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 minating 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 hydrocarbonimpacted 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 \leq 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 hydrocarbonmetabolizing 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 Agarose Gel Electrophoresis Chamber/Thermocycler $2 \mu l$ F-primer 70 µl dH20 $10 \,\mu$ l Taq buffer

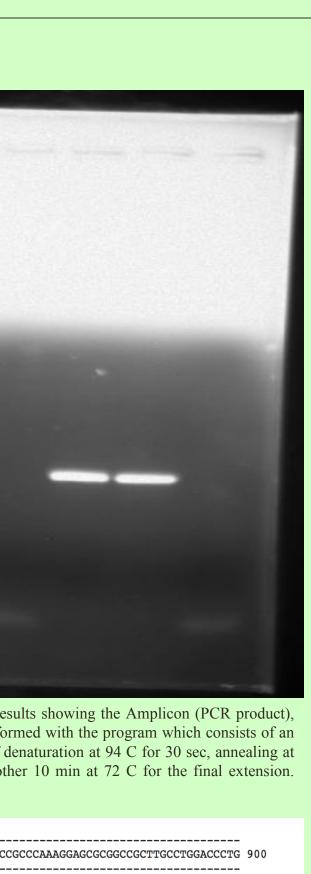
 $10 \,\mu l \,ddNTP$ $5 \,\mu$ l Arhodomonas sp. Seminole gDNA $2 \mu l R$ -primer $1 \mu l$ Taq Polymerase

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 μ l of dH2O, 10 μ l of Taq buffer, 10 μ l dNTP, 5 μ l Arhodomonas sp Seminole gDNA, 2 μ l Fprimer, 2 μ l R-primer, and 1 μ l of Taq polymerase to a small tube. After subjecting it to gel electrophoresis, we ran PCR and multiple 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.

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

PrimerQuest Assay Design / Results Assay Set Locations for Sequence 1	/ Help / About		
	/ Help / About		
ssay Set Locations for Sequence 1			
	408		
imerQuest			
Assay Design / Results	/ Help / About		
Sequence 1 Assay Set 1 Details			
Back to Results			
Parameter Set: General PCR (Primers only)			
Sequence Name: Sequence 1 Amplicon Length: 214			
	Start Stop Length Tm GC%		
Forward GGCAGATCGAGGTCTACCA (Sense) Reverse AACCACAGCGCTTTCCA (AntiSense)	96 115 19 63 57.9 293 310 17 62 52.9		
		\uparrow	
Base Sequence			
	AGGAGCGCCAGCTCGCCACCGCCGAGAGCCTCGAGGCCTGCCT		
301 <u>GCTGTGGTT</u> GCGCACTGGCGCCGGCGAAATCCCCAGTGATGAC 401 CGGCGCAA	GACGAGCCCACACTGCGCGAGGCCATGCGCCGGGTCGCCCAGCTCGGCGGCTTCCAG		
401 CGGCGCAA		and the second second second second	
Figure 1. The Assay Set Loc	ations for our Fused Contiguous	Eigung 2: The Distance of the Col El	lo otro a lo o
0	C	Figure 2: The Picture of the Gel El- indicated by the arrow. PCR amplifi	1
Sequences of DNA and the R	epresentation of 1 st Assay	initial denaturation at 94 C for 30 so	
Sequence Set for our Forwar	d and Reverse Primers, measured		5000
1	,	53 C for 30 sec extension at 72 C	for 1 n
ot http://www.idtdpo.com/pri		53 C for 30 sec, extension at 72 C	C for 1 m
at http://www.idtdna.com/pri	merquest/nome/index	53 C for 30 sec, extension at 72 C	C for 1 m
equences producing significant alignments:	merquest/nome/index		C for 1 m
equences producing significant alignments: elect: All None Selected:0	merquest/nome/index	5F Contig625_tail_Contig170rc_hea	C for 1 m
equences producing significant alignments: elect: <u>All None</u> Selected:0	Max Total Query E Ident Accession	5F	
equences producing significant alignments: elect: <u>All None</u> Selected:0 Alignments Bownload GenPept Graphics Description	Max Total Query E score score cover value Ident Accession	5F Contig625_tail_Contig170rc_hea	
equences producing significant alignments: elect: <u>All None</u> Selected:0 Alignments Download <u>GenPept Graphics</u> Description	Max Total Query E Ident Accession	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea	
aquences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Caldithrix abyssi]	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 2e-120 45% WP 00692874.1	5F Contig625_tail_Contig170rc_hea 5R_rc 5F	CCACCGAC
equences producing significant alignments: etect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 1e-120 45% WP_006929874.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928432.1	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea	CCACCGAC
equences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi]	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 2e-120 45% WP 00692874.1	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea	CCACCGAG CTGACCAG CTGACCAG CCGG CTGGTACC
	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 00692844.1 380 380 100% 2e-120 45% WP 006928432.1 380 380 100% 2e-118 45% WP 008869811.1	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea 5R_rc 5F	CCACCGAG CTGACCAG CTGACCAG
aquences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes]	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006928744.1 380 380 100% 2e-120 45% WP 006928744.1 380 380 100% 2e-110 45% WP 006928744.1 380 380 100% 2e-118 45% WP 006928744.1 380 380 100% 2e-118 45% WP 0015765537.1 323 323 99% 1e-101 39% WP 026159641.1 322 322 98% 2e-101 40%	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea 5R_rc 5F SF_Contig625_tail_Contig170rc_hea 5R_rc	CCACCGAC CTGACCAC CTGACCAC CTGGTACC CGAGGTC * .
equences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP 006920844.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 006928874.1 380 380 100% 2e-120 45% WP 006928874.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 006928432.1 363 363 99% 8e-114 46% WP 015765537.1 323 323 99% 1e-101 39% WP 026159641.1	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea	CCACCGAG CTGACCAG CTGACCAG CCGG CTGGTACG CGAGGTC
equences producing significant alignments: elect: All None Selected:0 Alignments Download CenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Caldithrix ab	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 006928744.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 006928432.1 363 363 99% 8e-114 46% WP 015765537.1 323 323 99% 1e-101 39% WP 026159641.1 322 322 98% 2e-101 40% WP 00609082.1	5F Contig625_tail_Contig170rc_hea 5R_rc 5F Contig625_tail_Contig170rc_hea 5R_rc 5F SF_Contig625_tail_Contig170rc_hea 5R_rc	CCACCGAC CTGACCAC CTGACCAC CTGGTACC CGAGGTC * .
equences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illini] transposase family protein [Thiorhodovibrio sp. 970]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP_006920844.1 380 380 100% 1e-120 45% WP_006920874.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_006928432.1 363 363 99% 8e-114 46% WP_015765537.1 363 363 99% 1e-101 39% WP_026159641.1 322 322 98% 2e-101 40% WP_006090082.1 320 320 99% 1e-100 36% WP_002773431.1 318 318 98% 4e-100 43% WP_009146961.1	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCACCGAC CTGACCAC CTGACCAC CTGGTACC CGAGGTC * . ACCGCACC
equences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Leptonema illini] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illini] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970] hypothetical protein [Leptonema illini]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 006928874.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 008928611.1 363 363 99% 8e-114 46% WP 015765537.1 323 323 99% 1e-101 39% WP 006102762.1 322 322 98% 2e-101 40% WP 006209082.1 320 320 99% 1e-100 36% <td><pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre></td> <td>CTGACCAO CTGACCAO CTGACCAO CTGGTACO CGAGGTC * . ACCGCACO</td>	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CTGACCAO CTGACCAO CTGACCAO CTGGTACO CGAGGTC * . ACCGCACO
aquences producing significant alignments: etect: All None Selected:0 Alignments Download GenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illini] transposase family protein [Thiorhodovibrio sp. 970] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP_006920844.1 380 380 100% 1e-120 45% WP_006920874.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_006928432.1 363 363 99% 8e-114 46% WP_015765537.1 363 363 99% 1e-101 39% WP_026159641.1 322 322 98% 2e-101 40% WP_006090082.1 320 320 99% 1e-100 36% WP_002773431.1 318 318 98% 4e-100 43% WP_009146961.1	5F Contig625_tail_Contig170rc_hea	CCACCGAG CCGACCAG CTGACCAG CTGACCAG CTGGTACG CGAGGTC * ACCGCAGAG GCCGAGAG GCCGAGAG
equences producing significant alignments: elect: All None Selected:0 Alignments Download CenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 1e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_006928432.1 363 363 99% 8e-114 46% WP_015765537.1 323 323 99% 1e-101 39% WP_026159641.1 322 322 98% 2e-101 40% WP_006102762.1 322 322 98% 3e-101 40% WP_006102762.1 320 320 99% 1e-100 36% WP_009146961.1 318 318 98% 4e-100 43%	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCGGCCGAGAC CCGGCCGAGAC CCGGCCGAGAC CCGAGGTCC * ACCGCACAC GCCGAGAC GCCGAGAC CCCGAGAC
equences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics phypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illini] transposase [Thiorhodovibrio sp. 970] hypothetical protein [Leptonema illini] hypothetical protein [Cacillatoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 1e-120 45% WP_006929874.1 380 380 100% 2e-120 45% WP_006929874.1 380 380 100% 2e-120 45% WP_006928474.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_008928744.1 363 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-111 46% WP_015765537.1 323 323 99% 1e-101 39% WP_006102762.1 322 322 98% 2e-101 40% WP_009142661.1 320 320 320 99% 1e-100 36% WP_002773431.1 318 318 98% 4e-100	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCACCGAG CCACCGAG CTGACCAG CTGACCAG CTGGTACC CGAGGTC * . ACCGCAGAG GCCGAGAG GCCGAGAG GCCGAGAG CCGGCGTC CTGGCGTC CTGGCGTC
equences producing significant alignments: elect: All None Selected:0 Alignments Download CenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Calditoria sp. PCC 10802] fransposase [Thiorhodovibrio sp. 970] fransposase [Thiocystis violascens] transposase [Thiocystis violascens]	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 006928874.1 380 380 100% 2e-120 45% WP 006928874.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 008928641.1 363 363 99% 8e-114 46% WP 015765537.1 323 323 99% 1e-101 39% WP 026159641.1 322 322 98% 2e-101 40% WP 00609082.1 320 320 99% 1e-100 36% WP 002773431.1 318 318 98% 4e-100 43% WP 009147165.1 319 319 99% 5e-100 36% w	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCACCGAG CCACCGAG CTGACCAG CTGACCAG CTGGTACC CGAGGTC * . ACCGCACC GCCGAGAG GCCGAGAG GCCGAGAG GCCGAGAG CTGGCGTC
aquences producing significant alignments: elect: All None Selected:0 Alignments Download CenPept Graphics bypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Oscillatoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] <	Max Total score Query cover E value Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 1e-120 45% WP_006928874.1 380 380 100% 2e-120 45% WP_0069288744.1 380 380 100% 2e-120 45% WP_006928674.1 380 380 100% 2e-118 45% WP_0015765537.1 323 323 99% 1e-101 39% WP_006102762.1 322 322 98% 3e-101 40% WP_006102762.1 320 320 99% 1e-100 <td< td=""><td><pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF SF Contig625_tail_Contig170rc_hea SR_rc</pre></td><td>CCTGCCGAG</td></td<>	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCTGCCGAG
equences producing significant alignments: elect: All None Selected:0 Alignments Download GenPept Graphics bypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illini] transposase family protein [Thiorhodovibrio sp. 970] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Cosillatoria sp. PCC 10802] transposase [Thiocystis violascens] transposase [Thiocystis violascens] transposase [Nitrolancetus hollandicus] hypothetical protein [Ralstonia pickettii]	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP_006929844.1 380 380 100% 1e-120 45% WP_006929874.1 380 380 100% 2e-120 45% WP_006928874.1 380 380 100% 2e-120 45% WP_006928874.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_00892811.1 363 363 99% 8e-114 46% WP_015765537.1 323 323 99% 1e-101 39% WP_026159641.1 322 322 98% 2e-101 40% WP_006102762.1 320 320 99% 1e-100 36% WP_002773431.1 318 318 98% 4e-100 43% WP_002769376.1 319 319 99% 4e-100 36% WP_002	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCTGGCGTC CTGGCGTC CTGGCGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CCTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC
equences producing significant alignments: elect: All None Selected:0 Alignments Download CenPept Graphics Description hypothetical protein [Caldithrix abyssi] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Leptonema illini] hypothetical protein [Calditoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Thiorystis violascens] transposase [Thiorystis violascens] transposase [Thiorystis violascens] hypothetical protein [Ralstonia pickettii] hypothetical protein [Ralstonia pickettii] transposase family protein [Nostoc sp. PCC 7524] transposase [Cyanothece sp. PCC 7822]	Max Total Cuery E Ident Accession 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 00892841.1 363 363 99% 8e-114 46% WP 015765537.1 323 323 99% 1e-101 39% WP 006102762.1 322 322 322 98% 3e-101 40% WP 009142762.1 320 320 99% 1e-100 36% WP 002773431.1 318 318 98% 4e-100 43% WP 002769376.1 319 319 99% 5e-100 36%	SF Contig625_tail_Contig170rc_hea	CCTGCCGAGAC CCTGCCGAGAC CCGGCGAGAC CCGAGGTC * . ACCGCACC GCCGAGAC GCCGAGAC GCCGAGAC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC
equences producing significant alignments: elect: <u>Ali None</u> Selected:0 Alignments Download CenPept Graphics Description bypothetical protein [Caldithrix abyssi] bypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes] bypothetical protein [Leptonema illin] transposase [Coleofasciculus chthonoplastes] bypothetical protein [Leptonema illin] bypothetical protein [Leptonema illin] bypothetical protein [Leptonema illin] bypothetical protein [Leptonema illin] bypothetical protein [Oscillatoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970] transposase [Inicystis violascens] transposase [Inicystis violascens] transposase [Inicystis violascens] transposase [Nitrolancetus hollandicus] hypothetical protein [Ralstonia pickettii] transposase [Cyanothece sp. PCC 7822] transposase [Cyanothece sp. PCC 7822] transposase Thio flekothermus marinus]	Max Total score Cuery cover E value Ident Accession 380 380 100% 1e-120 45% WP_006928874.1 380 380 100% 1e-120 45% WP_006928874.1 380 380 100% 2e-120 45% WP_006928874.1 380 380 100% 2e-101 40% WP_006102762.1 322 322 98% 2e-101 40% WP_006102762.1 320 320 99% 1e-100 36% WP_002773431.1 318 318 98% 4e-100 43% </td <td>SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF_rc SF Contig625_tail_Contig170rc_hea SF_rc SF Contig625_tail_Contig170rc_hea SR_rc SF</td> <td>CCTGGCGTC CTGGCGTC CTGGCGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CCTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC</td>	SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF_rc SF Contig625_tail_Contig170rc_hea SF_rc SF Contig625_tail_Contig170rc_hea SR_rc SF	CCTGGCGTC CTGGCGTC CTGGCGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CGAGGTC CCTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC
equences producing significant alignments: elect: Ali None Selected:0 Alignments Download CenPept Graphics Description bypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illin] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illin] hypothetical protein [Leptonema illin] hypothetical protein [Leptonema illin] hypothetical protein [Leptonema illin] hypothetical protein [Calditoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970] hypoth	Max Total Cuery E Ident Accession 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 00892841.1 363 363 99% 8e-114 46% WP 015765537.1 323 323 99% 1e-101 39% WP 006102762.1 322 322 322 98% 3e-101 40% WP 009142762.1 320 320 99% 1e-100 36% WP 002773431.1 318 318 98% 4e-100 43% WP 002769376.1 319 319 99% 5e-100 36%	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CTGGCGTQ CTGGCGAGAQ CCGAGGTC? * . ACCGCAGQ GCCGAGAQ GCCGAGAQ GCCGAGAQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ
evences producing significant alignments: elect: <u>Ali None</u> Selected:0 Alignments Download <u>GenPept Graphics</u> <u>Description</u> hypothetical protein [Caldithrix abyssi] hypothetical protein [Candidatus Accumulibacter phosphatis] transposase [Coleofasciculus chthonoplastes] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illin] transposase [Coleofasciculus chthonoplastes] hypothetical protein [Leptonema illin] hypothetical protein [Calditoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Chiorhodi Faladita]	Max score Total cover Cuery cover E value Ident Accession 380 380 100% 1e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_00692874.1 380 380 100% 2e-110 46% WP_015765537.1 323 323 99% 1e-101 39% WP_006102762.1 322 322 98% 3e-101 40% WP_006102762.1 320 320 99% 1e-100 36% WP_002769376.1 318 98% 4e-100 43% WP_002	5F Contig625_tail_Contig170rc_hea SR_rc 5F Contig625_tail_Contig170rc_hea SR_rc 5F Contig625_tail_Contig170rc_hea SR_rc 5F Contig625_tail_Contig170rc_hea SR_rc 5F Contig625_tail_Contig170rc_hea SR_rc 5F Contig625_tail_Contig170rc_hea SR_rc	CTGGCGTQ CTGGCGAGAQ CCGAGGTC? * . ACCGCAGQ GCCGAGAQ GCCGAGAQ GCCGAGAQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ
Vertical protein [Leptonema illin] Proteitical protein [Relstonia picketti] Transposase [Nitrola	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006929874.1 380 380 100% 2e-120 45% WP 00692874.1 380 380 100% 2e-120 45% WP 006928432.1 375 375 99% 2e-118 45% WP 006928432.1 363 363 99% 8e-114 46% WP 005928432.1 322 322 98% 2e-101 49% WP 00692821.1 322 322 322 98% 3e-101 40% WP 002773431.1 318 318 98% 4e-100 36% WP 00277637.1 319 319 99% 5e-100 36% WP 00277637.1 319 319 99% 2e-96 44% WP 00147801.1 309 309 98% 2e-97 42%	5F Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea 5F,rc SF Contig625_tail_Contig170rc_hea 5F,rc SF,rc 5F Contig625_tail_Contig170rc_hea 5F,rc SF,rc 5F Contig625_tail_Contig170rc_hea 5F,rc SF,rc 5F Contig625_tail_Contig170rc_hea 5F,rc SF,rc 5F Contig625_tail_Contig170rc_hea 5F,rc SF Contig625_tail_Contig170rc_hea 5F Contig625_tail_Contig170rc_hea	CTGGCGTQ CTGGCGAGAQ CCGAGGTC? * . ACCGCAGQ GCCGAGAQ GCCGAGAQ GCCGAGAQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ CTGGCGTQ
cuences producing significant alignments: elect: All None Selected:0 Alignments Download CenPept Graphics Inypothetical protein [Caldithrix abyssi] hypothetical protein [Leptonema illin] hypothetical protein [Leptonema illin] hypothetical protein [Leptonema illin] hypothetical protein [Leptonema illin] hypothetical protein [Caldithrix abyssi] hypothetical protein [Nosloc sp. PCC 70802] transposase [Thiorhodovibrio sp. 970] transposase [Nitrolancetus hollandicus] hypothetical protein [Nosloc sp. PCC 7524] transposase [Chloroflexi bacterium JGl 0000112-G22] hypothetical protein [Burkholderia sp.	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP 006930844.1 380 380 100% 1e-120 45% WP 006920874.1 380 380 100% 2e-120 45% WP 006920874.1 380 380 100% 2e-120 45% WP 006920821.1 375 375 99% 2e-118 45% WP 00692082.1 322 322 98% 2e-101 40% WP 006102762.1 322 322 98% 2e-101 40% WP 009140961.1 322 322 98% 4e-100 36% WP 009140961.1 318 318 98% 4e-100 36% WP 002769376.1 319 319 99% 5e-100 36% WP 002769376.1 319 319 99% 4e-100 36% WP 002769376.1 319 319 99% 4e-100 36% WP 0177	<pre>SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea SR_rc</pre>	CCTGGCGTC CTGGCGTC CTGGCGAGAC CCGAGGTC * . ACCGCAGAC GCCGAGAC GCCGAGAC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC CTGGCGTC
equences producing significant alignments: elect: Ali None Selected:0 Alignments Download CenPept Graphics protein [Caldithrix abyssi] hypothetical protein [Leptonema illin] hypothetical protein [Nosloc sp. PCC 1802] transposase [Thiorystis violascens] transposase [Nitrolancetus hollandicus] hypothetical protein [Ralstonia picketti] transposase [Nitrolancetus hollandicus] hypothetical protein [Ralstonia picketti] transposase [Chloroflexi bacterium JGI 0000112-G22] hypothetical protein [Burkholderia sp. BT03] hypothetical protein [Burkholde	Max Total Query E Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 1e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_00692827.1 323 323 99% 1e-101 39% WP_02615587.1 322 322 322 98% 2e-101 40% WP_00699082.1 320 320 390 98% 4e-100 36% WP_00277637.1 318 318 98% 4e-100 36% WP_00277637.1 319 319 99% 2e-98	SF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rcSF Contig625_tail_Contig170rc_hea SR_rc	CTGGCGAGAQ GCCGAGAQ GCCGAGAQ GCCGAGAQ GCCGAGAQ GCCGAGAQ GCCGAGAQ CTGGCGTQ
equences producing significant alignments: elect: Alj None Selected:0 Alignments Download CenPept Graphics Description bypothetical protein [Caldithrix abyssi] bypothetical protein [Leptonema illin] transposase [Coleofasciculus chthonoplastes] bypothetical protein [Leptonema illin] bypothetical protein [Calditoria sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Inlicystis violascens] transposase [Alitrolancetus hollandicus] bypothetical protein [Ralstonia pickettii] transposase [Chloroflexi bacterium JGI 0000112-G22] bypothetical protein [Burkholderia sp. BT03] bypothetical protein [Burkholderia sp. BT03] <td>Max Total Cuery E Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 2e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_006928432.1 363 363 99% 8e-114 46% WP_00512762.1 322 322 98% 2e-101 40% WP_00699082.1 322 322 92% 1e-101 3% WP_005102762.1 322 322 92% 4e-100 43% WP_008102762.1 320 320 99% 1e-101 3% WP_008102762.1 321 322 322 98% 4e-100 3% WP_008102762.1 321 322 322 98% 4e-100</td> <td>SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF_rcc SF Contig625_tail_Contig170rc_hea SF_rc SF SF SF SF SF <td>CCTGCCGAGA CCTGGCGAGA CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGAC CCGAGAA CCCGCAGAA CCCGCAGAA CCTGGCGTC CTGGCGTC CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG</td></td>	Max Total Cuery E Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 2e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_00692874.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_006928432.1 363 363 99% 8e-114 46% WP_00512762.1 322 322 98% 2e-101 40% WP_00699082.1 322 322 92% 1e-101 3% WP_005102762.1 322 322 92% 4e-100 43% WP_008102762.1 320 320 99% 1e-101 3% WP_008102762.1 321 322 322 98% 4e-100 3% WP_008102762.1 321 322 322 98% 4e-100	SF Contig625_tail_Contig170rc_hea SF_rcc SF Contig625_tail_Contig170rc_hea SF_rc SF SF SF SF SF <td>CCTGCCGAGA CCTGGCGAGA CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGAC CCGAGAA CCCGCAGAA CCCGCAGAA CCTGGCGTC CTGGCGTC CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG</td>	CCTGCCGAGA CCTGGCGAGA CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGTCC CCGAGGAC CCGAGAA CCCGCAGAA CCCGCAGAA CCTGGCGTC CTGGCGTC CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG CCTGCCAGG
Sequences producing significant alignments: Select: All None Selected:0 Alignments Download CenPept Graphics hypothetical protein [Caldithrix abyssi] hypothetical protein [Leptonema illin] hypothetical protein [Caldithrix abyssi] hypothetical protein [Relstonia sp. PCC 10802] transposase [Thiorhodovibrio sp. 970] transposase [Thiorhodovibrio sp. 970] transposase [Nitrolancetus hollandicus] hypothetical protein [Relstonia pickettii] transposase [Cuanothece sp. PCC 7524] transposase [Cuanothece sp. PCC 7524] transposase [Cuanothece sp. PCC 7524] transposase [Cuanothece sp. PCC 7524] hypothetical protein [Burkholderia sp. BT03] hypothetical protein [Burkholderia sp. BT03] hypothetical protein [Burkholderia sp. BT03] RNA-directed DNA polymerase [Cupriavidus necator] Figure 3: List of Reelated Proof Transposase is mosst prevaleer	Max Total Cuery E Ident Accession 380 380 100% 1e-12 45% WP_006930844.1 380 380 100% 1e-12 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928744.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-118 45% WP_006928432.1 322 323 99% 1e-101 39% WP_006928432.1 322 322 322 98% 2e-101 40% WP_00699082.1 320 320 99% 1e-101 3% WP_0027763.1 321 322 322 98% 4e-100 43% WP_0027763.1 318 318 98% 4e-100 3% WP_0027763.1 1 319 319 99% 2e-96 44% WP_0171800.1 1 306 306 99% <td>SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea</td> <td>CCTGCCACG CCTGCACG CCGAGGTC CCGAGGTC CCGAGGTC CCGAGGTC CCGAGGAC CCGAGAA CCCGCACAC CCTGCCAGC</td>	SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SF Contig625_tail_Contig170rc_hea SR_rc SF Contig625_tail_Contig170rc_hea	CCTGCCACG CCTGCACG CCGAGGTC CCGAGGTC CCGAGGTC CCGAGGTC CCGAGGAC CCGAGAA CCCGCACAC CCTGCCAGC
Sequences producing significant alignments: Select: All None Selected:0	Nax Total Query E Ident Accession 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 1e-120 45% WP_006930844.1 380 380 100% 2e-120 45% WP_006928432.1 375 375 99% 2e-110 45% WP_006928432.1 375 375 99% 2e-110 46% WP_006102762.1 322 323 99% 1e-100 36% WP_006102762.1 322 322 98% 2e-101 40% WP_006102762.1 322 322 98% 4e-100 36% WP_002773431.1 318 318 89% 4e-100 43% WP_00277637.1 314 314 99% 2e-96 44% WP_0177803.1 318 308 99% 2e-96 44% WP_0147801.1 309 309 98% 2e-96 44% WP_0147801.18.1	SF Contig625_tail_Contig170rc_hea SF_rcc SF Contig625_tail_Contig170rc_hea SF_rc SF SF SF SF SF <td>CCACCGAG CTGACCAG CTGACCAG CTGGTACC CGAGGTC * * ACCGCAGAG GCCGAGAA GCCGAGAA GCCGAGAA CTGGCGT CTGGCGT CTGGCGT CTGGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGCGT CTGCCGCGCGT CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCC CTGCCGC CTGCC CTGCCCGC CTGCCCGC CTGCC</td>	CCACCGAG CTGACCAG CTGACCAG CTGGTACC CGAGGTC * * ACCGCAGAG GCCGAGAA GCCGAGAA GCCGAGAA CTGGCGT CTGGCGT CTGGCGT CTGGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGT CTGCCGCGCGT CTGCCGCGCGT CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCCGCGC CTGCC CTGCCGC CTGCC CTGCCCGC CTGCCCGC CTGCC



GGAATTTT	8			
ACCGACACCCTGGAAGGGGGCCTGCGAGCGCCTTGT	950			
ATTCCGCAGCCGGACTTGAGCGTGCGGTGGCAGAT	35			
***。				
GCTG	15			
GCTGGCAGATCGAGGTCTACC	985			
	68			
* * * *				
CGCATCGAGGAGCGCCAGCTCGCCACC				
	1035			
CCGCATCGAGGAGCGCCAGCTCGCCACC	96			
* * * * * * * * * * * * * * * * * * * *				
GCCTGCCTTGGTGTCGATCTGGTGGTCGC	96			
	1085			
GCCTGCCTTGGTGTCGATCTGGMMMMMMMGTCGC				
**************************************	140			
••				
CCTCGCCCACCAGAGGCGCGAGGACCCGCAGGCAC	136			
CCTCGCCCACCAGAGGCGCGAGGACCCGCAGGCAC				
CCTCGCCCACCAGAGGCGCGAG-ACCCGCAGCA	187			

TCTCCGGGGAGCAGTGGAAAGCGCTGTGGTT				
TCTCCGGGGAGCAGTGGAAAGCGCTGTGGTTGCGC	1185			
TCCCT	206			
** *				
ATCCCCAGTGATGACGACGAGCCCACACTGCGCGA	1235			
CGCCCAGCTCGGCGGCTTCCAGCGGCGCAAGGCGG	1285			
	1205			
· · · · · · · · · · · · · · · · · · ·	.1			
compared with our Fused				

ntigs.) Measured at

ental error was not avoided as we puters to detect the chain of our before our analysis was complete. The forward and reverse primers protein, shown as transposase in illustrates the end of our analysis, ata, then, not only supports our ugh its genome, unless broken by e programs and the availability to 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

- 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.
- Oren, Aharon. "Chapter 11. Biotechnological Applications and Potentials of Halophilic Microorganisms." Halophilic
- Microorganisms and Their Environments. Dordrecht: Kluwer Academic, 2002. 369. Print. 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.
- 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.
- 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
- 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 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 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>.
- Ewing, B., and P. Green. "Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities." Genom Research 8.3 (1998): 175-85. PubMed. Web. 10 Nov. 2014. http://genome.cshlp.org/content/8/3/186.full.pdf+http://genome.cshlp.org/content/8/3/186.full.pdf+http://genome.cshlp.org/content/8/3/186.full.pdf+http://genome.cshlp.org/content/8/3/186.full.pdf+http://genome.cshlp.org/content/8/3/186.full.pdf+http://genome.cshlp.org/content/8/3/186.full.pdf+http://genome.cshlp.org/content/8/3/186.full.pdf

GRP#5

Section#1