presence of homoduplex, this ATP will act as "anti-DNA-binding" as shown in Figure 1 [1]. Thus, mismatch in the DNA acts a nucleotide exchanging factor and this is crucial in one of the models of repair [28], [29], [30]. This protein also interacts with other components of the MMR system like MutL and the beta replication clamp.

Figure 1. MutS ATPase Activity and Its Effects [1]
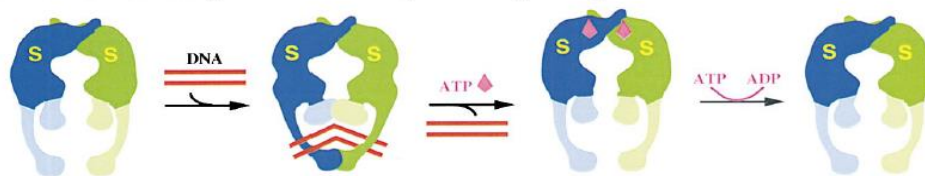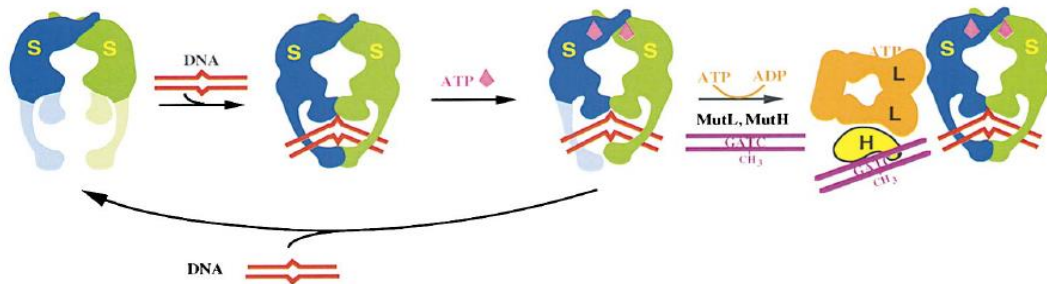
(A) When Muts binds to the DNA and finds the DNA as homoduplex (no mismatches), MutS will not proceed for MMR. (B) MutS recognizes the DNA as heteroduplex (containing a mismatch) then after binding of ATP to MutS will convert it into the sliding clamp which will complete the process of MMR. *Copyright ☐2001 by Cell Press*

1.2.3.2 MMR 'Matchmaker' Protein MutL:

Protein MutL is a 67kDa protein with 615 amino acids and is another ATPase that interacts with MutS. After MutS recognizes a mismatch, this complex forms a long-lived

sliding clamp on a double-stranded DNA that undergoes further ATP exchange and recruits MutH. The N-terminal of MutL have domains for binding ATP and beta clamp of the DNA polymerase while the C-terminal is important for dimerization and also recruitment of other protein components [31].

1.2.3.3 Nicking Endonuclease MutH:

MutH is a single-strand endonuclease which is recruited after a MutS/MutL complex formation and cuts near an adjacent d(GATC) on daughter strand. It is a 25kDa and exists in a monomeric form. The N-terminus regulates the activity of a cleft that binds the DNA helix. The C-terminal works in concert with the MutL which exerts a force on the bound DNA helix to shift it towards the catalytic site of the MutH [32] and consequently facilitates the endonuclease activity of the MutH.

1.2.3.4 Repair Helicase UvrD:

UvrD is a helicase that unwinds the mismatched DNA in 3' to 5' direction after MutH has cleaved it. UvrD participating in many cellular processes including DNA metabolism and mismatch repair [33].

1.2.3.5 Exonucleases and Polymerases for Excision and Re-Synthesis:

The action of the UvrD leads to the formation of single-stranded DNA(ssDNA) and this is digested by different exonucleases. These enzymes exhibit specificity towards the 3' or the 5' end of the DNA. ExoI, ExoVII, ExoX are specific to the 3' terminal while ExoVII and RecJ are specific for the 5' terminal of the ssDNA [34]. Single-Strand Binding proteins (SSBs) bind and stabilize the ssDNA.

DNA Polymerase III holoenzyme participates in *E. coli* DNA replication and is required for re-synthesizing the DNA strand [35]. The alpha catalytic subunit of the Pol III binds the MutS/MutL complex. Another component of the DNA Pol III is the beta clamp and a clamp loader. The clamp slides along the DNA Pol III on the DNA strand from the 3' and increases its processivity. The beta clamp has binding domains for multiple different components like DNA ligase, MutS/MutL complex. Specifically, MutL has clamp binding sites on the N-terminus as well as a certain region of the C-terminus [36].

1.2.4   Coordination of Mismatch Repair Across Long Patches of DNA:

Although the different types of MMR proteins and their function are known by *in vitro* studies, the mechanism by which they intracellularly coordinate excision and resynthesis of the DNA between the mismatch and a distant d(GATC) site is still unknown. However, there are three main proposed models for this coordination are:

1.2.4.1 The Stationary Model:

In this model, the DNA helix undergoes bending upon binding of MutS undergo the mismatch on the DNA strand. The looping of the DNA brings the two distant mismatch and the d(GATC) sites into close proximity. The ATP binding to the MutS help confirms the heteroduplex and further provides validation for the strand excision in the following steps. Studies by Junop *et. al* have shown that simultaneous binding of the ATP and the mismatch is required for the activation of MutS and subsequent cleavage by MutH [37].

1.2.4.2 The Translocation Model:

In this model, the binding of ATP to the MutS lowers its affinity for the mismatch in the DNA strand. This reduction in the affinity by the ATP hydrolysis causes the unidirectional movement of the MutS proteins along the DNA double helix. This movement leads to the formation of DNA loop that helps in the discriminating of the strand [38].

1.2.4.3 The "Molecular Switch" Model:

In this model, the binding of MutS to the DNA mismatch triggers a change in the confirmation that causes a release of the ADP and binding of the ATP also called the ATP exchange that again causes a conformation change that leads to the formation of a MutS sliding clamp. The subsequent excision events are triggered by ATP binding rather than the hydrolysis of the bound ATP [39].

How DNA mismatch repair is coordinated intracellularly is still the subject of debate. Molecular details of MMR process coordination *in vivo* with respect to its location on the DNA still needs to be established based on these above-proposed models. We are interested in finding these molecular details using newly developed biotechnologies.

1.2.5   Other Major Mechanisms of DNA Repair in *E. coli*:

A genetic mutation can occur in a cell by errors during replication, recombination, or DNA repair [40], but also from direct chemical damage to the DNA molecule. These other types of lesions are continuously repaired by other conserved and sophisticated

repair pathways that may also interact with MMR proteins. There are primarily two types [41],[42] of damage to DNA:

1.2.5.1 <u>Spontaneous or Endogenous Damage:</u>

Firstly, depurination or depyrimidination occurs when a base is lost, making the sugar/ phosphate backbone labile leading to strand breaks. The loss of base erases the genetic information from that site. Secondly, deamination of bases causes loss of amino group from the nucleotides. Due to this, cytosine converts to uracil which pairs with adenine instead of guanine. Thirdly, DNA replication and transcription create an overwinding problem which creates supercoiling of the DNA. This supercoiling needs to be relieved and this is done by enzymes like topoisomerases. These enzymes introduce a strand break to release the stress. Normally, this break is repaired by replication machinery but under some conditions like the presence of topoisomerase inhibitors or the failure in cell cycle this strand break is not repaired. Finally, oxidative damage is caused by the reactive oxygen species (ROS) produced by cellular respiration and metabolism that oxidize the nucleosides like Guanine (G) and Adenine (A) and converts them to derivatives like 8-oxo-dG and 8-oxo-dA.

1.2.5.2 <u>Environmental or Exogenous Damage:</u>

Ionizing radiation like gamma- rays or X-rays cause direct or indirect effects. When the radiation interacts directly with the DNA, the electrons cause phosphodiester bond dissociation due to the molecular resonance making the backbone labile causing strand breaks. The indirect interaction causes radiolysis of water creating indirect damage

leading to DNA damage. UV radiation forms cyclobutene pyrimidine dimers (CPDs) and 6-4 photoproducts. These dimers block the replication and transcription machinery.

Chemicals like alkylating agents disrupt the H-bonding of the bases by adding alkyl groups. Bulky adducts like Benzopyrene when metabolically activated distort the DNA double helix. The other compounds involved are intercalating agents and enzyme inhibitors.

These different type of DNA lesions can cause a variety of mutations and so it becomes critical for the cell to maintain the genome stability. In order to facilitate this, the cell has evolved various mechanisms to repair the DNA damage. Distinct types of DNA damage are repaired (Figure 2) by specific proteins or groups of specific proteins.

The first pathway is a direct reversal which does not need the strand synthesis but directly reverses the effect of the damage like CPDs by UV radiation and alkylated bases by alkylating agents. Enzymes like alkyltransferase [43] remove the alkyl groups from the DNA. Photoactivation process uses photolyases to reverse the CPDs [44].

The second pathway is Base excision repair (BER) repair the damages like oxidation, deamination or base loss which do not distort the DNA double helix. BER was discovered when Thomas Lindahl first identified *E.coli* uracil- DNA Glycosylase (ung). DNA Glycosylase recognizes the damaged base and removes it subsequently generating an abasic site. The AP endonuclease chews the short segment surrounding the abasic site and DNA Pol synthesizes the complementary sequence and the ligase then seals it [45], [46] After identification by glycosylases, there are two pathways: short patch repair [47]

which repairs single-nucleotide damage or long patch repair which repairs a stretch of 2-10 nucleotides [48].

The third pathway is Nucleotide Excision repair (NER) and was initially discovered in 1964 by Bill Carrier and Dick Setlow in *E.coli* [49] About two decades later, the uvr genes were cloned by Aziz Sancar and others [50], [51], [52] which further strengthened our understanding of the prokaryotic NER. The repair of CPDs and the photo- products that cause the distortion in the DNA double helix are models for studying NER.

In prokaryotes, the DNA damage is detected by UvrA along with UvrB, where UvrB identifies the precise location of the lesion and forms a site where the subsequent UvrC nuclease binds. UvrC cuts few nucleotides upstream and downstream from the site of the lesion so that UvrD can remove that portion out of the helix [53]. Further, DNA Pol recruits the complementary nucleotides with ligase sealing the nicks [54].

The fourth pathway is Mismatch Repair (MMR) which is described above. In addition to this repair which involves either no strand break or single-strand breaks, there is also the Double-Strand Break (DSB) repair mechanisms. DSBs are induced by ionizing radiation and chemical agents [55]. The two major pathways involved are Non-homologous End Joining (NHEJ) and Homologous Recombination (HR). NHEJ is an important pathway but is not present in *E. coli*. HR is functional and well understood in *E.coli*. It utilizes the homologous chromosome as a template to repair the DSBs. The 5' ends at the DSB undergo resection and then strand invasion to repair the DNA in an enzyme-dependent manner. Following the strand invasion step, the HR follows either

Double-stranded break repair (DSBR) pathway or Synthesis dependent strand annealing (SDSA). This is an error-free process, unlike the NHEJ [56].

In *E.coli* two similar pathways function RecBCD and RecF. RecBCD is the main pathway by which bacteria carry out DSB repair and if it fails due to some mutations then RecF pathway takes over. The RecBCD is an enzyme with helicase-nuclease domains and its activity is mediated by a regulatory sequence called the Crossover Hotspot Instigator (Chi) [57]. After reaching the Chi sequence, the nuclease domain cuts the DNA strand and RecA is recruited on the 3' end of the newly formed single strand. This binding of RecA is not a passive process but actively mediated by RecBCD enzyme complex [58]. RecA is a molecular machine responsible for finding the complementary sequence by strand invasion into the homologous chromosome [59]. Sometimes if HR is attempted between DNA that is not perfectly identical, incorrect HR is prevented by MMR proteins [60].

Figure 2. Different Types of DNA Damaging Agents and The Repair Pathways
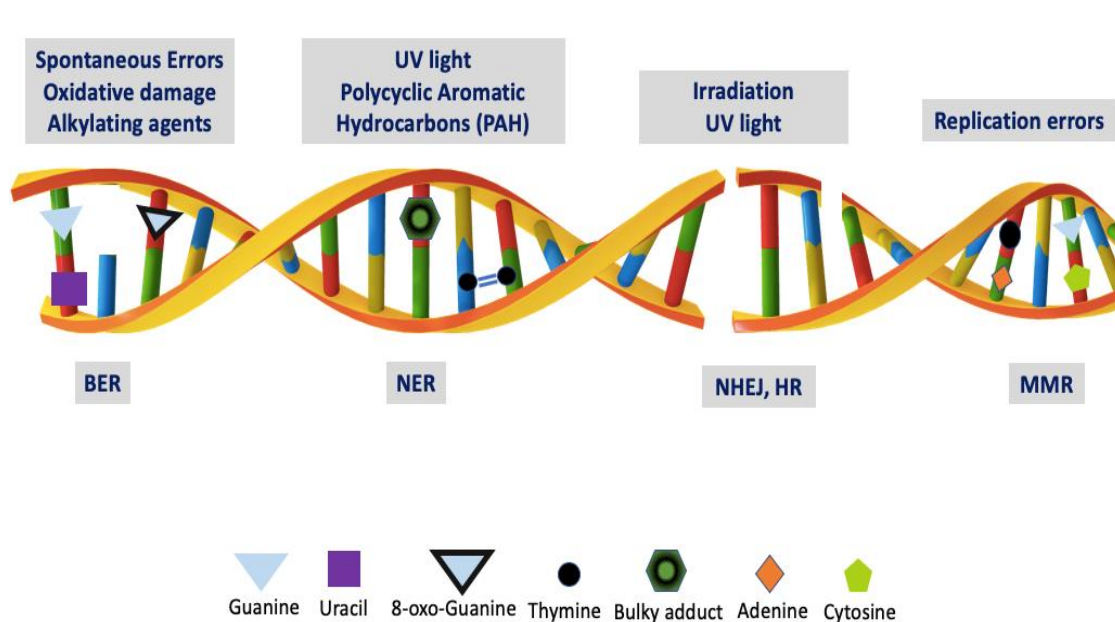
The spontaneous errors cause base loss leading the formation of abasic site and alkylating agents cause methylation of thymine leading to the formation of uracil which mispairs with guanine. The oxidative damage caused by cellular and metabolic processes oxidizes guanine to 8-oxo-guanine. These lesions are repaired by the base excision repair pathway. The UV light Causes the formation of cyclobutane pyrimidine dimers (CPD) and the aromatic hydrocarbons when metabolized form bulky addicts that distort the DNA. These are repaired by nucleotide excision repair. The irradiation leads to the formation of single-stranded and double-stranded breaks in the DNA Helix repaired by Non-homologous end-joining and Homologous recombination. The defects in the replication error lead to the formation of mispairing between the nucleotides, insertions and deletions. This is repaired by the mismatch repair pathway.

1.2.6   Mutation and Antibiotic Resistance:

Pathogenic bacteria acquire antibiotic resistance (ABR) through two

fundamentally different genetic mechanisms: Horizontal Gene Transfer (HGT) and

Spontaneous mutations [61]. The resistant genes which are encoded within the genome

to protect bacteria exist in the environment as a cassette, plasmid or a gene fragment

released from a dead bacterial cell which can be incorporated in the host via

transfection, transduction or HGT. These processes are primary means of spreading the

ABR, but the mutation is essential for evolution and diversification of the genes.

Spontaneous mutations occur during replication which makes bacteria resistant to the

given antibiotics. In *Mycobacterium tuberculosis,* the bacterial cause of tuberculosis*,*

the main mechanism of drug resistance is only spontaneous mutations. Genetic

mutations arise via oxidative damage, alkylation damage and DNA replication.

Replication errors are the result of a failure in base matching, proofreading and DNA

mismatch repair, which act sequentially to ensure the fidelity of DNA. MMR pathway

is crucial to maintain genome stability and avoiding mutations [62]. Bacteria with

defects in the MMR pathway or with reduced ability to correct a mismatch have a

higher probability of accumulating genetic mutations and recombination events, such

bacteria are known as "Hypermutatable" or "Hypermutator" phenotype [63]. These

hypermutator bacteria are indirectly selected for the favorable mutations with a

changing environment, such as antibiotic resistance. Studies from clinically isolated

bacteria have shown that 1% of the strains are mutators naturally which increase the

adaptive response in bacteria. There is a need to investigate how MMR process is

modulated in such bacteria to acquire rapid mutations.

1.3 Microscopic and Nanoscopic Studies of *E. coli* MMR Reaction Mechanism:

1.3.1 Single-Molecule Studies Of *E.coli* MMR Proteins:

In vitro studies of purified systems from bacteria have yielded important

insights into the MMR reaction. Reconstitution of *E.coli* systems *in vitro* was started in

1983 when, the essential protein components associated with MMR were identified

using biochemical analysis and heteroduplex DNA [64],[65]. By 1989, the whole MMR

system was reconstituted *in vitro* by purifying all necessary proteins to the system [66].

These studies were done using simple biochemical analysis like restriction enzymes,

electromobility shift assay and the protein structures were determined by

crystallography [67]. The MutS crystal structure was solved by removing 53 amino

acids from the 853 amino acid of the whole protein and it is known as MutSdelta800

[68],[67].

The interactions between individual MutS and MutL can be studied by a

technique called Single-molecule Föster Resonance Energy Transfer (smFRET) [69]

which uses a single donor and an acceptor fluorophore. This has made it possible to

elucidate the mechanism of MMR and provide the evidence for Switch Sliding clamp

of MMR. Use of smFRET or fluorescence Tracking (FT) into single-molecule

applications provides extremely accurate nanometer (nm) distance measures between

individual protein particles, for example, a donor molecule on MMR protein [70],[69]

and an acceptor molecule on a mismatch containing DNA. The advantage of

smFRET/includes the followings [71]: This helps in the ability to image molecules in

msecs and real-time based on the fluorescence. This helps determine intermediate factors or proteins associated with the process.

Quantum Dots (QDs) are also helpful in real-time tracking of these single proteins. This single-molecule tracking and smFRET has revealed the interaction of MutS with DNA and other proteins [72],[73]. Other techniques like AFM (Atomic Force Microscopy), small molecule Flow-Stretching (smFS), single-molecule Total Internal Reflection Fluorescence (smTIRF) [22],[2], [74],[75] have revealed the MutS induces a DNA conformation change when it encounters a mismatch. Thus, with AFM studies, it was possible to determine the DNA structure change and stoichiometries [75] because of MutS interaction with the DNA backbone at the site of the mismatch.

Also, the attraction of single-molecule studies was smFRET combined with TACKLE (Transition Analysis Combined with Kinetic Lifetime Examination) [76],[77], in which the idea was to separate two different fluorescently tagged molecules at a distance of few bps in the same DNA. Whenever the DNA undergoes any conformational change two molecules will be in closed proximity with each other and emit fluorescence using resonance energy transfer. With the help of this technique, they examined the kinetics of conformational changes of DNA induced by MutS protein by real-time monitoring. These studies also suggested quantitatively that MutS kinetics and dynamics will differ in the interaction with DNA mismatches based on the type of mismatches [76] (e.g. dG-dT, dA-dC or dT-dT).

Moreover, single-molecule studies recently have shown with (smTIRF) [22],[2] studies on glass slides reveal more information upon the interaction of MMR proteins. Using this realtime microscopy technique they have shown that *E.coli* MutS mismatch recognition results in the formation of stable ATP-bound sliding clamp that diffuses along with the DNA randomly, which acts as a platform to recruit MutL and forms MutS-MutL search complex. MutL binding to ATP will form a long-lived sliding clamp of MutS-MutL on a mismatch that together recruits MutH introducing ssDNA breaks at newly replicated DNA strand. Further analysis reveals that long-lived MutL sliding clamps associates with UvrD and increases its ability to unwind the DNA at the break.

This additionally suggests few more information about *E.coli* MMR like MutH and UvrD share overlapping interaction sites with MutL and also supports the previous finding that requirement of exonuclease digestion for MMR excision in *E.coli* is very rare and the majority of the strand excision is done by MutL-UvrD complex in the absence of ssDNA exonucleases.

Figure 3 describes the entire process of MMR in *E.coli* studies revealed by smTIRF. These studies support the molecule switch/sliding clamp model [78] as an originally proposed model for *E.coli*. These studies reveal some difference to the original model that includes the scission of d(GATC) sites by MutS-MutL/MutH complex surrounding the mismatch and more detailing about the unwinding [78] of DNA by MutL-UvrD. This also finds unnecessary involvement of multiple slide clamps

[79] of MutS-MutL and MutH generated DNA fragments displacement during MMR process, which increases the complexity of the system analysis [80] and makes it difficult for a researcher to determine the precise mechanism as repair will only be completed when mismatch containing portion will be removed and resynthesized as shown in Figure 3.
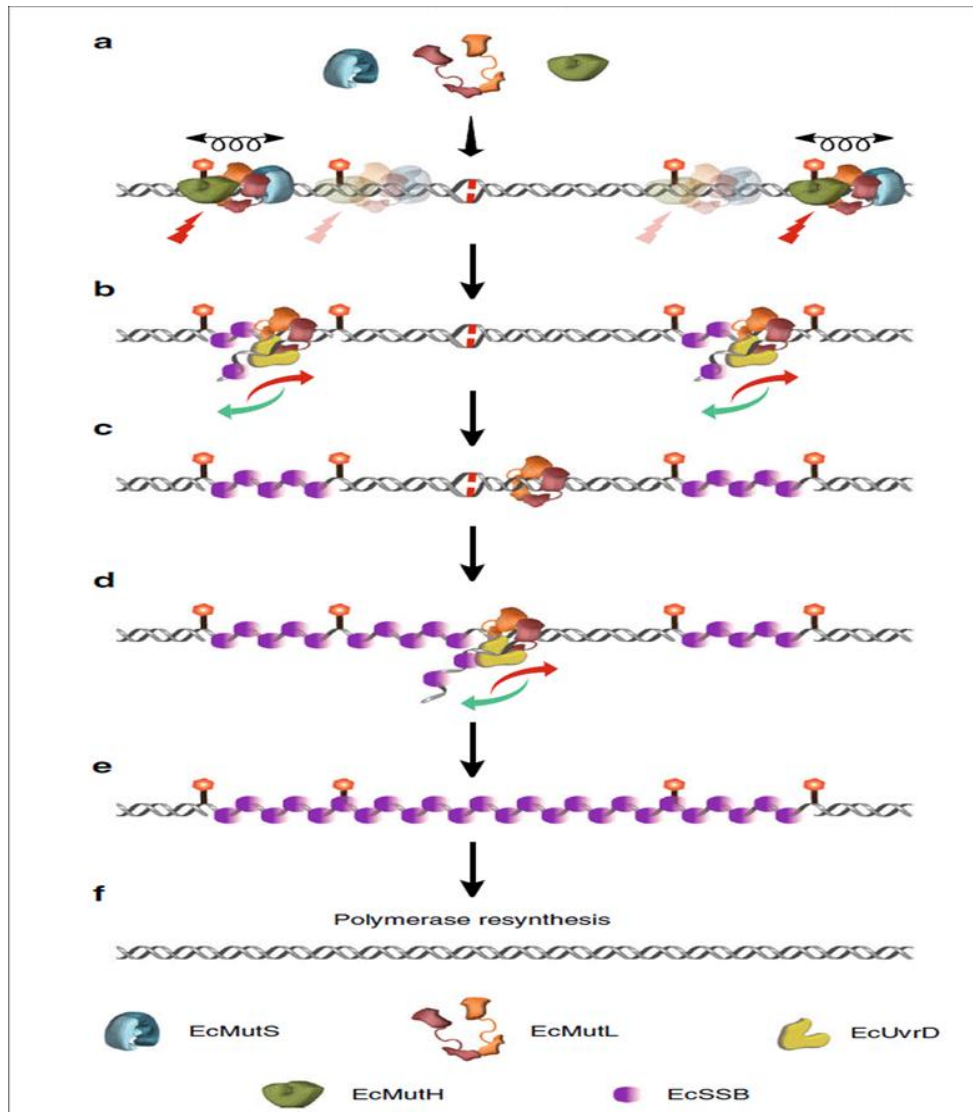
(a) MutS (blue) and MutL (brown) cascading clamps recruit the MutH (green) protein to generate multiple strand cuts (red lightning bolts) on mismatched DNA at d(GATC) sites. (b) MutL loads the UvrD (yellow) protein on the cuts on mismatched DNA where it unwinds (red arrow) and rezips (green arrow) the DNA that again processed by SSBs. (c) when MutL-UvrD unzipping reaches the MutH cut site near adjacent d(GATC) site, the SSB bound DNA fragment is generated between these cut sites. (d,e) even though the strand displacement between two adjacent d(GATC) is random, MMR will be completed only after removal of a mismatch DNA strand. (f)

The DNA polymerase and ligase complete the MMR by resynthesizing the removed DNA strand by MutL-UvrD complex. *Copyright © 2019, Springer Nature*

However, we will note that these reactions are performed in the absence of actively replicating DNA, so the conditions that this reaction might occur in a cell remain unclear.

1.3.2 Single *E.coli* Cell Studies of MMR Activity:

Study of mismatch repair *in vivo* is crucial. One of the studies of *E.coli* K-12 strain [81] has shown how the absence of MMR proteins will accelerate the mutation acquisition in bacteria and also drug resistance in the absence of MMR components. However, the nature of MMR and its associated machinery have made it difficult to study this process *in vivo*. In addition to above-described techniques, the MMR process was also studied using a microfluidic chip to see how the dynamics of mutations take place in a cell over time.

Uphoff [82] designed a microscopy-based approach to study the real-time mutagenesis in response to an alkylating agent (methylmethane sulfonate) and antibiotic treatments using microfluidics in a single *E.coli* cell. They found the stochastic expression of MMR genes with single-cell studies, describing how *E.coli* cells modulate mutagenesis. The chronology of mutagenesis and responses to genotoxic stress in (Figure 4) [3] shows the genomic stress in presence of methylmethane sulfonate (MMS).

Figure 4. Real-Time Mutagenesis in E.coli [3]
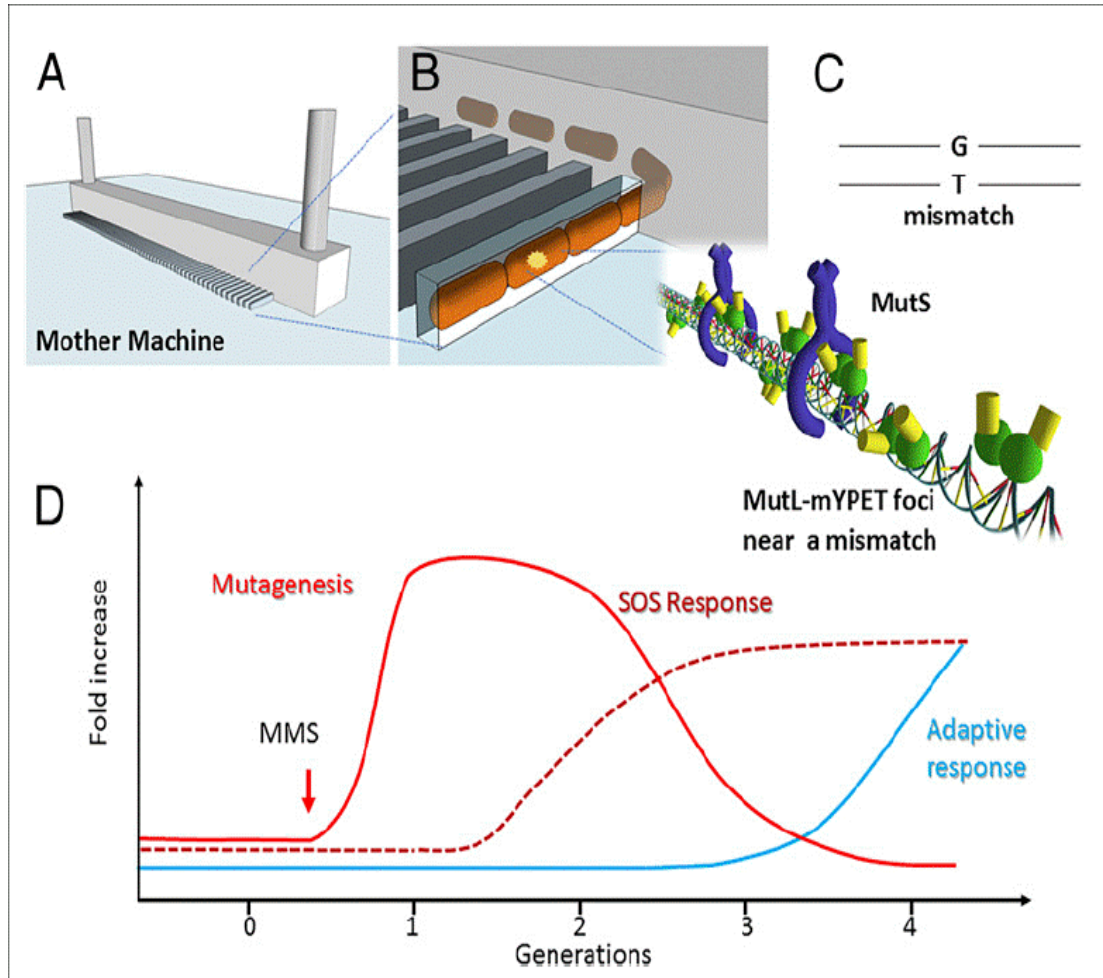
(A) Mother machine (microfluidic chip) (B) parent cell and progeny cell coming out of the capillary, (C) G-T mismatch near the MutL-mYPET to track the response of a cell. (D) The events of mutagenesis and genomic stress in response to methylmethane Sulfonate (MMS). *Van Houten and Kad PNAS | July 10, 2018 | vol. 115 | no. 28 Copyright (2018) National Academy of Sciences*

Another protocol to visualize the dynamics and effects of mutations on fitness is developed by Robert et al. [83],[4], which has implemented mutation visualization (MV) and microfluidic mutation accumulation (μMA), which allow us to follow the occurrence of mutations created by DNA replication errors using *E.coli* in which MutL was tagged

with Yellow Fluorescent Protein (YFP). MV allows precise characterization of dynamics of point mutations created by replication errors in *E.coli*. Additionally, it also states the difference in the rate at which replication errors occur in a 'single-cell' and the 'per-cell' error rate of different subpopulations of cells because the MutL is associated with error repair and allows mutation detection.

Figure 5 [4] describes at what frequency mutation occurs and they get fixed in *E.coli* cell. When an error occurs during the replication, only one of the four dsDNA molecules will carry the mutation. Therefore, in *E.coli* in minimal media conditions, a single replication cycle will generate 1/4 mutated progeny of a single cell progeny and this proportion can be lower in the rich media when doubling time is shorter than the time necessary to replicate the chromosome. As shown in Figure 5c, when there are six replication forks generated in the minimal media 1/4 of the progeny will have mutations. Figure 5d shows that in the rich medium this frequency is lower to 1/8 progeny to receive the mutation.

Single-cell visualization can also be used to understand the replication dynamics in cells and its interactions with MMR proteins [84],[85] by engineering fluorescent domains on to DNA repair and replisome associated proteins, researchers have shown how leading and lagging strands are replicated [86], how replisomes and MMR proteins interact, single protein dynamics and DNA interactions [87], DNA replication fork dynamics with the help of nanoscopic fluorescence microscopy [88],[87],[84].

Most of these studies are performed and give some important information on how MMR is coordinated in *E.coli*.
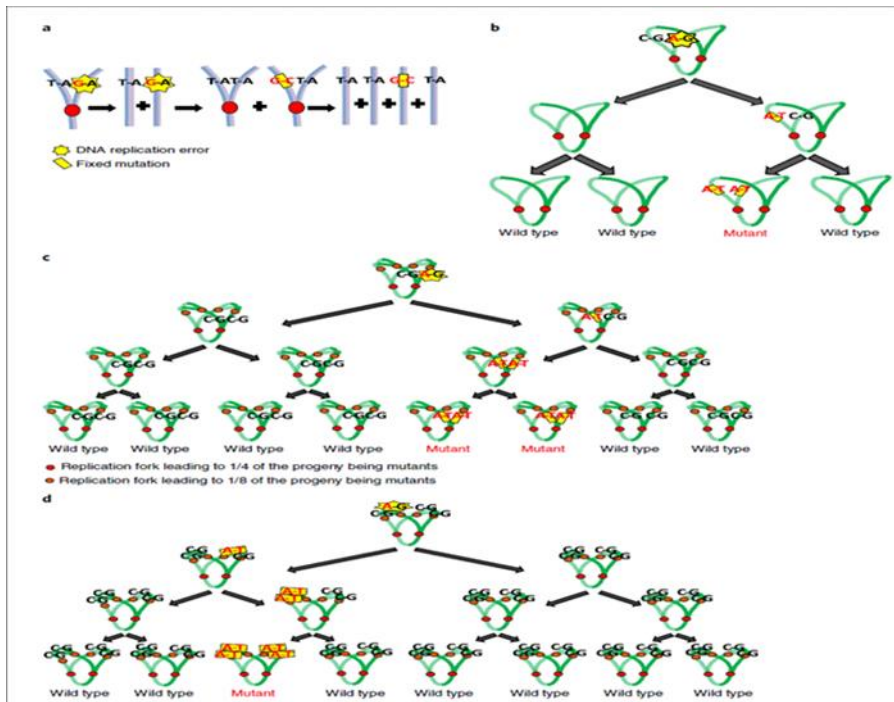
(a)A mismatch (yellow star) base is inserted during one round of replication (red dot indicates the replisome) and is converted to a permanent mutation in the genome in the next round of replication leaving one mutated DNA among four. (b,c) when replication time is smaller than the doubling time of a cell then replication error was tracked as MutL-YFP spot in ¼ of the progeny in the minimal media. (d) When the cells are provided with rich medium and doubling time is shorter than the time required for a chromosome to replicate, frequency of progeny becomes 1/8 compare to minimal media.

1.4 Outstanding Questions About Mismatch Repair Activity *in vivo*:

While *in vitro* experiments using purified MMR proteins can provide insights into

their biophysical interactions using highly defined experimental conditions, these

experiments are typically performed in the absence of the replication fork or on non-

replicating DNA. In contrast, studies of MMR in living cells need to rely on the rare mismatches that are generated randomly through the cell or require additional mutagens to increase the probability of a mismatch. In those cases, the biochemical conditions are uncontrolled and often unknown. Additionally, *in vivo* (here referred to as in bacterial cell) studies of MMR use as their measure the overall cellular mutation rate, rather than MMR activity itself. There is a need to elaborate the molecular details beyond these interactions, and to test whether the behavior observed *in vitro* occurs during intracellular MMR. For example:

1. How is repair coordinated during replication? *In vitro* repair can be coordinated using d(GATC) sites from either side of the mismatch, but a study from Hasan and Leach [40] suggests that mismatch repair in cells is coordinated by the direction of replication.

2. What is the stoichiometry of the repair complexes during repair in cells? For example, *in vitro* MutS can form tetramers, but Mendillo et al [89]. have shown that mutations that disrupt MutS tetramerization only modestly affect mutation rates during a rifampicin resistance assay.

3. Under what conditions are each of the components of the MMR pathway necessary in the cell? For example, on the lagging strand, there are many breaks in the DNA, so MutH or d(GATC) sites might not be necessary to repair lagging strand mismatches if repair is coupled to replication.

4. Are all types of different mismatches repaired in the cell according to the same mechanism?

25

5.   What is the kinetics of the MMR reaction in living cells? How important is the coupling to replication?

6.   How well conserved is this mechanism across different organisms with divergent MMR proteins?

Investigating these questions will likely require a novel technique to quantify the MMR *in vivo* in a more controlled manner than spontaneous reporter assays.

1.5 Oligonucleotide Recombination:

1.5.1   Introduction:

A biotechnological tool for gene editing, oligonucleotide recombination is one technique where we can introduce mismatched nucleotides into specific places along the genome of a living cell. During oligonucleotide recombination, a short single-stranded DNA (ssDNA) molecule called an oligonucleotide that is always complementary to one strand of the *E. coli* genome except for one or more mismatches can become integrated into the genome during replication.

After the oligonucleotide is introduced into a cell, during replication it can get incorporated in the genome at the rate of 10^{-7} to 10^{-5} per transfected cell. If the beta recombinase from the λ phage is also expressed during transfection, oligonucleotide recombination efficiency increases to 10^{-5} to 10^{-3}. Which is also known as lambda (λ) Red- mediated recombination [90].

If those mismatches are not repaired by MMR, they become permanent mutations. Thus, oligonucleotide recombination provides a way to study replication-

coupled MMR processes in living *E. coli* cells while maintaining exact biochemical control over where that mismatch occurs.

1.5.2   λ Red-Mediated Recombination:

Lambda red mediated oligonucleotide recombination is also known as 'Recombination Engineering' or 'Recombineering'. Bacteriophage λ-red mediated recombination of dsDNA requires exo, bet and gam functions [90],[91] while recombineering with single-stranded DNA requires only beta-protein function which is shown in Figure 6 [5]. Ellis *et al.* [91], has shown the ability of λ protein beta to promote recombination between the chromosome and a synthetic ssDNA donor by confirming the hypothesis that λ beta is sufficient to perform oligonucleotide recombination. First-ever gene replacement eliminating the requirement of RecA [92],[93] was achieved using a multicopy plasmid expressing these proteins of bacteriophage λ in *E.coli* following the transfection of linear DNA substrates.

This gene replacement was 15 to 130 times higher than *recBC*, *sbcBC* or *recD* recombinant strains. Furthermore, *E.coli* strain was constructed to replace recBCD gene with $P_{lac}^{-bet\ exo}$ operon and a kanamycin resistance cassette. This new strain was used to replace the gene higher rate of recombinants were observed compared to plasmid-encoded λ-red. This has proven that cloning of the gene of interest before replacement is not necessarily required [90].

λ Beta is a ssDNA binding protein which promotes annealing of complementary single strands and mediates the strand exchange *in vitro* [94],[6]. It has also shown that this beta protein of phage λ protects the ssDNA oligo from the nuclease degradation

while performing reannealing of ssDNA with genomic DNA [95]. The Beta protein itself

cannot mediate strand invasion of a dsDNA but it combines ssDNA to duplex DNA by

strand annealing [6],[92]. Other proteins in the Red operon include λ Gam that inhibits

the RecBCD function and its nuclease activity which protects electroporated DNA

substrate [96],[97]. λ Exo is a ds-DNA dependent exonuclease which provides the 3'-

overhangs for the as a substrate for efficient recombination by digesting 5'-> 3' direction

[98].

Figure 6. Model of Bacteriophage λ Mediated Single Strand Annealing [5]
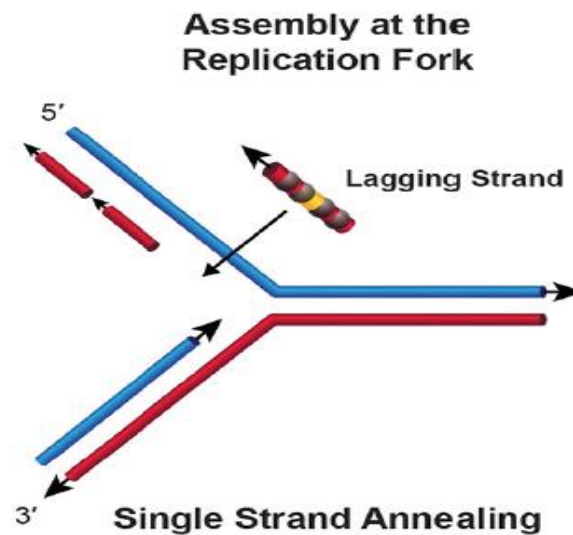
A ssDNA oligonucleotide recombination at replication fork
on a lagging strand. This shows that the oligonucleotides
are annealed to complementary ssDNA. *Copyright © 2016,*
American Society for Microbiology

Additionally, it is also suggested that this recombination with ssDNA technology can be applied to a wide range of organisms concluded after successful oligonucleotide recombination in *Saccharomyces cerevisiae* [99] and *E.coli* [91].

Figure 7 [6] shows the λ Beta protein-mediated strand exchange. In this experiment M13 DNA was used having radiolabeled 43-mer thermally annealed to it. The ssDNA oligo having 63-mer was introduced into the cell. The 63-mer DNA oligonucleotide had an additional 20 nucleotide and four mismatches at its 5'-side compare to 43-mer and its homology with 43-mer radiolabeled DNA. The strand displacement experiment was carried out and it showed the strand exchange of radiolabeled DNA by beta protein.

Figure 7. The Displacement of Radiolabeled (^{32}p) DNA by Beta-Protein [6]
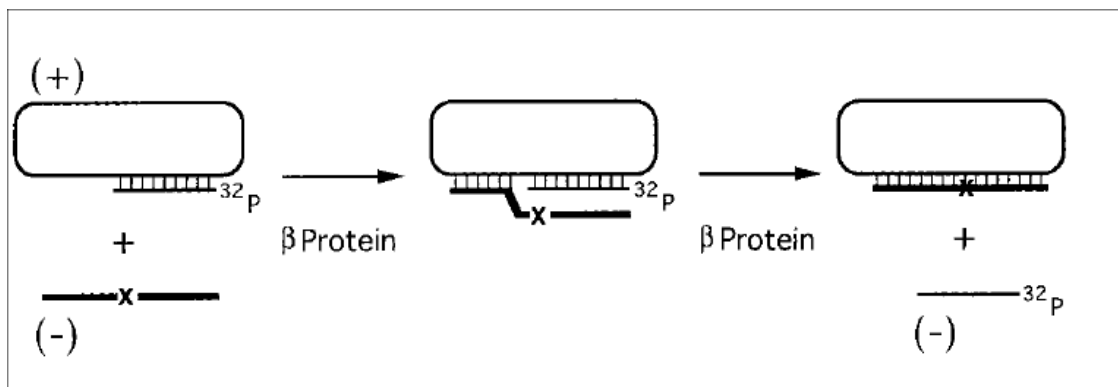
This Figure shows the M13 genomic DNA where the radiolabeled 43-mer is thermally annealed. (-) strand is the 63-mer with mismatches (**X**). This process shows the strand exchange by beta protein in the end where radiolabeled DNA is removed from the M13 DNA and 63-mer was annealed with mismatches(**X**). *Copyright © 1998 Academic Press.*

When either of the strands of the genomic DNA was targeted by ssDNA oligo to a specific location on the chromosome the highest efficiency of recombination was achieved in the oligos that had complementarity to the lagging strand of the DNA [91],[100]. The studies have shown the highest efficiency of oligonucleotide recombination near Okazaki fragment means near the replication fork as shown in Figure 8 [7]. This suggests that oligonucleotide recombination takes place at the DNA replication fork during replication. MMR deficient strains have shown the increased recombination [101] which suggest that mismatches introduced by oligonucleotides are repaired by the MMR mechanism [101].
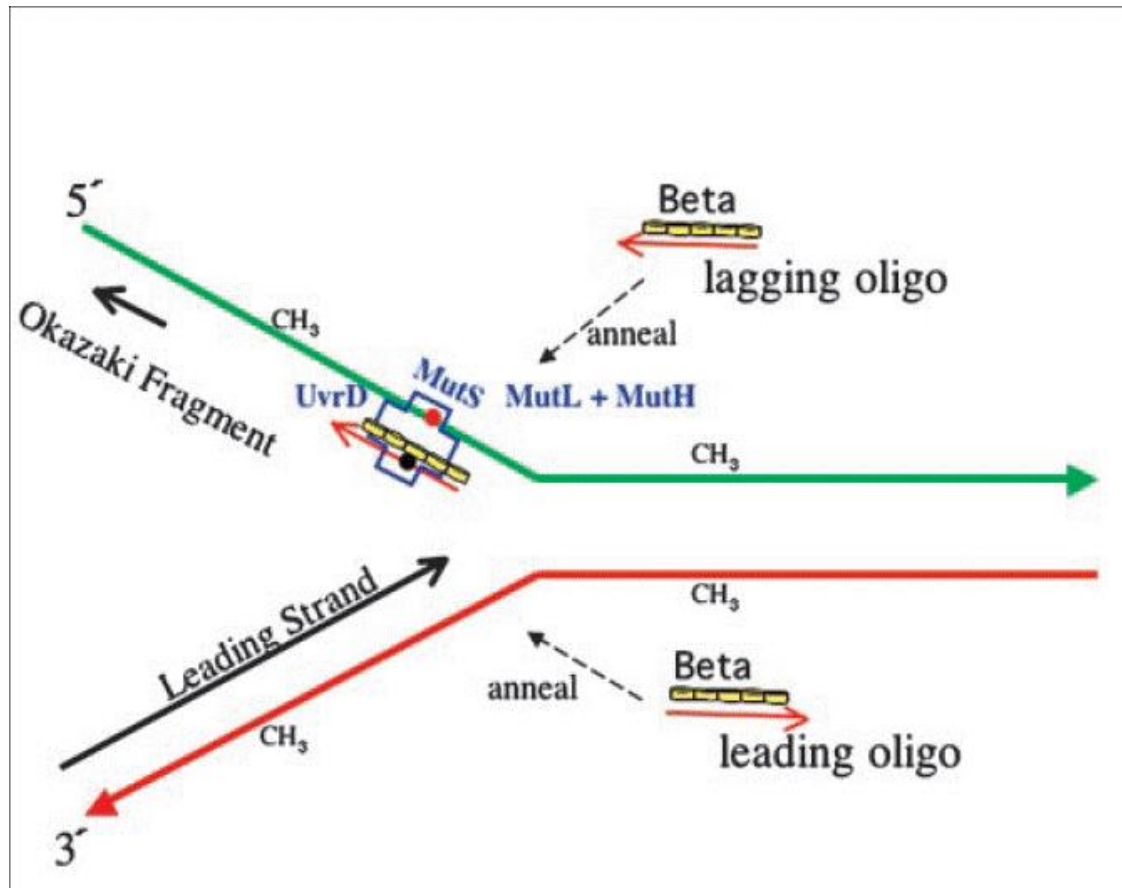
The illustration of MMR at a replication fork. The Okazaki fragment is shown on the lagging strand and the MutS binding to the mismatch generated after incorporation of ssDNA oligonucleotide (Red arrow)on the lagging strand. Methyl (CH_3) groups are shown on the parental strand. The beta protein binds to the ssDNA oligos is shown for the illustration and this event of leading and lagging strand oligos is not occurring at the same time in one cell. This is the schematic of how beta mediated recombination occurs near the replication fork either on leading or lagging strand. *Costantino and Court PNAS _ **December 23, 2003** _ vol. 100 Copyright (2003) National Academy of Sciences*

The illustration of mismatch containing oligonucleotide incorporation and their replication in *E.coli* genome is shown in Figure 9.

Figure 9. Beta Mediated Oligonucleotide Recombination E.coli
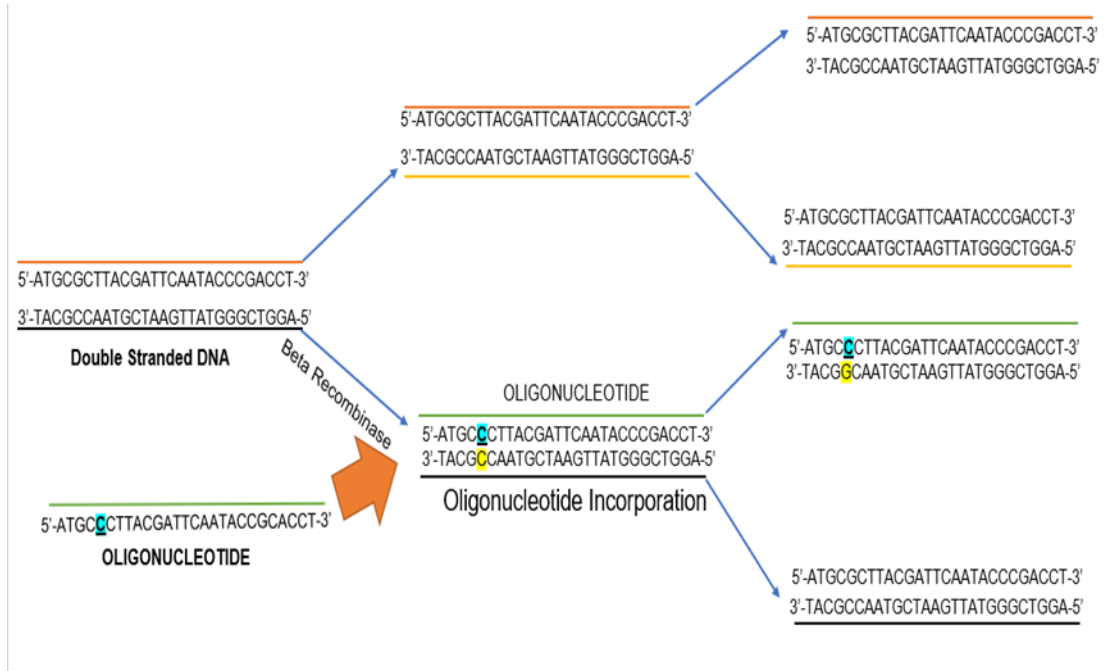
This Figure represents, how oligonucleotides are incorporated into the genome with beta recombinase to the complementary strand during replication. Then, once incorporated, this oligonucleotide will be replicated every time and thus some fraction of the cells will have with the mutation. Mutations here are shown with light blue boxes on the oligonucleotide.

Moreover, three important points are noticed that supports the hypothesis that beta-mediated oligonucleotide recombination occurs at replication fork [7] during DNA replication [102], [5], [103]:

1.   The incorporation rates depend on whether the oligo is designed to bind to the lagging or leading strand template [91],[104],[86] because there are 30 times larger gaps present [7] in the lagging strand compare to the leading strand, so it is likely to see the high recombination events on the lagging strand. This is evidence that incorporation

occurs during replication with the highest efficiency of oligonucleotide recombination

occurring on the lagging strand [104].

2.   Mutations to different proteins involved in replication alter the strand-

specific frequency of oligonucleotide recombination [105],[106].

3.   Oligonucleotides are subject to partial degradation from either 5'-side or

3'-sides after annealing by exonucleases that are associated with DNA polymerase I and

III. Thus, this evidence suggests that oligonucleotide recombination is affected by

exonucleases present at the time of DNA replication[102].

There is also significant evidence that these mismatches introduced during

oligonucleotide recombination are repaired specifically by MMR proteins. There is an

increased efficiency of oligonucleotide recombination when MMR proteins are

inactivated or when the mismatch introduced is a dC-dC mismatch [107] or insertions

and deletions (IDLs) more than 3bps, which are not recognized by MMR proteins [19]. It

has been also reported that MutS cannot recognize the multiple mismatches as efficiently

as a single mismatch [108] and high oligonucleotide recombination was achieved by

changing 6-20 nucleotides in a row [109],[110]. This recombineering has been combined

with other tools to develop new tools like MAGE [111],[112] (Multiplex Automated

Genome Engineering) and TRMR [113] (Trackable Multiplex Recombineering). MAGE

is unique in targeting many locations on the chromosome for modification in a single cell

or across the cell population simultaneously. This has proven improved metabolic

engineering in *E.coli*.

Oligonucleotide recombination has been demonstrated to be an effective gene-editing technique in different organisms and cells [7],[114],[115]. Research carried in the past few years has demonstrated the ability of this recombineering tool to modify different cell types like archaea [116],[117],[118] human embryonic stem cells [119] and mycobacteria [120]. Oligonucleotide recombination in mycobacteria has also combined with another system like Bbx1 integrase and RecT to generate the library of insertion, deletions or fusions of a bacterial chromosome which is known as ORBIT [121] (oligonucleotide mediated recombineering followed by Bbx1 integrase targeting). Oligonucleotide recombination can also be combined with the CRISPR-Cas9 [122] tool to perform the chromosomal engineering by using Cas9 to introduce double-strand breaks at DNA without the mutation, increasing the probability of isolating a mutant. Thus, recombineering has advanced the genetic manipulation, genomic studies and metabolic engineering[123],[111] of bacteria and increases the possibility of eukaryotic cell genetic manipulations too.

1.6 SPORE (Semi Protected Oligonucleotide Recombination) Assay:

However, a challenge of using oligonucleotide recombination directly to study MMR is that incorporation rate is low and depends on the specific oligonucleotide sequence, where it is being integrated, and other factors.

A modified form of oligonucleotide recombination is known as called a SEMI PROTECTED OLIGONUCLEOTIDE RECOMBINATION (SPORE) assay, was developed specifically to study MMR processes by allowing researchers to separate incorporation efficiencies from mismatch repair efficiencies during oligonucleotide

recombination and allows us directly to evaluate MMR in the living cells [8], [9] (Figure

10).

(A) The oligos designed complementary to the lagging strand so that it will correct a synthetic STOP codon (TAG) in the galactose kinase gene *galK*, allowing bacteria with a successfully transformed oligonucleotide to survive in media with galactose as the only sugar source. Non-Template Strand (NT) and Template(T) strand.
(B) The 'control' mismatch that corrects the *galK* gene is chemically protected from LPR by phosphorothioate bonds (Indicated by "*"). 'probe' mismatch is unprotected and located on one side of the control mismatch, so we can see how the repair is performed by either the 5'- or -3' sides.

(C) Protocol for transfection of SIMD50 competent cells, which express the beta recombinase, with oligo 5'-GT & 3'-GT (where the probe mismatch is located on either the 5' or 3' side of the control mismatch, respectively). After transfection, the cells are grown on a galactose selective media which indicates a mutation of stop (TAG) codon. Then the grown cells are checked for the probe mutation by sequencing and then we can quantify the efficiency of repair with respect to MMR-deficient (mutS knockout, KO) strain.
(D) mutS Wild type vs mutS KO sequencing results of the repair of G-T probe mismatch. *Copyright © 2017, Oxford University Press*

During a SPORE assay [8],[9], a synthetic oligonucleotide is designed to contain multiple mismatches on the same oligo that are separated spatially. Among these mismatches, one is control mismatch which is chemically protected by phosphorothioate bonds (*) and the other two are unprotected, MMR-active 'probe' mismatches (Figure 10). Those control mismatches are mismatches like dC-dC which do not excite the MMR response, and they are flanked by phosphorothioate bonds (*), the function of which is to protect the oligonucleotides against the degradation by exonucleases. These bonds also block long patch repair (LPR), which helps in separating the repair occurring from either 5' or 3' end based on the location of the probe site relative to the control site. The hypothesis is that when we select for control mismatches which should be incorporated regardless of whether repair occurs, we can then look at the presence or absence of the probe mismatch to determine whether mismatch repair occurred.

Figure 10 [8] is the demonstration of a SPORE assay which was based on the phenotypic selection of the mutation introduced by the control mismatch in the bacteria. SPORE assay gives insight to the following molecular details of MMR process.

1.   G-T and T-T mismatches are repaired differently on the lagging strand. The G-T mismatches are recognized strongly by MutS while the T-T mismatches are weakly recognized. The repair is coordinated from the 3'-side of oligonucleotide (5'-side of the replicating DNA fragment. This also suggests the repair of G-T mismatches are two orders of magnitudes greater than T-T mismatches.

2.   The asymmetry was observed in G-T vs T-T mismatches with mutS mutant suggests that 3'- repair is coordinated by MutS tetramerization [124], while MutS dimer can coordinate the repair from 5'-side. One of the reasons for this conformational bias of MutS can be asymmetry in the replicated DNA [40]. The observed bias in the repair of T-T and G-T mismatch was unexpected as MutS crystal structure [125] has shown common binding mode across different mismatches, but this result [8] has proven that MMR may coordinate differently for the different types of mismatches. The quantitative analysis based on the SPORE assay data shows that tetramerization of MutS increased the 3'-repair rates 10-fold compared to the mutant form, which was incapable of forming tetramer. But 5'-repair rates remain the same. A mutation to disrupt MutS tetramerization but allow dimerization only affected mismatch repair from 3'-side but did not affect on 5'-coordinated repair [126].

3.   Another published study has proven the mechanism for lagging strand MMR with the help of SPORE assay [9] which shows the evidence that MutH independent MMR is coordinated from 3'-side in the direction of lagging strand while MMR from 5'-side of the Okazaki fragment requires the presence of endonuclease.

4.   Researchers hypothesized that there were two different mechanisms of lagging strand repair in the same study:

a)   A tetrameric MutS assisted endonuclease independent mismatch repair which is coordinated from 3'-side of a mismatch.

b)   A dimeric MutS assisted endonuclease dependent mismatch repair.

The above SPORE assay results are contrasting to what has been studied extensively with non-replicating DNA [127] *in vitro* because those studies found that MMR is bidirectional or can repair DNA equally from both 5' and 3' sides.

A two-step model (Figure 11) [9] summarizes the relative repair efficiencies (RE), nature and directionality of MutH dependent and independent repairs from the SPORE assay. This two-stage model suggests for the repair of dG-dT mismatches on the lagging strand. Another possibility is that if replication-coupled MMR process is coordinated from 3'-of a mismatch fails, the second attempt of repair is made. In this second attempt (Figure 11 (B)) MutS dimer will become a diffusive sliding clamp on the DNA which will recruit MutL and then MutH and find for the nearest d(GATC) site to initiate the repair process.

This model was suggested from the quantitative data of the SPORE assay. Which describes the repair of G-T on 5' mismatches on oligos even if 3'-LPR is blocked. This suggests that while 3'-LPR is attempted first, there exist an efficient "backup" mechanism which restarts the repair from 5'-side of DNA.

Figure 11. A Two-Stage Model of dG-dT Repair on The Lagging Strand of galK Gene [9]
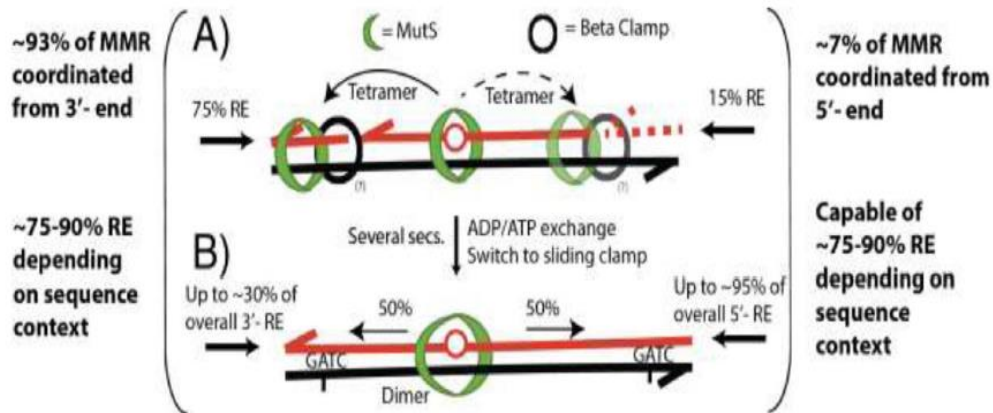
(A) A tetramer MutS assisted endonuclease independent MMR attempted at 3'side which was initiated by MutS dimer, capable of tetramerizing later.
(B) If this replication-associated MMR fails then MutS dimer will be converted into a sliding clamp which will diffuse randomly with MutLH complex to search for d(GATC) site to initiate the repair process. *© 2018 Elsevier B.V. All rights reserved*

These two published studies [8],[9] of SPORE assay has given details about the following advantages of this technique over 'recombineering':

1.   It does not require a spontaneous phenotypic reporter assay [128],[129],[130] like rifampicin resistance phenotypic assay, which can result in false-positives and provides evidence of overall mutation rate rather than MMR activity.

2.   It requires amplification of only targeted sites to identify the rare mutations, eliminating the need for whole-genome sequencing [64]. Also, this targeted site amplification will allow the researcher to perform next-generation deep sequencing [132], which can increase the sensitivity of quantification of probe mismatches.

3.   It reduces the false-positive results by careful design of oligos having higher melting temperature with phosphorothioate bonds, this can be improved to <<0.1% by designing different oligonucleotides [8].

4.   We can target different chromosomal locations in the organisms that have shown the success of oligonucleotide recombination [133],[134],[135],[136].

5.   Introduction of phosphorothioate bonds [8],[9] few base pairs away from either 5' or 3' direction on the oligonucleotides protects the control mismatch from exonuclease digestion associated with DNA polymerase [102].

6.   This assay allows us to monitor oligonucleotide recombination efficiency and MMR repair efficiency of individual probe mismatch, of LPR quantitatively from 5' and 3' direction [9].

7.   This method is highly sensitive in a manner that can quantify the repair of T-T mismatch which are poorly recognized by MutS [8].

8.   It has allowed the construction of a simple mathematical model to quantify the effects of MMR protein mutations different nucleotide mismatches and compare it to biochemical and spontaneous reporter assay [8].

In SPORE assay could likely be further improved by using Locked Nucleic Acids (LNAs) [137],[138] that contains a modified sugar residue with an addition of 2'-C, 4'-C-oxymethylene linker [137] which maintains the ribose ring to 3'-endo conformation. DNA: LNA duplexes [139], [140] as the control mismatch. LNAs demonstrate increased stability and are more resistance to nucleases.

LNAs have shown alteration of three nucleotides in *E.coli* and mouse embryonic stem cells successfully without activation of MMR [138]. LNAs hinders the MutS binding to the mismatch *in vivo* which can be useful in protecting oligonucleotide for specific mutation selection. This has provided another type of protection which can be useful to perform oligonucleotide recombination which may give an insight of MMR *in vivo* [138],[8],[9].

1.7 Developing an Improved SPORE Assay with Genotypic Selection of the Control Mismatch:

We have seen above the unique features of SPORE assay to study the MMR in a cell. This assay is limited by its requirement for phenotypic selection of a bacteria which were selected to grow on a certain media as a result of the 'control' mutation, and it becomes very limiting when some gene selection gives false-positive results. For example, a published study [141] has demonstrated that more than once a kind of spontaneous mutation in *E.coli* bacteria help them to grow on rifampicin. In cases where the rate of spontaneous mutation approaches the rate of oligonucleotide incorporation, we will get false-positive results. When the false-positive rate approaches the rate of oligonucleotide incorporation, as in leading strand MMR, it cannot be used.

Also, the phenotypic selection will create challenges in adapting the SPORE assay for studies in the Mycobacterium genus. Our lab is interested in elucidating a putative MMR-like process in *Mycobacterium tuberculosis*, the bacterial cause of tuberculosis. [142]. Antibiotic resistance in *M. tuberculosis* emerges as a result of genetic mutations [61], [143]. However, the mechanisms of mutation in *M. tuberculosis* are not well

understood. Unlike most other bacteria, *M. tuberculosis* and related organisms do not

possess proteins with homology with canonical MutS/MutL proteins. Rather, in these

bacteria, only one NucS/EndoMS [144] protein has been recently identified which is

believed to play a role in an independently evolved mismatch repair-like process. We

know that (canonical) MMR proteins inhibits recombination events and limits the

mutation rate. We hypothesized that bacteria are modulating their gene expression of

MMR proteins in the presence of antibiotic to acquire a drug resistance or there can be

more recombination events occur which help them becoming resistant. We want to find

the other proteins and involved in the MMR of these bacteria which are still unknown.

These finding of MMR proteins and other unknown proteins *in vivo* will help the

scientific community to design a better drug to inhibit one of the proteins involved in

drug resistance of these bacteria.

For my master's thesis, we decided to develop a SPORE assay that uses a purely

genotypic selection of mutant DNA. This would allow us to more easily study an unusual

MMR process in Mycobacteria. Improving the sensitivity of the SPORE assay would also

allow us to quantify the MMR on the leading strand where incorporation is much rarer,

and also across different chromosomal locations without any need for phenotypic

selection of bacteria. SPORE assay can also give us an idea about how different

mismatches (dG-dT, dA-dC, dG-dA, dT-dC) affect the repair process and how they are

coordinated.

We were motivated by reports of the Depletion of Abundant nucleotide

Sequences by Hybridization) DASH technique [10] of selective PCR of mutated DNA

using restriction enzymes from a highly concentrated background of unmutated DNA. There are general problems associated with a clinical diagnosis of disease using next-generation sequencing, so to develop a highly sensitive and cost-effective molecular diagnostic tool for clinical setup they used CRISPR-Cas9 technology to enrich samples containing rare mutations. Cas9 is an endonuclease from *Streptococcus pyogenes* which is guided by RNA composed of tracr-RNA and crRNA transcribed from the CRISPR (Clustered regularly interspaced short palindromic repeats) sequence [145],[146]. CRISPR-Cas9 uses these RNAs to degrade phage DNA with a complementary sequence to the crRNA from the host without harming the bacteria's own genome. Their central idea (Figure 12) was to deplete 'unmutated' sequences using the gene-editing tool CRISPR-Cas9 [145],[146] before performing PCR so that, during a subsequent PCR amplification step they could selectively enrich the only DNA for KRAS mutant (c.35G>A,p.G12D) sequences.

With pre-treatment by a Cas9 targeting the unmutated 'wild-type' DNA but not the mutated DNA, using DASH they saw a 60-fold increase of mutant sequences (Figure 13) in 1:1000 ratio of mutant: wild type sequence ratio and no detection of mutated KRAS allele with only wild type allele with 0:1.

Figure 12. Depletion of Abundant Sequences by Hybridization [10]
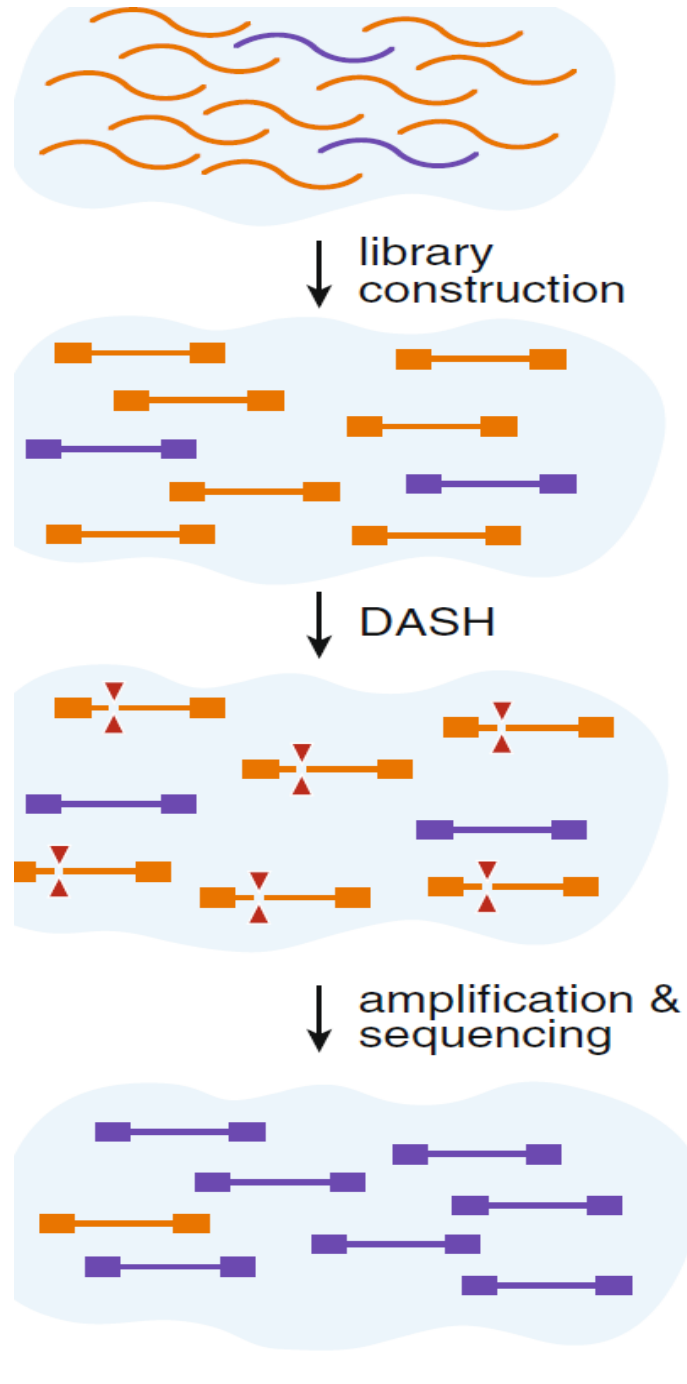
In this Figure, they have shown the target DNA (Orange) and Non-target DNA (Purple). The proportion of the Target to Non-Target sequence has shown very high in the next-generation sequencing library. The strategy of the technique is, adapters can only bind to the intact non-target sequences as they cannot be cleaved by CRISPR Cas9. The target sequences which are cleaved by CRISPR Cas9 cannot be amplified as the adapter cannot bind to them. So, in the end, only enriched non-targets will be present in the final sequencing results. *Copyright © 2016, Springer Nature*

Figure 13. Enrichment of Mutated Sequences from the Samples Using a DASH Technique [10]
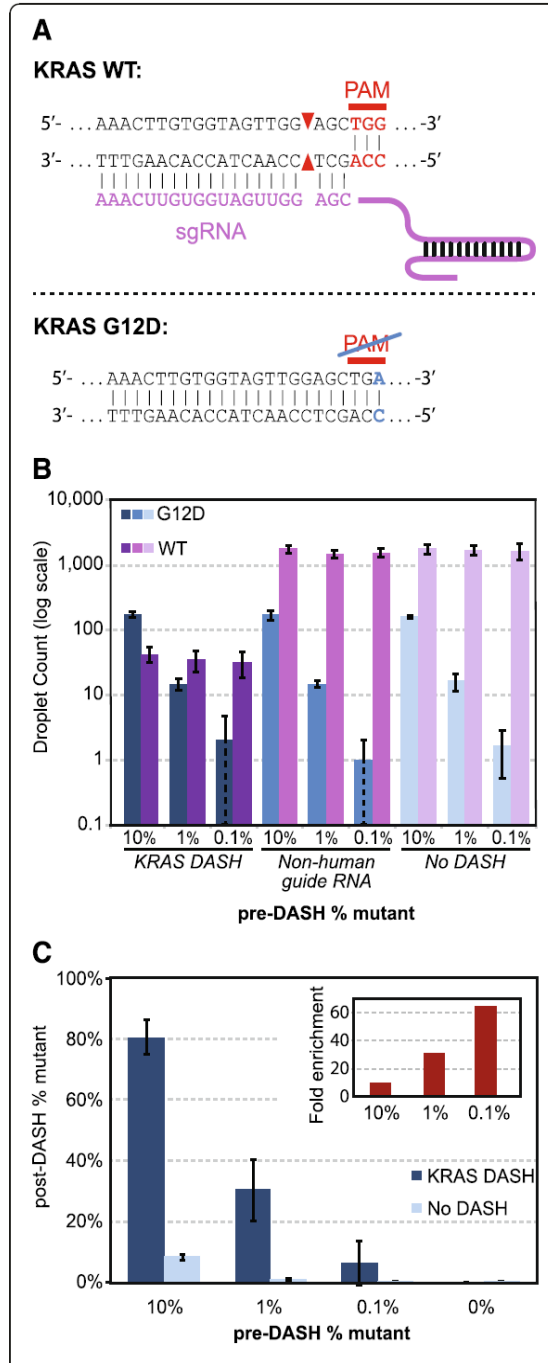
(A) Wild type allele of KRAS at glycine12 position and KRAS G12D mutated glycine 12 position (c.35G>A) which is responsible for malignancies.

(B) Human genomic DNA with different ratios of wildtype to mutant KRAS treated with KRAS-Cas9, Non-human guide RNA (Negative Control) and KRAS-no Cas9. The graph presented as a percentage count of WT and mutant allele sequence was carried out by Digital droplet PCR (ddPCR). Purple bars show the WT allele sequence and Blue bars indicate the mutant allele of KRAS G12D. The data shows that WT allele numbers are decreased after addition of Cas-9 to DASH (mentioned as KRAS DASH) compare to the other control where Cas-9 was not added (mentioned as DASH).

(C) *Inset* represents the same data as Figure 13 B but as fold enrichment of mutant sequences of KRAS treated with Cas9 (Dark Blue) and KRAS not treated with Cas9 (Light Blue). Here, we can see that they could enrich the mutant allele detection from 10% to 81% (8.1-fold), 1% to 30% (30-fold) and 0.1% to 6% (60 fold). 0% indicates no enrichment of mutated sequence as it was only WT KRAS allele. *Copyright © 2016, Springer Nature*

We hypothesized (Figure 14) that if a restriction enzyme could continuously digest 'unmutated' DNA to prevent their enrichment during PCR, rather than just digesting one time before PCR, this technique would be significantly more sensitive and we could identify mutations from the SPORE assay that was much rarer than even 1:1000. This would allow for more difficult experiments, like using the SPORE assay to study MMR on the leading strand or SPORE in mycobacteria. In this study, we used the ApeKI restriction enzyme, derived from hyperthermophilic organism Aeropyrum pernixKI (ApeKI) [147], which is highly thermostable restriction enzyme is most active at temperatures where PCR occurs (70-75 degrees C) [148]. The mutations introduced by our oligos using oligonucleotide recombination make DNA uncleavable by ApeKI restriction enzyme and thus we can do genotypic selection to study MMR using SPORE assay. In other words, the unmutated DNA would contain the recognition site for ApeKI (GCWGC, where W = "A" or "T") while successful incorporation of oligos will have these sites mutated to GCTCC and GCACC.

During my master's project, we measured the sensitivity of ApeKI enzyme under PCR conditions *in vitro* with plasmid DNA containing 'mutated' and 'wild-type' sequences. Then we check the ApeKI efficiency with genomic DNA after performing oligonucleotide recombination to mutate ApeKI sites. Finally, we have performed a SPORE assay *in vivo* and performed genotypic screens to demonstrate that we could use this technique to study MMR processes in living cells with very high sensitivity and without requiring a phenotypic screen.

Figure 14. Our Approach
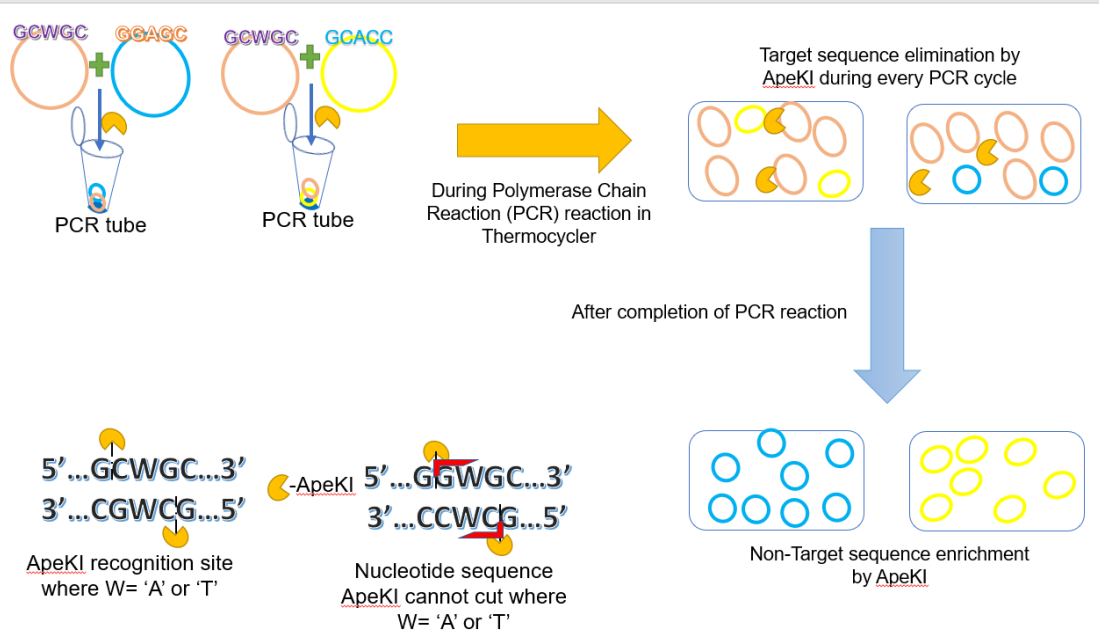
The elimination of un-mutated or targeted site cleavage by ApeKI with every cycle of PCR and simultaneous enrichment of mutated site or sequence that not recognized by of ApeKI will result in DNA for sequencing that is highly enriched for DNA containing the selected genotype.

CHAPTER II

AIMS

1.   Development and optimization of a method to selectively enrich the DNA
of bacteria transformed by oligonucleotide recombination, without phenotypic selection:

According to our hypothesis, we wanted to see the enrichment of selected
mutations (GCWCC, W= A or T) on the genomic DNA after performing PCR by adding
the ApeKI restriction enzyme to PCR reaction. The central idea of this experiment has
been shown in the hypothesis (Figure 14). We first checked the efficiency of ApeKI
restriction enzyme during PCR reaction using the plasmid DNA containing mutations on
the recognition site for this restriction enzyme. So during the PCR reaction, this enzyme
will only cut the DNA containing its recognition site (GCWGC, W = A or T) but it
would not cut the plasmid containing a mutation in this recognition site. After performing
this experiment we found ApeKI restriction highly efficient in its function during a PCR
cycle to enrich selective mutations across the plasmid DNA. Thus, this confirms the
efficiency of ApeKI restriction enzyme and its ability to selectively enrich the mutations
present on the DNA.

Then we perform the oligonucleotide recombination assay on two different genes to mutate the naturally present ApeKI recognition site (GCWGC, W = A or T) in the genome: *galK* and *araD* on *E.coli* genome. The oligos were designed to induce a single mutation in ApeKI recognition site (GCWCC, W= A or T) on the respective gene and they were complementary to the lagging strand of the DNA strand in the genome. The oligo targeted for *galK* gene was capable of inducing one mutated ApeKI recognition site on it while oligo targeting *araD* gene was capable of inducing mutations in two distinct ApeKI recognition sites separated by 4bps, as the *E.coli* naturally contains two ApeKI recognition sites on *araD* gene. We performed the genotypic selection to screen for the specific mutations on the genomic DNA after performing oligonucleotide recombination for both different gene locations. We found that oligo having two mutated ApeKI recognition sites had a better signal to noise ration after Sanger sequencing compared to the *galK* gene having one mutated ApeKI recognition site. Thus, we confirmed to move forward with oligos having two ApeKI mutated sites and we decided to carry the SPORE analysis on *araD* gene to study the MMR.

2.   A SPORE assay to study the MMR process using newly developed genotypic selection method to screen for the presence or absence of mismatches:

We confirmed that ApeKI gives the best enrichment of selected mismatches on the genomic DNA with two mutated ApeKI recognition sites after performing the genotypic selection.

SPORE assay [8], [9] was mainly designed to have protected control mismatches by phosphorothioate bonds so that it will not elicit the MMR action and also saves the oligonucleotides from being degraded by exonucleases. It was based on the phenotypic selection to screen for bacterial mutants after performing oligonucleotide recombination on the *E.coli* genome. Here, we studied of MMR process with Semi Protected Oligonucletoide Recombination with newly developed genotypic selection method was primarily attempted. We tried this assay on the *araD* gene but with two different oligos. Both the oligos were same in length as oligo was used for *araD* gene oligonucleotide recombination but these oligos contained the phosphorothioate bonds (*) at two control mismatches which induce the mutations in two ApeKI recognition sites. The other difference between these two oligos is that one of them contains the 5'dA-dC mismatch and 3'dC-dC mismatch while the second oligo contains 5' dC-dC mismatch and 3'dA-dC mismatch. These oligos will help us to look at the direction that repair of dA-dC mismatches. This will also give quantitive data based on the repair efficiency of MMR of these mismatches.

CHAPTER III

MATERIALS, METHODS, RESULTS AND DISCUSSION

First, we validated that including ApeKI in a PCR reaction could allow us to selectively enrich DNA with mutated ApeKI target sites then measured the sensitivity of ApeKI enzyme *in vitro* with plasmid DNA. We checked the efficiency of ApeKI on genomic DNA by performing oligonucleotide recombination and selectively enriching successful recombinants. We optimized the technique to increase the sensitivity further. Lastly, we have performed a SPORE assay to study the MMR *in vivo*.

3.1 Materials:

Escherichia coli strain SIMD50 (W3110 *galKtyr145UAG ΔlacU169* [λ cI857 *Δ(cro-bioA) (int-cIII<>bet)*]) obtained as a generous gift of the laboratory of Don Court (National Cancer Institute, Frederick, MD, USA). Phusion polymerase (Pfu) Master Mix was obtained by New England Biolabs (Ipswich, MA, USA) and used for all PCR reactions. Gene Pulser(R)/MicroPulser(tm) Electroporation Cuvettes, 0.1 cm gap were obtained from Bio-Rad Laboratories.

Oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA, USA) with standard desalting and used without further purification. Genome Isolation kits, Polymerase Chain Reaction (PCR) purification kits, and gel extraction kits were obtained by New England Biolabs (Ipswich, MA, USA).

3.2 Experimental Procedure:

3.2.1 <u>Testing the Efficiency and Sensitivity of ApeKI Using Plasmid DNA:</u>

To begin, we first wanted to check whether ApeKI was active during PCR reaction conditions and, whether it would allow us to enrich not just mutated ApeKI sites by itself, but also if other flanking mutations (labelled (#) in Figure 15) on the same plasmid are also enriched. The enrichment of side mutations outside the ("control mutation) ApeKI site is a pre-requisite for SPORE assay that we will perform. We also wanted to measure how sensitive this technique was, and how rare the mutated ApeKI sites could be.

We performed this experiment using a "Target" plasmid having ApeKI restriction site ("GCWGC"- W= 'A' or 'T'), and a plasmid having mutated ApeKI site that we labelled "Non-Target" ("GGAGC") plasmid that also contained 7 different nucleotides across about 100 bp outside the mutated ApeKI site [Figure 15 (Bottom Figure)], depending on the mutation.

Samples were prepared by mixing Target plasmid and Non-Target plasmid to achieve Target: Non-Target of 100:1, 10,000:1 and 1,000,000:1 in reactions with a total concentration of 50nM plasmid.

For each mixture 50nM of plasmid DNA concentration was used as template DNA in a PCR reaction with primers 5'-CACTGGAGTTGTCCCAATTCTTG-3' and 5'-TCGCTGGGATTACACATGGC-3' in 50μl of the PCR reaction, which was performed with two sets:

Set 1- Target and Non-target without adding 1μl (5 units/μl) of ApeKI to the PCR reaction.

Set-2 Target and Non-Target plasmid by adding 1μl (5 units/μl) of ApeKI to the PCR reaction. The PCR cycles were used for 25x cycles and unpurified samples were sent out for the Sanger sequencing at GENEWIZ (South Plainfield, NJ).

3.2.1.1 Quantitative Analysis:

The quantitative analysis of the Sanger sequencing results was performed using BioEdit software and MATLAB (Mathworks, Inc). The raw chromatogram data from the sequencing results were imported in BioEdit software and extracted into .txt files and imported into MATLAB. A code was written in MATLAB to extract the signal strength for each nucleotide channel at each called sequence peak. The signals for each channel were normalized by the mean signal intensity of that nucleotide channel across the region of interest, and the fraction of the signal intensity under each peak was assigned to the Target or Non-target plasmid for the samples treated with ApeKI and without ApeKI. This will give sequence comparison quantitatively that how much (%) of the sequence mutated or repaired mismatched were enriched after adding ApeKI.

Figure 17 shows the enrichment of the Non-Target plasmid with the mutated

ApeKI site as well as the flanking mutations on the same Non-Target plasmid DNA after

completion of PCR reaction.

Figure 15. The Compared Sequence of Two Distinct Plasmids
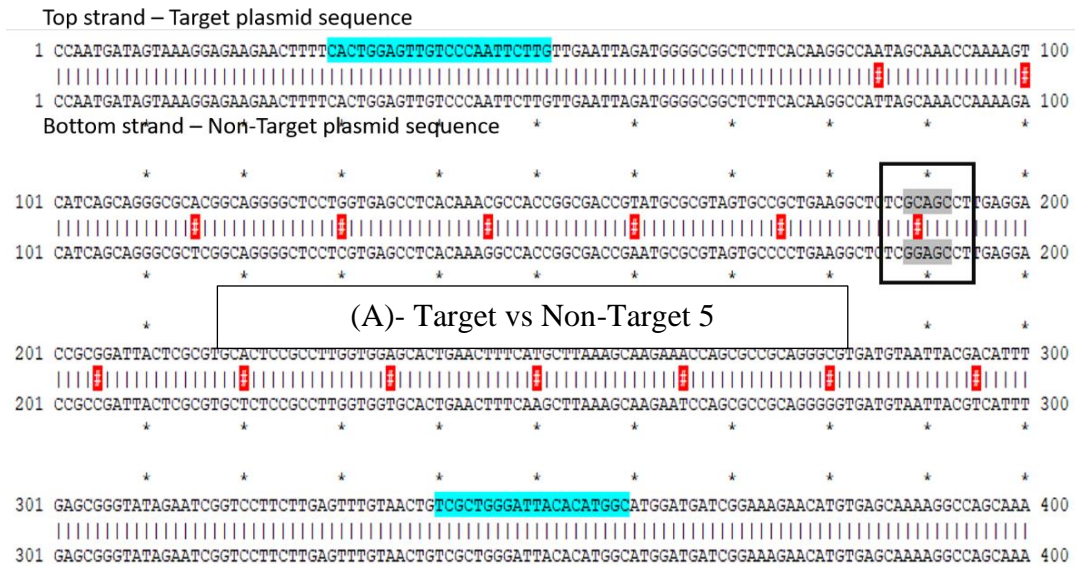
The sequence alignment of a "Target" plasmid DNA with an ApeKI restriction site (top strand), and a "Non-Target" plasmid (bottom strand) with a mutated ApeKI site. Black square box indicates the focused mutation that we are attempting to selectively enrich from the "Non-target" plasmid. (A) Sequence alignment between the "Target" plasmid and plasmid "Non-Target-5". We also want to determine if ApeKI will enrich the flanking differences outside the restriction. (#) indicates the different mutated site present between plasmids. [Highlighted blue sequence indicate the primer sides for the plasmid]

Figure 16. Gel Image of PCR Amplified Products
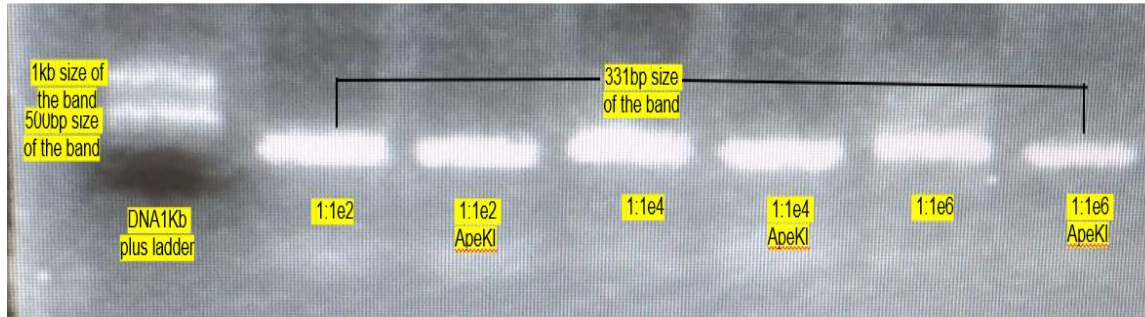
Plasmid DNA after PCR amplification with and without adding ApeKI with different ratios of Non-Target to Target plasmid DNA. This shows the DNA band of 331bps after PCR which shows the enrichment of desired sequence having side mutations as well as the main ApeKI site mutation enrichment which is further confirmed by Sanger Sequencing.

Figure 17. Quantitative Enrichment of Non-Target (NT) Sequence Fraction from the Mixed Sequences

(B) 1e4:1 T:NT

nucleotide position (ApeKI mutation at 107)

(C)
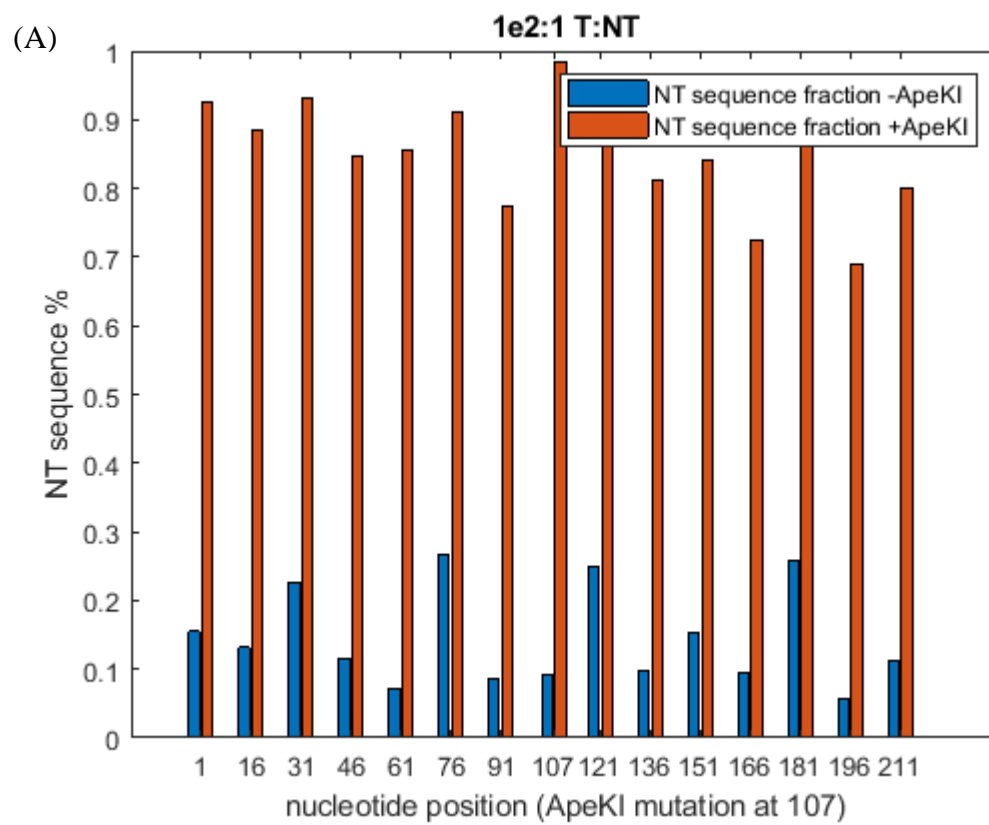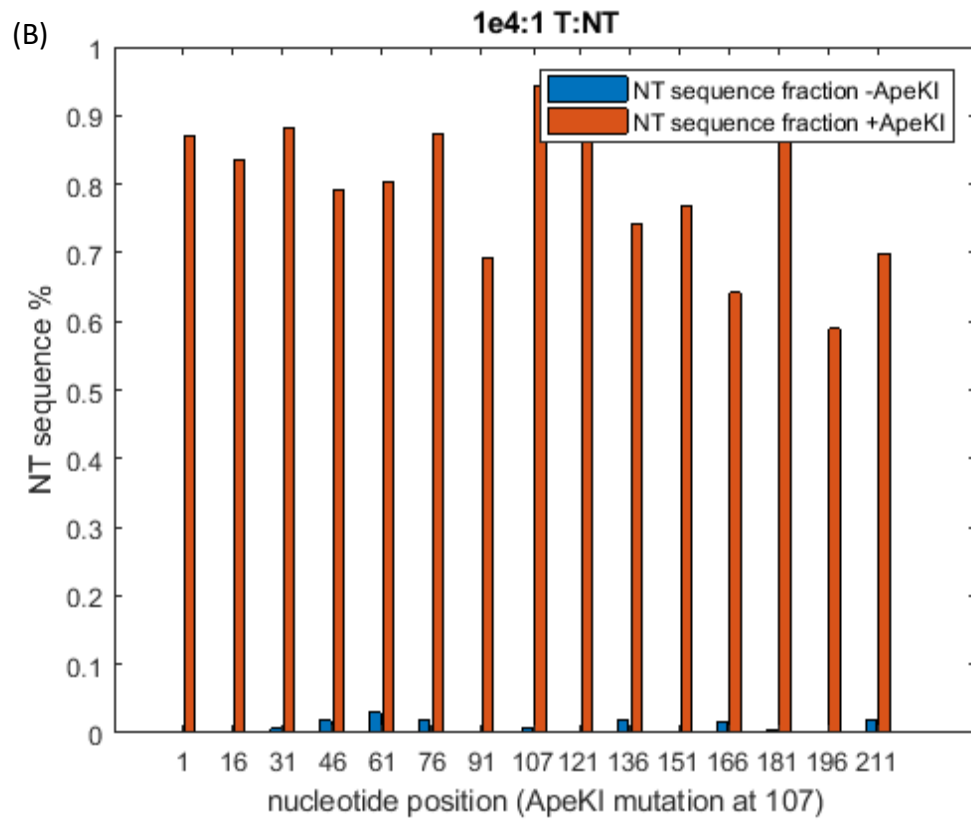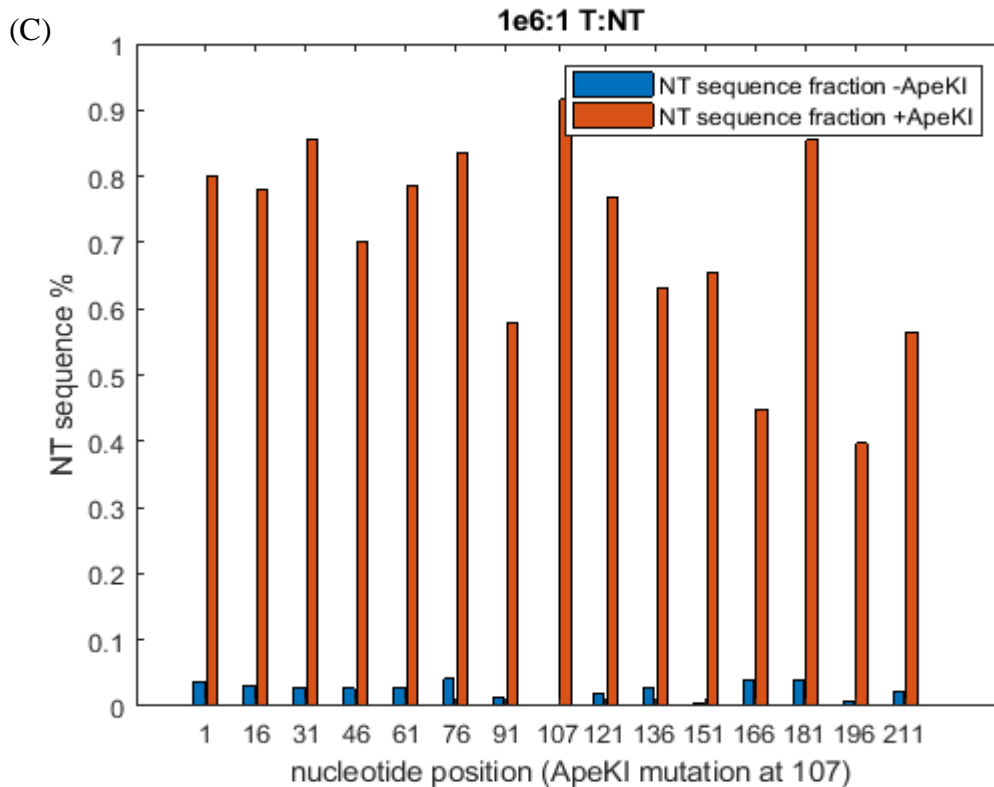
This represents the enriched fraction of NT sequence before and after adding ApeKI to the PCR reaction. Blue bars indicate the fraction of NT sequence when ApeKI was not added in the PCR and Red bars represent the fraction of NT sequence from the pool of mixture after adding ApeKI to the PCR reaction. Above data, ratios of Target (T): Non-Target (NT) in each reaction. (A) 1e2:1, (B) 1e4:1 and (C) 1e6:1.

Here, we can see quantitatively that the Non-Target (NT) sequence is significantly enriched over Target (T) plasmid sequence after PCR with ApeKI in all the ratios. After ApeKI enrichment, Sanger sequencing revealed that the signals initially mixed as 1e2:1 Target: Non-Target plasmids were on average 85.47% Non-Target sequence with a standard deviation of 8.27% across all sites that differ in sequence.

This is a significant increase compared to the previous report of DASH using Cas9 digestion before PCR, which resulted in 30% untargeted sequences after initial

1e2:1 digested:non-targetted mixture (Figure 13). 1e4:1 dilutions were, after enrichment, 79.32% non-target with a standard deviation 10.29% across all differing sites, and 1e6:1 dilutions were, after enrichment 70.48% Non-Target sequence with a standard deviation of 15.52% across all sites. In summary, for 1e2:1, 1e4:1, and 1e6:1-fold dilutions, that is an enrichment of ~85-fold (compared to 30-fold using Cas9 DASH (Figure 13)), 7900-fold, and 705000-fold enrichment, relatively.

Conclusion: We have confirmed that ApeKI could be used to successfully enriching not just the mutated ApeKI site but the entire flanking sequence on the plasmid surrounding the site with high efficiency. This is a prerequisite for using this approach to study the MMR with genotypic selection in the SPORE assay.

3.2.2 Testing the Efficiency of ApeKI to Enrich Successful Oligonucleotide Recombination Mutations in the *E. coli* Genome:

To perform oligonucleotide recombination, *E. coli* strain SIMD50, which expresses the beta recombinase in a heat-inducible manner at 42⁰C, was transformed with ssDNA oligos and screened according to the new genotypic selection protocol using oligos complementary to the lagging strand of *galK* and *araD* genes of *E.coli*. We used sterile techniques to make electrocompetent *E.coli* cells.

Bacterial colonies of SIMD50 were grown in 10ml Luria Broth (LB) overnight at 30⁰C with shaking at 200rpm and next day, 0.5ml of growth solution was added to 35ml of the sterile LB in 50ml centrifuge tubes and grown further at 30⁰C at 200rpm until 0.4-0.6nm O.D. of the bacterial culture was achieved. Then the tubes were heat-shocked at 42⁰C in a water bath for 15mins with agitation and then immediately cooled on ice for 10

mins. The tubes were then spun in a centrifuge at 6500g for 7mins at 4ºC and the LB gently decanted after centrifuge. Bacterial pellets were resuspended in 1ml of sterile and pre-chilled water and then pre-chilled 30ml of sterile water was added to the resuspended pellet and again centrifuged for 7mins at 4ºC and 6500g in a centrifuge. The tubes were immediately removed and the supernatant was gently removed and bacterial pellets were again resuspended in the 1ml of water and transferred to the sterile pre-chilled 1.5ml microcentrifuge tubes and spun for 30secs at 4ºC. then after discarding the supernatant the pellet again resuspended in 1ml of prechilled sterile water and spun for 30 secs at maximum (g) at 4ºC. Then after discarding a supernatant pellet was resuspended in 200μl of prechilled sterile 15% glycerol and divided the competent cells into 1.5 microcentrifuges each having 50μl of competent cells and stored at -80ºC or used fresh by mixing 2μl of 100mM ssDNA oligos gently. The mixtures were electroporated at 1.8kV using GenePulser electroporation system and immediately mixed with 1ml of prewarmed LB media after electroporation and grown for 30mins at 30ºC with shaking in a shaking incubator. 30mins after recovery each tube was divided into two tubes with 500μl of bacteria in each tube and grown for 15mins at 200rpm in 30ºC shaker incubator and then after 15 mins, the samples were put on room temperature for 10mins more and then genome isolation protocol was performed using New England Biolabs (Ipswich, MA, USA) kit. The genotypic selection protocol shown in Figure 18 was performed after isolating the genome DNA from bacteria.

Figure 18. The Optimized Genotypic Selection Protocol
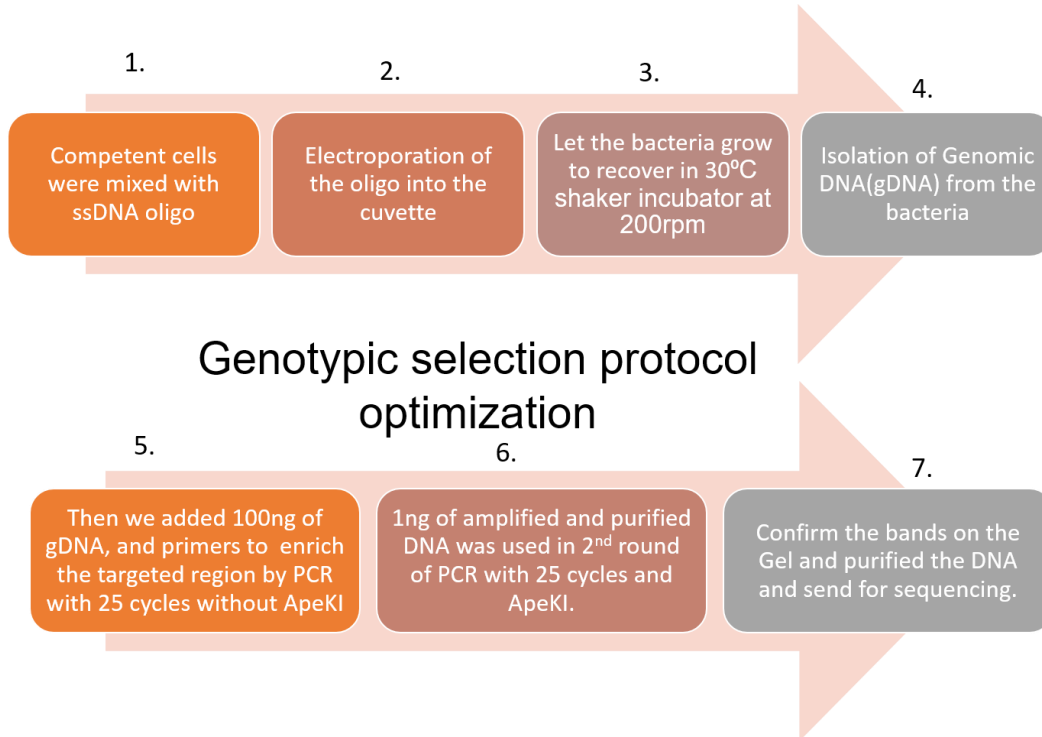
After transfection with this oligo, we isolated *E.coli* genomic DNA. 100ng of genomic DNA were added to a PCR reaction, and primers used to enrich the targeted region by PCR with 25 cycles without adding ApeKI. Then a single band was confirmed on a gel and purified for the second round of PCR. 1ng of purified DNA was used for the 2nd round of PCR with 25 cycles.

To validate that our approach could be used to enrich a genomic mutation introduced using the oligonucleotide recombination technique, we first performed oligonucleotide recombination using oligos we designed to have complementarity with lagging strand of *galK* gene. The oligo sequence 5'-CGT GCG TCA TAT ACT GAC TGA AAA CGC CCG CAC C**C**T TGA A**GC TC**C CAG CGC GCT **C**GA GCA AGG CGA CCT GAA ACG TA T GGG CGA G (Contains the mutated ApeKI recognition site C-C mismatch (showed as <u>C</u>) which is going to bind and mutate the original sequence

with ApeKI site= (GCTGC, W= A or T) on the *galK* gene. Underline nucleotides on the sides **'C'** represents the flanking mismatches. Recall that dC-dC mismatches are not recognized by MutS so we do not expect this oligonucleotide to excite MMR.

As a baseline test of selection for successful recombinants, we made oligos strategically that can induce the mutation to the *galK* gene which mutates the ApeKI recognition site (yellow) from 'GCTGC' to 'GCTCC' through dC-dC mismatches (red, above). We then added flanking dC-dC mismatches (blue) outside of the ApeKI site to test whether those other mutations would be enriched along with the DNA with mutations in the ApeKI site. This time we did not select for a mutation with a premature stop codon on the *galK* gene as described in Figure 10 (used previously for phenotypic selection), but rather we are performing genotypic selection on the middle C.

The protocol shown in Figure 18 was followed and then we prepared the following two sets of PCR samples with 5'-CTGTTGCGCATGAACTGGAC-3' and 5'-GCGCATAGAGGCATGAGACT-3' primers.

Reaction 1 (Figure 19 (TOP)): 1ng of amplified DNA to the 50μl PCR reaction for 25 cycles without adding ApeKI as control.

Reaction 2(Figure 19 (Bottom)): 1ng of amplified DNA to the PCR reaction for 25 cycles with ApeKI during the PCR cycle.

After completion of PCR, the PCR sample was loaded on the agarose gel. We purified the amplified DNA and then we send it to GENEWIZ (South Plainfield, NJ) for Sanger sequencing. The results are shown in Figure 19.

Figure 19. Sanger Sequencing Chromatographs of galK Gene
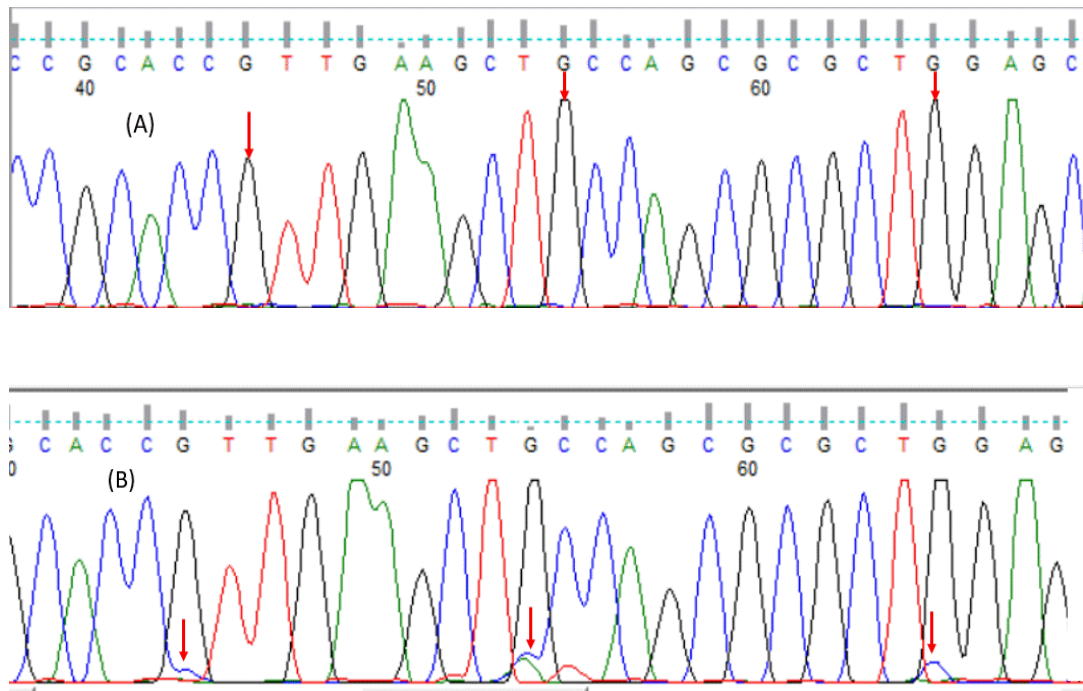
Isolation of rare genetic mutation induced on *galK* gene by oligonucleotide recombination. (TOP) Sequencing of a gDNA amplicon without adding ApeKI during PCR cycle. (Bottom) Sequencing of a gDNA amplicon after adding the ApeKI during the PCR cycle. Red arrows show and compare two Figure's signal difference.

3.2.2.1 Quantitative Analysis:

The raw chromatographs of the Sanger sequencing were processed as before to find the signal strength for each nucleotide channel present at the called sites of the introduced mutations. The code extracts the nucleotide numbers from the raw chromatograph data based on each position showed in Table 1.

63

Table 1. Comparison of Quantitative Values from Sanger Sequencing Chromatographs of galK Gene Before and After Adding ApeKI

	Without adding ApeKI-The mean (%) of mutated sequence on the galK gene after oligonucleotide recombination		After adding ApeKI- The mean (%) of mutated sequence on galK gene after oligonucleotide recombination	
	G -Wild Type (n=2)	C- Mutated (n=2)	G – Wild Type (n=2)	C - Mutated (n=2)
5'	95.02 + 2.54	0.37 + 0.19	82.02 + 3.10	13.56 + 4.13
ApeKI RS (Restriciton site)-1	98.7 + 1.14	0.43 + 0.61	69.94 + 3.75	18.22 + 3.18
3'	96.82 + 1.85	1.55 + 0.97	87.27 + 0.96	11.14 + 1.17

The quantitative analysis of selectively enriched mutations selected for different positions on the oligonucleotides where one ApeKI recognition site was mutated. The data represents the mean value and standard deviation of the signal strength for the G and C nucleotide channels during Sanger sequencing at each targeted site without adding ApeKI (left) and after adding ApeKI enzyme (right) to the sample after oligonucleotide recombination. A G- means the ApeKI site is unmutated while a C is a mutation in the ApeKI site.

Figure 20. Bar Graph Obtained After Quantitative Analysis of galK Gene Chromatographs
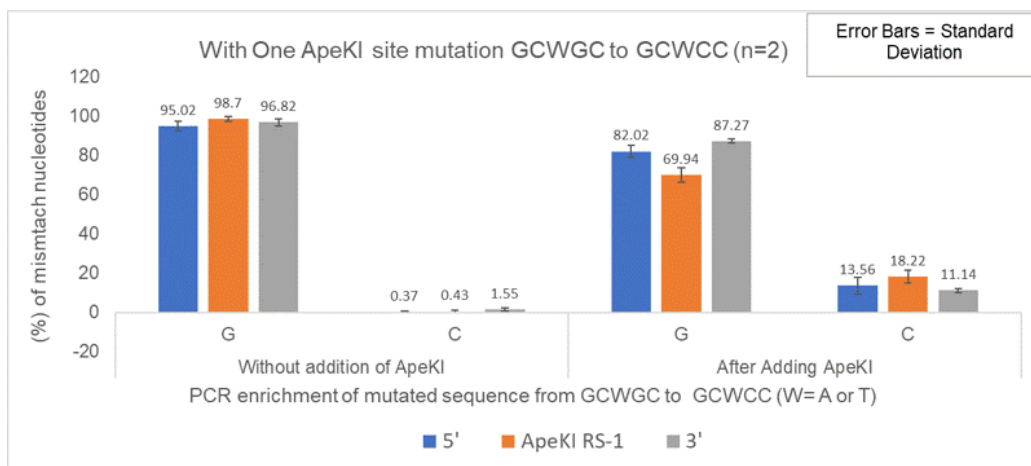
Enrichment of GCTCC sequence from GCTGC (mean + standard deviation, n= 2) after successful oligonucleotide recombination into the *E.coli* genome. This

64

shows that enriched with dG-to-dC mutations after addition of ApeKI during genotypic selection method.

Oligonucleotide recombination with phenotypic selection typically results in mutational efficiencies on the order of 1-in-100, although to the best of our knowledge the true mutational frequency from genotypic screens from oligonucleotide recombination is not known but expected to be a lot lower. Figure 20 shows the enrichment of dG-to-dC mutations at the different positions targeted for mutation by the oligonucleotide either with or without ApeKI in the PCR reaction. In Table 1 we can see that the maximum enrichment of genomes with dG-to-dC mutations to between 11.14% to 18.23%.

These results show the lower than expected enrichment of dG-to-dC mutations after oligonucleotide recombination. Inspection of the chromatograph of Sanger sequencing in Figure 19 revealed that a significant fraction of the enriched population had unintended mutations in the ApeKI mutated site, which likely limit.

Conclusion: While we see selective enrichment of DNA with mutated ApeKI sites and the flanking mutations, Sanger sequencing also revealed other mutations in the ApeKI site that was not intentionally created by the incorporation of the oligo. These might have been present from natural variants in the *E.coli* population or, more likely, created from errors by the DNA polymerase during PCR.

We hypothesized that if we have two ApeKI sites mutated then we can enrich the signal more efficiently. We found the *araD* gene in *E. coli* genome, where there are two ApeKI recognition sites separated by 4bp are present. We tested this hypothesis by

designing oligonucleotides that would mutate these two ApeKI sites using the same

oligonucleotide.

3.2.3 Oligonucleotide Recombination With TWO ApeKI Sites in the Genome:

The oligo complementary to *araD* lagging strand was designed with two dC-dC

mismatches at the ApeKI restriction sites in the genomic DNA and two flanking dC-dC

mismatches. The oligonucleotide sequence is CGT CAG TTA GCG CCG CAG TTA

CC**C** GAT AT<u>G CAC C</u>AA AC<u>G CTC C</u>TG GAT AAA CAC TAT CT**C** CGT AAG

CAT GGC GCG AAG GC where the mutations are shown with <u>C</u> and <u>C</u> nucleotides and

flanking mutations are presented as **<u>C</u>**. This oligo resembles the oligo used in the SPORE

assay with 'control mismatch (yellow in Figure 21), two probe mutations (C).

Figure 21. Oligonucleotide Recombination on araD Gene in E.coli

(B)

LEADING ORIGINAL 5'-CGTCAGTTAGCGCCGCAGTTACCGGATATGCAGCAAACGCTGCTGGATAAACACTATCTGCGTAAGCATGGCGCGAAGGC-3'

OLIGO 5'-CGTCAGTTAGCGCCGCAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCATGGCGCGAAGGC-3'

LAGGING ORIGINAL 3'-GCAGTCAATCGCGGCGTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGTACCGCGCTTCCG-5'

AFTER OLIGONUCLEOTIDE INCORPORATION

AFTER OLIGO INCORPORATION 5'-CGTCAGTTAGCGCCGCAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCATGGCGCGAAGGC-3'
3'-GCAGTCAATCGCGGCGTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGTACCGCGCTTCCG-5'

NEXT ROUND OF REPLICATION

5'-CGTCAGTTAGCGCCGCAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCATGGCGCGAAGGC-3'
3'-GCAGTCAATCGCGGCGTCAATGGGCTATACGTGCTTTGCGAGGACCTATTTGTGATAGAGGCATTCGTACCGCGCTTCCG-5'

5'-CGTCAGTTAGCGCCGCAGTTACCGGATATGCAGCAAACGCTGCTGGATAAACACTATCTGCGTAAGCATGGCGCGAATTC-3'
3'-GCAGTCAATCGCGGCGTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGTACCGCGCTTCCG-5'

Enriched sequence by ApeKI in the PCR

Enriched sequence after PCR 5'-CGTCAGTTAGCGCCGCAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCATGGCGCGAAGGC-3'
3'-GCAGTCAATCGCGGCGTCAATGGGCTATACGTGCTTTGCGAGGACCTATTTGTGATAGAGGCATTCGTACCGCGCTTCCG-5'

(A)Integration of an oligonucleotide to introduce mutations to two ApeKI sites during replication. It is the representation of how oligonucleotide incorporates into the genome that is complementary to the lagging strand. The lagging strand is extended in the opposite direction of the replication fork. The oligonucleotide (grey color) which is complementary to the lagging strand (light blue). Yellow color indicates the ApeKI restriction sites. Underline nucleotides represents the mutations that we are focusing on. (B) shows how oligonucleotide will bind to the complementary sequence in the genome of bacteria. Then after incorporation into the genome, it will stay permanent inside the genome and replicated as indicated above in every round of replication with that mutation.

Oligonucleotide recombination was performed as mentioned earlier (Figure 18) and the same PCR with primers 5'-CAGGCGAAAGCCTTGTTCAC-3' and 5'-CATGGGGCAAAAATGCCGAA-3', enrichment protocol was followed by genomic DNA isolation. The DNA samples after gel extraction were sent to the GENEWIZ (South Plainfield, NJ) for Sanger sequencing.

Figure 22. Gel Image of PCR Amplified Products of araD Gene
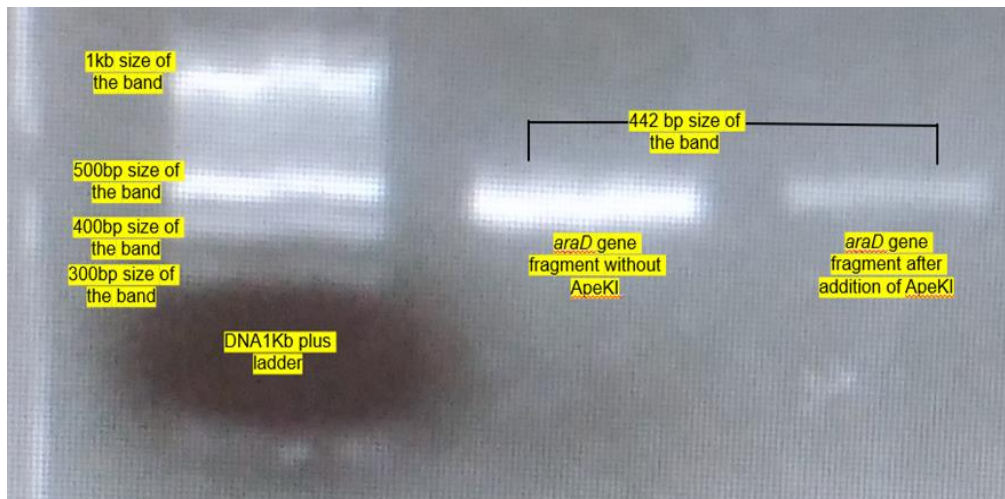
After 2nd round of PCR with and without the addition of ApeKI to the *araD* gene fragment. The gel was stained with SYBR Gold Stain and the single band was confirmed at 442 bp size. Here, the single bands from the gel were cut for further Sanger sequencing and analysis.

Figure 23. Sanger Sequencing Chromatographs of araD Gene
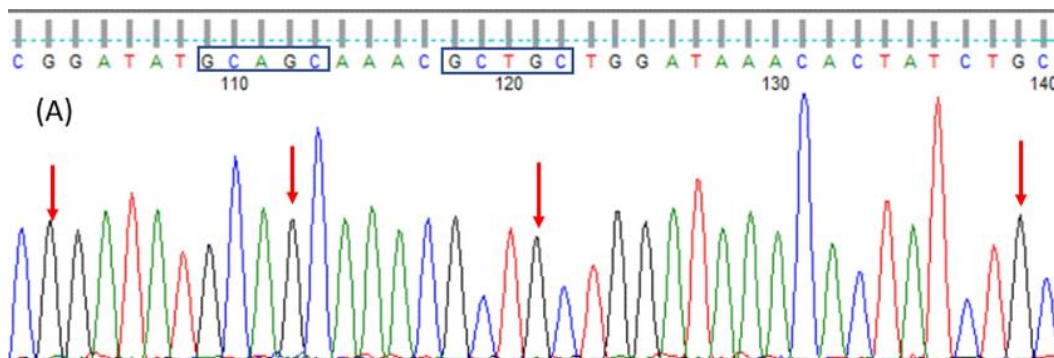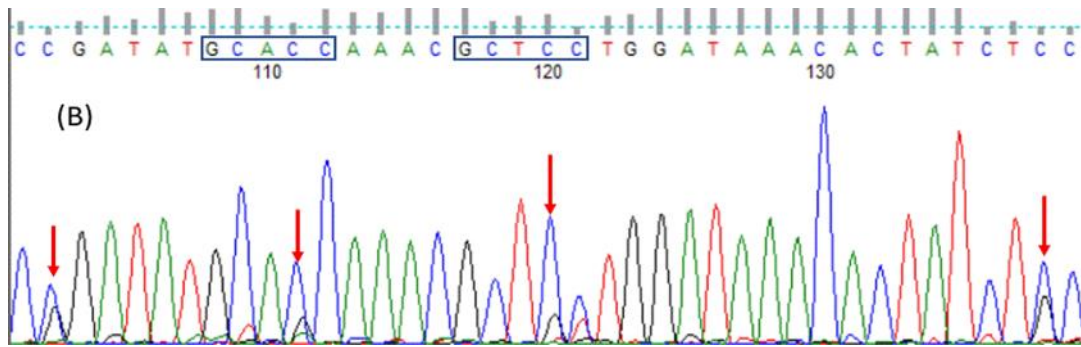
(A)Sequencing of a gDNA amplicon without adding ApeKI during PCR cycle. (B) Sequencing of a gDNA after adding the ApeKI during the PCR cycle. Red arrows show and compare two Figure's signal difference. Sequence difference can be seen between (A) and (B) with the highlighted nucleotides in the box.

3.2.3.1 Quantitative Analysis:

The raw chromatographs of the Sanger sequencing were processed as before with the results shown in Table 2. The data presented in the table is also represented in the form of graphical analysis which is shown in Figure 24.

Table 2. Comparison of Quantitative Values from Sanger Sequencing Chromatographs of araD Gene Before and After Adding ApeKI

	Without adding ApeKI-The mean (%) of mutated sequence on the araD gene after oligonucleotide recombination		After adding ApeKI- The mean (%) of mutated sequence on araD gene after oligonucleotide recombination	
	G -Wild Type (n=2)	C- Mutated (n=2)	G – Wild Type (n=2)	C - Mutated (n=2)
5'	96.08 + 0.34	0.52 + 0.73	23.15 + 9.76	67.5 + 0.43
ApeKI RS (Restriction site)-1	96.21 + 1.92	1.9 + 0.76	8.7 + 7.6	80.8 + 2.98
ApeKI RS (Restriction site)-2	94.65 + 1.56	2.12 + 2.36	9.33 + 9.02	81.03 + 0.71
3'	96.8 + 0.22	1.17 + 0.94	23 + 16.62	67.02 + 8.31

The quantitative analysis of selectively enriched mutations selected for different positions on the oligonucleotides. The data represents the mean value and standard deviation of mutation at each site on the oligonucleotide without adding ApeKI and after adding ApeKI enzyme to the sample.

Figure 24. Bar Graph Obtained After Quantitative Analysis of araD Gene
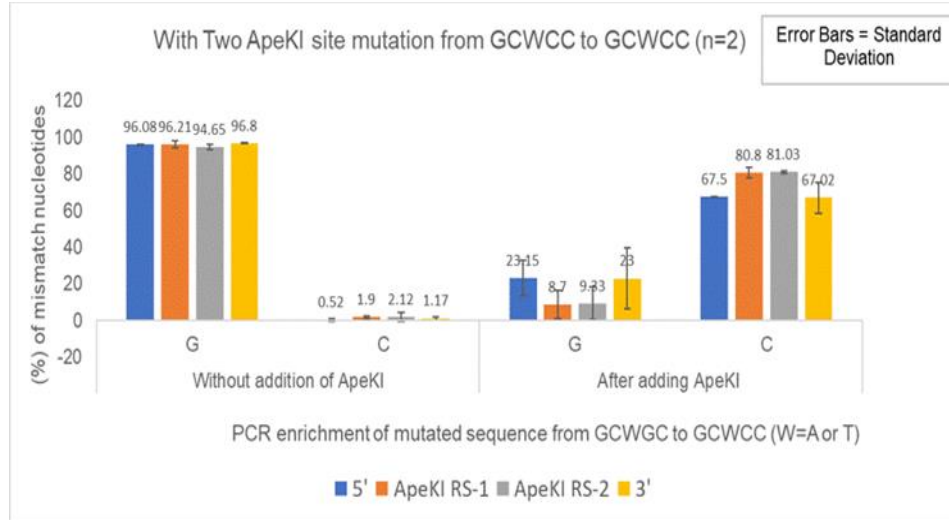
Enrichment of GCWCC sequence from GCWGC (W= A or T) (mean $\pm$ standard deviation, n =2) after successful oligonucleotide recombination into the *E.coli* genome. This shows that enriched dC-dC mismatches after addition of ApeKI during genotypic selection method.

Figure 24 shows the enrichment of dG-to-dC mutations at different positions targeted by the oligonucleotide after addition of ApeKI compared to when ApeKI was not added into the PCR reaction. The comparison can be made between Figure 23 (A) before adding ApeKI where it shows the GCAGC and GCTGC which is highlighted with the box and Figure 23 (B) the Sanger sequencing chromatographs which shows GCACC and GCTCC sequence after adding ApeKI. The sequence below the graph represents the 5'-end & 3'-end, on the oligo it follows the order of nucleotide enrichment from 5' to 3' direction on the oligonucleotide.

This also represents the enrichment of dG-to-dC mutations is higher at ApeKI RS (restriction site) -1,2 positions which serve as a 'control mismatch' in the spore assay.

While enrichment of dG-to-dC mutations at 5' and 3' represents the 'probe mismatch' of SPORE assay.

From Table 1 we can see that enrichment of dG-to-dC mutations to 67% for flanking mutations and around 81.07% for mutations to the ApeKI restriction sites. When this data we compare with the oligo containing one ApeKI mutated site we can see the better enrichment of dG-to-dC mutations on the genome which is a little bit less compare to the 85% $\pm$ 8.7% enrichment of plasmid DNA. This indicates that two mutated ApeKI recognition sites are the best to perform the genotypic selection on the genome which gives the highest enrichment of mismatches on the genomic DNA compared to the one mutated ApeKI recognition site.

Conclusion: Above results indicate that the strategy to screen simultaneously for two mutated ApeKI recognition site mutations has worked significantly better than the screen using only one, and we got more efficient signals from the sequencing. We can see significantly stronger enrichment of all four mutations (two flanking the ApeKI sites) introduced by this oligonucleotide.

This suggests that using two ApeKI separate recognition sites for genotypic screening helps to enrich the targeted mutations. We have reproduced these results two times. Thus, we confirmed that these oligonucleotides are the best to design the oligonucleotides for SPORE assay.

3.2.4 SPORE Assay using Genotypic Selection:

With the above findings that we could successfully enrich mutations introduced by oligonucleotide recombination using our genotypic selection strategy, we performed

the SPORE assay. As we know from Figure 10 that SPORE assay [8] is a technique that allows us to study MMR process inside the cell. SPORE assay also helps to study the directionality of MMR [9] described in Figure 25. Here we attempted to test whether we could use ApeKI for a genotypic screening of SPORE assay.

Figure 25. The Directionality of MMR on Lagging Strand in E.coli [9]
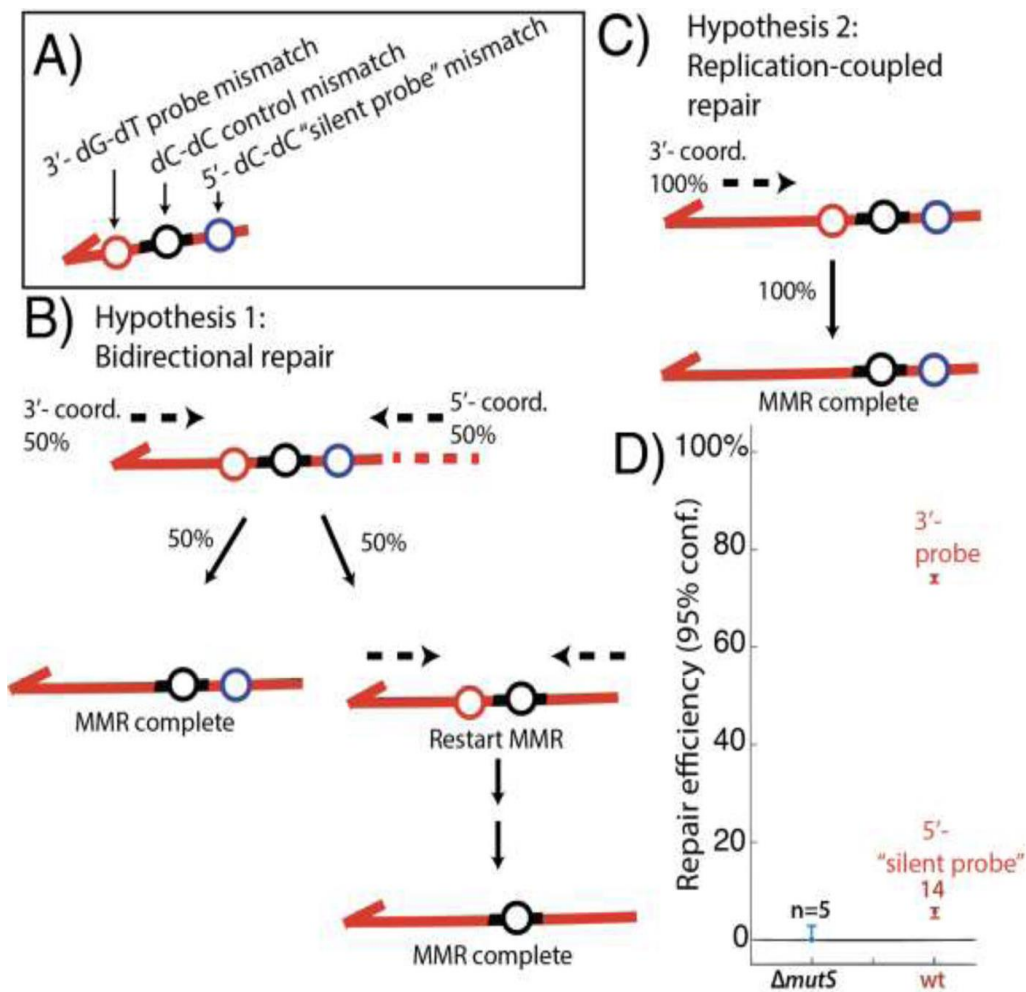
(A)The probe mutation dG-dT (red color) excites MMR, while the control mutation protected by phosphorothioate bonds (black color) and a silent mutation dC-dC (blue color), does not excite the MMR.

(B) If the MMR is bidirectional then there is a 50% chance the silent mismatch (blue circle) will be repaired inadvertently by long-patch repair (LPR) as often as the probe mismatch (red circle).

(C) If MMR is coupled to replication and unidirectional, then we would not expect the silent mismatch to be repaired during repair of the probe mismatch.

(D) Represents the data that MMR is bidirectional and Long Patch Repair (LPR) is coordinated from 3'-side of DNA where probe mutation is repaired around 80% and 5'-silent mutation repaired with low efficiency.

In this study, we had two different oligos (Figure 26): Oligo 1 which one had 3'-site "MMR active (underlined)" probe mismatch, two protected (with phosphorothioate bonds labelled *) "control mismatches" and one unprotected "silent mismatch (underlined)" at 5'-site. Another oligo (oligo 2) had a 5'- site "MMR active" probe mismatch (underlined), two protected "Control mismatches" and one unprotected "Silent mismatch (underlined)" at 3'-site.

SPORE Oligo 1: 5'-CGT CAG TTA GCG CCG CAG TTA CC<u>C</u> GAT AT*G*C*A*C CAA AC GCTC*C*T*G* GAT AAA CAC TAT CT<u>A</u> CGT AAG CAT GGC GCG AAG GC-3'. SPORE Oligo 2: 5'-CGT CAG TTA GCG CCG CAG TTA CC<u>A</u> GAT AT*G*C*A*C CAA AC GCTC*C*T*G* GAT AAA CAC TAT CT<u>C</u> CGT AAG CAT GGC GCG AAG GC-3'

The oligos used for this study is different than they were used in Figure 25, where researcher used dG-dT mismatches as the probe and here we have used dC-dA as a probe mismatch to study the MMR.

Figure 26. SPORE Assay on araD Gene
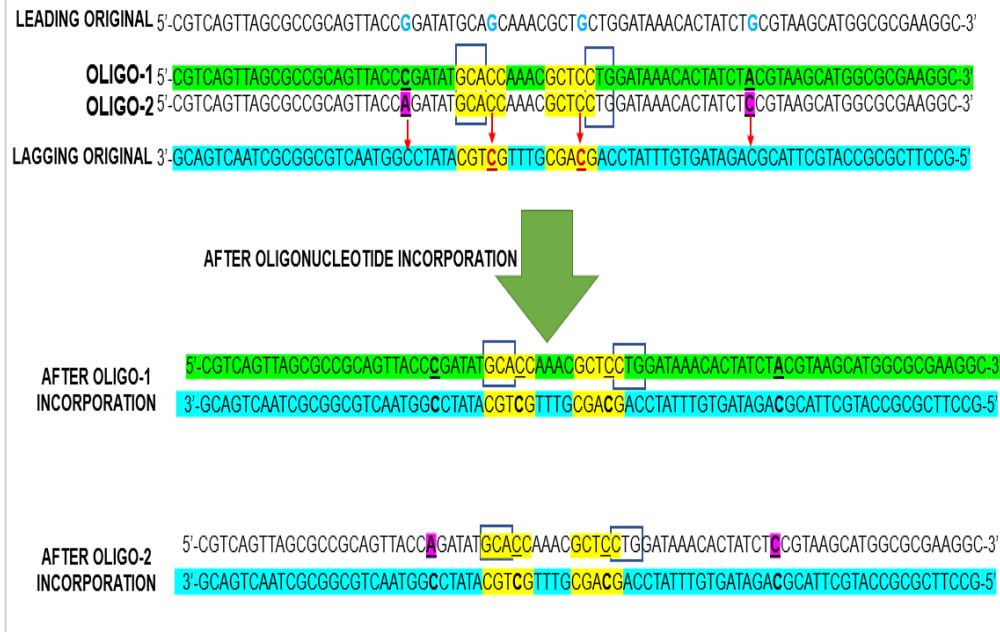
Representation of genomic lagging strand which is complementary to the genomic leading strand (no color), oligo-1(green) and oligo-2(no color) also. If no repair occurs. Blue boxes on the oligonucleotide indicate the phosphorothioate bonds to protect the control mismatches recognized by MMR proteins. Probe mutations are indicated with bold and underlined letters. Oligo-2 probe mismatches are indicated with purple color box. Red arrows indicate the targeted mismatches to the lagging strand. This Figure also shows the genomic DNA sequence after oligonucleotide incorporation. We have discussed the hypothesis and results in the next section after this on how this oligonucleotide will be repaired. [Blue boxes shows the presence of phosphorothioate bonds as protection for control mismatch, purple boxes shows the probe mismatch in oligo 2 while in oligo 1 probe mismatches are underlined outside of the phosphorothioate bonds. Each yellow box indicates the ApeKI recognition site mutation.]

76

We have strategically introduced dA-dC mismatches on the 3' side of the oligonucleotide and dC-dC mismatch on the 5' side of it. Now, if the MMR starts from the 5' side and stops then only dC-dC will be converted to G-C base pair. On the other hand, if the MMR proteins repair only from 3' side then dA-dC will be repaired to G-C base pair in the next round of replication. If MMR proteins work bidirectionally in the genome then both dA-dC and dC-dC will be converted to G-C base pair on the sides as showed in the Figure. [Blue boxes show the presence of phosphorothioate bonds as protection for control mismatch, oligo 1 probe mismatches are underlined outside of the phosphorothioate bonds. Each yellow box indicates the ApeKI recognition site mutation.]

Figure 28. The Outcomes of Different Hypotheses of How the Repair of Oligo-2 could Occur After Incorporation
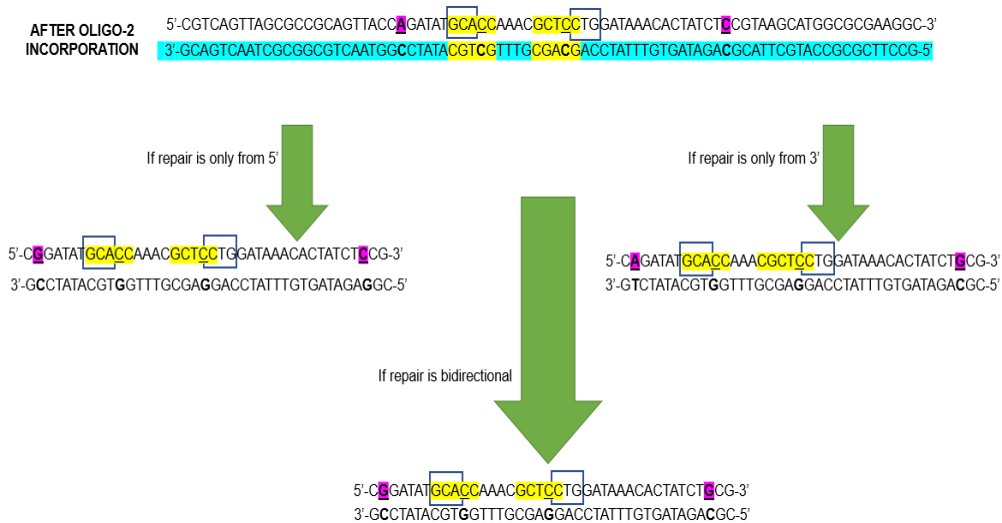
In the oligo-2 we have strategically introduced dA-dC mismatch on the 5' side of the oligonucleotide and dC-dC mismatch on the 3' side of it. Now, if the MMR starts from the 3' side and stops then only dC-dC will be converted to G-C. On the other hand, if the MMR proteins repair only from 5' side then dA-dC will be repaired to G-C in the next round of replication. If MMR proteins work bidirectionally in the genome then both A-C and C-C will be converted to G-C on the sides as showed in the Figure. [Blue boxes show the presence of phosphorothioate bonds as protection for control mismatch, purple boxes shows the probe mismatch in oligo 2 outside of the phosphorothioate bonds. Each yellow box indicates the ApeKI recognition site mutation.]

Figure 29. The Gel Image of a SPORE Assay Gene Fragment of Oligo 1 and Oligo 2
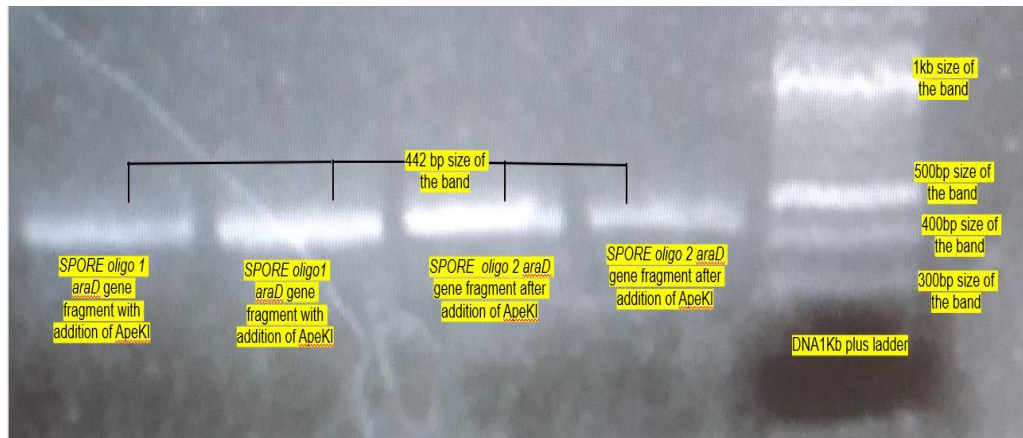
The Gel image was taken after the 2nd round of PCR with the addition of ApeKI only. These bands with 442bp size were cut out and sent for the Sanger sequencing.

Figure 30. Sanger Sequencing Chromatographs of SPORE Oligonucleotide After Genotypic Selection
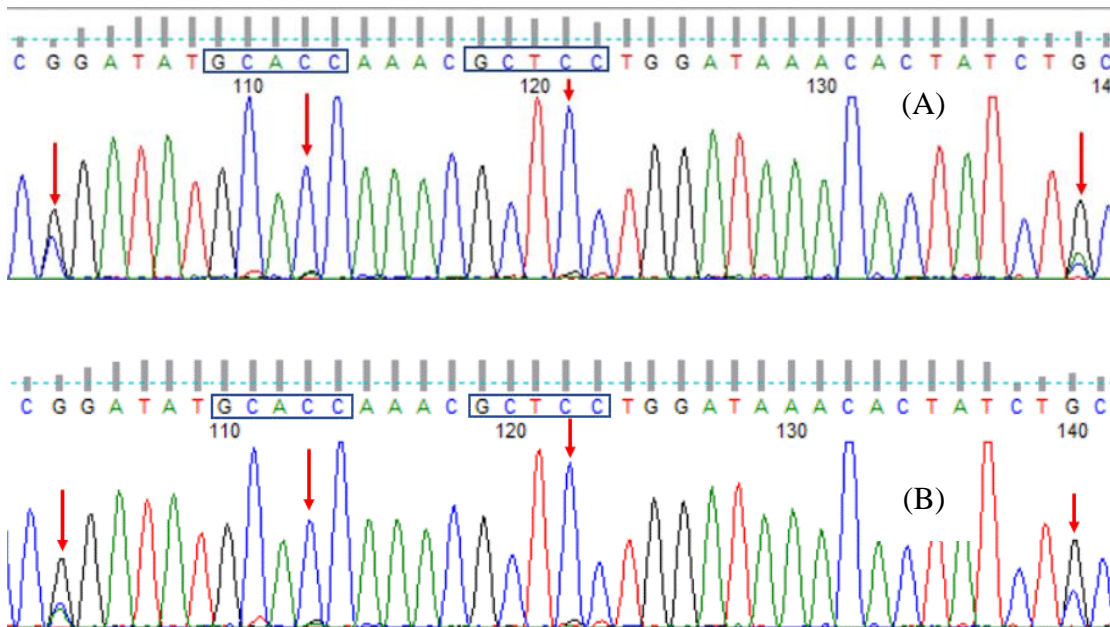
Red arrows with numbers show the mutation that we selected for. (A) When oligo 1 was inserted into the genome they have got repaired in which MMR activating mismatch (dA-dC) was on the 3'- and silent mismatch (dC-dC) was on 5'- side on the

oligonucleotide (B) When oligo 2 was inserted into the genome they got repaired by MMR. Here, MMR activating mismatch (dA-dC) was on the 5'- side and silent mismatch (dC-dC) was on the 3'-side on the oligonucleotide.

Table 3. Quantitative Values of Repaired and Mutated Mismatch on Oligo 1

Fraction of repaired (dC-dC, dA-dC) and mutated (dC-dC) mismatch on OLIGO 1			
	G – Wild Type (n=3)	C - Mutated (n=3)	A - Mutated (n=3)
5'	64.6 + 1.90	31.44 + 5.97	3.34 + 5.09
ApeKI RS (Restriction site)-1	4.06 + 2.24	90.84 + 1.22	4.61 + 0.56
ApeKI RS (Restriction site)-2	3.73 + 0.51	93.85 + 0.17	1.45 + 0.45
3'	67.23 + 2.50	14.39 + 5.67	16.13 + 5.76

The quantitative analysis of selectively enriched mutations selected for ApeKI restriction sites-1,2 positions on the oligonucleotides.
This also represents the percentage of dC-dC mismatch repair on the 5'-side and dA-dC mismatch repair on the 3' side. The data represents the mean value and standard deviation of mutation at each site on the oligonucleotide without adding ApeKI and after adding ApeKI enzyme to the sample.
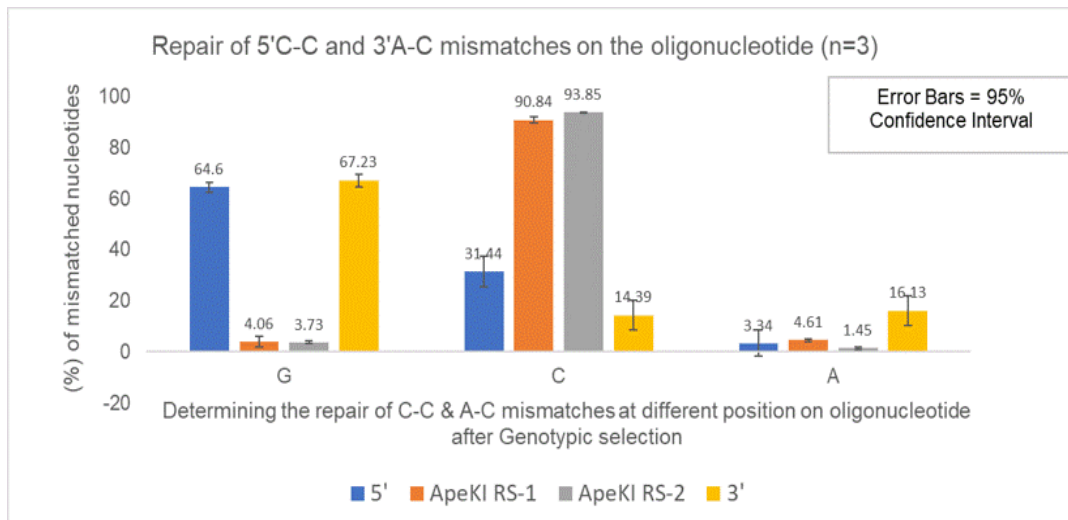
Using oligo-1, Enrichment of dG-to-dC mutations (control) at the middle position while the repair of dA-dC and dC-dC (probe) mismatch on the 3' and 5' side of the oligonucleotide respectively (mean $\pm$ 95% confidence interval, n = 3) after successful oligonucleotide recombination into the *E.coli* genome. This shows that enriched mutations at the control sites after addition of ApeKI during genotypic selection method.

The sequence present in Figure 26 is a representation of Sanger sequencing chromatograph from Figure 30 (A) only. where it indicates the repair of dC-dC mismatch on the 5' and 3' end of the oligo. Moreover, the graph in Figure 31 represents the enrichment of dG-to-dC mutations is higher at ApeKI restriction sites -1,2 because they are protected as shown in the oligonucleotide sequence below the graph.

From Table 3 we can see that maximum enrichment of dG-to-dC mutations in the middle with 92.34 $\pm$ 2.40 (95% confidence interval).

On the 3'-side the repair of dA-dC mismatch (showed as "G" on the 3' bar of the graph) is an average of 67.2% with 2.5% confidence interval. While the 16.13% presence

of "A" nucleotide at 3' bar indicate that dA-dC mismatch was occurred in the genome by our oligonucleotide but only 67.2% of the incorporated dA-dC mismatch was corrected at this site. Moreover, this data also indicate that the dC-dC mismatches (indicated as "G" at 5' bar of the graph), which presumably were repaired by Long-patch repair originating from the 5'- side in an attempt to repair the dA-dC mismatch, occurred 64.6% (1.9% confidence interval).

Conclusion: The above SPORE assay data represent that *E.coli* MMR is coordinated in a bidirectional fashion and it has, presumably unintentionally through long-patch repair, repaired silent dC-dC mismatches on the 5' side of the oligo as well as dA-dC mismatches on the 3' side of the oligo-1. Thus, we can compare it with our hypothesis in Figure 27 and with published evidence (Figure 25) [9] which shows the sequence of oligonucleotide if the MMR was attempted from both the side. Moreover, if we compare this quantitative result with published Figure 25 (D) which indicate the repair of dG-dT probe mismatch around 80% at 3' side of the oligonucleotide. While we see the repair of dA-dC mismatch at 3' side which is on average $67\% \pm 2.5\%$ of the incorporated nucleotides. This shows the difference in the repair of two different mismatches dG-dT and dA-dC.

Table 4. Quantitative Values of Repaired and Mutated Mismatch on Oligo 2

Fraction of repaired (dC-dC, dA-dC) and mutated (dC-dC) mismatch on OLIGO 2		
	G – Wild Type (n=6)	C - Mutated (n=6)	A - Mutated (n=6)
5'	71.7 + 7.99	16.8 + 4.88	10.55 + 4.35
ApeKI RS (Restriction site)-1	7.54 ± 1.60	86.96 ± 1.68	4.93 ± 0.75
ApeKI RS (Restriction site)-2	5.3 ± 1.29	92.3 ± 1.62	1.48 ± 0.62
3'	79.04 ± 6.11	19.12 ± 6.39	0.5 ± 0.45

The quantitative analysis of selectively enriched mutations selected for ApeKI restriction sites-1,2 on the oligonucleotides. This also represents the percentage of dC-dC mismatch repair on the 3'-side and dA-dC mismatch repair on the 5' side. The data represents the mean value and standard deviation of mutation at each site on the oligonucleotide without adding ApeKI and after adding ApeKI enzyme to the sample.
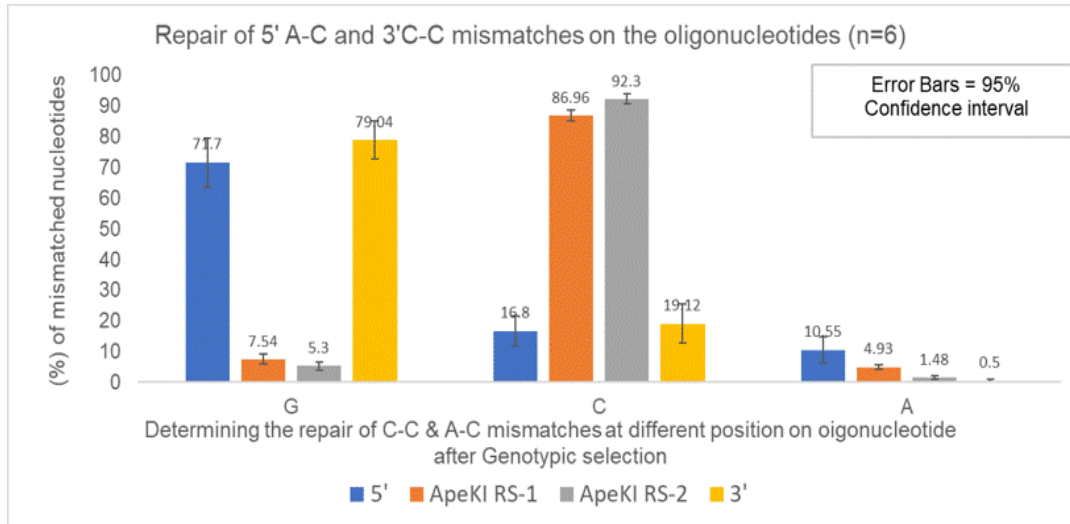
Using oligo-2, Enrichment of dG-to-dC mutations (control) at the middle position while the repair of dA-dC and dC-dC (probe) mismatch on the 5' and 3' side of the oligonucleotide respectively (mean $\pm$ 95% confidence interval, n = 6) after successful oligonucleotide recombination into the *E.coli* genome. This shows that enriched dC-dC mismatches after addition of ApeKI during genotypic selection method.

Figure 32 shows the enrichment of dC-dC mismatches at ApeKI restriction sites-1,2 positions on the oligonucleotide after addition of ApeKI and this data can be compared to the Sanger sequencing chromatographs which show GCACC and GCTCC sequence after adding ApeKI. This data indicates the MMR activity on the SPORE oligo 2 which has dA-dC probe mutation on 5' side while the dC-dC silent probe mutation on 3'side.

In Figure 28 we can see two <u>C</u> nucleotide between two boxes, that represents the protection of these nucleotides and these nucleotides will not be recognized by MMR proteins as they are protected. Also, this sequence is a representation of Sanger sequencing chromatograph from Figure 30 (B) only. where it indicates the incorporation

of dC-dC mismatch on the genome while the repair of dA-dC and dC-dC mismatch on the 5' side and 3' side respectively. These data also indicate the total repair of dC-dC mismatches (indicated as "G" at 3' bar of the graph) on an average of 79% with 6.11% confidence interval at the 3' side of the oligonucleotide. On the other 5'-side the repair of dA-dC mismatch (showed as "G" on the 5' bar of the graph) is an average of 71.7% with 7.9% confidence interval. While the 10.55% presence of "A" nucleotide at 5' bar indicate that dA-dC mismatch was occurred in the genome by our oligonucleotide but the only average of 71.7% of the incorporated dA-dC mismatch was corrected at this site.

Conclusion: The above SPORE assay data represent that *E.coli* MMR is bidirectional *in vivo* and repair can be initiated from both 3'- and 5'- each with high efficiency, as it has repaired dC-dC silent mismatches on the 3' side of the oligo as well as dA-dC mismatches on the 5' side of the oligo. Thus, we can compare it with our hypothesis in Figure 28 and with published evidence (Figure 25) which shows the sequence of oligonucleotide if the MMR was attempted from both the side.

3.2.4.1 Overall Conclusions from the SPORE Assay Experiments:

For both oligos, dA-dC probe mismatches were repaired with high efficiencies from both 5'- and 3'-, which is in agreement with earlier results from SPORE assays using dG-dT mismatches that those mismatches could be repaired with MMR coordinated from either direction *in vivo*. The repair of dA-dC mismatches was also less than dG-dT mismatches, which is expected given previous studies and the relative affinities of MutS with those mismatches.

What was unexpected was the significant repair of the 'silent probe' mismatches, which suggests repair of the dA-dC mismatches is being initiated from both 5'- and 3'-sides of the mismatch with approximately equal probabilities, in contrast to results that found that dG-dT mismatches were repaired primarily from the 3'- side. We consider possibilities for these differences between our results of the coordination of the probe mismatch:

1. One possibility is that dA-dC mismatches are repaired by a different mechanism *in vivo* than dG-dT mismatches. Further investigation is necessary.

2. Another possibility is that our results are based on the genotypic selection of a mutation while the previous assay used phenotypic selection on selective minimal media. It may be possible that the silent mismatch, although it introduced a 'silent mutation' where the sequence changed but the same amino acid was coded for, introduced a selective pressure during the phenotypic assay that interfered with the results. This is one reason we are developing a genotypic screen in the first place.

3. Another possibility is that the earlier study was performed within the *galK* gene while this experiment was performed to mutate sequences in the *araD* gene. These genes both have different locations on the *E.coli* genome and positions relative to nearby d(GATC) sites. So, a repair could have been coordinated different based on those two factors as well. Further investigation is necessary.

From Figure 30, which shows the chromatographs of SPORE assay results performed with oligo 1 and oligo 2. Upon examination, we noticed some unusual signals at the dA-dC mismatch sites. We observed a noticeable signal for a dG-to-dC mutation at

the dA-dC mismatch sites which could not be explained by long-patch repair. Generally,

dA-dC mismatch should get repaired to dG-dC base pairing but in our results, but there

was a signal indicating a dC-dG basepair. This activity was observed only with the repair

of dA-dC mismatch on both oligos at 3' side on oligo 1 and 5' side of oligo 2. We assume

that one of the possibilities of this signal after repair can be human error, or errors at the

manufacturing level so the oligos could be contaminated. This repair interferes with our

data of quantifying repair of dC-dC mismatch on the oligonucleotide and suggests there

is some level of contamination between oligo 1 and oligo 2 that could have occurred

either at the production level, purification level, or at transfection level. So, for this, we

recalculated the frequency of dC signal at dA-dC sites, which would indicate the level of

contamination, to estimate the range of possible repairs at the 'silent mutation' site, since

contamination would have increased the apparent level of repair at that site by the same

amount.

     For oligo 1 the recalculated estimate of the repair of the silent mutation is from

49.9% (4.28, 95% confidence) To the original value of 67.23% (2.50%, 95% confidence)

while for oligo 2 the estimated repair of the silent mutation ranged from 62.2% (10.81,

95% confidence interval) To the original value of 71.7% (7.99%,95% confidence). This

observation requires further investigation to conclude new details of this MMR.

Table 5. The Estimated Range of Silent Mutation Repair Efficiencies

Estimated range of silent mutation repair efficiencies (X% to Y%)
SPORE Oligo 1
45.56% to 65.69%
51.75% to 62.63%
53.29 % to 65.45%

Estimated range of silent mutation repair efficiencies (X% to Y%)
SPORE Oligo 2
47.49% to 71.63%
51.38% to 71.73%
58.83% to 77.27%
59.19% to 78.56%
73.08% to 83.33%
83.20% to 91.75%

These values are calculated from the quantitative analysis of dC-dC mismatch on both oligos 1 & 2. This is recalculated by subtracting the dC-dC mismatch repair at the site of dA-dC mismatch repair. The data represents the range of dC-dC mismatch repair from (X%-values after correction to Y% values before correction).

Above all the results indicate that MMR repair can be initially coordinated bidirectionally on the lagging strand as we can see the repair of probe mutations on 5'-side as well as 3'-side approximately equally under certain conditions. We have further noticed the difference in the repair of dA-dC mismatch when it is present on the 3'side and 5' side of the oligonucleotide. The same difference we see in the repair of dC-dC mismatch too. As we see the difference in dA-dC probe mismatch compared to the published studies showed repair of dG-dT mismatch in Figure 25 (D)[9].

We know that dG-dT and dA-dC mismatch can excite the MMR process in the cell but from the quantitative analysis, we see the difference in their repair. We assume that one of the possibility of MutS binding to the different mismatches. MutS binds to different mismatches from highest binding to weakly binding according to dG-dT, dA-dC, dA-dA, dG-dG > dT-dT, dT-dC, dA-dG >> dC-dC. In this, we have seen that mismatch repair was attempted from the opposite side of the 'probe mutation' in both the cases which is an addition to the previous study, [9] where it is shown that LPR is coordinated from the same side where 'probe mutation' was present as indicated in Figure 25. Now, with the help of this study, we can quantify the repair efficiency of dA-dC mismatch repair which was not measured previously.

With these experiments, our trial was to develop new biotechnology to study the DNA MMR in cells. We see that our new method of genotypic selection with mutation enrichment using ApeKI during PCR reaction has worked well and has proven the hypothesis shown in Figure 25(B) studies of SPORE assay.

Thus, we can conclude that we can pull out the rare genomic mutation with this new technique as well as we can study the mismatch repair process quickly without growing bacteria on the selective media.

CHAPTER IV

SUMMARY AND FUTURE DIRECTIONS

Here, we have shown the development of a genotypic selection to screen for the rare genomic DNA mutations without growing bacteria on selective media, using PCR and a thermotolerant restriction enzyme. We have seen the maximum of 705,000-fold enrichment of the selected plasmid DNA even in 10^{-6} times dilution. We have shown that this method can also be used to screen for mutations introduced by oligonucleotide recombination with high frequency using three different oligos. This method can be combined with SPORE assay to study the MMR quantitatively. The results we found provided interesting validations as well as discrepancies when we compared it with previously published work that used phenotypic screens and are worthy of further investigation.

While the experimental work for this dissertation was unfortunately cut short by the outbreak of the COVID-19 pandemic, future work would be focused on:

1.   Further optimization and validation of the sensitivity of this technique and its use for studying MMR-like processes.

2.   Using SPORE assays with genotypic screens and strains of *E. coli* with mutations to MMR proteins to understand if replication-independent and replication-dependent MMR processes are coordinated similarly or differently.

3.   Using genotypic screens to allow for time-dependent studies of MMR as it occurs in living cells. Phenotypic screens require several generations for successful screening, while the genotypic screen can be performed before a single replication event has occurred after oligonucleotide incorporation.

4.   The SPORE assay in other organisms where oligonucleotide recombination success might be poor or phenotypic screens might be difficult, like actinomyces which have important applications to human health and biotechnology.

The new biotechnology developed during this thesis paves the way for those exciting applications.

BIBLIOGRAPHY

[1]   M. S. Junop, G. Obmolova, K. Rausch, P. Hsieh, and W. Yang, "Composite Active Site of an ABC ATPase: MutS Uses ATP to Verify Mismatch Recognition and Authorize DNA Repair," *Molecular Cell*, p. 12.

[2]   J. Liu *et al.*, "MutL sliding clamps coordinate exonuclease-independent Escherichia coli mismatch repair," *Nat Commun*, vol. 10, no. 1, p. 5294, Dec. 2019, doi: 10.1038/s41467-019-13191-5.

[3]   B. Van Houten and N. M. Kad, "Single-cell mutagenic responses and cell death revealed in real time," *Proc Natl Acad Sci USA*, vol. 115, no. 28, pp. 7168–7170, Jul. 2018, doi: 10.1073/pnas.1808986115.

[4]   L. Robert, J. Ollion, and M. Elez, "Real-time visualization of mutations and their fitness effects in single bacteria," *Nat Protoc*, vol. 14, no. 11, pp. 3126–3143, Nov. 2019, doi: 10.1038/s41596-019-0215-x.

[5]   L. C. Thomason, N. Costantino, and D. L. Court, "Examining a DNA Replication Requirement for Bacteriophage λ Red- and Rac Prophage RecET-Promoted Recombination in *Escherichia coli*," *mBio*, vol. 7, no. 5, pp. e01443-16, /mbio/7/5/e01443-16.atom, Nov. 2016, doi: 10.1128/mBio.01443-16.

[6]   "The Beta Protein of Phage λ Promotes Strand Exchange | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S0022283697915722?token=7CE04A2BAD861 E649882EEAA4E284787D28A6BAB199A2FB751BE1A1979799BD8E719099DBE627 F09995FA81F2B8DF4F5 (accessed Feb. 29, 2020).

[7]   N. Costantino and D. L. Court, "Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants," *Proc Natl Acad Sci U S A*, vol. 100, no. 26, pp. 15748–15753, Dec. 2003, doi: 10.1073/pnas.2434959100.

[8]   E. A. Josephs and P. E. Marszalek, "A 'Semi-Protected Oligonucleotide Recombination' Assay for DNA Mismatch Repair in vivo Suggests Different Modes of Repair for Lagging Strand Mismatches," *Nucleic Acids Res*, vol. 45, no. 8, p. e63, May 2017, doi: 10.1093/nar/gkw1339.

[9]   E. A. Josephs and P. E. Marszalek, "Endonuclease-independent DNA Mismatch Repair Processes on the Lagging Strand," *DNA Repair (Amst)*, vol. 68, pp. 41–49, Aug. 2018, doi: 10.1016/j.dnarep.2018.06.002.

[10]   W. Gu *et al.*, "Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications," *Genome Biol*, vol. 17, no. 1, p. 41, Dec. 2016, doi: 10.1186/s13059-016-0904-5

[11]   F. R. Blattner, "The Complete Genome Sequence of Escherichia coli K-12," *Science*, vol. 277, no. 5331, pp. 1453–1462, Sep. 1997, doi: 10.1126/science.277.5331.1453.

[12]   J. S. Lewis, S. Jergic, and N. E. Dixon, "The E. coli DNA Replication Fork," in *The Enzymes*, vol. 39, Elsevier, 2016, pp. 31–88.

[13]   J. Cairns, "The bacterial chromosome and its manner of replication as seen by autoradiography," *Journal of Molecular Biology*, vol. 6, no. 3, pp. 208-IN5, Mar. 1963, doi: 10.1016/S0022-2836(63)80070-4.

[14]   R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, and A. Sugino, "Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains," *PNAS*, vol. 59, no. 2, pp. 598–605, Feb. 1968, doi: 10.1073/pnas.59.2.598.

[15]   J. M. Kaguni, "Replication initiation at the Escherichia coli chromosomal origin," *Curr Opin Chem Biol*, vol. 15, no. 5, pp. 606–613, Oct. 2011, doi: 10.1016/j.cbpa.2011.07.016.

[16]   I. J. Fijalkowska, R. M. Schaaper, and P. Jonczyk, "DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair," *FEMS Microbiol Rev*, vol. 36, no. 6, pp. 1105–1121, Nov. 2012, doi: 10.1111/j.1574-6976.2012.00338.x.

[17]   J.-G. Tiraby and M. S. Fox, "Marker Discrimination in Transformation and Mutation of Pneumococcus," *Proc Natl Acad Sci U S A*, vol. 70, no. 12 Pt 1–2, pp. 3541–3545, Dec. 1973.

[18]   R. S. Lahue, K. G. Au, and P. Modrich, "DNA Mismatch Correction in a Defined System," *Science*, vol. 245, no. 4914, pp. 160–164, 1989.

[19]	J. A. Sawitzke *et al.*, "Probing cellular processes with oligo-mediated recombination; using knowledge gained to optimize recombineering," *J Mol Biol*, vol. 407, no. 1, pp. 45–59, Mar. 2011, doi: 10.1016/j.jmb.2011.01.030.

[20]	S. S. Su and P. Modrich, "Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs.," *Proceedings of the National Academy of Sciences*, vol. 83, no. 14, pp. 5057–5061, Jul. 1986, doi: 10.1073/pnas.83.14.5057.

[21]	C. Dohet, R. Wagner, and M. Radman, "Methyl-directed repair of frameshift mutations in heteroduplex DNA.," *Proc Natl Acad Sci U S A*, vol. 83, no. 10, pp. 3395–3397, May 1986.

[22]	J. Liu *et al.*, "Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair," *Nature*, vol. 539, no. 7630, pp. 583–587, Nov. 2016, doi: 10.1038/nature20562.

[23]	V. Dao and P. Modrich, "Mismatch-, MutS-, MutL-, and Helicase II-dependent Unwinding from the Single-strand Break of an Incised Heteroduplex," *J. Biol. Chem.*, vol. 273, no. 15, pp. 9202–9207, Apr. 1998, doi: 10.1074/jbc.273.15.9202.

[24]	C. Ramilo *et al.*, "Partial Reconstitution of Human DNA Mismatch Repair In Vitro: Characterization of the Role of Human Replication Protein A," *Molecular and Cellular Biology*, vol. 22, no. 7, pp. 2037–2046, Apr. 2002, doi: 10.1128/MCB.22.7.2037-2046.2002.

[25]	P. Modrich and R. Lahue, "Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology," *Annual Review of Biochemistry*, vol. 65, no. 1, pp. 101–133, 1996, doi: 10.1146/annurev.bi.65.070196.000533.

[26]	R. Bruni, D. Martin, and J. Jiricny, "d(GATC) sequences influence Escherichia coli mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch.," *Nucleic Acids Res*, vol. 16, no. 11, pp. 4875–4890, Jun. 1988.

[27]	M. M. Hingorani, "Mismatch binding, ADP-ATP exchange and intramolecular signaling during mismatch repair," *DNA Repair (Amst)*, vol. 38, pp. 24–31, Feb. 2016, doi: 10.1016/j.dnarep.2015.11.017.

[28]	S. Acharya, P. L. Foster, P. Brooks, and R. Fishel, "The Coordinated Functions of the E. coli MutS and MutL Proteins in Mismatch Repair," *Molecular Cell*, vol. 12, no. 1, pp. 233–246, Jul. 2003, doi: 10.1016/S1097-2765(03)00219-3.

[29]   A. Joshi, S. Sen, and B. J. Rao, "ATP-hydrolysis-dependent conformational switch modulates the stability of MutS–mismatch complexes," *Nucleic Acids Res*, vol. 28, no. 4, pp. 853–861, Feb. 2000.

[30]   G. Obmolova, C. Ban, P. Hsieh, and W. Yang, "Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA," *Nature*, vol. 407, no. 6805, pp. 703–710, Oct. 2000, doi: 10.1038/35037509.

[31]   C. Ban and W. Yang, "Crystal Structure and ATPase Activity of MutL: Implications for DNA Repair and Mutagenesis," p. 12.

[32]   K. G. AuS and K. Welsh, "Initiation of Methyl-directed Mismatch Repair," p. 7.

[33]   W. Yang, "Lessons Learned From UvrD Helicase : Mechanism For Directional Movement," *Annu Rev Biophys*, vol. 39, pp. 367–385, 2010, doi: 10.1146/annurev.biophys.093008.131415.

[34]   M. Viswanathan, V. Burdett, C. Baitinger, P. Modrich, and S. T. Lovett, "Redundant Exonuclease Involvement in *Escherichia coli* Methyl-directed Mismatch Repair," *J. Biol. Chem.*, vol. 276, no. 33, pp. 31053–31058, Aug. 2001, doi: 10.1074/jbc.M105481200.

[35]   M. W. Olson, H. G. Dallmann, and C. S. McHenry, "THE □½□ COMPLEX FUNCTIONS BY INCREASING THE AFFINITY OF □ AND □ FOR □½□J TO A PHYSIOLOGICALLY RELEVANT RANGE," p. 9.

[36]   F. J. L. de Saro, M. G. Marinus, P. Modrich, and M. O'Donnell, "The β Sliding Clamp Binds to Multiple Sites within MutL and MutS," *J. Biol. Chem.*, vol. 281, no. 20, pp. 14340–14349, May 2006, doi: 10.1074/jbc.M601264200.

[37]   M. S. Junop, G. Obmolova, K. Rausch, P. Hsieh, and W. Yang, "Composite Active Site of an ABC ATPase: MutS Uses ATP to Verify Mismatch Recognition and Authorize DNA Repair," *Molecular Cell*, p. 12.

[38]   J. Jiang, L. Bai, J. A. Surtees, Z. Gemici, M. D. Wang, and E. Alani, "Detection of High-Affinity and Sliding Clamp Modes for MSH2-MSH6 by Single-Molecule Unzipping Force Analysis," *Molecular Cell*, vol. 20, no. 5, pp. 771–781, Dec. 2005, doi: 10.1016/j.molcel.2005.10.014.

[39]   M. L. Mendillo, D. J. Mazur, and R. D. Kolodner, "Analysis of the Interaction between the *Saccharomyces cerevisiae* MSH2-MSH6 and MLH1-PMS1 Complexes with DNA Using a Reversible DNA End-blocking System," *J. Biol. Chem.*, vol. 280, no. 23, pp. 22245–22257, Jun. 2005, doi: 10.1074/jbc.M407545200.

[40]   A. M. M. Hasan and D. R. F. Leach, "Chromosomal directionality of DNA mismatch repair in Escherichia coli," *Proc Natl Acad Sci U S A*, vol. 112, no. 30, pp. 9388–9393, Jul. 2015, doi: 10.1073/pnas.1505370112.

[41]   A. A. Freitas and J. P. de Magalhães, "A review and appraisal of the DNA damage theory of ageing," *Mutation Research/Reviews in Mutation Research*, vol. 728, no. 1–2, pp. 12–22, Jul. 2011, doi: 10.1016/j.mrrev.2011.05.001.

[42]   A. Ciccia and S. J. Elledge, "The DNA Damage Response: Making It Safe to Play with Knives," *Molecular Cell*, vol. 40, no. 2, pp. 179–204, Oct. 2010, doi: 10.1016/j.molcel.2010.09.019.

[43]   A. E. Pegg, "Multifaceted Roles of Alkyltransferase and Related Proteins In DNA Repair, DNA Damage, Resistance to Chemotherapy and Research Tools," *Chem Res Toxicol*, vol. 24, no. 5, pp. 618–639, May 2011, doi: 10.1021/tx200031q.

[44]   E. C. Friedberg, "DNA damage and repair," *Nature*, vol. 421, no. 6921, pp. 436–440, Jan. 2003, doi: 10.1038/nature01408.

[45]   Y. Liu *et al.*, "Coordination of Steps in Single-nucleotide Base Excision Repair Mediated by Apurinic/Apyrimidinic Endonuclease 1 and DNA Polymerase β," *J. Biol. Chem.*, vol. 282, no. 18, pp. 13532–13541, May 2007, doi: 10.1074/jbc.M611295200.

[46]   T. Lindahl, "An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues," *Proceedings of the National Academy of Sciences*, vol. 71, no. 9, pp. 3649–3653, Sep. 1974, doi: 10.1073/pnas.71.9.3649.

[47]   G. Dianov, A. Price, and T. Lindahl, "Generation of single-nucleotide repair patches following excision of uracil residues from DNA.," *Mol. Cell. Biol.*, vol. 12, no. 4, pp. 1605–1612, Apr. 1992, doi: 10.1128/MCB.12.4.1605.

[48]   G. Frosina *et al.*, "Two Pathways for Base Excision Repair in Mammalian Cells," *J. Biol. Chem.*, vol. 271, no. 16, pp. 9573–9578, Apr. 1996, doi: 10.1074/jbc.271.16.9573.

[49]   R. B. Setlow and W. L. Carrier, "THE DISAPPEARANCE OF THYMINE DIMERS FROM DNA: AN ERROR-CORRECTING MECHANISM," *Proc Natl Acad Sci U S A*, vol. 51, no. 2, pp. 226–231, Feb. 1964.

[50]   A. Sancar, K. A. Franklin, and G. B. Sancar, "Escherichia coli DNA photolyase stimulates uvrABC excision nuclease in vitro.," *Proc Natl Acad Sci U S A*, vol. 81, no. 23, pp. 7397–7401, Dec. 1984.

[51]   A. Sancar, B. M. Kacinski, D. L. Mott, and W. D. Rupp, "Identification of the uvrC gene product.," *Proc Natl Acad Sci U S A*, vol. 78, no. 9, pp. 5450–5454, Sep. 1981.

[52]   A. Sancar and W. D. Rupp, "A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region," *Cell*, vol. 33, no. 1, pp. 249–260, May 1983, doi: 10.1016/0092-8674(83)90354-9.

[53]   I. G. Minko, Y. Zou, and R. S. Lloyd, "Incision of DNA-protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair," *Proceedings of the National Academy of Sciences*, vol. 99, no. 4, pp. 1905–1909, Feb. 2002, doi: 10.1073/pnas.042700399.

[54]   P. R. Caron, S. R. Kushner, and L. Grossman, "Involvement of helicase II (uvrD gene product) and DNA polymerase I in excision mediated by the uvrABC protein complex.," *Proc Natl Acad Sci U S A*, vol. 82, no. 15, pp. 4925–4929, Aug. 1985.

[55]   P. Pfeiffer, "Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations," *Mutagenesis*, vol. 15, no. 4, pp. 289–302, Jul. 2000, doi: 10.1093/mutage/15.4.289.

[56]   P. Sung and H. Klein, "Mechanism of homologous recombination: mediators and helicases take on regulatory functions," *Nat Rev Mol Cell Biol*, vol. 7, no. 10, pp. 739–750, Oct. 2006, doi: 10.1038/nrm2008.

[57]   A. S. Ponticelli, D. W. Schultz, A. F. Taylor, and G. R. Smith, "Chi-dependent DNA strand cleavage by RecBC enzyme," *Cell*, vol. 41, no. 1, pp. 145–151, May 1985, doi: 10.1016/0092-8674(85)90069-8.

[58]   D. G. Anderson and S. C. Kowalczykowski, "The Translocating RecBCD Enzyme Stimulates Recombination by Directing RecA Protein onto ssDNA in a □-Regulated Manner," p. 10.

[59] M. M. Cox and I. R. Lehman, "recA protein of Escherichia coli promotes branch migration, a kinetically distinct phase of DNA strand exchange.," *Proc Natl Acad Sci U S A*, vol. 78, no. 6, pp. 3433–3437, Jun. 1981.

[60] C. M. George and E. Alani, "Multiple cellular mechanisms prevent chromosomal rearrangements involving repetitive DNA," *Crit Rev Biochem Mol Biol*, vol. 47, no. 3, pp. 297–313, May 2012, doi: 10.3109/10409238.2012.675644.

[61] R. C. MacLean and A. San Millan, "The evolution of antibiotic resistance," *Science*, vol. 365, no. 6458, pp. 1082–1083, Sep. 2019, doi: 10.1126/science.aax3879.
[62] M. G. MARINUS, "DNA Mismatch Repair," *EcoSal Plus*, vol. 5, no. 1, Nov. 2012, doi: 10.1128/ecosalplus.7.2.5.

[63] J. L. Martinez and F. Baquero, "Mutation Frequencies and Antibiotic Resistance," *Antimicrob. Agents Chemother.*, vol. 44, no. 7, pp. 1771–1777, Jul. 2000, doi: 10.1128/AAC.44.7.1771-1777.2000.

[64] A. L. Lu, S. Clark, and P. Modrich, "Methyl-directed repair of DNA base-pair mismatches in vitro.," *Proc Natl Acad Sci U S A*, vol. 80, no. 15, pp. 4639–4643, Aug. 1983.

[65] D. L. Cooper, "Methyl-directed Mismatch RepairIs Bidirectional," p. 7.

[66] R. S. Lahue, S. S. Su, and P. Modrich, "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction.," *Proc Natl Acad Sci U S A*, vol. 84, no. 6, pp. 1482–1486, Mar. 1987.

[67] G. Natrajan, M. H. Lamers, J. H. Enzlin, H. H. K. Winterwerp, A. Perrakis, and T. K. Sixma, "Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates," *Nucleic Acids Res*, vol. 31, no. 16, pp. 4814–4821, Aug. 2003.

[68] M. C. Monti *et al.*, "Native mass spectrometry provides direct evidence for DNA mismatch-induced regulation of asymmetric nucleotide binding in mismatch repair protein MutS," *Nucleic Acids Res*, vol. 39, no. 18, pp. 8052–8064, Oct. 2011, doi: 10.1093/nar/gkr498.

[69] C. Jeong *et al.*, "MutS switches between two fundamentally distinct clamps during mismatch repair," *Nat Struct Mol Biol*, vol. 18, no. 3, pp. 379–385, Mar. 2011, doi: 10.1038/nsmb.2009.

[70]   "ATP Alters the Diffusion Mechanics of MutS on Mismatched DNA | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S0969212612001761?token=10D70BA870D523 812610BB034D10D96C2C6796332095D6BC42C8BDA6B1FE2C30DE125AAC6CCA2 F296FB2AF4C633C871C (accessed Mar. 14, 2020).

[71]   J. Hanne, J. Liu, J.-B. Lee, and R. Fishel, "Single-molecule FRET Studies on DNA Mismatch Repair," *International Journal of Biophysics*, vol. 3, no. 1A, pp. 18–38, 2013.

[72]   J. Gorman *et al.*, "Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6," *Mol Cell*, vol. 28, no. 3, pp. 359–370, Nov. 2007, doi: 10.1016/j.molcel.2007.09.008.

[73]   I. Bonnet *et al.*, "Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA," *Nucleic Acids Res*, vol. 36, no. 12, pp. 4118–4127, Jul. 2008, doi: 10.1093/nar/gkn376.

[74]   H. Wang *et al.*, "DNA bending and unbending by MutS govern mismatch recognition and specificity," *Proc Natl Acad Sci U S A*, vol. 100, no. 25, pp. 14822–14827, Dec. 2003, doi: 10.1073/pnas.2433654100.

[75]   Y. Yang, L. E. Sass, C. Du, P. Hsieh, and D. A. Erie, "Determination of protein–DNA binding constants and specificities from statistical analyses of single molecules: MutS–DNA interactions," *Nucleic Acids Res*, vol. 33, no. 13, pp. 4322–4334, 2005, doi: 10.1093/nar/gki708.

[76]   L. E. Sass, C. Lanyi, K. Weninger, and D. A. Erie, "Single-molecule FRET TACKLE reveals highly dynamic mismatched DNA-MutS complexes," *Biochemistry*, vol. 49, no. 14, pp. 3174–3190, Apr. 2010, doi: 10.1021/bi901871u.

[77]   J. W. Gauer *et al.*, "Single-Molecule FRET to Measure Conformational Dynamics of DNA Mismatch Repair Proteins," *Methods Enzymol*, vol. 581, pp. 285–315, 2016, doi: 10.1016/bs.mie.2016.08.012.

[78]   "doi:10.1016/S1097-2765(03)00219-3 | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S1097276503002193?token=908CFF1A3BF20E BD7C174FE542B913705904386BA87F14901350E46DB87AF28ECED9C71060B51B5 30BC7F750EEC9DFD2 (accessed Mar. 14, 2020).

[79]   J. Liu, J.-B. Lee, and R. Fishel, "Stochastic Processes and Component Placticity Governing DNA Mismatch Repair," *J Mol Biol*, vol. 430, no. 22, pp. 4456–4468, Oct. 2018, doi: 10.1016/j.jmb.2018.05.039.

[80]   R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, and A. Sugino, "Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains.," *Proc Natl Acad Sci U S A*, vol. 59, no. 2, pp. 598–605, Feb. 1968.

[81]   H. Long *et al.*, "Antibiotic treatment enhances the genome-wide mutation rate of target cells," *Proc Natl Acad Sci USA*, vol. 113, no. 18, pp. E2498–E2505, May 2016, doi: 10.1073/pnas.1601208113.

[82]   S. Uphoff, "Real-time dynamics of mutagenesis reveal the chronology of DNA repair and damage tolerance responses in single cells," *Proc Natl Acad Sci USA*, vol. 115, no. 28, pp. E6516–E6525, Jul. 2018, doi: 10.1073/pnas.1801101115.

[83]   L. Robert, J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez, "Mutation dynamics and fitness effects followed in single cells," *Science*, vol. 359, no. 6381, pp. 1283–1286, Mar. 2018, doi: 10.1126/science.aan0797.

[84]   Y. Li, J. W. Schroeder, L. A. Simmons, and J. S. Biteen, "Visualizing bacterial DNA replication and repair with molecular resolution," *Current Opinion in Microbiology*, vol. 43, pp. 38–45, Jun. 2018, doi: 10.1016/j.mib.2017.11.009.

[85]   H. L. Klein *et al.*, "Guidelines for DNA recombination and repair studies: Mechanistic assays of DNA repair processes," *Microb Cell*, vol. 6, no. 1, pp. 65–101, Jan. 2019, doi: 10.15698/mic2019.01.665.

[86]   K. E. Duderstadt *et al.*, "Simultaneous Real-Time Imaging of Leading and Lagging Strand Synthesis Reveals the Coordination Dynamics of Single Replisomes," *Molecular Cell*, vol. 64, no. 6, pp. 1035–1047, Dec. 2016, doi: 10.1016/j.molcel.2016.10.028.

[87]   Y. Liao, J. W. Schroeder, B. Gao, L. A. Simmons, and J. S. Biteen, "Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair," *Proc Natl Acad Sci USA*, vol. 112, no. 50, pp. E6898–E6906, Dec. 2015, doi: 10.1073/pnas.1507386112.

[88]   X. Lei, Y. Zhu, A. Tomkinson, and L. Sun, "Measurement of DNA mismatch repair activity in live cells," *Nucleic Acids Res*, vol. 32, no. 12, p. e100, 2004, doi: 10.1093/nar/gnh098.

[89]   M. L. Mendillo, C. D. Putnam, and R. D. Kolodner, "*Escherichia coli* MutS Tetramerization Domain Structure Reveals That Stable Dimers but Not Tetramers Are Essential for DNA Mismatch Repair *in Vivo*," *J. Biol. Chem.*, vol. 282, no. 22, pp. 16345–16354, Jun. 2007, doi: 10.1074/jbc.M700858200.

[90]   K. C. Murphy, "Use of Bacteriophage λ Recombination Functions To Promote Gene Replacement in Escherichia coli," *J Bacteriol*, vol. 180, no. 8, pp. 2063–2071, Apr. 1998.

[91]   H. M. Ellis, D. Yu, T. DiTizio, and D. L. Court, "High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides," *Proc Natl Acad Sci U S A*, vol. 98, no. 12, pp. 6742–6746, Jun. 2001, doi: 10.1073/pnas.121164898.

[92]   A. Kuzminov, "Recombinational Repair of DNA Damage in Escherichia coli and Bacteriophage □," *MICROBIOL. MOL. BIOL. REV.*, vol. 63, p. 63, 1999.

[93]   K. C. Murphy and M. G. Marinus, "RecA-independent single-stranded DNA oligonucleotide-mediated mutagenesis," *F1000 Biol Rep*, Jul. 2010, doi: 10.3410/B2-56.

[94]   "12636.full.pdf." Accessed: Mar. 14, 2020. [Online]. Available: https://www.jbc.org/content/256/24/12636.full.pdf.

[95]   G. Karakousis, N. Ye, Z. Li, S. K. Chiu, G. Reddy, and C. M. Radding, "The beta protein of phage λ binds preferentially to an intermediate in DNA renaturation," *Journal of Molecular Biology*, vol. 276, no. 4, pp. 721–731, Mar. 1998, doi: 10.1006/jmbi.1997.1573.

[96]   "7377.full.pdf." Accessed: Mar. 14, 2020. [Online]. Available: https://www.jbc.org/content/250/18/7377.full.pdf.

[97]   K. C. Murphy, "A Gam Protein Inhibits the Helicase and x-Stimulated Recombination Activities of Escherichia coli RecBCD Enzyme," *J. BACTERIOL.*, p. 14.

[98]   "679.full.pdf." Accessed: Mar. 14, 2020. [Online]. Available: https://www.jbc.org/content/242/4/679.full.pdf.

[99]   A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, and C. Cullin, "A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae.," *Nucleic Acids Res*, vol. 21, no. 14, pp. 3329–3330, Jul. 1993.

[100]   Y. Zhang, J. P. Muyrers, J. Rientjes, and F. Stewart, "Phage annealing proteins promote oligonucleotide-directed mutagenesis in Escherichia coli and mouse ES cells," *BMC Molecular Biology*, p. 14, 2003.

[101]   "doi:10.1016/S1568-7864(02)00245-8 | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S1568786402002458?token=3A43003E89FB7E 7437672981873A53103F61C26BB4C56D4653C982968022691248B71E1F3D26529E9 898ABF99CB7028B (accessed Mar. 15, 2020).

[102]   X. Li, L. C. Thomason, J. A. Sawitzke, N. Costantino, and D. L. Court, "Bacterial DNA polymerases participate in oligonucleotide recombination," *Molecular Microbiology*, vol. 88, no. 5, pp. 906–920, 2013, doi: 10.1111/mmi.12231.

[103]   M. S. Y. Huen, X. Li, L.-Y. Lu, R. M. Watt, D.-P. Liu, and J.-D. Huang, "The involvement of replication in single stranded oligonucleotide-mediated gene repair," *Nucleic Acids Research*, vol. 34, no. 21, pp. 6183–6194, Dec. 2006, doi: 10.1093/nar/gkl852.

[104]   X. Li *et al.*, "Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli," *Nucleic Acids Res*, vol. 31, no. 22, pp. 6674–6687, Nov. 2003, doi: 10.1093/nar/gkg844.

[105]   "main.pdf." Accessed: Mar. 15, 2020. [Online]. Available: https://pdf.sciencedirectassets.com/272198/1-s2.0-S1097276500X00640/1-s2.0-S1097276503001102/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEJL%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVz LWVhc3QtMSJHMEUCIQC9Jx7luMSMMNiGmML8%2FoxGKhpy2gOTnmye9PKzzb I3UQIgFiEepYmRFgxbXppEb9JZVBz3AP5qtnoM3AfldGLnKU4qtAMIehACGgwwNT kwMDM1NDY4NjUiDGBQdDMAxWfUMH%2BojiqRA79nE%2BAy%2FM%2FJJhm Tu1w6sXAeT1l0XxoIuL1h4gIss%2BwsSBlZROkf7Agn19ZYQ47d9zMigc2OwuGRKdl mIoNxlnwkKnrPD5EhEMGDUwKmOwpUILUMgfzVIroGkgcUSI%2F8dIWXLFnXSc eUcvrUwwxtiPFCJPjSCx9XgaL1mso2FW9VnwwKDVRYJjzGxm9W5rXQ4QUGCVsc XoE%2BF3DBI%2BfB5Iqa7xvC8a5KolB%2BWYyUPsQnNokNdNKMdzvpV%2Bl9D qu4Tfhn0z1TGs%2Fa434fUiVrH8GTR9mS5YV3RS8INfh9Vxsku5p22PFsbvN634gvY1 GwPcB1sFexxxsbvn5IYQ0RrXLTN69SniJ2a%2FVU99%2BsHoAbwsCfY30CUVr7CX gsxdizmS4%2BTFkXhplYS1ivRv2xn%2BKla4us2piM8TVP3lwL2HbhT88nmzaJHGkd u3043w1bvGOMTpaLJykiKlEpVaRAqh0gax%2Bm30UrQRr5dqznhRXCBDKVIWc47 %2Blut%2BNLVB00h0lx%2BceXOVpcuqG14fIlHZn9MNfBufMFOusBhuFmGv8%2B i9ywcu%2Fwam9vF1ZP4oFua7zvO%2FTwxRiXP7SkF0VjcYJkvq%2BmNAqlNkfE6X S3jKoEjeW2GkQhJaOHqzBnPD%2FASWmtXxo0g2BVUwlzA3jy9mVlBwNmddk4C4 Gjn1zUqQRFrfp%2Fs6gWnb5GZPskS6GvKxTvVcnnybC3QsYAyyNkQEaocP5K2%2B mEDW6NVHYmc2gt3yB7QApV4z%2F2dxvQgtyzzrfYWVkndUnkIpCldeqSM0fe1SbG

qydNs8iIbxZA4HSQ0TH44lOvnZxbCTVl2Y0FHsuUYUU6VUK1qitPf1f0hE9nU9v4R
Q%3D%3D&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-
Date=20200315T172301Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-
Amz-Credential=ASIAQ3PHCVTYUEOKVMUV%2F20200315%2Fus-east-
1%2Fs3%2Faws4_request&X-Amz-
Signature=f1a7bf456fff46afec41f8a1fa50dab84cb1ef118a462d8f0ec54214d0b743c1&ha
sh=287a86579188a2d9570b8453f7b2b8bdc3865970b6120f9461b97fee527ddfd6&host=
68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S1097
276503001102&tid=spdf-46dc2630-1e8f-461a-97a1-
aee9de7b1a68&sid=fc68033a19994342561a4fa9271cf35477c7gxrqa&type=client.

[106]   X. Li and K. J. Marians, "Two Distinct Triggers for Cycling of the Lagging
Strand Polymerase at the Replication Fork," *J. Biol. Chem.*, vol. 275, no. 44, pp. 34757–
34765, Nov. 2000, doi: 10.1074/jbc.M006556200.

[107]   "6829.full.pdf." Accessed: Feb. 26, 2020. [Online]. Available:
https://www.jbc.org/content/263/14/6829.full.pdf.

[108]   A. Joshi and B. J. Rao, "MutS recognition: Multiple mismatches and sequence
context effects," *J. Biosci.*, vol. 26, no. 5, pp. 595–606, Dec. 2001, doi:
10.1007/BF02704758.

[109]   Y. Yang, "A simple two-step, 'hit and fix' method to generate subtle mutations in
BACs using short denatured PCR fragments," *Nucleic Acids Research*, vol. 31, no. 15,
pp. 80e–880, Aug. 2003, doi: 10.1093/nar/gng080.

[110]   S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, and D. L. Court,
"Recombineering: a homologous recombination-based method of genetic engineering,"
Nat Protoc, vol. 4, no. 2, pp. 206–223, Feb. 2009, doi: 10.1038/nprot.2008.227.

[111]   H. H. Wang *et al.*, "Programming cells by multiplex genome engineering and
accelerated evolution," *Nature*, vol. 460, no. 7257, pp. 894–898, Aug. 2009, doi:
10.1038/nature08187.

[112]   M. T. Bonde, M. S. Klausen, M. V. Anderson, A. I. N. Wallin, H. H. Wang, and
M. O. A. Sommer, "MODEST: a web-based design tool for oligonucleotide-mediated
genome engineering and recombineering," *Nucleic Acids Research*, vol. 42, no. W1, pp.
W408–W415, Jul. 2014, doi: 10.1093/nar/gku428.

[113]   J. R. Warner, P. J. Reeder, A. Karimpour-Fard, L. B. A. Woodruff, and R. T. Gill,
"Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides,"
Nat Biotechnol, vol. 28, no. 8, pp. 856–862, Aug. 2010, doi: 10.1038/nbt.1653.

[114]   F. Castro-Peralta and L. P. Villarreal, "The use of oligonucleotide directed cleavage of DNA and homologous recombination in the production of large recombinant adenoviral vectors," *Gene Ther*, vol. 7, no. 7, pp. 583–586, Apr. 2000, doi: 10.1038/sj.gt.3301136.

[115]   B. Swingle, E. Markel, and S. Cartinhour, "Oligonucleotide recombination: A hidden treasure," *Bioengineered Bugs*, vol. 1, no. 4, pp. 265–268, Jul. 2010, doi: 10.4161/bbug.1.4.12098.

[116]   "Oligonucleotide recombination in corynebacteria without the expression of exogenous recombinases | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S0167701214002140?token=7516930CF696C31 90BF1C83A3075CFDDB06E524CB7D272DE524F11AD7653F4045700A9F8A860762 E9BFC130D17945186 (accessed Mar. 15, 2020).

[117]   T. Su *et al.*, "Improved ssDNA recombineering for rapid and efficient pathway engineering in Corynebacterium glutamicum," *Journal of Chemical Technology & Biotechnology*, vol. 93, no. 12, pp. 3535–3542, 2018, doi: 10.1002/jctb.5726.

[118]   S. Binder, S. Siedler, J. Marienhagen, M. Bott, and L. Eggeling, "Recombineering in Corynebacterium glutamicum combined with optical nanosensors: a general strategy for fast producer strain generation," *Nucleic Acids Res*, vol. 41, no. 12, pp. 6360–6369, Jul. 2013, doi: 10.1093/nar/gkt312.

[119]   M. Dekker, "Targeted gene modification in mismatch-repair-deficient embryonic stem cells by single-stranded DNA oligonucleotides," *Nucleic Acids Research*, vol. 31, no. 6, pp. 27e–227, Mar. 2003, doi: 10.1093/nar/gng027.

[120]   K. C. Murphy, K. Papavinasasundaram, and C. M. Sassetti, "Mycobacterial Recombineering," in *Mycobacteria Protocols*, T. Parish and D. M. Roberts, Eds. New York, NY: Springer New York, 2015, pp. 177–199.

[121]   K. C. Murphy, S. J. Nelson, S. Nambi, K. Papavinasasundaram, C. E. Baer, and C. M. Sassetti, "ORBIT: a New Paradigm for Genetic Engineering of Mycobacterial Chromosomes," vol. 9, no. 6, p. 20, 2018.

[122]   M. E. Pyne, M. Moo-Young, D. A. Chung, and C. P. Chou, "Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli," *Appl. Environ. Microbiol.*, vol. 81, no. 15, pp. 5103–5114, Aug. 2015, doi: 10.1128/AEM.01248-15.

[123]   K. C. Murphy, "λ Recombination and Recombineering," *EcoSal Plus*, vol. 7, no. 1, Jan. 2016, doi: 10.1128/ecosalplus.ESP-0011-2015.

[124]   E. A. Josephs, T. Zheng, and P. E. Marszalek, "Atomic force microscopy captures the initiation of methyl-directed DNA mismatch repair," *DNA Repair*, vol. 35, pp. 71–84, Nov. 2015, doi: 10.1016/j.dnarep.2015.08.006.

[125]   M. H. Lamers, A. Perrakis, J. H. Enzlin, H. H. K. Winterwerp, N. de Wind, and T. K. Sixma, "The crystal structure of DNA mismatch repair protein MutS binding to a G·T mismatch," *Nature*, vol. 407, no. 6805, pp. 711–717, Oct. 2000, doi: 10.1038/35037523.

[126]   L. Manelyte, C. Urbanke, L. Giron-Monzon, and P. Friedhoff, "Structural and functional analysis of the MutS C-terminal tetramerization domain," *Nucleic Acids Res*, vol. 34, no. 18, pp. 5270–5279, Oct. 2006, doi: 10.1093/nar/gkl489.

[127]   M. Grilleyt and J. Griffithg, "Bidirectional Excision in Methyl-directed Mismatch Repair," p. 8.

[128]   "14820.full.pdf." Accessed: Mar. 17, 2020. [Online]. Available: https://www.jbc.org/content/268/20/14820.full.pdf.

[129]   L. Garibyan, "Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome," *DNA Repair*, vol. 2, no. 5, pp. 593–608, May 2003, doi: 10.1016/S1568-7864(03)00024-7.

[130]   M. G. Reynolds, "Compensatory evolution in rifampin-resistant Escherichia coli.," *Genetics*, vol. 156, no. 4, pp. 1471–1481, Dec. 2000.

[131]   H. Lee, E. Popodi, H. Tang, and P. L. Foster, "Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing," *Proceedings of the National Academy of Sciences*, vol. 109, no. 41, pp. E2774–E2783, Oct. 2012, doi: 10.1073/pnas.1210309109.

[132]   M. W. Schmitt, S. R. Kennedy, J. J. Salk, E. J. Fox, J. B. Hiatt, and L. A. Loeb, "Detection of ultra-rare mutations by next-generation sequencing," *Proceedings of the National Academy of Sciences*, vol. 109, no. 36, pp. 14508–14513, Sep. 2012, doi: 10.1073/pnas.1208715109.

[133]   L. Ferrara and E. B. Kmiec, "Targeted gene repair activates Chk1 and Chk2 and stalls replication in corrected cells," *DNA Repair*, vol. 5, no. 4, pp. 422–431, Apr. 2006, doi: 10.1016/j.dnarep.2005.11.009.

[134]   M. Aarts and H. te Riele, "Parameters of oligonucleotide-mediated gene modification in mouse ES cells," *J Cell Mol Med*, vol. 14, no. 6b, pp. 1657–1667, Jun. 2010, doi: 10.1111/j.1582-4934.2009.00847.x.

[135]   L. Ferrara, H. Parekh-Olmedo, and E. B. Kmiec, "Enhanced oligonucleotide-directed gene targeting in mammalian cells following treatment with DNA damaging agents," *Experimental Cell Research*, vol. 300, no. 1, pp. 170–179, Oct. 2004, doi: 10.1016/j.yexcr.2004.06.021.

[136]   J. E. DiCarlo, A. J. Conley, M. Penttilä, J. Jäntti, H. H. Wang, and G. M. Church, "Yeast Oligo-Mediated Genome Engineering (YOGE)," *ACS Synth. Biol.*, vol. 2, no. 12, pp. 741–749, Dec. 2013, doi: 10.1021/sb400117c.

[137]   S. K. Singh, A. A. Koshkin, J. Wengel, and P. Nielsen, "LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition," *Chem. Commun.*, no. 4, pp. 455–456, 1998, doi: 10.1039/a708608c.

[138]   T. W. van Ravesteyn *et al.*, "LNA modification of single-stranded DNA oligonucleotides allows subtle gene modification in mismatch-repair-proficient cells," *Proc Natl Acad Sci USA*, vol. 113, no. 15, pp. 4122–4127, Apr. 2016, doi: 10.1073/pnas.1513315113.

[139]   A. A. Koshkin *et al.*, "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition," *Tetrahedron*, vol. 54, no. 14, pp. 3607–3630, Apr. 1998, doi: 10.1016/S0040-4020(98)00094-5.

[140]   S. Obika *et al.*, "Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2′-O,4′-C-methyleneribonucleosides," *Tetrahedron Letters*, vol. 39, no. 30, pp. 5401–5404, Jul. 1998, doi: 10.1016/S0040-4039(98)01084-3.

[141]   P. L. Foster, H. Lee, E. Popodi, J. P. Townes, and H. Tang, "Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing," *Proc Natl Acad Sci USA*, vol. 112, no. 44, pp. E5990–E5999, Nov. 2015, doi: 10.1073/pnas.1512136112.

[142]   J.-M. Reyrat and D. Kahn, "Mycobacterium smegmatis: an absurd model for tuberculosis?," *Trends in Microbiology*, vol. 9, no. 10, pp. 472–473, Oct. 2001, doi: 10.1016/S0966-842X(01)02168-0.

[143] S. H. Gillespie, "Evolution of Drug Resistance in Mycobacterium tuberculosis: Clinical and Molecular Perspective," *AAC*, vol. 46, no. 2, pp. 267–274, Feb. 2002, doi: 10.1128/AAC.46.2.267-274.2002.

[144] A. Castañeda-García *et al.*, "Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies," *Sci. Adv.*, vol. 6, no. 7, p. eaay4453, Feb. 2020, doi: 10.1126/sciadv.aay4453.

[145] P. D. Hsu, E. S. Lander, and F. Zhang, "Development and Applications of CRISPR-Cas9 for Genome Engineering," *Cell*, vol. 157, no. 6, pp. 1262–1278, Jun. 2014, doi: 10.1016/j.cell.2014.05.010.

[146] J. A. Doudna and E. Charpentier, "The new frontier of genome engineering with CRISPR-Cas9," *Science*, vol. 346, no. 6213, p. 1258096, Nov. 2014, doi: 10.1126/science.1258096.

[147] Y. Sako *et al.*, "Aeropyrum pernix gen. nov., sp. nov., a Novel Aerobic Hyperthermophilic Archaeon Growing at Temperatures up to 100 C," *International Journal of Systematic Bacteriology*, vol. 46, no. 4, pp. 1070–1077, Oct. 1996, doi: 10.1099/00207713-46-4-1070.

[148] J. Huan, K. Wan, Y. Liu, W. Dong, and G. Wang, "Removing PCR for the elimination of undesired DNA fragments cycle by cycle," *Sci Rep*, vol. 3, no. 1, p. 2303, Dec. 2013, doi: 10.1038/srep02303.

	Without adding ApekI-The mean (%) of mutated sequence on 5'-side of the oligonucleotide		After adding ApekI- The mean (%) of mutated sequence on 5'-side of the oligonucleotide	
	galK (n=2)	*araD* (n=2)	*galK* (n=2)	*araD* (n=2)
G	**95.02 + 2.54**	**96.08+ 0.34**	**82.02+ 3.10**	**23.15+ 9.76**
A	3.54 + 1.18	2.64 + 0.28	2.58 + 0.40	3.72 + 1.49
T	1.05 + 1.16	0.72 + 0.80	1.83 + 0.62	5.54 + 7.83
C	0.37 + 0.19	0.52 + 0.73	13.56 + 4.13	67.5 + 0.43

```
5'-CCCTTGAAGCTCCCAGCGCGCTCGAGCAA-3'       galK gene oligo
3'-GGGAACTTCGAGGGTCGCGCGAGCTCGTT-5'
```

```
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCTAGGACCTATTTGTGATAGAGGCATTCGT-5'
```

araD gene oligo

Description: This table indicate the difference in nucleotides after and before adding ApeKI. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide.

	Without adding ApekI-The mean (%) of mutated sequence on ApeKI restriction site-1 of the oligonucleotide		After adding ApekI- The mean (%) of mutated sequence on ApeKI restriction site-1 of the oligonucleotide	
Gene	*galK* (n=2)	*araD* (n=2)	*galK* (n=2)	*araD* (n=2)
G	98.70 $\pm$ 1.14	96.21 $\pm$ 1.92	69.82 $\pm$ 3.75	8.70 $\pm$ 7.65
A	0.46 $\pm$ 0.40	1.58 $\pm$ 0.74	0.7 $\pm$ 0.01	3.50 $\pm$ 4.96
T	0.39 $\pm$ 0.12	0.29 $\pm$ 0.41	0.92 $\pm$ 0.56	6.96 $\pm$ 9.63
C	0.43 $\pm$ 0.61	1.90 $\pm$ 0.76	1.55 $\pm$ 3.18	80.8 $\pm$ 2.98

5'-CCCTTGAAGCTCCCAGCGCGCTCGAGCAA-3'
3'-GGGAACTTCGAGGGTCGCGCGAGCTCGTT-5'

galK gene oligo

5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'

araD gene oligo

Description: This table indicate the difference in nucleotides after and before adding ApeKI. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide.

Without adding ApekI-The mean (%) of mutated sequence on ApeKI restriction site-2 of the oligonucleotide			After adding ApeKI- The mean (%) of mutated sequence on ApeKI restriction site-2 of the oligonucleotide	
Gene	*galK* (n=2)	*araD* (n=2)	*galK* (n=2)	*araD* (n=2)
G	**Not-present**	**94.65 ± 1.56**	**Not-present**	**9.33 ± 9.02**
A	Not-present	2.09 ± 0.96	Not-present	7.75 ± 9.06
T	Not-present	1.25 ± 0.16	Not-present	1.87 ± 0.66
C	Not-present	2.12 ± 2.36	Not-present	81.03 ± 0.71

5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'

araD gene

Description: This table indicate the difference in nucleotides after and before adding ApeKI. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide.

	Without adding ApekI-The mean (%) of mutated sequence on 3'-side of the oligonucleotide		After adding ApekI- The mean (%) of mutated sequence on 3'-side of the oligonucleotide	
Gene	*galK* (n=2)	*araD* (n=2)	*galK* (n=2)	*araD* (n=2)
G	96.82 ± 1.85	96.80 ± 0.22	87.27 ± 0.96	23.00 ± 16.62
A	0.7 ± 0.71	1.16 ± 0.71	1.15 ± 0.13	0.60 ± 0.85
T	0.92 ± 0.15	0.85 ± 0.01	0.42 ± 0.34	9.37 ± 9.16
C	1.55 ± 0.97	1.17 ± 0.94	11.14 ± 1.17	67.02 ± 8.31

5'-CCCTTGAAGCTCCCAGCGCGCTCGAGCAA-3'
3'-GGGAACTTCGAGGGTCGCGCGAGCTCGTT-5'
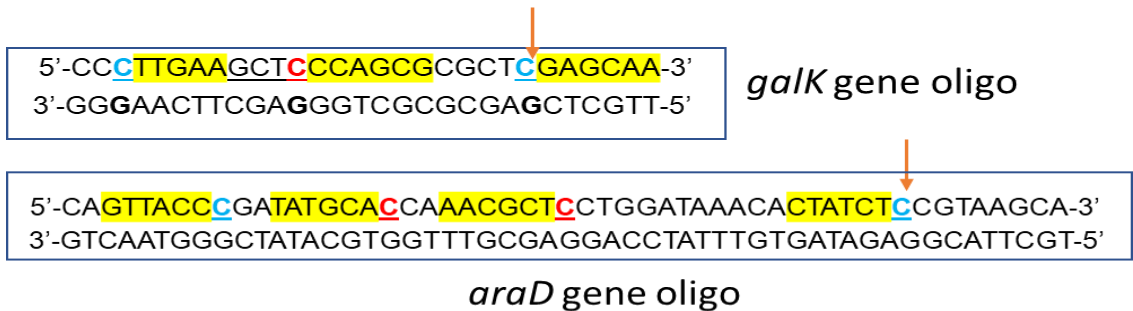
galK gene oligo

5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'

araD gene oligo

Description: This table indicate the difference in nucleotides after and before adding ApeKI. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide.

oligo	*araD* (n=2)	Oligo 1 (n=3)	Oligo 2 (n=6)
G	23.15 ± 9.76	64.6 ± 1.70	71.7 ± 9.99
A	3.72 ± 1.49	3.34 ± 4.50	10.5 ± 5.43
T	5.54 ± 7.83	0.61 ± 0.33	0.89 ± 0.41
C	67.5 ± 0.43	31.44 ± 5.28	16.8 ± 6.11

After adding ApeKI-The mean (%) of mutated sequence on 5'-side of the oligonucleotide

araD
```
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'
```

OLIGO 1
```
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTACGTAAGCA-3'
3'-GTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGT-5'
```

OLIGO 2
```
5'-CAGTTACCAGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGT-5'
```

Description: This table indicate the difference in nucleotides on the SPORE oligonucleotide after doing ApeKI treatment. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide. These values are indicated with 95% confidence interval.
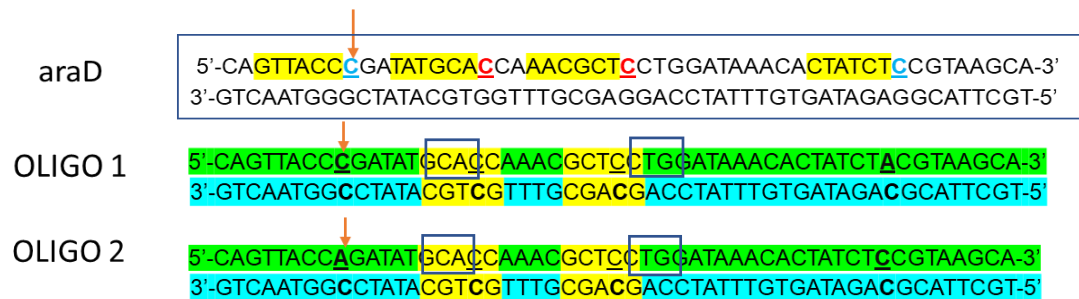
After adding ApeKI-The mean (%) of mutated sequence on ApeKI restriction site-1 of the oligonucleotide			
oligo	*araD* (n=2)	Oligo 1 (n=3)	Oligo 2 (n=6)
G	**8.70 ± 9.76**	**4.06 ± 1.70**	**7.54 ± 9.992**
A	3.50 ± 4.96	4.60 ± 0.49	4.93 ± 0.93
T	6.96 ± 9.63	0.47 ± 0.57	0.58 ± 0.34
C	80.80 ± 2.98	90.84 ± 1.07	86.96 ± 2.10

araD
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'

OLIGO 1
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTACGTAAGCA-3'
3'-GTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGT-5'

OLIGO 2
5'-CAGTTACCAGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGT-5'

Description: This table indicate the difference in nucleotides on the SPORE oligonucleotide after doing ApeKI treatment. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide. These values are indicated with 95% confidence interval.

oligo	*araD* (n=2)	Oligo 1 (n=3)	Oligo 2 (n=6)
After adding ApeKI-The mean (%) of mutated sequence on ApeKI restriction site -2 of the oligonucleotide			
G	**9.33 ± 9.02**	**3.73 ± 0.455**	**5.30 ± 1.62**
A	7.75 ± 9.06	1.44 ± 0.40	1.48 ± 0.78
T	1.87 ± 0.66	0.96 ± 0.15	0.99 ± 0.47
C	81.03 ± 8.31	93.85 ± 5.01	92.30 ± 7.98

araD

5'-CAGTTACC**C**GATATGCA**C**CAAACGCT**C**CTGGATAAACACTATCT**C**CGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'

OLIGO 1

5'-CAGTTACC**C**GATATGCA**C**CAAACGCT**C**CTG GATAAACACTATCT**A**CGTAAGCA-3'
3'-GTCAATGG**C**CTATA CGT**C**GTTTGCGA**C**GACCTATTTGTGATAGA**C**GCATTCGT-5'

OLIGO 2

5'-CAGTTACC**A**GATATGCA**C**CAAACGCT**C**CTG GATAAACACTATCT**C**CGTAAGCA-3'
3'-GTCAATGG**C**CTATA CGT**C**GTTTGCGA**C**GACCTATTTGTGATAGA**C**GCATTCGT-5'

Description: This table indicate the difference in nucleotides on the SPORE oligonucleotide after doing ApeKI treatment. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide. These values are indicated with 95% confidence interval.

After adding ApeKI-The mean (%) of mutated sequence on 3'-side of the oligonucleotide

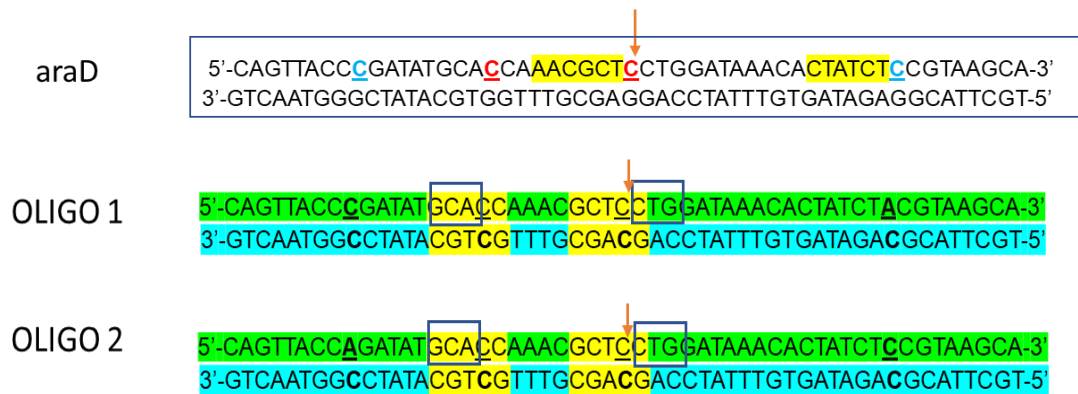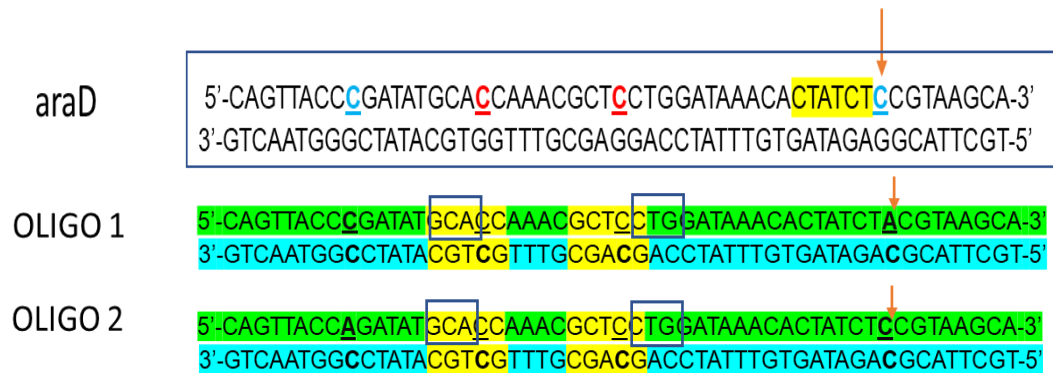| oligo	*araD* (n=2)	Oligo 1 (n=3)	Oligo 2 (n=6)
G	23.00 ± 9.76	67.23 ± 2.21	79.04 ± 7.63
A	0.60 ± 0.85	16.13 ± 0.40	0.49 ± 0.78
T	9.37 ± 9.16	2.23 ± 0.15	1.32 ± 0.34
C	67.02 ± 8.31	14.39 ± 5.01	19.12 ± 7.98

araD
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGGCTATACGTGGTTTGCGAGGACCTATTTGTGATAGAGGCATTCGT-5'

OLIGO 1
5'-CAGTTACCCGATATGCACCAAACGCTCCTGGATAAACACTATCTACGTAAGCA-3'
3'-GTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGT-5'

OLIGO 2
5'-CAGTTACCAGATATGCACCAAACGCTCCTGGATAAACACTATCTCCGTAAGCA-3'
3'-GTCAATGGCCTATACGTCGTTTGCGACGACCTATTTGTGATAGACGCATTCGT-5'

Description: This table indicate the difference in nucleotides on the SPORE oligonucleotide after doing ApeKI treatment. The sequences below the table belong to the oligonucleotides incorporated in the genome using oligonucleotide recombination and red arrow indicate the nucleotide position for which the quantitative values were calculated for each nucleotide. These values are indicated with 95% confidence interval.