Identifying a gap using DNA Sequencing in an Arhodomonas Related Protein

Group 7
Section 001

Patricia Canaan Ph.D., Robert Pokoo, Shanell Shoop, Cody Price, Madeline West, MacKenzie Wedman, and Carlee Luttrell

ABSTRACT

Arhodomonas sp. Seminole is a bacterium found in soil around oil rigs and other high-salt conditions. Arhodomonas' ability to adapt to such conditions is intriguing and should be better understood. We are attempting to solve the bacterium's genome for a better grasp on how Arhodomonas survives in such extreme conditions. We focused on one small region of DNA and replicated it using a Polymerase Chain Reaction (PCR). After completing the PCR we received a reverse sequence that aligned with the contig upstream of our gap. From this information we were able identify that the contig and sequence match up. This is a step in the right direction. If we were to go back and redesign the primers we would most certainly sequence the gap with both a forward and reverse DNA sequence. Allowing us to close yet another gap in the genome of Arhodomonas sp. Seminole.

INTRODUCTION

When the identification of Arhodomonas sp. Seminole's genomic sequence first began there were millions of broken DNA contigs, we have now progressed to approximately 750 gaps remaining in the DNA sequence. Arhodomonas sp. Seminole is a bacterium that can survive in extremely high salinity, especially around oil rigs where the bacterium breaks down crude oil in the soil. We are attempting to solve the bacteriums genome for a better grasp on how Arhodomonas thrives in such extreme conditions so we can attempt to produce an artificial equivalent to fight the environmentally harmful effects of oil. We hypothesize that we can align a related protein sequence with two Arhodomonas artificially fused contigs and, after PCR and DNA sequencing, we will have a new forward and reverse sequence that will contain the sequence to our gap between the two original contigs. Within the current study, we were able to successfully identify our related protein as well as make a new sequence with our designed primers. The issue lied within our primers. The reverse primer did its job, just not over the gap and there was not a forward primer product. The important part was that we saw the primer worked just too far upstream on the contig. This can be fixed by redesigning the primers and allowing more knowledgeable people to run it. There were several occasions where, as a group, we felt lost but Dr. Canaan did a great job explaining such complex ideas to us.

MATERIALS AND METHODS

In order to find what the missing gap of DNA was composed of, we ran a PCR on the piece of DNA, then ran a gel of the PCR product. First, we checked with the Blastx program to ensure that the two pieces of DNA were adjacent to eachother and determined which was the head and tail contig. In order to run the PCRs, we used the PrimerQuest program on the IDT website to design our forward and reverse primers (http://idtdna.com/primerquest/home/index). For the PCR, we used the following materials:

- 50 ng of DNA,
- 0.6 µM each of the forward and reverse primers (derived from the PrimerQuest program)
- 200 µM of PCR nucleotide mix (Fisher Bioreagents, Fisher Scientific, Pittsburgh, PA)
- 1.75 mM MgCl2
- 2.5 units Taq polymerase in Buffer A (M1865, Promega Chemicals, Madison, WI)

Using a pipette, we measured out the specific quantities of each and put them into a small tube which we then put into the PCR machine.
After we ran the PCR, we then pipetted some of the product into a gel and ran it.

RESULTS

We began with two full length contig sequences that ranged from 23,084 to 45,172 bases. These were compared to our fused contig on the following webpage: http://blast.ncbi.nlm.nih.gov/Blast.cgi . Results from the blastx show a partial alignment on the head end of our protein(Table 1). Then based on the alignment of our of contigs to the related protein we determined the contigs were fused(Table 2). According to blastx there was only a gap of 1%. We consider our first step a great success.

Table 1

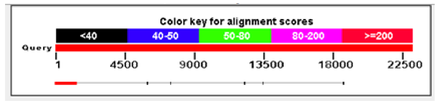
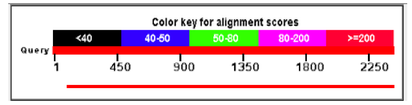


Table 2



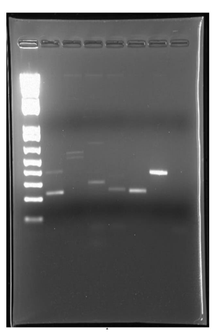
We worked toward verifying the gap in the DNA sequence by designing forward and reverse primers to be used for PCR amplification. The ideal primer consisted of being centered, intermediate in size, and begin approximately 100 bases away from the gap. Our primer was described exactly as such. We hypothesized that our primer would work because of how well it fit all the qualifications.

Parameter Set: General PCR (Primers only)
Sequence Name: Sequence 1
Amplicon Length: 261

		Start	Stop	Length	Tm	GC%
Forward	AGGCCATTAAGATGGTCTACAG (Sense)	109	131	22	62	45.5
Reverse	GCTCCTTGAGGTGGTAGC (AntiSense)	352	370	18	61	61.1

Base	Sequence
1	TCGGCGAGACCGTCGTGCAGGGCTCGGTGAACCCCGACGAGTTCTACGTGCACAAGCCCACCCTGGAGGCCGGGCGCCCGGCGATCCTACGCCGCATCCT
101	CGGGGGCAAGGCCATTAAGATGGTCTACAGCGACGAAACCGGTCACGGCAAGTCGGTTCAGACCGTCGAGGTCGACGACGCCGACCGCAACCGCTACTCC
201	CTCACCGACGAGGACGTGGAGAGCCTCGCCCGCCAGGCGCTCNNNNNNTGGACATCGAATGGGGCGAAGGACGGCGAGACCGGCCGCCTGTATATCCTCC
301	AGGCCCGCCCCGAGACGGTGAAAAGCCGTGACGGCAGCCAGATCCTGCAGCGCTACCACCTCAAGGAGCGCGGGCGCGTGGTGGCCGCCGGCCGCCGCAT
401	CGGCCAGCGTATCGGCGCCGGCCGGGCC

The PCR was prepared and ran through a thermocycler in order for the DNA to be synthesized. The PCR was denatured, then annealed at 53 degrees Celsius. While some groups temperatures had to be adjusted the actual PCR mixture was left the same.
This was followed with an agarose gel which was successful after the first try. According to the gel, the PCR is 300 base pairs in length. Even though we had a successful PCR in the gel, our first attempt at sequencing failed. After another attempt with slight adjustments we were able to recover a reverse sequence, but no forward. This was expected because even though the PCR showed up on the agarose, it was a fairly weak band. We had predicted a low success rate.
We ran yet another blast to identify the related protein. Phosphoenolpyruvate synthase. After this discovery we briefly looked into this protein to better understand the connection.



DNA Analysis:
We ran the reverse sequence with the fused contig in hopes of a complete alignment. Our result was not so cooperative. We found that our reverse primer aligned with a portion of the contig upstream of the gap. This gives us a new alignment and proof that this pair can match up and could potentially align completely with some redesigning of primers. Experimentally we took the reverse compliment of the primer to get the correct strand then manually searched the sequence for the primer.

DISCUSSION

Although we only had a reverse primer, we analyzed the DNA by running it through the clustalw website with our fused contig and found that the reverse primer matched up with the contig upstream of the gap. Because of this new match we know there's a match between Phosphoenolpyruvate synthase and Arhodomonas sp. Seminole. We can redesign the primers to produce a forward and reverse and find the DNA sequence to fill the gap. Filling this given gap will get us one step closer to solving the Arhodomonas sp. Seminole genome.

The findings from this research are tremendously important. If we are able to use PCR and DNA analysis to fill the gaps of this DNA sequence we are one step closer to breeding a bacteria that we can use to reduce or eliminate the environmental waste coming from oil rigs. This would allow for the continued practice of drilling for oil with exponentially less environmental impact, creating a much cleaner energy source.

Reasons of importance:

- Full Arhodomonas sp. Seminole DNA sequence understanding
- Better understanding of bacterium extreme abilities to adapt
- Reduce the crude oil waste exponentially from oil rigs
- Begin studying ways to use bacteria as an environmental cleaner

REFERENCES

• Dalvi, S., Nicholson, C., Najar, F., Roe, Ba., Canaan, P., Hartson, SD., Fathepure, BZ., Arhodomonas sp. Strain Seminole and Its Genetic Potential To Degrade Aromatic Compounds under High-Salinity Conditions. *National Center for Biotechnology Informations.* (Nov. 2014).

• Website: blast.ncbi.nlm.nlh.gov/Blast.cgl

• Website: http://www.ebi.ac.uk/Tools/msa/clustalw2/