MOLECULAR CHARACTERIZATION OF AN UNKNOWN P-ELEMENT INSERTION IN *DROSOPHILA MELANOGASTER*

Cyndi Chung, Shanchita Ghosh, Sarah Han, Forrest Lai, Theresa Laverty, Carol Li, Chris McGrath, Sajid Mehmood, Brett Mensinger, Mona Sadegh, Tina Varghese, Chrys Webb Woodbury

Advisor: Dr. Heather Cook
Assistant: Kirsten Ruch

ABSTRACT

The goal of this team project was to discover the unknown location of a genetically engineered transposon that is inserted in the genome of *Drosophila melanogaster*. If the transposon is inserted in the middle of a gene, the transposon may disrupt it, resulting in a homozygous lethal phenotype. The change in phenotype could provide valuable information regarding the gene's function. Before starting the experiment, **P{lacW}**, a transposon that contains a plasmid sequence was integrated into the fruit flies' genome. Using the plasmid rescue method, we isolated the DNA, cleaved it by means of restriction enzymes, promoted intramolecular ligation, and transformed the genetic material into *Escherichia coli (E. coli)*. In this final stage, the experiment was discontinued because there was no growth of *E. coli* containing the mutated DNA. Despite our inability to complete the experiment, we gained valuable lab experience and knowledge about plasmid rescue and its role in mutagenesis.

INTRODUCTION

Genes control all of the biological processes in organisms, ranging from reproduction to the everyday functions of organs and tissues. They are made up of deoxyribonucleic acid (DNA), which is composed of four different subunits called nucleotides or bases. The sequence of the bases codes for the proteins which govern the activities of cells in an organism. In order to better understand how specific genes lead to specific phenotypes, scientists began researching and mapping genomic sequences early in the twentieth century. [1]

The Human Genome Project (HGP) was a mass attempt to successfully sequence and understand all of the genes in the human genome. In the process, scientists hoped to develop new, more efficient methods for DNA sequencing that could be utilized by public industries. Robert Sinsheimer, the chancellor of the University of California in Santa Cruz, conceived the idea of creating the Human Genome Project in 1984. The project was initiated in 1986 by Charles DeLisi, the director of the U.S. Department of Energy's Health and Environmental Research Programs. [2] In 1990, this project was formally funded [3] as a joint effort between numerous countries, including the United States, France, Great Britain, Germany and Japan. [2]

The DNA that was analyzed came from various volunteers who donated small samples of blood and tissue. [4] The final sequencing of the entire human genome was completed in April of 2003, two years earlier than expected. In October of 2004, researchers approximated that the human genome contains twenty to twenty-five thousand genes, which was much lower than the original estimate of a hundred thousand. By mapping and sequencing the human genome, scientists can now study the relationship between a person's genes and a particular inherited disease or trait.

In addition to the mapping of the human genome, numerous companies and laboratories expanded their efforts in order to sequence the genomes of other biologically relevant organisms. Celera, a biomedical diagnostic company, investigated *Drosophila melanogaster*, the common fruit fly. By the fall of 1999, all one hundred sixty-five million bases of the fruit fly genome had been sequenced. [2] Knowing the sequence of the genome gives us a better understand of how closely humans are related to *Drosophila.* This project has shown that humans share a significant number of genes with *Drosophila*. Through mutagenesis, scientists can further locate and study mutations within the fly's genomic DNA and examine its biological relevance to humans. [6] The sequence of *Drosophila* was particularly helpful for our team project because we were able to know the sequence of the gene in which the transposon was inserted.

Drosophila Melanogaster

Drosophila melanogaster is a three millimeter long fruit fly that has been one of the most valuable model organisms in the field of biology for almost a century. [5] Geneticists use the fruit fly as a model organism because many of the fruit flies' genes are similar to the genes in humans. By studying *Drosophila's* genes, we can hopefully learn more about human genes. The *Drosophila* genome contains approximately 14,000 genes and four pairs of chromosomes. These are the X/Y sex chromosomes and autosomes two, three, and four. [6] Of the known proteins, sixty-one percent of those in the fly have similar amino acid sequences to those in humans. [5] Also, fifty percent of the fly's protein sequences have mammalian analogues. [6]

Figure 1: Physical Characteristics of *Drosophila melanogaster* [6]

As shown in Figure 1, *Drosophila* have red eyes, a yellow-brown body color, and black rings on their abdomen. The eye of the fruit fly contains eight hundred unit eyes each containing eight photoreceptor cells, support cells, pigment cells, and a cornea. [6]

The life cycle of the *Drosophila* is relatively short, which is one of the reasons why scientists choose to work with fruit flies. At twenty-five degrees Celsius, the flies live only two weeks. [6] The female fly can lay up to five hundred eggs in ten days. It takes about twenty four hours for the embryo to develop and hatch into larvae, which grow for five days and molt after one, two, and four days. These three molting periods are known as instars. Two days after the third instar, the larvae molt one more time to form pupa. The fly larvae feed off of microorganisms and sugar in fruit, and after a five day metamorphosis, the adult flies emerge. [5] Twelve hours after emerging as an adult, the female fruit fly is ready to mate. [6]

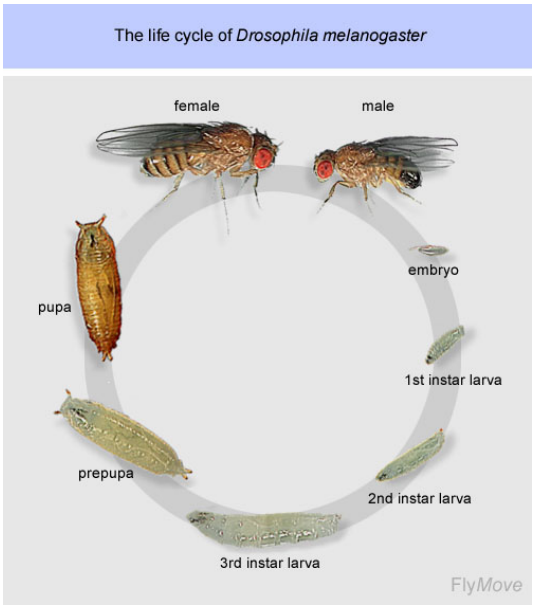

Figure 2: Life Cycle of *Drosophila melanogaster* [7]

Drosophila melanogaster is used regularly in genetics and developmental biology for many reasons. It is small, inexpensive, easy to grow in a laboratory, and has a short life cycle. These factors were significant for our team project because we had a limited amount of time, space, and money to complete our experiment. [5] In terms of ethics, *Drosophila* is an ideal organism to use in scientific research because there is no opposition to using a fruit fly for genetic engineering. Moreover, *Drosophila* is currently being used as a model for many contemporary diseases, including Parkinson's disease and Huntington's disease. [6]

Mutagenesis: Discovering Gene Functions

Although *Drosophila's* genome was sequenced in 2000, very little is known about the majority of gene functions. Approximately twenty percent of the genes have been described in the scientific literature, but only ten percent of the genes have actually been scientifically analyzed. Thus, researchers are working to explore and identify the functions of the other ninety percent of the genome. One way scientists can determine the functions of the unknown genes is by genetic analysis using mutagenesis. [5]

Mutagenesis is the process of producing genetic mutations, alterations in the sequence of DNA. [5] This process is used to analyze the function of a gene. If the mutation causes a change in a gene, this can change the function of a protein. This in turn can affect the phenotype of an organism. By observing these changes, scientists can infer the potential function of the mutated gene. For example, when looking for genes required for brain development, one would mutate many flies and then search for changes in the brain configuration. There are different types of mutagenesis including spontaneous mutagenesis, induced mutagenesis like site-directed and insertional mutagenesis [8]

Spontaneous mutagenesis refers to DNA damage which results from naturally occurring alterations in the DNA. Most are thought to be produced from DNA replication errors and spontaneous damage to DNA. [8] However, some mutations can be induced intentionally on DNA for research; for instance, in our project we used flies whose genome had been intentionally mutated.

Induced mutations come in many forms and are caused by mutagens. [9] Various types of radiation or chemical exposure can alter DNA and create mutations. X-ray exposure, for instance, can cause changes in the DNA and produce dramatic genetic rearrangements, such as deletions, insertions, and translocations. Ultraviolet radiation can also induce the joining of adjacent pyrimidine bases, resulting in a point mutation. Additionally, chemical exposure can induce mutations if the chemicals react with the DNA. [8]

In order to induce specific, predetermined alterations into a cloned gene, site-directed mutagenesis is utilized, cloning and altering specific genes. As a result, the amino acid sequence that specifically corresponds to the altered gene will be changed. Previously, protein biochemists used chemical methods to alter amino acids in the proteins they wanted to study in order to observe the effects of these changes. However, the chemical methods used were imprecise and the researchers were unsure how the amino acid sequence had been altered. Thus, site-directed mutagenesis became more popular among scientists because it was precise and exact. [10]

Drosophila's genome had been altered for our team project by using transposable DNA elements, called P-elements, in order to integrate mutations into the genomic DNA

of *Drosophila*. While site-directed mutagenesis enables one to know the location of the mutation, insertional mutagenesis requires one to test the location of the insertion site. [8]

Insertional Mutagenesis

Insertional mutagenesis was performed on our fruit flies before the start of our experiment using P-elements. Transposons are sometimes referred to as "jumping genes" because they can move within a genome. [9] The simplest type of transposon is the insertion sequence, which contains sequences only for the proteins needed for the movement of that transposon. Transposase is the group of proteins that catalyze the movement of transposons [10], but other molecules such as DNA polymerase are often needed from the cell for replication. More complicated transposons, called composite elements, have additional genes, such as a gene for resistance to a certain antibiotic. Some transposons remain in their original spots after being inserted into a new site, while others are removed from the original site. [9]

Transposons are an extremely useful tool for mutagenesis because they can be inserted into a genome. A transposon can be inserted into a gene and disrupt the structure of a protein, or the transposon can insert into a regulatory region and affect the level of protein expression. Both of these can cause a change in phenotype. By looking at phenotypic changes, scientists can predict what role the mutated gene plays. [5]

The transposon that was inserted into our fly line, P{lacW}, contained two key elements: the *white* gene and the plasmid, pBR322. The *white gene,* which codes for red eyes, serves as a visual marker to distinguish which flies have been mutated by the transposon. pBR322 contains a gene for ampicillin resistance.

Being homozygous for the mutant form of the gene causes lethality. Further investigation into the cause of death could reveal the exact function of the gene. Once the sequence of the gene and the putative protein is identified, scientists can compare these sequences to protein sequences in known databases. If the protein sequence is similar to a protein sequence of a known function, scientists can infer the unknown function of the protein.

Plasmid Rescue Method Overview

The plasmid rescue method was used to identify the genomic location of the P{lacW} transposon. We utilized the plasmid rescue method to locate the transposon in the genomic DNA of the *Drosophila melanogaster*. This procedure identifies the insertion site of a transposon by recovering the genomic DNA sequences adjacent to the transposon. [5]

In the first step of the plasmid rescue method, we isolated the genomic DNA of *Drosophila*, which has one P{lacW} insertion in its genome. We cleaved, or cut, the genomic DNA with restriction enzymes. We then performed ligation on the digested DNA under dilute conditions to allow for intramolecular ligation reactions so that the

previously linear DNA would become circular. The next step was transformation of the ligated DNA into *E. coli* bacteria, which was then cultured on ampicillin-coated plates. [5] This ensured that only the bacteria with the proper genomic DNA segment would be able to reproduce.

METHODS AND PROCEDURES

Isolation of Genomic DNA

Genomic DNA was isolated from twenty five flies as described in Practical Uses in Cell and Molecular Biology. [11] The flies were ground in 100 µl grinding buffer (5% sucrose; 80mM NaCl; 100 mM Tris, ph 8.5; 0.5% SDS; 50 mM EDTA), and the DNA was precipitated with potassium acetate and isopropanol. The pellet was resuspended in 100 µl TE buffer with RNase. Then, 2.5 µl proteinase K was added to eliminate proteins from the DNA sample. To remove the proteinase K and remaining lipid molecules, a 1:1 mix of phenol:chloroform extract was used. Ethanol precipitation was added to concentrate the nucleic acids. The DNA was resuspended in 25 µl TE buffer. [5]

Restriction Enzyme Digestion of DNA

2 µl of DNA was digested with one of six separate restriction enzymes (BamHI, BglII, EcoRI, PstI, SacII, or XbaI) for 1.5 hours at 37º C. [5]

Ligation

The cut DNA was ligated under dilute conditions as follows: 40 µl of the digested DNA was ligated in 601 µl mixture of 10x ligase buffer, sterile H_2O, and T4 DNA ligase. To concentrate the DNA, the ligation reaction was precipitated with ethanol. The precipitated DNA was resuspended in 20 µl TE buffer. [5]

Transformation

The *E. coli* bacteria (DH5α) were treated with rubidium chloride to make the cells competent. To transform the DNA into *E. coli*, 10 µl purified ligation was added to the 100 µl of $RuCl_2$ bacteria, incubated on ice for 30 min, and heat shocked at 42ºC for 45 sec. The cells were plated on an ampicillin-containing agar medium in order to select the bacteria with the pBR322 plasmid. [5]

RESULTS

After performing the *E. coli* cell transformation, no bacteria grew on any of the six ampicillin-containing agar plates. At this point, we were forced to end our experiment. During the transformation process, four controls were used to check for possible flaws in the procedure: (1) *E. coli* bacteria that had not undergone transformation were plated on ampicillin medium, (2) a digestion reaction without DNA followed by a ligation and an *E. coli* transformation, (3) *E. coli* were transformed with pBluescript

ligation, (4) *E. coli* were transformed with pBluescript that was neither digested nor ligated.

As a result of the first control, no growth occurred on the plate, indicating that ampicillin was effective and that the bacteria without the plasmid were not resistant to the antibiotic ampicillin. In the second control there was no bacterial growth, confirming that enzymes were not contaminated with extraneous DNA during the digestion or ligation. In the third and fourth controls, pBluescript was a control plasmid that contained an ampicillin-resistant gene. Because the pBluescript in the third control was in a ligated, circular form (as shown in Figure 3), it promoted bacterial growth, demonstrating that the bacteria were competent and that the ligation was successful. Finally, in the fourth control, the plasmid was still in a circular form, allowing *E. coli* to grow because of the plasmid's ampicillin-resistant nature. Unfortunately, we should have tested digestion as a separate control in order to ensure that the ligase worked, but we did not do so. The controls were particularly helpful in identifying probable sources of error during isolation, digestion, ligation, or transformation.

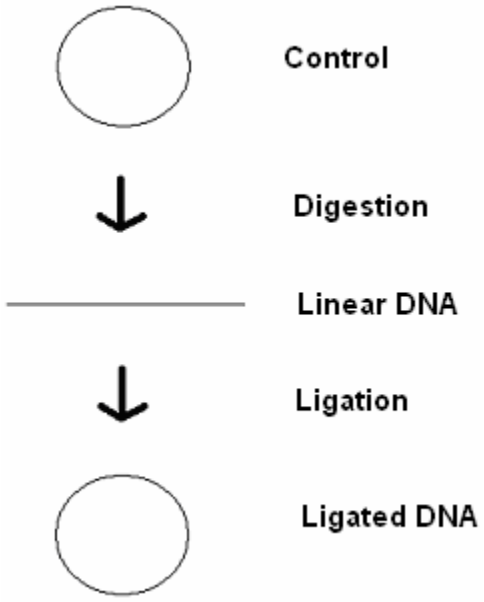

Figure 3: Flowchart demonstrating the third control

Expected Results

If the transformation had been successful, the next step would have been to isolate the plasmid. After recovering pBR322, restriction mapping would have been used to characterize the plasmid. The restriction mapping process uses restriction enzymes to cut DNA into linear fragments. Through gel electrophoresis, these fragments are separated based on size, or number of base pairs. [5] The size of the plasmid fragment would indicate whether or not we obtained both the plasmid and genomic DNA.

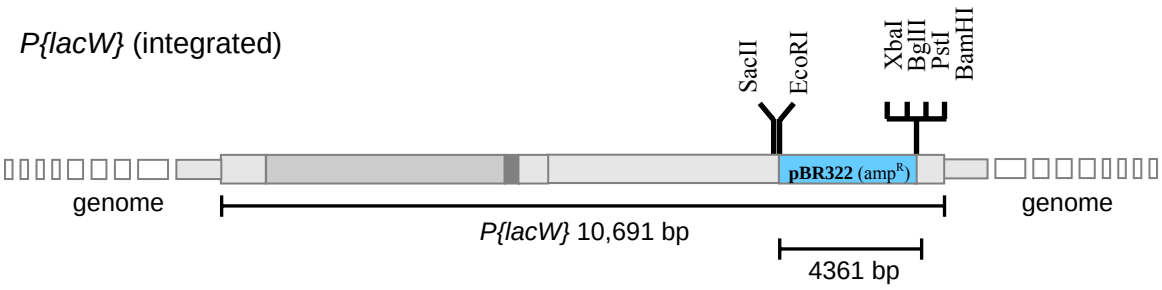

Figure 4: Restriction enzyme cleavage sites on P{lacW} transposon [5]

As seen in Figure 4, because the restriction enzymes EcoRI and SacII cut to the left side, a fragment size that was larger than 4,361 base pairs would indicate the presence of the plasmid and genomic DNA. However, the other restriction enzymes, XbaI, BglII, BamHI, and PstI, cut to the right side. Thus, a fragment size of more than 10,691 base pairs would indicate that we obtained genomic DNA with our transposon. Once each of the plasmids had been characterized by restriction mapping, the experimental part of our project would have ended. During the final stage of our project, the *Drosophila* DNA sequences would have been downloaded from the National Center for Biotechnology Information (NCBI) database [5]. At this point, the sequences of the recovered genomes would have been analyzed using the bioinformatic database.

DISCUSSION

Based upon the controls from the transformation, there are a few potential causes for our results. The results of the controls indicate that the *E. coli* bacteria were competent and that transformation was effective because the *E. coli* transformed using pBluescript were resistant to the ampicillin and grew successfully. Therefore, we propose that the possible error(s) most likely occurred in earlier steps. Four probable causes of error include: (1) not enough DNA was isolated for ligation, (2) improper lab technique, (3) the restriction enzymes cut too far away from the known recognition sites (4) digestion of the isolated genomic DNA was not complete. Improper lab technique is probably not the cause of failure because it is unlikely that six separate lab groups all performed the procedure incorrectly. The most likely cause of failure was incomplete digestion of the *Drosophila* DNA with the six restriction enzymes. A short digestion time did not allow for thorough cleavage of the genome, resulting in DNA fragments that were either too large to be ligated or once ligated, too large to be taken up by the *E. coli*. Quantifying the isolated genomic DNA and allowing for a longer period of digestion may result in successful transformation in future attempts.

CONCLUSION

Although we did not accomplish our original goals, we gained invaluable knowledge and experience in molecular biology research and techniques. We learned how to work with vital biological organisms, *E. coli* and *Drosophila melanogaster*. Also, exploring themes in microbiology, such as the role of transposons and mutations in

insertional mutagenesis, taught us to appreciate the complexity of genetics. We learned how to prepare bacteria for transformation and how to apply ligation, digestion, and transformation to plasmid rescue. We were also taught proper lab techniques, such as micropipetting, bacterial plating, and sterile handling of reagents. Most importantly, we learned that in research there is no guarantee for success and that undesired outcomes can still be significant to the experiment.

REFERENCES

[1] http://en.wikipedia.org/wiki/Genes

[2] http://www.answers.com/topic/human-genome-project

[3] http://www.ornl.gov/sci/techresources/Human_Genome/project/hgp.shtml

[4] http://encarta.msn.com/encyclopedia_761580681_2/Human_Genome_Project.html

[5] Cook H. Molecular characterization of an unknown P-element insertion in *Drosophila melanogaster*, Laboratory Manual, 2005.

[6] http://en.wikipedia.org/wiki/Drosophila_melanogaster

[7] http://www.anatomy.unimelb.edu.au/researchlabs/whitington/img/life_cycle.jpg

[8] http://www.els.net/els/subscriber/article/article_main.asp?sessionid=1865649073522 724048&srch=qsearch&aid=A0000559&term=mutagenesis

[9] Lewin B. Genes VII. Oxford: Oxford University Press; 2000. 14-19 p.

[10] Weaver RF. Molecular Biology. Second edition. New York: The McGraw-Hill Companies, Inc, 2002. 112-115 p.

[11] Goldstein and Fryberg, ed. *Drosophila melanogaster*: Practical Uses in Cell and Molecular Biology. Academic Press, 1994. 90 p.