Research Proposal:

“Determining 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.2 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. From one article, it states that “S. sanguinis possesses a remarkable abundance of putative surface proteins, which may permit it to be a primary colonizer of the oral cavity and agent of streptococcal endocarditis.”2 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 gene causes such mechanism to occur. Therefore, it is imperative to know which genes cause such expression.

From an article, the entire genome of S. sanguinis was determined (FIG 1)2. It was determined that the bacteria had more than 100 transcriptional regulators. 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 conditions1. 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 a general depiction of said mechanism.

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 known that when these genes are expressed; mostly saying under what conditions do the genes in the bacteria become expressed and which 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 bacteria as the article mentions it being2. A statistical concept of determining significance levels will be used to determine whether or not the gene is essential to S. sanguinis.

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

1. Experiment

It is understandably tedious to annotate functions for many hypothetical genes that are identified. One useful strategy mentioned by Adam et al. is to use mutagenesis to predict gene function based on phenotypes8. 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. In a sense, it is a statistical, computational testing. 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 phenotypes2.

1. Bacterial Growth: S. sanguinis is grown in a typical culture with necessary bacterial feed. In addition to this, a transposable element is a DNA sequence that is added to the bacteria and it is able to change its position in the genome. 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.
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 was subjected to the same fitness assay condition for the wild-type bacteria. The assays would be grown in a microplate reader that measured absorbance. The reader would record the amount of inhibitor (penicillin) while observing the decline of growth for the wild type strain. The less amount of the population decline, the more significant the gene is whereas the more decline in population, the less significant the gene is. This reader determines the percentage of trials that show the bacteria surviving.
3. Analysis: 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 Adam et al. technique, gene fitness will be calculated – which is the average mutant strain for each gene8. 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 experiment8. (Adam 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. For further correlation of t values and p values, FIG 4 is shown.



FIG 4: Table entry test in converting T score to P score. The P values are the numbers ranging from .25 to .0005 and the T values are values in the middle. The df value is calculated as n-1 (number of experiments performed minus 1).

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. So, in overall experimental view, an experiment was performed in which 21 experiments were performed in targeting one specific gene. The df would be 20 and the t and p values will follow suit.

1. Discussion

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 experiment. 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 Adam et al, have shown to integrate statistical 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.

1. References
2. 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.
3. Ping, Xu et al. (2007). “Genome of the opportunistic pathogen Streptococcus sanguinis.” *Journal of Bacteriology* 189.8:3166-3175. *NCBI*.
4. 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.*
5. Lee, Chih and Chun-His Huang. (2012). “Searching for transcription factor binding sites in vector spaces.” *BMC Bioinformatics* 13: 215. *NCBI*.
6. 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.*
7. 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*.
8. “Signal Transduction”. Faculty of Biology. Ludwing-Maximilians University, Munich.
9. Adam Deutschabauer et al. “Towards an Informative Mutant Phenotype for Every Bacterial Gene.” *Journal of Bacteriology* 196.20 (2014): 3642-3655. Web.