The Quest for a Missing Sequence in the Genome of Arhodomonas sp. Seminole

GRP#5 Section#1

P. Canaan, S. Shoop, B. Andrews, R. Pokoo, A. Anderson, E. Blair, M. Yang, S. Burchfield, and J. Gauss

ABSTRACT

The bacterium Arhodomonas sp. Seminole was observed in Seminole Co, OK to have traits favorable to eliminating oil pollution, in an eco-friendly way, in soil near well sites. Their traits come from the activities of their enzymes, but our main focus lies with where they came from: the genome, or DNA. Within genomic DNA sequencing, we tried to decode a region of DNA we were unable to identify initially. As a result, we helped further complete our knowledge about the gap but also learned what proteins/enzymes are associated with it. Our group discovered that our fused contiguous DNA codes for transposase, which move "mobile genetic materials…to another genomic position"[1]. This is important in identifying as transposases have a direct influence on transposons, which "can be widely utilized for the creation of random mutants, which might be exploited and applied further for the…functions of genes" [1] or can cause "preparatory genetic modification"[1]and cause genetic change.

INTRODUCTION

The bacterium Arhodomonas sp. Seminole was observed in Seminole Co, OK to be aerobic and halophilic, enriched from salty, crude-oil-impacted soil. Halophilic bacteria are highly favorable because they all have the potential to "degrade oil hydrocarbons up to quite high salt concentrations"[2] and be "intrinsically stable"[3]. Our species specifically has been proven, in recent studies, to have a "cluster of genes"[4] involved in "the natural attenuation of hydrocarbon-impacted hypersaline environments" because its enzymes breakdown "benzene or toluene to acetyl coenzyme A (acetyl-CoA) and pyruvate"[4]. These enzymes essentially break down the chemicals to what is needed for aerobic transpiration, specifically the Kreb's cycle, which gives the cell chemical energy needed for function and reproduction. When samples of Arhodomonas sp. Seminole bacteria were enriched to make a "pure culture of salt-loving, aromatic hydrocarbon-metabolizing bacteria"[5] and sequenced, the computer system did not analyze segments between "750 pieces"[5] of contiguous segments of DNA. Our group was presented with the task of analyzing one of these unidentified segments in a collaborative effort with other groups to uncover more of the genome. In our study, our analysis involved forward and reverse primers, PCR, comparisons to relatable proteins, and DNA sequencing in order to identify the unidentified nucleotide bases. Our objective was to accurately identify these nucleotides bases in order to see, we hypothesized, if "these two contigs (the tail of one and the head of another) might actually be joined together on the chromosome of this bacterium"[6].

MATERIALS AND METHODS

Materials:

P20 Micropipettor	10 μl ddNTP
Agarose Gel	5 μl Arhodomonas sp. Seminole gDNA
Electrophoresis Chamber/Thermocycler	2 μl F-primer
70 μl dH20	2 μl R-primer
10 μl Taq buffer	1 μl Taq Polymerase

Methods:

We first used the blastx program to identify if the two contigs. (contiguous pieces of DNA) were adjacent by comparing it to a similar, known protein. We isolated 100 bases from both ends of the sequence gap and used blastx and PrimerQuest to create a forward and reverse primer to match the sequence we isolated. Afterwards, we used a P20 micropippetor to add the reagents 70 μl of dH2O, 10 μl of Taq buffer, 10 μl dNTP, 5 μl Arhodomonas sp Seminole gDNA, 2 μl F-primer, 2 μl R-primer, and 1 μl of Taq polymerase to a small tube. After subjecting it to gel electrophoresis, we ran PCR and mutliple rounds of the mixture in the thermocycler to amplify this gap region (make millions of copies) for DNA sequencing in the HBRC. In the HBRC, the DNA sequencing involved "template denaturation, primer annealing and primer extension"[7] and "the incorporation of fluorescently labeled dideoxynucleotides (ddNTPs)"[7] into our PCR product. Our sequence was determined through "high-resolution electrophoretic separation"[7] by putting the bases through a laser which "coupled [the bases] to [a] four-color detection of emission spectra"[8]. Software "is then used to convert the gel image to an inferred base sequence (or read) for each template" [8] while also "improv[ing] the accuracy and completeness of assembly"[9]. We compared our sequence with the fused contigs. in clustalw2 .

RESULTS

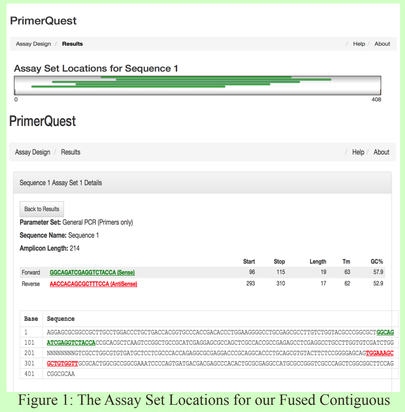

Figure 1: The Assay Set Locations for our Fused Contiguous Sequences of DNA and the Representation of 1st Assay Sequence Set for our Forward and Reverse Primers, measured at http://www.idtdna.com/primerquest/home/index

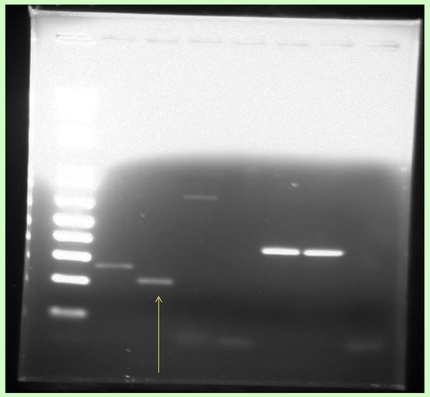

Figure 2: The Picture of the Gel Electrophoresis Results showing the Amplicon (PCR product), indicated by the arrow. PCR amplification was performed with the program which consists of an initial denaturation at 94 C for 30 sec, 30 cycles of denaturation at 94 C for 30 sec, annealing at 53 C for 30 sec, extension at 72 C for 1 min, another 10 min at 72 C for the final extension.

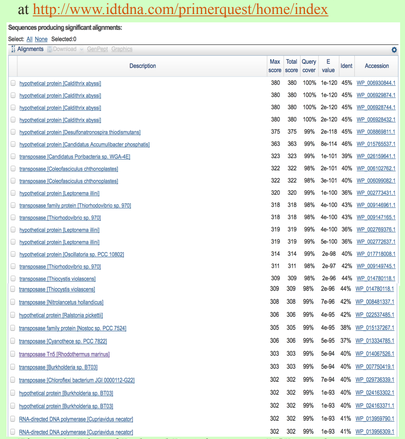

Figure 3: List of Related Proteins to our PCR product; Transposase is most prevalent. Measured at http://blast.be-md.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

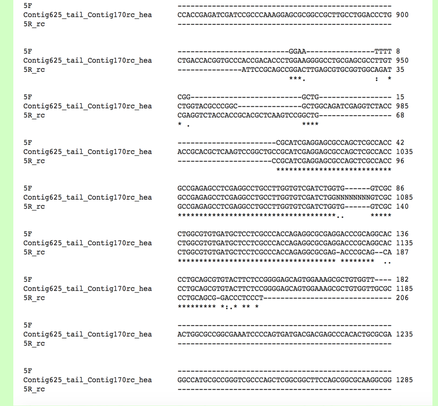

Figure 4: DNA Sequence Product as compared with our Fused Contiguous Sequences of DNA (Contigs.) Measured at http://www.ebi.ac.uk/Tools/msa/clustalw2/

Our results match our expected results in terms of methodology and our end result. However, experimental error was not avoided as we had complications with DNA sequencing. Our PCR product was faint and so was hard for the computers to detect the chain of our DNA sequence product. The process of DNA sequencing had to be done twice through the computer before our analysis was complete. There are too many ways our initial PCR product could have been faint and are unnecessary to address. The forward and reverse primers in Figure 1 were essential for the PCR of our sequence as shown in Figure 2. The analysis of our related protein, shown as transposase in Figure 3, was an additional way of achieving our experimental objectives. Our final figure, Figure 4, illustrates the end of our analysis, trying to see if our fused contigs. matched with our PCR product through DNA analysis. Our data, then, not only supports our expectations but also supports scientific theory and principles, that DNA is connected all the way through its genome, unless broken by some unforeseen mutation or environmental condition. The strengths of our design were using the online programs and the availability to compare our fused contigs. with proteins from a database built from scientists all over the world. Thus, our project was not achieved by narrow means. The weaknesses of our design were the multitude of possible errors concerning the preparation and process of PCR.

DISCUSSION

Our findings and observations with Arhodomonas sp. Seminole supports our hypothesis that our two contiguous sequences of DNA were together as the head of one met with the tail of the other. This is important because we discovered that our contigs. code for a transposase, the vehicle of which transposons are moved from one place of DNA to the other and is directly involved in genetic variation from generation to generation. Our results indicate that we were successful in finding the forward and reverse primers necessary to perform PCR, find a relatable protein, and see for certain whether our hypothesis was supported or unsupported with the DNA sequencing. This is important for the advancement of understanding not only the genetic characteristics of the bacterium but also for increasing our knowledge about its practical use in the real world. Because this bacterium lives off of salt and toxic chemicals, especially runoff from oil well sites, it provides a sound, ecological approach to cleaning up active oil well sites or derelict well sites. Once we are able to get a hold of the whole genome, we can process this species at a mass level to help with this bioremediation because cleaning up oil well sites is a very complicated process and sometimes is unfeasible to remedy. This means that while we are still dependent on oil reserves, we are preserving the integrity of the land: water quality, soil quality, and agriculture. Similar projects to ours include PCR used for other purposes than ours. It can help amplify a desired trait from an organism or help understand the quality of those traits under further experimentation. Of course, Arhodomonas sp. Seminole is not a universal cure to bioremediation of oil well sites; there are other factors which complicate this process. Factors such as soil quality, geographical and biogeographical location, density of bacteria, growth, natural selection, sustainability, time, and competition have short term and long term consequences to all affected by its ecology. Our bacteria has the potential to remedy these oil well sites but at the cost of years and perhaps decades to fully renew the site back to normal. The question is, too, what will happen to the bacteria when the bioremedial process is complete? How should companies/scientists monitor saline levels of the bacteria's environment? How should this process be carried out when there may be ethical or legal barriers to cross and how long should these activities last? Our group suggests a comprehensive plan between companies, scientific institutions, and judicial parties to make our bacterium successfully implemented in old well sites, where desired. Therefore, the significance of our research lies with the better understanding of the bacterium's genome for future analysis and replication as well as picking out desirable traits for future generations of the bacteria. Our research implies these outcomes as well as trying to make this bacteria as biofriendly as possible and effective for practical, industrial use. Thus, future research could include analyzing the enzymes and proteins that have the potential to benefit the bacterium's environment or which ones can negatively impact its environment. Other research could include comparing bacteria to ours and seeing how they affect one another in a colonized setting, to see if they are compatible in a single, controlled environment. For our bacterium, the quest has only begun.

REFERENCES

1. Choi, Kyoung-Hee, and Kang-Ju Kim. "Applications of Transposon-Based Gene Delivery System in Bacteria." *Journal of Microbiology and Biotechnology* 19.3 (2009): 217. *The Korean Society for Microbiology and Biotechnology*. 23 Jan. 2009. Web. 9 Nov. 2014. <http://www.jmb.or.kr/journal/download.php?Filedir=../submission/Journal/019/&num=4719>.
2. Oren, Aharon. "Chapter 11. Biotechnological Applications and Potentials of Halophilic Microorganisms." *Halophilic Microorganisms and Their Environments*. Dordrecht: Kluwer Academic, 2002. 369. Print.
3. Gunde-Cimerman, Nina, Aharon Oren, and Ana Plemenitaš. "Halophilic Archaea and Bacteria as a Source of Extracellular Hydrolytic Enzymes." *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*. Dordrecht: Springer, 2005. 339. Print.
4. Dalvi, S., Nicholson, C., Najar, F., Roe, B., Canaan, P., Hartson, S., & Fathepure, B. (2014). Arhodomonas sp. Strain Seminole and Its Genetic Potential To Degrade Aromatic Compounds under High-Salinity Conditions. *Applied and Environmental Microbiology (AEM), 80*(21), 6664-6664. Retrieved November 10, 2014, from PubMed.
5. Canaan, P. "The Project Bacterium: Arhodomonas Sp. Seminole." *OC-Oklahoma State University*. Web. 11 Nov. 2014. <https://oc.okstate.edu/d2l/le/content/1113489/viewContent/4088873/View>.
6. Canaan, Patricia. *Identify and Examine the Sequence Gap Using Alignment to a Related Protein*. 2 Sept. 2014. Digital Notebook exercise for Arhodomonas sp. Seminole. Oklahoma State University, Stillwater
7. Shendure, Jay, and Hanlee Ji. "Next-generation DNA Sequencing." *Nature Biotechnology* 26.10 (2008): 1135. *ProQuest*. Web. 10 Nov. 2014. <http://search.proquest.com/docview/222237350?pq-origsite=summon>.
8. Ewing, B., L. Hillier, M. C. Wendl, and P. Green. "Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment." *Genome Research* 8.3 (1998): 175. *PubMed*. Web. 10 Nov. 2014. <http://www.ncbi.nlm.nih.gov/pubmed/9521921>.
9. Ewing, B., and P. Green. "Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities." *Genome Research* 8.3 (1998): 175-85. *PubMed*. Web. 10 Nov. 2014. <http://genome.cshlp.org/content/8/3/186.full.pdf+html>.