Research Proposal:

“Using Mutagenesis to Determine the Significance of a Two-Component Regulatory Response Gene in Bacteria Streptococcus *sanguinis”*

1. Introduction

Bacteria are microorganisms that are known to have a wide variety of functions. Some are known to cause living beings harm, and some are known to be beneficial for survival. Over the course of time, it has become imperative to understand the nature of virulent bacteria. In other words, it is essential to that researchers know which genes of the bacteria are the cause of the virulence and which genes allow the bacteria to survive critical conditions. One certain bacteria that is known for being virulent and stress tolerant in its environment is Streptococcus *sanguinis* (S. sanguinis).

Streptococcus *sanguinis* is a gram positive bacteria that is normally found in the human mouth6. Normally, streptococcus bacteria are non-pathogenic to humans and the bacteria develops in the mouth and intestines of humans. However, there are some that cause harm to humans such as *Streptococcus sanguinis,* causing cardiac diseases.5 After infecting the mouth, the strain will then move down the throat and pass through the endothelial cells, reaching its way to the heart. Antibiotics have been used to subdue the bacterial growth; however, the strain shows to be highly tolerant to treatments. S. sanguinis binds to the saliva-coated teeth and it has been revealed that the bacteria becomes an attachment mechanism in which other oral bacteria cells bind to the tooth surface and form plaques. However, it is not known which proteins causes an expression.

From an article, the entire genome of S. sanguinis was determined (FIG 1)5. After sequencing the genome, it was determined that the bacteria had more than 100 transcriptional regulators. Some of these regulators are part of what are called two-component regulatory systems. A major gene that is part of the two-component regulatory system may be the target of interest that allows S. sanguinis to express tolerance. A two component response regulatory system is basically a mechanism in which organisms are able to respond and adapt to the different environmental conditions6. This action occurs through signal transduction where a phosphate group from ATP molecules bind to histidine kinases through environmental stimulus. The phosphorylated kinase then sends the phosphate to a response regulator enzyme which changes its structural conformation and finally leads to the stimulation of gene expression. FIG 2 shows one depiction of the many said mechanism. The figure shows that a signal received from an environment causes the Histidine Kinase to react with ATP. The kinase is phosphorylated and then phosphorylates a response regulator enzyme. This response regulator then causes a change in the gene expression due to a change in the regulator’s conformation which binds to the gene’s binding site.

FIG 1: The genome of S. saguinis. (Ping et al.)

Now, it is understandable that S. *sanguinis* has a two component response regulatory system, and in Ping Xu et al. article, there were 29 genes that contained 14 regulatory systems. However, it is not known when these genes are expressed/under what conditions these genes are expressed. That is the main question for the experiment being performed for this proposal. In order to focus more on a single two component regulatory system, gene 1565 will be the target of interest and to determine if it is significant or not to the bacteria5. Ping et al. stated that gene 1565 may be an important gene to the bacteria under unfavorable conditions, such as being under antiobitcs. A system of using transposons and microplate readers will be used to hopefully determine the significance of the gene.

FIG 2: General Depiction of the 2 Component Regulatory Pathway with Asparagine as a specific conformation change7. (“Signal Transduction”)

1. Experiment

There are many hypothetical genes that are identified which means scientists may not exactly know if the function of the gene is accurate. One useful strategy mentioned by Deutschbauer et al. is to use mutagenesis to predict gene function based on phenotypes2. This experiment involves the use of collecting mutant strains, and in this case, it would be Streptococcus *mutans,* and comparing each strain to another through microarray data. Microarray data is data that contains the measurement of expression levels of genes in a region of a genome. A fitness assay was made for the bacteria by allowing it to be affected by penicillin antibiotics. This would cause the bacteria to express certain phenotypes1. In this case, bacterial population survival is being measured through the amount cell colonies of the bacteria when it is under antibiotic conditions.

1. Bacterial Growth: S. sanguinis is grown in a typical culture with necessary bacterial feed. In addition to this, a transposable element is added to the gene. A transposon is a DNA sequence that is added to the bacteria and it is able to affect the sequence of the bacteria in the genome9. Class II versions of these transposons will be used in which a DNA tag will be inserted to the wild type bacteria strain and through growth, mutants will appear. Class II versions differ from Class I in that Class I affects DNA through RNA reverse transcription and Class II binds directly to target sites in the DNA, without having to use RNA. Transposons can act as mutagens which damage the DNA and then the entire organism will negatively affected. However, the focus is that one gene is affected at a time, which would create the mutants of the bacteria. The transposons will be inserted to the binding site of response regulators to the genes in order to damage the “interaction”, but not as much so that the nucleotide sequence completely kills of the bacteria. The transposon will then, hopefully, disable the gene in which it will affect the gene regulators. Due to the damaged genes, the regulators will not be able to bind to the genes properly and this would mean that there would be no upregulation or downregulation of the gene. In summary, a damaged binding site would cause the response regulatory system to not function properly, causing the targeted gene to become useless under stress.
2. Mutant Assay Comparison: A fitness assay was made for both the wild-type bacteria strain and the mutant bacteria strain. The DNA tag was measured with microarray data that contained sequence complements. This process will be performed multiple times in order to obtain many mutant strains. That way, it will be beneficial to have multiple results that may or may not depict similar gene expression. Each mutant strain is subjected to the same fitness assay condition for the wild-type bacteria. The assays would be created in a microplate reader that measured absorbance. The microplate measures absorbance through the amount of light emitted from the target gene regulator. A system in the microplate illuminates the target gene with an optical system where the regulator will start to emit its own light. A second system in the reader will absorb the emitted light (fluorescence) of the gene and measure it. The amount of light detected is equivalent to the concentration of the gene of interest. Due to the effects of the transposons, the gene would have improper levels of the regulator. The regulators would not be able to bind properly to the gene and there would be a decline in the amount of regulators, meaning less absorbance. The less amount of light would mean that the gene is expressed lightly (due to low levels of regulator concentration). The different bacteria are then grown to form colonies and the amount of colonies are determined for each mutant.
3. Analysis:

In a phenotypic scenario where the bacteria is growing in the penicillin environment, it is possible to see if the bacteria survives or not. If the gene regulator is in low concentrations and the bacteria population is declining, then that means that the gene regulator might be significant. If the gene regulator is in low concentration but the population of the bacteria is stable, then the gene regulator is not necessary for the bacteria’s survival in that condition. In this case, it would mean that the certain gene regulator is not responsible for the bacteria’s survival when the bacteria is under the effects of the penicillin. It could be that another gene regulator has the function to allow the bacteria to survive under antibiotic conditions. This is through microplate reader observations. Using another technique, it would be possible to determine a gene’s significance through a statistical viewpoint.

For each experiment involving the different mutants, it was essential to calculate the consequences for inhibiting a gene involved with the insertion of the transposons while affected with penicillin. By using Deutschbauer et al. technique, gene fitness will be calculated – which is the average mutant strain for each gene2. The gene fitness values reflect the mutant fitness of the mutant strains and the typical gene of the wild type is set to a standard of zero. There will be a strong correlation level of the genes for the mutant strains because the transposon insertions may happen to the same gene.

A test statistic was calculated for each gene level. This test statistic is used to take into account the consistency of the measurements being made. This calculations that are used appear in FIG 3. The number of experiments means the number of trials performed.



FIG 3: The calculation to determine the test statistic for a gene. X is the amount of genes affected, $μ$ is the average number of genes for each experiment, n is the number of experiments, V is the difference in strain fitness, $Ψ$ is the standard deviation of the amount of genes for each experiment2. (Deutschbauer et al.)

The test statistic was transformed into the p value with the number of experiments performed determining the df value. The P values determine the significance of a gene’s fitness for one experiment. The P values range from 0 to 1 and this value determines the level of a gene’s fitness. The higher the p value, the more significant the gene is to the bacteria.

1. Discussion

One great limitation that may be seen for the experiment section is if the microplate reader date can actually be equated to the amount of response regulators detected. Is may be possible that the maximum amount of regulators for the gene is 1 and that the concentration may be similar to if the gene not having the regulators binding to it. The major assumption that is being made in the experiment is that if a regulator does appear to the gene, then that means is a concentration of regulator whereas there would be no concentration at all if there was no binding involved. Another limitation that can be seen from this is that if the transposons do not affect the gene regulation at all! What if all of the results that were procured are all the same and the mutants are no different. It is known that transposons would affect the gene in a negative way9. However, the antibiotic may not work against the mutated gene and maybe the penicillin will not work at all. Since all of this experiment is hypothetical with biological reasoning, a real experiment must be performed in order to see if results are obtained from the methods.

Statistically, if the experiment goes according to plan and the calculations measure up to the way they are intended for, then hopefully a value will be obtained for the Gene 1565. If the P value of the gene is significantly above zero, then it appears that the gene is considered essential to the bacteria’s survival. If it is closer to zero, then it would appear that the gene is considered to be essential and that it may not be an actual two component response regulator or it may have a different function. It may also be apparent that the p values could be greater than just .25, and there may be multiple factors that have not been taken into consideration as well. These factors could influence the way the values are calculated.

However, there is a huge limitation that can be seen for this analysis. That limitation is if the mathematical reasoning is correct. The calculations derived from Adam et al. are created through the basics of statistical concepts. The original t value is different in which it determines the distribution of a population. In this experiment, it measures the distribution of the reactivity between the bacterial genes with penicillin. All of the calculations being made are hypothetical and not many articles, besides Deutschbauer et al, have shown to integrate concepts of distribution into determining gene significance levels. Statistical concepts may have been used to determine population decline or growth, but not for gene significance testing. The absorbance levels made during the microplate reader may not have been interpreted properly for this experiment either. The microplate reader indirectly determines the ratio of flourishing bacteria to the total amount of bacteria being grown. This may not have been the right choice of collecting values for the statistical part of the experiment but it was a simple try. As mentioned before, this concept is all hypothetical.

This experiment may not be the best way to determine if a gene is significant; however, it is an attempt at trying. As researchers, it is useful to observe something at different angles/points of view. So, one can say that this may be a first step in trying to analyze a gene. It may be possible that a certain gene is essential to the bacteria; it may be possible that there are multiple genes that are essential to a bacteria, but how does someone know that is a question that should be attempted to be solved. The mathematical reasoning might be vague, but it is an attempt nevertheless through combining biological experimentation with observational and statistical analysis.

1. References
2. Caufield, Page et al. (2000). “Natural History of Streptococcus sanguinis in the oral cavity of Infants: Evidence for a Discrete Window of Infectivity.” *Infection and Immunity* 68.7: 4018-4023. *NCBI*.
3. Deutschbauer et al. (2014). “Towards an Informative Mutant Phenotype for Every Bacterial Gene.” *Journal of Bacteriology* 196.20: 3642-3655. Web.
4. Lee, Chih and Chun-His Huang. (2012). “Searching for transcription factor binding sites in vector spaces.” *BMC Bioinformatics* 13: 215. *NCBI*.
5. Marx, Patrick et al. (2010). “Identification of genes for small non-coding RNAs that belong to the regulation of the two-component regulatory system CiaRH in Streptococcus.” *BMC Genomics* 11:661. *NCBI.*
6. Ping, Xu et al. (2007). “Genome of the opportunistic pathogen Streptococcus sanguinis.” *Journal of Bacteriology* 189.8:3166-3175. *NCBI*.
7. Seaton, K et al. (2014). “Regulation of competence and gene expression in Streptococcus mutans by the RcrR transcriptional regulator.” *Molecular Oral Microbiology* 10.1: 1-13. *NCBI*. Web.
8. “Signal Transduction”. Faculty of Biology. Ludwing-Maximilians University, Munich.
9. Todeschini, Anne-Laure, Adrein Geroges, and Reiner Veitia. (2014). “Transcription factors: specific DNA binding and specific gene regulation.” *Trends in Genetics* 30.6: 211-219. *ScienceDirect.*
10. Twetman, Svante et al. (2014). “A mariner transposon vector adapted for mutagenesis in oral streptococci.” *MicrobiologyOpen* 3.3: 333-400. *NCBI*.