

TUFTS UNIVERSITY

Mutations Investigation Experiment (MIE)

Bioinformatics Inquiry Through Sequencing (BIOSEQ)

Tabitha Amondi

An honors thesis for the Department of Chemistry

Table of Contents	Pages
Abstract	
Chapter 1: Introduction	
1.1: BIOSEQ	4
1.2: NGS	6
1.3: Thesis overview	9
4.3: References for Introduction	10
Chapter 2: Experimental Design	
2.1: Background	11
2.2: Methodology	13
2.3: References for Experimental Design	15
Chapter 3: Results	
3.1: Experiment #1	16
3.2: Experiment #2	21
Chapter4: Discussion	
4.1: Observations and deductions	24
4.2: Summer class project	28
4.3: References for discussion	31
Chapter5: Protocols and materials	
5.1: Transformation into XL-1 Red cells	32
5.2: Blue-white screening	33
5.3: Plasmid miniprep	34
5.4: List of materials	35

Abstract

In response to the increasing bulk data from sequencing, the Bioinformatics field emerged integrating statistics, mathematics and computer science. The Bioinformatics Inquiry through sequencing (BIOSEQ) project aims to introduce the field of bioinformatics to high-school and college students through educational sequencing projects developed by undergraduate students. One of such projects is the Mutation Investigation Experiment project (MIE). The MIE project investigates random mutagenesis through the propagation of a reporter plasmid in a DNA repair deficient *E.coli* strain and sequences the mutations using the MiSeq Illumina next generation sequencing platform.

This thesis provides data from two pilot experiments under MIE and discusses the findings. From the two experiments, a total of 23 distinct point mutations were observed, ten from the first experiment and 13 from the second. The mutations were either deletions or transition single nucleotide polymorphisms. The MIE project will be part of the BIOSEQ summer course at Tufts University for 2014 and will provide an avenue to discuss mutations, aspects of genomics, and bioinformatics with high-school students in an open-ended inquiry manner.

List of Figures	Pages
Figure 1: Illumina Genome Analyze sequencing	8
Figure 2: Illustration of experimental design	14
Figure 3: Images of <i>E-coli</i> XL-1 blue plated cells from experiment #1	17
Figure 4: Image of 0.4ng plasmid input of <i>E-coli</i> XL-1 blue plated cells from experiment #1	18
Figure 5: Image of 0.4ng plasmid input of <i>E-coli</i> XL-1 blue plated cells and culture volume 50µl for experiment #2	21
Figure 6: Relative mutation frequencies and number of observed mutations	26
Figure 7: Image of cell culture showing floating cell debris Observed from cycle 5 onwards	27
Figure 8: Details tab for sample 10 from experiment #1	28
List of Tables	
Table 1: Table of Qubit DNA quantification	16
Table 2: Data summary for experiment #1	19
Table 3: Data summary for experiment #2	22

Chapter 1: Introduction

1.1: BIOSEQ

The Bioinformatics Inquiry Through Sequencing (BIOSEQ) project is a collaborative initiative involving Tufts University, Somerville and Medford high schools among others and the National Institutes of Health (NIH). Through the Walt Lab in the Tufts Chemistry department, Tufts computer department and with funding from NIH, the BIOSEQ project has been set up to introduce the field of Bioinformatics to a wider audience specifically targeting high school and college students. The project has three major aims, namely: setting up a sequencing center at Tufts Medford campus, developing a semester long research-based bioinformatics course and developing modular activities to communicate the impact of Bioinformatics on science and medical research (1). The project is expected to span a five year period, with collaborative work involving students (high school & college), staff, graduate students and faculty. The first aim involving setting up of a sequencing set has been achieved with the setting up of a microbial genome sized sequencer – the MiSeq from Illumina. The second aim of designing and implementing a bioinformatics course is currently on track.

The field of Bioinformatics is defined by the integration of statistics, mathematics, and information and computer science. Sparked by the need to respond to the increased data on DNA sequence alignment and gene prediction algorithm, the field has metamorphosed and has necessitated the need to shift the education paradigm. The paradigm shift would involve teaching bioinformatics as a science that explains computational ideas and illustrates how the concepts pertain to life sciences problems (5). While masters programs have been designed to address the need for specialists in the Bioinformatics field, the need to improve the quantitative and analytical skills for life sciences students at the undergraduate level cannot be negated.

Undoubtedly, the increasing interdisciplinary nature of science research especially in life sciences and the vast amounts of data acquired necessitates the need to train life science students on Bioinformatics. Introducing Bioinformatics through an integrated biology and computer course to high school students based on hands-on inquirer experiments would provide a platform to emphasize core concepts of data management and analysis that is usually ignored in Biology curriculum which includes genetics and genetic testing (6). The BIOSEQ project's major impact would be to formally introduce the field of Bioinformatics to an audience that would not normally have access to such technology but to also enable inquiry-based research that enhances retention and understanding of concepts by students.

1.2: Sequencing Technology

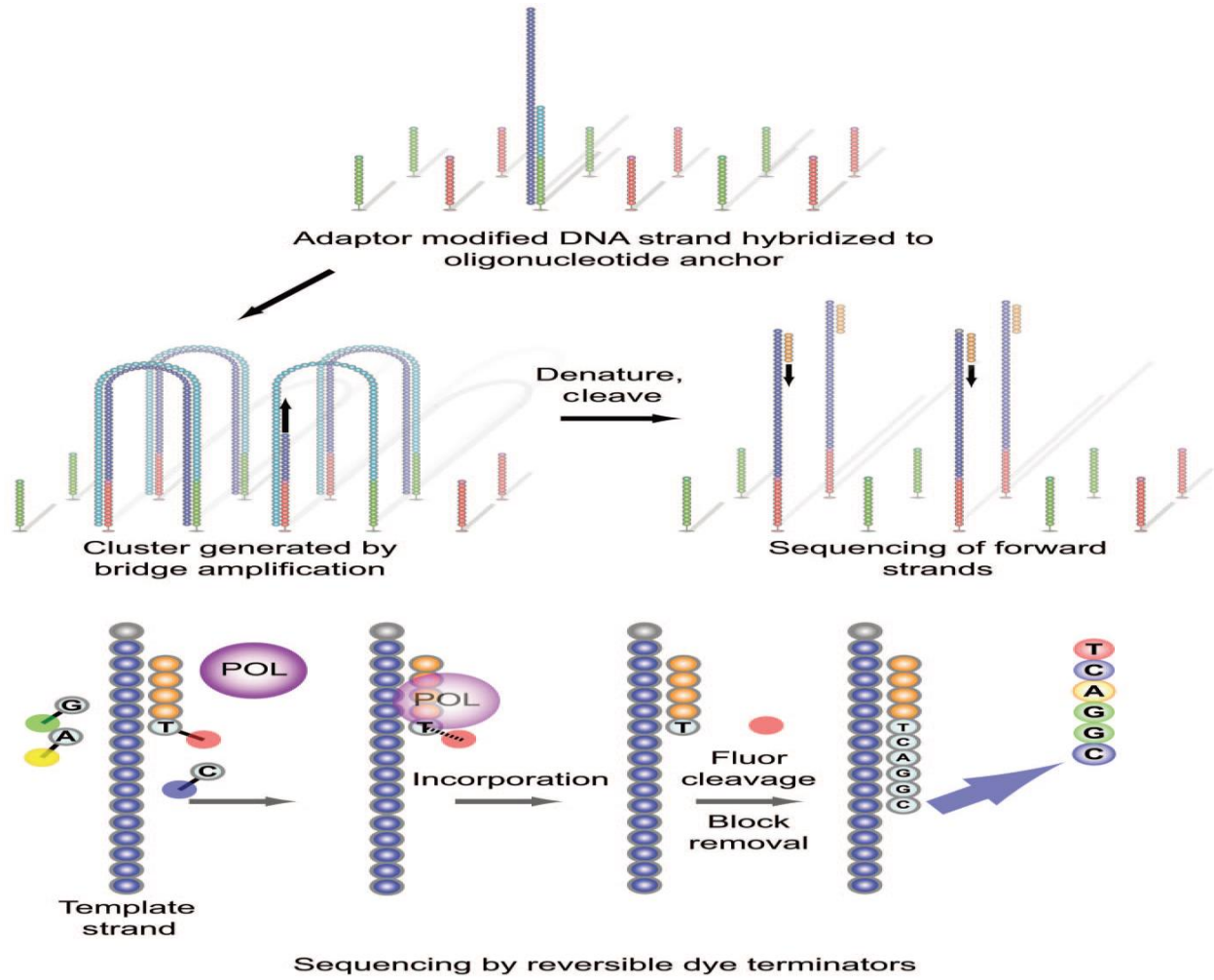
Sequencing, which refers to identifying the precise order of bases in a nucleic acid chain, has historically been performed using the Sanger method. The Sanger method involves the use of chain terminating analogs resulting in base specific termination of primer DNA, was used to sequence the first human genome in 2008 (3). The Sanger method has proved to be prohibitively expensive for projects on the scale of the genome, prompting development of newer platforms that are more versatile, affordable and feasible for modern genomic research. The Next - generation sequencers (NGS) are the newer platforms that have enabled genomic research to be disseminated to a small scale level and to more institutions (3). Some of the commercially available NGS's include: Illumina/Solexa, Roche 454, Applied Biosystems and Helicos Biosciences. Most of the NGS technologies are characterized by massively parallel sequencing of clonally amplified or single DNA molecules separated in a flow cell or attached to beads yet they are distinct from each other in their template preparation and sequencing methods (3). Template preparation methods include single DNA molecule templates and clonally amplified templates originally from single DNA molecules (2).

Roche 454 applies Pyro-sequencing, a form of sequencing by synthesis (SBS), which entails the addition of a single nucleotide to a DNA fragment and detection of the nucleotide based on pyrophosphate release, which generates a fluorescence signal (3). Roche 454 uses a template preparation method involving fragmentation by nebulization or sonication of the template DNA, end-pairing of the fragments and ligation to adapter oligonucleotides. This is followed by dilution of the library to single-molecule concentration, denaturation, and hybridization to individual beads containing complementary adapter oligonucleotides. The sequencing chemistry occurs within picoliter-scale sequencing reaction wells, which are etched onto a plate surface (3).

The Applied Biosystems or Solid platform has a similar sample preparation method to Roche 454 (2). The sequencing chemistry occurs in a picoliter well plate, where beads are immobilized onto a glass flow-cell (3). However, Solid uses a ligation reaction to add nucleotides to the sequencing strand (4). Base calling is achieved by detection of fluorescence from a dye attached to the ligated probe in each cycle (4).

The other NGS platform that is common and commercially available is the Illumina/Solexa platform, which is used for the BIOSEQ project. The Illumina MiSeq sequencer uses a sequencing-by-synthesis mechanism that is not based on pyrophosphate detection. Template preparation involves fragmentation of a DNA sample to short pieces with lengths of hundreds of base pairs and blunt ends; indices can be added for multiplexing if need be. Multiplexing refers to the sequencing of different DNA samples and distinguishing the different samples after base calling by using a unique index sequence. The preparation process results in single-stranded DNA fragments that can be annealed to complementary oligonucleotide adapters on the flow cell (3). Bridge amplification then follows, where the immobilized DNA strands arch over to adjacent adapters and multiple amplification cycles convert the molecule to clusters (3). Next, the clusters are denatured, and some strands are washed away leaving behind only the forward strands. Sequencing is then initiated by hybridizing a primer complementary to adapter sequences followed by polymerase addition of a complementary single base into a clonal cluster. Excess reagents are washed away, and the fluorescence is recorded, marking the end of the first cycle of sequencing by synthesis. Then, the cycle is repeated (3). The following image illustrates the above discussed chemistry of the Illumina/Solexa platform

Figure 1: Illumina Genome Analyze sequencing:



(Voelkerding et al. 2009)

1.3: Thesis Overview

Chapter 2 provides a detailed explanation of the experimental design and methodology for the Mutation Investigation Experiment (MIE) project under the BIOSEQ umbrella. The Mutation Investigation Experiment is a project designed to investigate random mutagenesis through the use of DNA repair deficient *E-coli* XL-1 Red strain and sequencing of a reporter plasmid. Chapter 3 discusses the results for two trial runs of the MIE project, analyzing the sequencing data and the data from the preparation of the library of mutant plasmids. Chapter 4 provides an in-depth discussion of the results from chapter 3, analyzing the data and inferring from the data the impact on the project. The future directions of MIE are also discussed, with a focus on the 2014 summer class for high school students where MIE project will be used as part of the six week course. The last chapter, 5, provides the protocols and materials for the two experiments and for the summer class.

1.4: References for Introduction

1. Walt, David. "Bioinformatics Inquiry through Sequencing (BIOSEQ)." *NIH BIOSEQ Proposal*. David Walt, 06 11 2012.
<[https://www.dropbox.com/sh/z60le29ua7pj8x5/IVJ4LvRRAw/Pages from WALT - NIH BIOSEQ Proposal 06.11-2.pdf](https://www.dropbox.com/sh/z60le29ua7pj8x5/IVJ4LvRRAw/Pages%20from%20WALT%20-%20NIH%20BIOSEQ%20Proposal%2006.11-2.pdf)>.
2. Metzker, Michael. "Sequencing technologies-the next generation." *Nature Reviews*. (8 December 2009): 31-46. 10.1038/nrg2626
3. Voelkerding, Karl, Shale Dames, and Jacob Durtschi. "Next-Generation Sequencing: From Basic Research to Diagnostics." *Clinical Chemistry*. no. 4 (2009): 641-658.
<http://www.clinchem.org/content/55/4/641.short>
4. Heewook, Lee, and Haixu Tang. "Next-generation sequencing technologies and fragment assembly algorithms.." *PubMed*. (2012): 155-74. 10.1007/978-1-61779-582-4_5.
5. Pevzner, P., & Shamir, R. (2009). Computing Has Changed Biology-- Biology Education Must Catch Up. *Science*, 325(5940), 541-542.
6. Hack. (2005). Bioinformatics: Current Practice and Future Challenges for Life Science Education. *Biochemistry and Molecular Biology Education*, 33(2), 82-85

Chapter 2: EXPERIMENTAL DESIGN

2.1: Background

Mutation Investigation Experiment (MIE) project explores mutagenesis through the use of DNA repair deficient *E-coli* XL-1 Red strain. Mutagenesis refers to the spontaneous or induced process of changing the genetic information of a species resulting in mutations (1). Mutated genes are characterized by their stability resulting in changes being replicated hence forming a new hereditary line (3). Mutations enable variations to be introduced into the gene pool thus driving natural selection which is a key factor in evolution. Spontaneous mutagenesis occurs due to intracellular changes such as DNA lesions, errors during DNA replication and the movement of transposable elements (2).

Sites of DNA damage such as mismatches, single stranded breaks, inter-strand crosslinks and abasic sites are generally classified as DNA lesions. Transposon elements constitute a significant change in DNA sequence and are enhanced by DNA damage (2). DNA synthesis fidelity during replication is highly dependent on DNA polymerase enzyme nucleotide selection, proofreading of replication errors and post replication mismatch repair (2). Spontaneous mutagenesis arises through various pathways but one major route has been through the DNA repair mechanism.

Mutations can be induced through different strategies that can be broadly classified as cell based and or enzymatic and chemistry approaches (4). Protocols capitalizing on the high error rates of either engineered or natural DNA polymerases enzymes constitute enzymatic strategies. The use of DNA modifying reagents such as ethyl methane sulfonate (EMS) are classified as chemical methods. Methods based on the increased rates of DNA repair errors constitute the bulk of cell-based strategies. The defective repair function of cells as with *E-coli* XL-1 Red cells strain is an example of a cell based mutagenesis method used in this experiment.

As a result of mutations in three of the primary DNA repair pathways in the *E. coli* XL1-Red strain, the rate of random mutations is 5000-fold higher than that of wild-type strain, providing for a means to create a mutant library for screening (4). The three genes that are altered in the *E. coli* XL1-Red strain are: *mutD*, *mutT* and *mutS*. The *mutD* protein is involved with the correction of mismatched bases during the selection of the complementary nucleotide to the template by DNA polymerase. The *mutT* protein is involved in the removal of a non-complementary base by an editing nuclease (8). *MutT* hydrolyzes 8-oxo-Dgtp (a potent mutagenic substrate for DNA synthesis) to the monophosphate form (5). The *mutS* protein is involved in the correction of a miss-incorporated base by mismatch repair system (6).

MIE project entails propagation of a plasmid through the DNA repair deficient *E. coli* XL-1 Red strain resulting in a library of mutant plasmids. The pUC18 plasmid was chosen as the mutagenesis reporter, meaning that the plasmid will be sequenced instead of the entire *E. coli* genome. The pUC18 plasmid is a circular vector mostly used in cloning (7). It has 2686 base pairs and is characterized by the following regions: an ampicillin resistance gene cassette, the *ori* region, which is an origin of replication, and a suite of multiple cloning sites that has the *lacZ* gene in the cassette (7).

With a possibly mutated plasmid library, a simple test for the presence of mutations in the plasmid would involve Blue-white screening method. Blue-white screening is based on the enzymatic activity of *E. coli* β -galactosidase encoded by *lacZ* gene. The *lacZ* gene can be divided into two segments, the products of which can combine to create a functional β -galactosidase protein. In this case, the N-terminal segment is part of the reporter plasmid and the other segment is included in genome of the *E. coli* XL1-Blue cells. When the *lacZ* gene fragment is intact, that is, there are no mutations in the N-terminal segment of *lacZ* found on the reporter plasmid, then

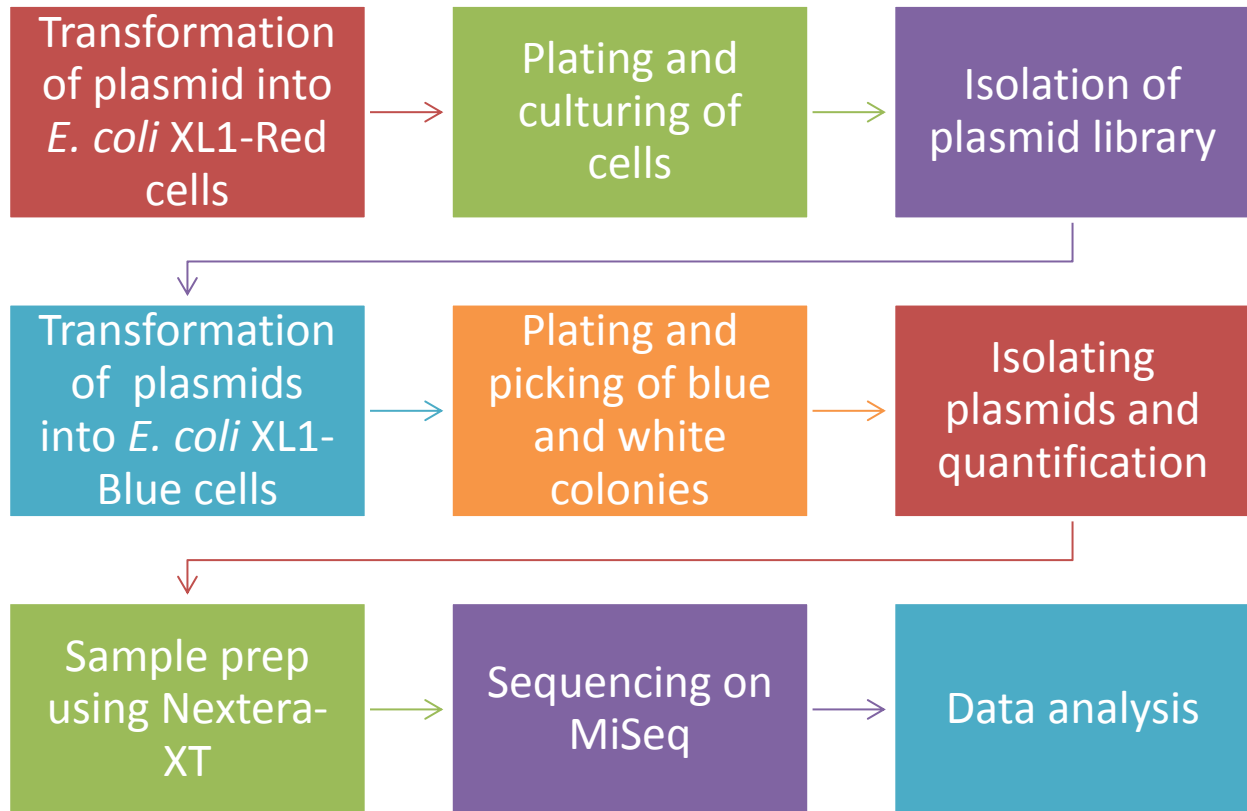
the β -galactosidase protein is formed and in the presence of X-gal/IPTG results in a deep blue color change of colonies. If a mutation occurs in the *lacZ* gene cassette region that affects the formation of β -galactosidase, then colonies will be white or pale blue (7). Point mutations on the *lacZ* cassette for white colonies and other mutations in other regions of the plasmid are expected from the plasmid library. The mutagenesis process is probed and detailed steps are discussed further in the methodology section.

2.2: Methodology

The Mutation Investigation Experiment (MIE) project involves preparing a library of mutated plasmids and sequencing for the random mutations on MiSeq Illumina. To induce random mutations, a plasmid (PUC19/PUC18) is transformed into competent *E.coli* XL-1 Red cells, the cells are plated after transformation and allowed to grow for over 41 hours. Colonies are then picked from the plates and inoculated in LB medium containing ampicillin (LB+Amp) to allow for bacterial growth and replication resulting in random point mutations. The cells are cultured in LB+Amp for 10 cycles of 24 hours growth period with a 1:500 dilution after each cycle. The cells are spun down and plasmids isolated from the cells using the QIAGEN miniprep protocol. The DNA is then quantified using Qubit Fluorometer. Next, the library of mutant plasmids is further transformed to *E.coli* XL-Blue cells to allow for blue-white screening. The cells are plated on LB agar plates with IPTG and X-gal, and incubated at 37°C for 24 hours. Next blue and white colonies are selected, inoculated in LB+Amp liquid medium and grown for a further 24 hours period. The plasmids are then isolated as before and this provides the DNA input for sequencing. The plasmid library is then prepared for sequencing using the Nextera XT sample preparation kit from Illumina. The sequencing is done using the MiSeq nano reagent kit and the small genome

resequencing workflow. The data is obtained from the MiSeq Reporter or Illumina's BaseSpace cloud computing environment. The entire experimental design is illustrated in figure 2:

Figure 2: Illustration of experimental design:



2.3: References for Experimental Design

1. Beale, G. (1993). "[The Discovery of Mustard Gas Mutagenesis by Auerbach and Robson in 1941](#)". *Genetics* **134** (2): 393–399. [PMC 1205483](#)
2. Kunz BA, Ramachandran K, Vonarx EJ (April 1998). "[DNA sequence analysis of spontaneous mutagenesis in *Saccharomyces cerevisiae*](#)". *Genetics* **148** (4): 1491–505. [PMC 1460101](#).[PMID 9560369](#)
3. [Auerbach, C.](#); Robson, J.M.; Carr, J.G. (March 1947). "[Chemical Production of Mutations](#)".*Science* **105** (2723): 243–7. [doi:10.1126/science.105.2723.243](#)
4. Rasila, Tiina, Maria Pajunen, and Harri Savilahti. "Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment."*Elsevier*. no. 1 (May 2009): 71-80.
5. Tajiri, Tatsuro, Hisaji Maki, and Mutsuo Sekiguchi. "Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in Escherichia coli." *Elsevier*. no. 3 (May 1995): 257-267
6. Wu, T H, and M G Marinus. "Dominant negative mutator mutations in the mutS gene of Escherichia coli.." *Journal of Bacteriology*. no. 17 (September 1994): 5393-5400.
<http://jb.asm.org/content/176/17/5393.full.pdf.html>
7. Pasternak, Jack J. *An Introduction to Human Molecular Genetics*. 2. New Jersey: John Wiley & Sons, 2005. 116-118. eBook.
<http://books.google.com/books?id=ITUNkbzWso4C&printsec=frontcover&source=gbs_ge_summary_r&cad=0
8. Echols, H, C Lu, and P M Burgers. "Mutator strains of Escherichia coli, mutD and dnaQ, with defective exonucleolytic editing by DNA polymerase III holoenzyme." *PNAS*. no. 8 (April 1983): 2189-2192

Chapter 3: RESULTS

3.1: Experiment # 1

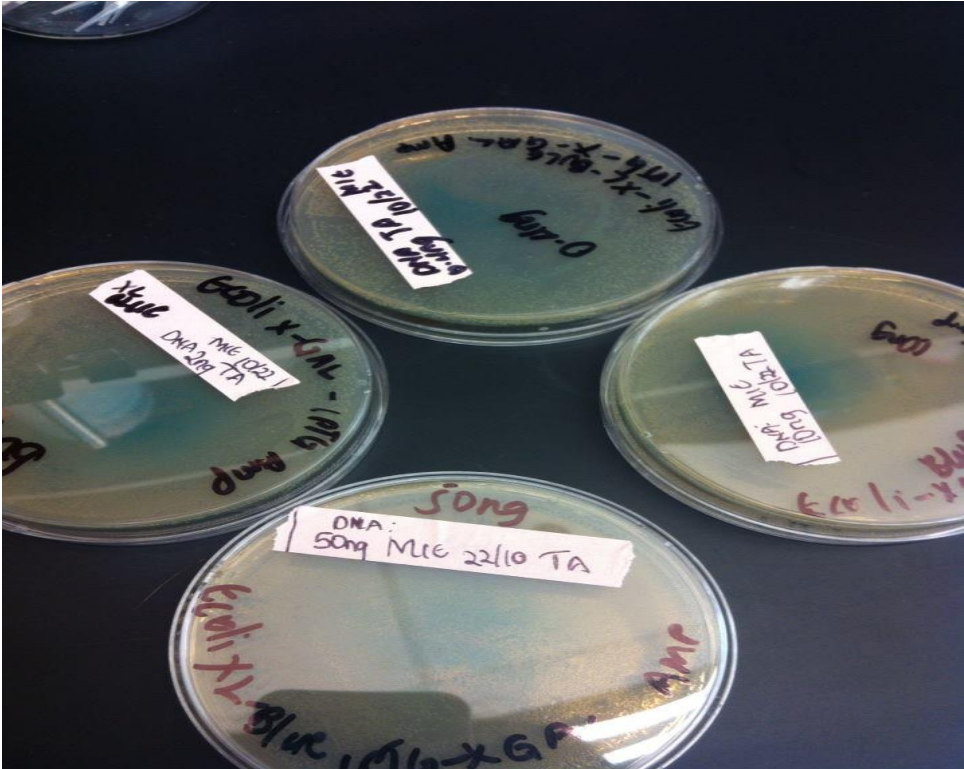
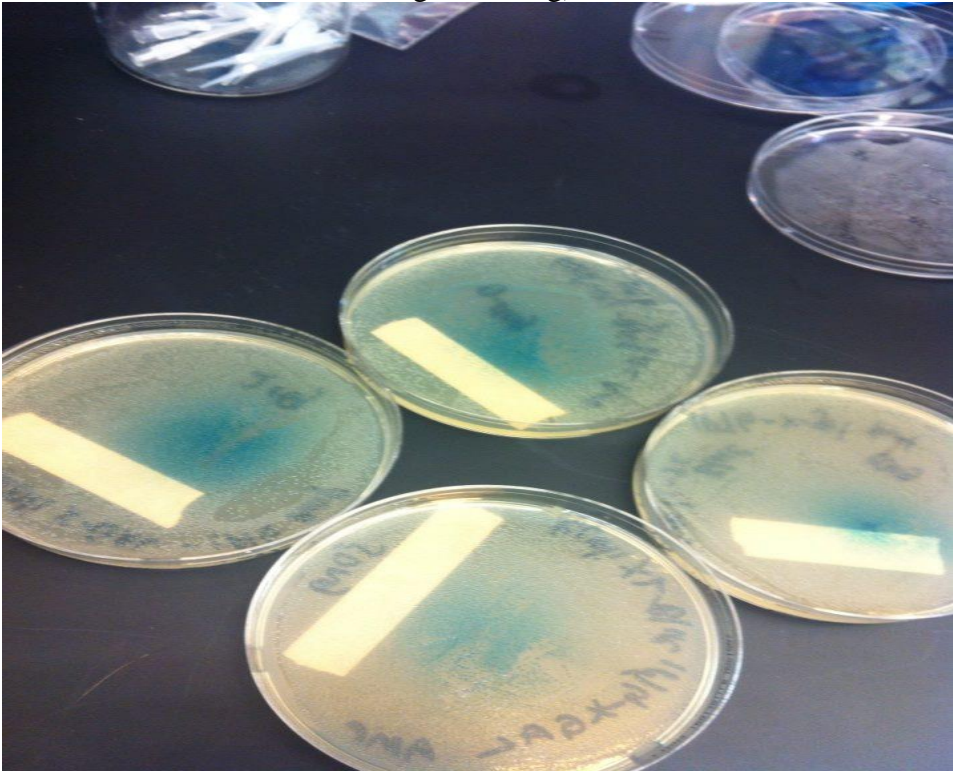
In the first experiment, two transformations of 100 μ l of XL-1 Red competent cells in Falcon tubes were made; the transformed cells were plated onto two agar + ampicillin plates (A1 and B1) and grown overnight. Two more plates (A2 and B2) of the transformed cell mixture were plated the following day, making the total number of plates equal to four. Colonies were picked from the plates and inoculated overnight in 5ml of LB broth containing ampicillin. The plasmids were isolated from the cultures using the QIAGEN miniprep kit and the DNA quantified using a Qubit fluorometer. The concentration of plasmid recovered is shown in Table 1:

Table 1: Table of Qubit DNA quantification

Qubit DNA quantification		
Plates	Tubes	Concentration(μ g/ml)
First	A1	16.4
	A2	8.47
Second	B1	11.3
	B2	3.05

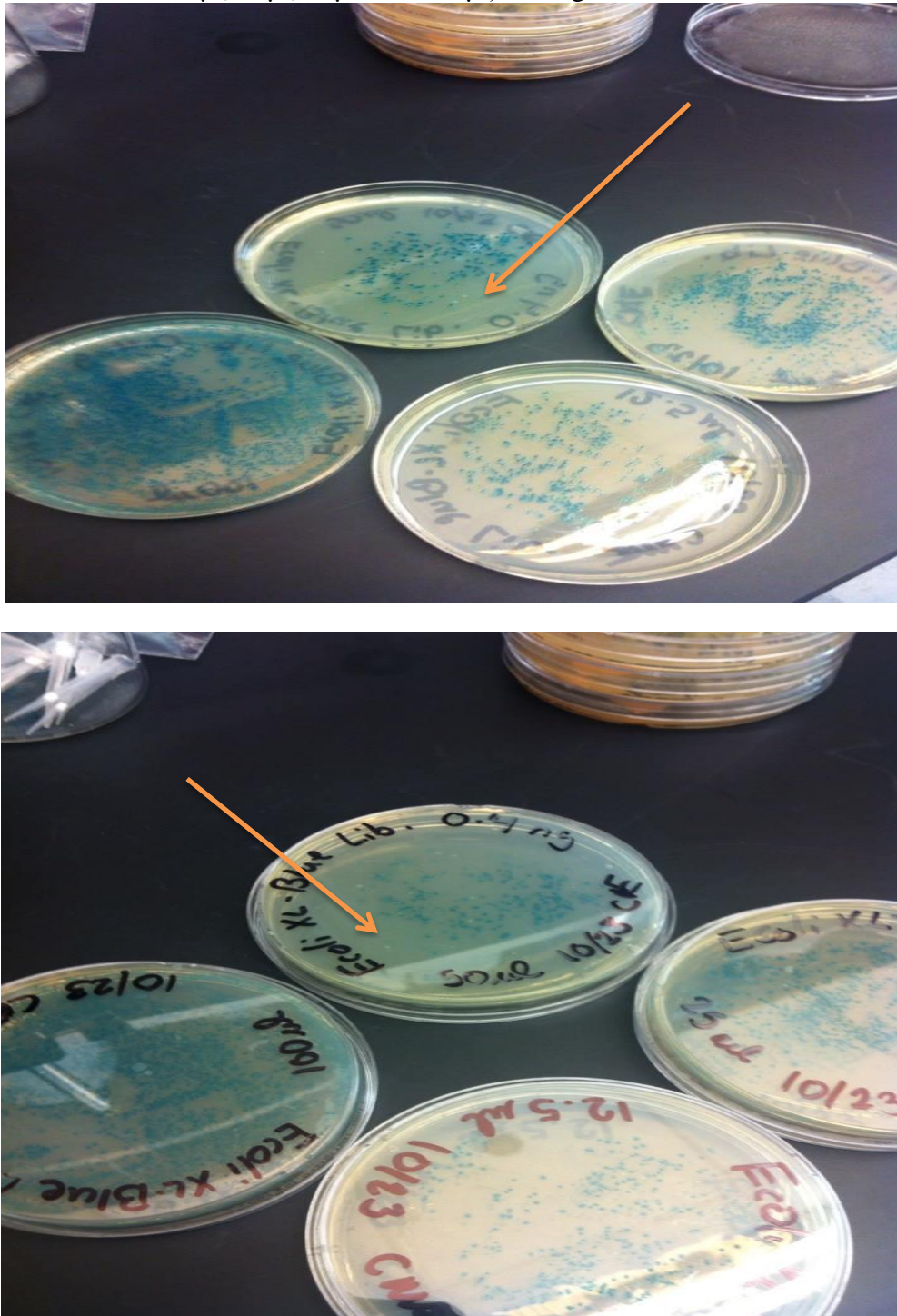
The plasmid sample with the highest DNA concentration (A1) was diluted to a concentration of 0.2ng/ μ l and used for transformation into *E-coli* XL-1Blue cells. After the second transformation, the cells were spread on LB agar plates containing ampicillin, IPTG and X-gal and allowed to grow overnight for color screening. The plasmid concentration for the second transformation was: 50ng, 10ng, 2ng and 0.4ng. The plates of the four samples are shown in figure 3:

Figure 3: Images of E-coli XL-1 blue plated cells from experiment # 1 (Dna input: 50ng, 10ng, 2ng and 0.4ng)



The 4ng DNA input was further plated on four more plates with different culture volumes of: 100 μ l, 50 μ l, 25 μ l and 12.5 μ l. The colonies observed are shown in figure 4:

Figure 4: Image of 0.4ng plasmid input of *E-coli* XL-1 blue plated cells from experiment #1 (Culture volume: 100 μ l, 50 μ l, 25 μ l and 12.5 μ l). Orange arrows indicate white colonies



Twenty samples of white and blue colonies from the 0.4ng DNA input (12.5µl and 50µl spread) plates were inoculated in 5ml of LB + Amp and grown overnight. The plasmids from the 20 samples were recovered using miniprep protocol. The twenty samples were then prepared for sequencing using the Nextera-XT sample prep kit and sequenced on MiSeq using the small genome resequencing workflow. The data on observed point mutations from the twenty samples is shown in table 2.

Table 2: Data summary for experiment #1 (grey part indicating the white colonies numbered one to ten and the blue section indicating blue colonies numbered eleven to twenty)

Position	309	340	365	366	463	1280	1306	1308	1406	2345	2345	2494
Variant type	SNP	SNP	SNP	SNP	Indel	SNP	SNP	SNP	Indel	SNP	SNP	Indel
Call	C→[T/T]	T→[C/ C]	C→[T/T]	C→[T/T]	TG/T-	G→[A/ A]	G→[A/ A]	G→[A/ A]	G-/GT	C→[C/ T]	C→[T/T]	C-/CT
Sample 1				X	X			X			X	
sample 2				X	X			X		X		
sample 3					X		X					
Sample 4	X				X			X	X			
sample 5					X			X				
sample 6				X	X			X			X	
sample 7			X		X			X				X
Sample 8					X			X				
Sample 9		X		X	X			X			X	
Sample 10				X	X			X		X		
Sample 11					X			X				
Sample 12					X			X				
Sample 13					X			X				
Sample 14					X			X				
Sample 15					X			X				
Sample 16					X			X				
Sample 17					X			X				
Sample 18					X			X			X	
Sample 19					X			X			X	
Sample 20					X	X		X				

3.2: Experiment #2

The experiment followed the outline on the experimental design section except that the transformed *E-coli* XL-1 Red cells were grown in 24 hour cycles for 10 days in LB +Amp. The successive cycles were diluted 1:500 ratio to allow the cells to continue to multiply. A library of mutant plasmids was isolated from each of the two cultures from cycle day 10. The DNA was quantified using Qubit fluorometer and transformed into *E-coli* XL-1 Blue competent cells to enable blue-white screening. A total of sixteen colonies, eight white and eight blue, were selected, cultured and plasmids isolated. Seventeen samples including the untransformed pUC18 plasmid as a control were prepared for sequencing using the Nextera-XT sample prep kit and sequenced using the Illumina MiSeq small genome resequencing workflow.

Figure 5: Image of 0.4ng plasmid input of E-coli XL-1 blue plated cells and culture volume 50μl for experiment #2. Orange arrows indicate white colonies

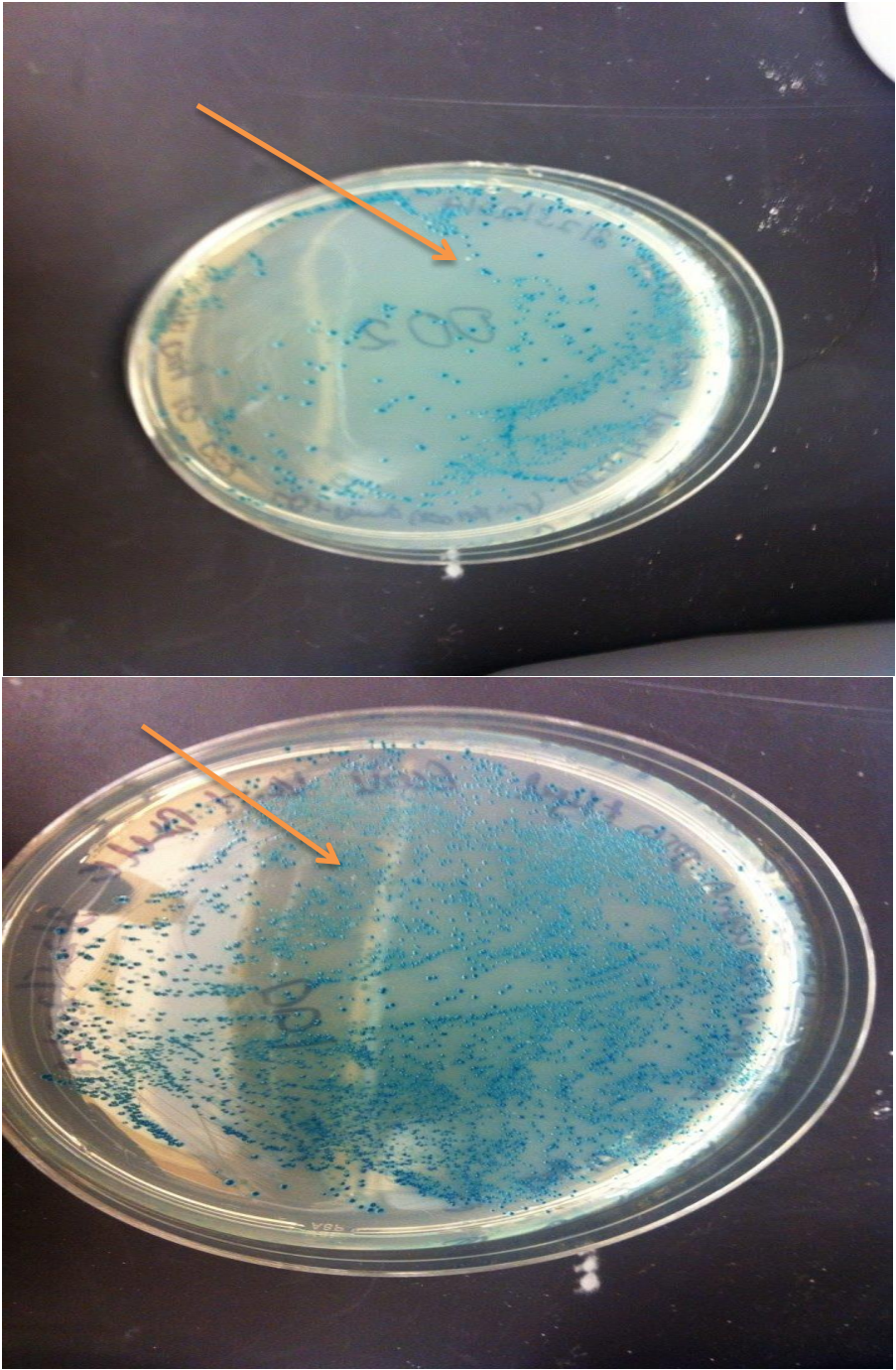


Table 3: Data summary for experiment #2 (grey part indicating the white colonies numbered one to ten and the blue section indicating blue colonies numbered eleven to sixteen, with a control sample)

Position	317	341	367	463	573	879	1308	1328	1325	1621	1919	2326	2530	2565	2530
Variant type	Indel	SNP	SNP	Indel	SNP	SNP	SNP	SNP	SNP	Indel	SNP	SNP	SNP	Indel	Indel
Call	AG/A-	A→[G/G]	A→[A/G]	TG/T-	C→[C/T]	T→[T/C]	G→[A/A]	C→[T/T]	G→[G/A]	G-/GT	T→[T/C]	A→[T/T]	T→[T/C]	GA/G-	C-/CT
Sample 1				X			X								
sample 2				X			X				X				
sample 3			X	X		X									
Sample 4				X	X		X								
sample 5	X			X			X				X		X	X	
sample 6			X	X			X								
sample 7		X		X			X								X
Sample 8				X			X	X							
Sample 9				X			X								
Sample 10				X			X			X					
Sample 11				X			X								
Sample 12				X			X								
Sample 13				X			X								
Sample 14				X			X								
Sample 15				X			X	X							
Sample 16				X			X		X						
Control				X			X								

Chapter 4: DISCUSSION

4.1: Observations and deductions

From experiment #1, the highest amount of DNA plasmid was recovered when the plated cultures were grown for a longer period of time after the first transformation. From table 1, the two plates labeled A, which were grown for forty one hours instead of the thirty hours as stated in the Stratagene protocol, generally had higher a yield than B plates.

For the second transformation to *E-coli* XL-1 blue cells and plating segment, colonies were more spread out in the culture volumes of 12.5 μ l and 50.0 μ l compared to that of 25.0 μ l and 100 μ l. Most blue and white colonies were picked from the plates in which the colonies were more dispersed. In addition, the protocol was adjusted so that the optimal culture volume spread for blue-white screening was set to be 50.0 μ l.

Experiment #2 was modified so that the *E-coli* XL-1 Red cells were cultured for ten cycles of 24 hours with a 1:500 dilution for each cycle. From day five onwards, some evidence of cells death was observed in the culture tubes. The sequence sample pool of sixteen in experiment #2 with eight white and eight blue colonies was smaller than that of experiment # 1. In addition, a control, which was the plasmid pUC18 was also sequenced along with the samples. The control showed that the input plasmid for the transformations differed from the published reference genome in positions 463 and 1308 which are located within the *LacZ* cassette and Ori regions respectively. These two variations were in the pUC18 plasmid and were not as a result of random mutagenesis through propagation of DNA repair deficient *E-coli* strain.

It was observed that there were more mutations within the white colonies group than the blue ones in both experiments. White colonies were expected to have at least a single mutation within

the *LacZ α* gene encoding cassette. The *LacZ α* encoding gene cassette is within the 235-395 bp plasmid sequence length hence a mutation within this region would ascertain the mutation indicated by white color change (2). The other mutations were in the Ori and ampicillin resistance cassette regions. The Ori (origin of replication site) is within the 852-1466bp sequence length while the ampicillin resistance cassette is within the 1626-2486bp sequence length (3). In addition to being dependent on the *LacZ α* gene encoding cassette being intact, the expression of the β -galactosidase enzyme is also dependent on the *LacZ* mRNA promoter (begins at position 507bp), the Lac repressor binding site (507-487) and the CAP protein binding site(591-554). Thus, any mutation within the region of 236-591bp sequence length could be responsible for the white color colonies. In table 3, the mutation at position 573 could account for its white color, which results from no formation or formation of a non-function β -galactosidase enzyme.

A total of thirteen mutations at different positions were observed in experiment #2 compared to ten from experiment #1. For experiment # 1, all the ten white samples except sample 3, 5 and 8 in the white colonies group had a point mutation within the 236-591bp sequence region. The number of observed mutations within the three regions: 236-591bp region, Ori and ampicillin cassette was 4, 3 and 3 respectively. For experiment #2, samples 1, 2, and 8 which were isolated as white colonies did not have any mutations within the 236-591bp gene encoding region as would be expected for white set. The other four samples from the pool did have at least a single point mutation within that region. The number of observed mutations within the three regions: 236-591bp region, Ori and Ampicillin cassette was 3, 4 and 6 respectively. The count for the three regions was calculated as a sum of mutations at different positions and number of occurrences in each sample.

All the mutations observed from the two experiments were either deletions or single nucleotide changes which were transitions. No transversion or insertion mutations were observed. The random mutagenesis using the *E.coli* mutator strain should result in only point mutations and this was the case for both experiments. This method of random mutagenesis has been shown to have an increased 5000-fold mutation rate compared to the wild type thus enabling generation of single point mutations within a cloned gene of interest overnight (1). The total number of point mutations generated from the experiments reported here indicated that the mutation rates were comparable to the previously reported 5000-fold increased rate.

The number of generated mutations was expected to increase with successive growth cycles however, the difference in the number of mutations generated between observed between the two experiments was only three. Rasila et al. (2) in the evaluation of random mutagenesis using six different methods including the *E.coli* XL1-Red strain observed only a single mutation or none within the 500bp region which included the *LacZ α* cassette in their sequenced samples. In addition, all the mutations they detected after sequencing were only from white and pale blue colonies and no mutations were observed from the blue colonies (2). Besides sequencing, they (Rasila et al.) analyzed the frequency of mutations by determining the proportion of white and pale blue colonies from the total number of colonies. From the physical count of colonies and their frequencies per 1000bp, they observed that the number of white and pale blue colonies increased significantly after day 8. The data is shown in figure 6.

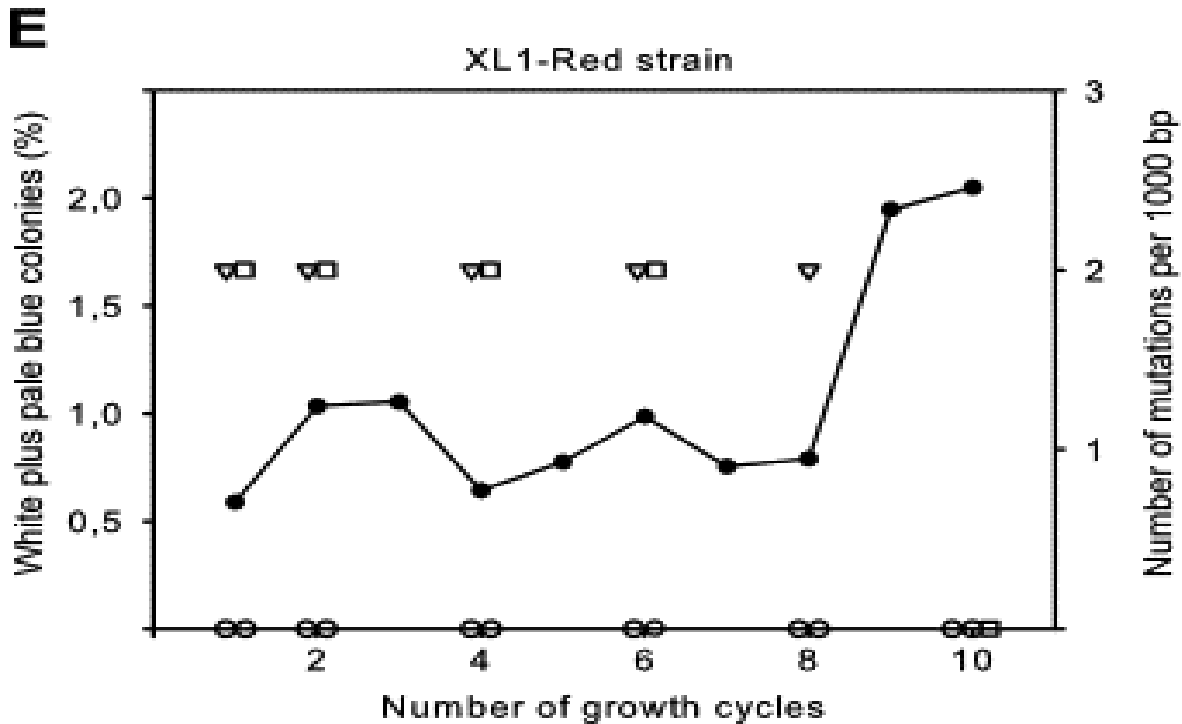


Figure 6: Relative mutation frequencies and number of observed mutations. Right scale indicates number of mutations per 1000bp. Number of mutations from blue colonies shown by open circles, those from pale blue colonies shown by open triangles and those from white colonies shown with open squares (2).

Data from the MIE experiment and Rasila et al. indicates that propagating the plasmid through longer cycles of 24hours does not result in a significant increase in the total number of mutations. If a great number of mutations were desired, then a different approach besides the increased number of successive growth cycles would be needed. One possible way of doing this would be to transform the plasmid into the XL-1 Red component cells, culture the cells up to five cycles of 24hrs with a 1:500 dilution. After day 5, isolate the plasmids and transform the plasmid library into fresh XL-1 Red component cells and allow the cells to grow for a further five day cycle. In the MIE experiment #2 after day 5 of the successive growth cycles, cell debris was

observed to be floating at the top of the liquid medium as shown in figure 7. This observation was consistent all through to cycle ten. This debris was mostly likely dead cells, which meant that over the growth cycles, plasmids that may have had more mutations may have not been recovered and analyzed.

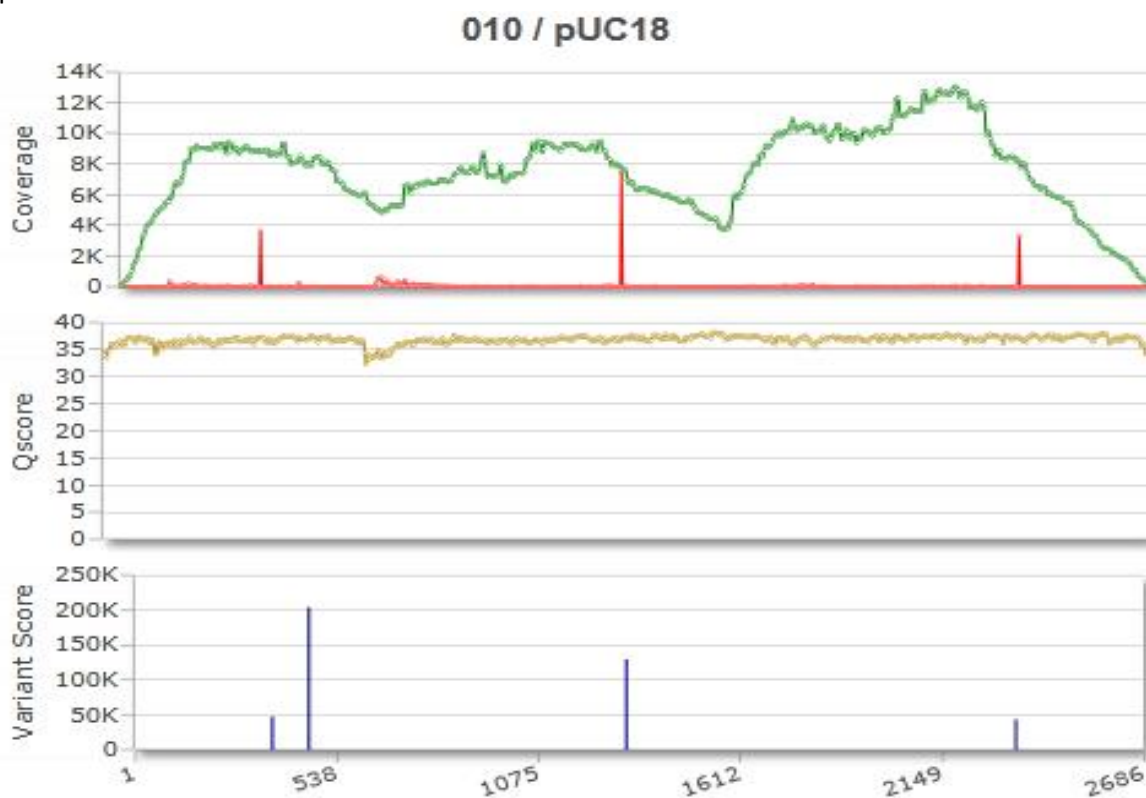
Figure 7: Image of cell culture showing floating cell debris observed from cycle 5 onwards.



The other unexpected observation from the sequencing data involved the coverage graph for the two experiments. Since the input DNA for sequencing was a plasmid, the coverage was expected to be relatively even through the entire base pair sequence. However, the coverage declined significantly at the ends of the sequence length as shown in figure 8. This implies that during alignment, the algorithm aligns the segments not as a circular entity but as linear. As a result, mutations that may have been present at the ends of the sequence were truncated. To address

this, the base calls could be aligned using different secondary software or the alignment could be reshuffled so that the segments are mixed up allowing the ends to be aligned as the central focus.

Figure 8: Details tab for sample 10 from experiment #1. Focus is on the coverage of the sequence with a deep at the ends of the plasmid sequence. This was observed for all samples including those from experiment #2.



After successfully running two pilot experiments for the MIE project and with the observations made, the MIE project will be used during the summer as a module for introducing the field of Bioinformatics to students. From the two MIE experiments the questions that arose from the observations made emphasize the nature of life science research. Providing students at the high school level with such an experience that is open ended and engaging while introducing the field of bioinformatics and research would hopefully get them interested in science as a career option.

4.2: BIOSEQ Summer class

During the summer, the MIE project will be part of the BIOSEQ summer class for high school students who have completed honors or AP biology. The BIOSEQ summer class would expose the students to the field of Bioinformatics and research. The MIE project will be part of the research and lab section that will familiarize the students with molecular biology and biochemistry lab techniques. Some of the techniques they will acquire from this project would include: bacterial transformation, cellular culturing, plasmid prep and DNA sample prep and sequencing. The MIE project is open-ended and would allow the students to reflect on what the data from sequencing means and its implications. The students will be challenged by the questions that will arise from the observations made. The whole project will be inquiry-based and would encourage the students to think of their own proposals for research. The BIOSEQ summer course will be a great opportunity to implement the MIE project and for the students to become acquainted with the Bioinformatics field.

4.3: References for the Discussion

1. A Greener, M. Callahan, B. Jerpseth, "An efficient random mutagenesis technique using an *E.coli* mutator strain," *Mol. Biotechnol.*, 7 (1997), pp. 189-195
2. Rasila, Tiina, Maria Pajunen, and Harri Savilahti. "Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment." *Elsevier*. no. 1 (May 2009): 71-80
3. Yanisch-Perron, C., et al., Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors, *gene*, 33, 103-119, 1985 <<http://www.thermoscientificbio.com/uploadedFiles/Resources/pUC18-pUC19-map.pdf>>
4. Drake, John W., Brian Charlesworth, Deborah Charlesworth, and James F. Crow. "Rates of Spontaneous Mutation." *Genetics*. 148.4 (April 1, 1998): 1667-1686. Web. 18 Apr. 2014. <<http://www.genetics.org/content/148/4/1667.full>>

Chapter 5: METHODS AND PROTOCOLS

5.1: Transformation into XL-1 Red cells

The purpose of this is to transform the pUC18 plasmid into *E.coli* XL-1 Red competent cells so as to create a library of mutated plasmids. The protocol is adapted from Stratagene *E.coli* XL-1 Red manual

- 1) Set up the water bath to 42°C
- 2) Pre-chill 14ml BD Falcon polypropylene tubes on ice. Place 1ml of SOC medium into eppendorf tubes and pre warm to 42°C on the heat block.
- 3) Thaw the XL1-Red competent cells on ice.
- 4) Gently mix the XL1-Red competent cells by hand. For each transformation reaction, aliquot 100 µl of the XL1-Red competent cells into a prechilled 14-ml BD Falcon polypropylene round-bottom tube.
- 5) Add 1.7 µl of the β-mercaptoethanol to each 100-µl of XL1-Red competent cells, giving a final concentration of 25 mM.
- 6) Swirl the contents of the tubes gently. Incubate the tubes on ice for 10 minutes, swirling the tubes gently every 2 minutes.
- 7) Add 1 µl of pUC18 plasmid DNA to each aliquot of XL1-Red competent cells and swirl gently.
- 8) Incubate the tubes on ice for 30 minutes.
- 9) Heat-pulse the tubes in 42°C water bath for 45 seconds. The duration of the heat pulse is very critical for obtaining the highest efficiencies.
- 10) Incubate the tubes on ice for 2 minutes.
- 11) Add 0.9 ml of SOC medium (pre-warmed to 42°C) to each tube and incubate the tubes at 37°C for 1 hour with shaking at 225–250 rpm.
- 12) Spread 100 µl of the XL1-Red transformation reaction on an LB–ampicillin agar plate
- 13) Incubate the plates at 37°C for 30–40 hours.
- 14) Using a sterile wire loop or a sterile toothpick, select ≥ 200 colonies at random from the transformation plates and inoculate cultures of 5–10 ml of LB broth containing ampicillin antibiotic. Grow these cultures overnight at 37°C.
- 15) Grow the cultures over 10 days with dilution cultures of 1:500 factor every 24 hours to allow further growth after lag phase stage.
- 16) Prepare miniprep DNA from 1.5 ml of the overnight culture in order to isolate the randomly mutated plasmid DNA.

5.2: Blue-white screening (Transformation into XL-1 Blue cells)

The purpose of this is to transform the mutated plasmid library into *E. coli* XL-1 Blue competent cells to allow for blue-white screening (adapted from the Stratagene manual).

1. Pre-chill a number 14-ml BD Falcon polypropylene round-bottom tubes on ice. Preheat SOC medium to 42°C.
2. Thaw the supercompetent cells on ice. When thawed, gently mix and aliquot 100 µl of cells into each of the pre-chilled tubes.
3. Add 1.7 µl of β-mercaptoethanol to each aliquot of cells.
4. Swirl the contents of the tubes gently. Incubate the cells on ice for 10 minutes, swirling gently every 2 minutes.
5. Add 0.4 ng of the mutated plasmid library to aliquot of cells and swirl the tubes gently. Dilute the stock plasmid library to final concentration of 0.2ng/ml in TE buffer (10 mM Tris (hydroxymethyl) aminomethane, 1 mM EDTA, pH 8.0) before use.
6. Incubate the tubes on ice for 30 minutes.
7. Heat-pulse the tubes in 42°C water bath for 45 seconds. The duration of the heat pulse is critical for maximum efficiency.
8. Incubate the tubes on ice for 2 minutes.
9. Add 0.9 ml of preheated SOC medium and incubate the tubes at 37°C for 1 hour with shaking at 225–250 rpm.
10. Plate 50 µl of the transformation mixture on LB agar plates containing ampicillin as well as IPTG and X-gal to allow for color screening..
11. Incubate the plates at 37°C. Colonies containing plasmids with mutations that disrupt the expression of the β-galactosidase α-peptide will remain white, while colonies containing plasmids without mutations that disrupt the expression of the α-peptide will be blue.

Preparation of Agar Plates for Blue-White Color Screening (adapted from Stratagene)

Spread 100µl of LB broth containing 10mM IPTG and 2% X-gal on solidified LB +amp agar plates 30 minutes prior to plating of transformations.

5.3: Plasmid miniprep (Recovering plasmid from cells)

The purpose of this is to isolate plasmids from *E.coli* cells (Adapted from QIAprep Spin Miniprep Kit manual)

1. Pipet 1.5ml of the bacterial overnight culture into a clean eppendorf tube. Centrifuge the samples at (>8000rpm) for 3 min at room temperature. Carefully use a pipet to remove liquid above the pellet and repeat this step to collect all the cells from your culture. If your culture volume was 5ml, you would do this three times.
2. Then add 250µl Buffer P1 into each eppendorf tube and pipet mix.
3. Add 250 µl Buffer P1 and mix thoroughly by inverting the tubes 4-6 times until the solution becomes blue. (Remember that the blue color change will only be observed if LyseBlue reagent was added to Buffer P1 otherwise the solution appears clear).
4. Centrifuge the samples for 10mins at 13000rpm on the bench top micro-centrifuge and at room temperature.
5. Pipette the supernatant after centrifuging into blue QIAprep spin column. Centrifuge the QIAprep spin column (which contains the supernatant) for 1min at 13000rpm and discard the flow through.
6. Wash the spin column by adding 500 µl Buffer PB. Then centrifuge for 1 min at 13000rpm and discard the flow through.
7. Wash the spin column by adding 750 µl Buffer PE. Then centrifuge for 1 min at 13000rpm and discard the flow through.
8. Centrifuge the spin the column again for 1min to ensure the removal of residual wash buffer.
9. Take the QIAprep spin column and place the column in a clean 1.5ml eppendorf tube. Add 50 µl Buffer EB to the center of the column. Allow the column to stand for 1min and then centrifuge for 1min at 13000rpm. This step elutes the DNA. The flow through should contain your DNA plasmid.
10. Discard the spin column and waste in the right place.

5.4: List of Materials

Transformation of Ecoli-XL Red cells:

XL-1 Red competent cells kit from stratagene which includes:

- Red-orange tube of XL-1 Red cells 5 x 0.2ml aliquots
- Blue tube of XL-1 Blue cells 5 x 0.2ml aliquots
- pUC18 control plasmid 0.1ng/μl in TE buffer 10μl
- 1.42 M β-Mercaptoethanol 50μl

\$408 per kit

Transformation using blue-white screening competent cells and sample prep for sequencing:

- 14ml Polypropylene Round-Bottom Falcon BD tubes
- SOC Medium. 10mL. 10 pack. Sterile. Cat. No. S1602 (Teknova)- **\$64.83**
- LB Agar plates with 100 μg/mL Ampicillin. 100mm, 20 plates, Sterile Cat.No. L1004-
\$27.56 *2
- TE Buffer, Isopropanol, 70% ethanol
- X-gal 2%
- 10mM IPTG
- LB liquid media- Description : 2% Tryptone, 1% Yeast extract. 1.0% NaCl
- Bioanalyzer DNA kit or Qubit material for quantification
- MiSeq Reagent Kit v2 Nano