DISCOVERY AND CHARACTERIZATION OF A NEW GROUP OF IS10 INSERTION SEQUENCES

by

Rachel Kinzelman

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Biological Sciences

at

The University of Wisconsin-Milwaukee

August 2016

ABSTRACT

DISCOVERY AND CHARACTERIZATION OF A NEW GROUP OF IS10 INSERTION SEQUENCES

by

Rachel Kinzelman

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Dr. Charles Wimpee

Insertion sequences (ISs) are small mobile genetic elements that can have significant impact on the genotype and phenotype of a host organism. Previous work in this laboratory revealed an insertion sequence that disrupted the *luxA* gene in *Vibrio harveyi* strain BCB451, knocking out light production. Phylogenetic analysis of this insertion sequence, dubbed IS451, reveals that it is in the IS10 family, but represents a novel variant that is only 79% identical to other known IS10 sequences. Twelve copies of IS451 were isolated from a genomic library and sequenced, and were found to be essentially identical, but located in dispersed chromosomal locations. We find that of all the copies sequenced, *luxA* is the only structural gene disrupted by IS451. All other copies are inserted into intergenic regions. Screening of 1,107 additional *Vibrio* strains isolated from the same geographical location indicates that IS451 is rare. Of the two isolates that contain IS451, both have an abnormally large number of IS10s within their genomes. While isolating IS451 from positive isolates, three other novel IS10 sequences were discovered. These insertion sequences, designated IS226-A, IS226-B, and IST2-8, are only 88%, 79%, and 84% similar to other known ISs, respectively. Together with IS451, these insertion sequences are included in the subgroup of IS10s which includes the IS10s that make up the Tn10 transposon. The three isolates that contain these ISs are all within the genus *Vibrio*, but represent different species.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	amino acid
ASW	Artificial Seawater
BLAST	Basic Local Alignment Search Tool
bp	base pair
CTAB	cetyltrimethylammonium bromide
IR	inverted repeat
IS	insertion sequence
LB	Lysogeny Broth
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NCBI	National Center for Biotechnology Information
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
SWC	Seawater Complete

ACKNOWLEDGEMENTS

I want to first and foremost thank Chuck for taking a chance on me. I would not be where I am now without his help. He gave me this interesting project to run with, and despite some times where it was frustrating, I have discovered something that no one else has ever seen before, and that is a neat feeling.

I also would not be where I am without the help of my parents and my sister. Throughout all of my life, they were always there to support me, and without the late nights texting, telling stories, and sharing thoughts with Ally, I would have gone crazy.

Introduction

Insertion sequences

Insertion sequences can be thought of as the ultimate selfish gene. As far as we know, they exist only for their own propagation. They encode one protein, a transposase, with the singular function of being able to cut out the insertion sequence from DNA and re-insert it into a different portion of DNA. They are among the most common protein-encoding genes in nature, with a survey of sequenced genomes finding that approximately 1.1% of viruses, 66% of archaea, 86.9% of bacteria, and 58.6% of eukaryotes contain at least one IS (Aziz et al. 2010). Of protein-encoding genes, ISs make up 0.825%, with an average 38.42 insertion sequences per genome that contains at least one IS (Aziz et al. 2010). It is estimated that the human genome itself is made up of 40% mobile genetic elements such as insertion sequences (cited in Aziz et al. 2010).

Structure

Most insertion sequences are flanked by two inverted repeats, which are flanked by two direct repeats. The inverted repeats are part of the insertion sequence and are the target site of the transposase for excision. The direct repeats are a consequence of insertion into host DNA. When the transposase cuts the host DNA to insert the insertion sequence, a staggered cut of two single strands DNA is formed (cited in Mahillon and Chandler 1998). The insertion sequence is inserted into the staggered cut, and repairing the single strand host DNA forms the direct repeats on either end of the insertion sequence (Figure 1).

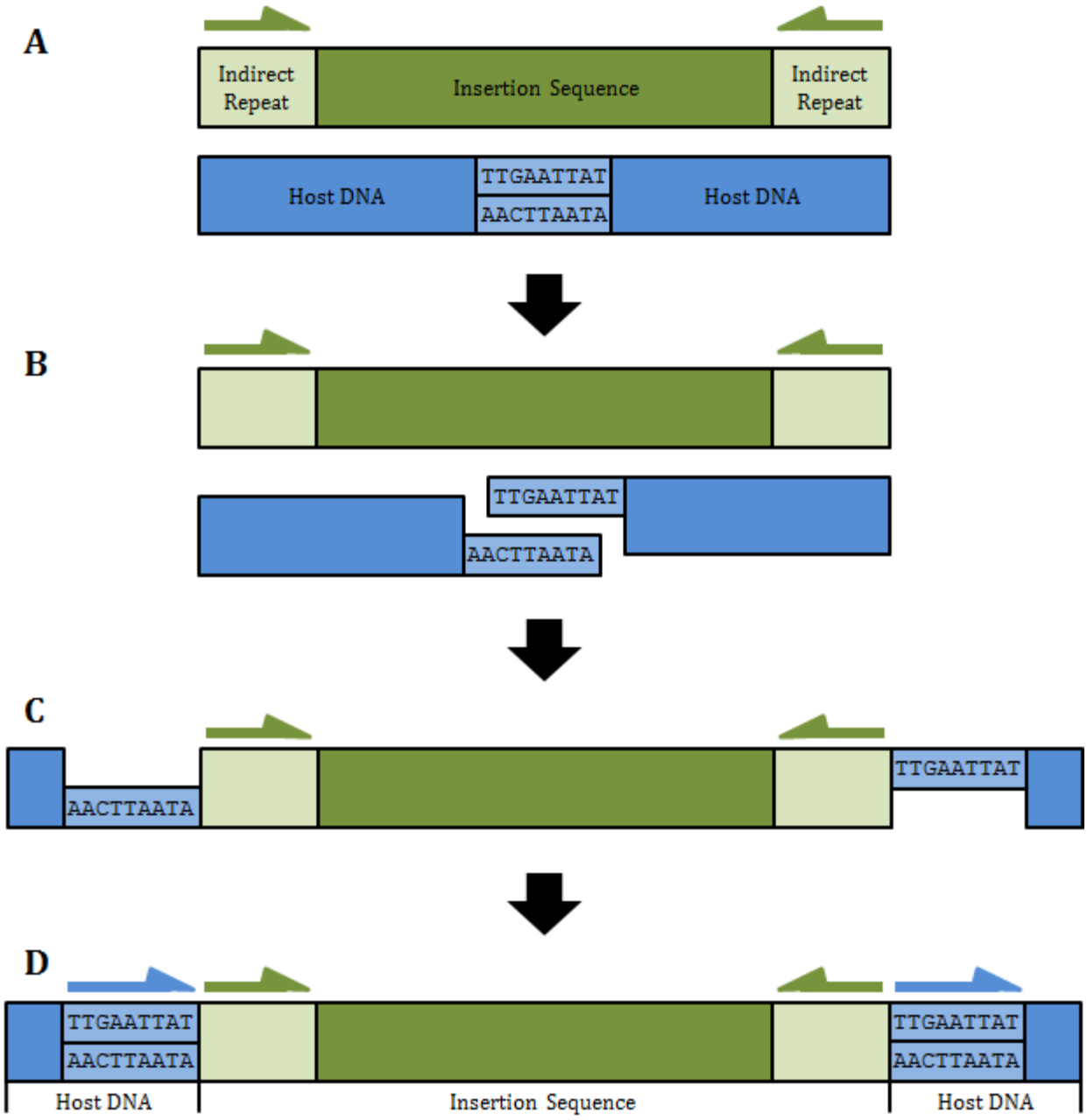

Figure 1: Insertion of an Insertion Sequence.
Host DNA is in blue and IS DNA is in green. A) Insertion sequence has been removed by transposase. B) A staggered cut is made in host DNA by transposase. C) Insertion sequence is inserted into host DNA. D) DNA polymerase repairs single-stranded host DNA, forming direct repeats flanking the IS.

Diversity in insertion sequences

Insertion sequences are diverse. There are >20 families of ISs, most of which are further subdivided into smaller groups (Siguier et al. 2006). The sizes of ISs vary greatly, and can range from 700 to 5400 bp (Siguier et al. 2006). These families have different transposase configurations as well as mechanisms of transposition. Transposases in the IS1 and IS3 families, for example, are made up of 2 ORFs that form different proteins dependent on frame shifting (cited in Mahillon and Chandler 1998). The IS4 (which includes IS10), IS6, and most of the IS5 families instead contain one ORF that encodes a transposase (cited in Mahillon and Chandler 1998).

Transposition

Mechanisms of transposition differ among families of insertion sequences (Figure 2). The two main methods are known as “copy-paste” and “cut-and-paste”. One type of copy-paste mechanism inserts the IS into a new location while still attached to the original locus. The IS then splits into single stranded DNA and is filled in by DNA polymerase in a semi-conservative replication mechanism (Figure 2A). Another copy-paste mechanism works through reverse transcription, where the insertion sequence is first transcribed into RNA, reverse transcribed into DNA, and the copied IS can then be inserted into another location within the host genome (Figure 2B). A cut-and-paste method excises the insertion sequence from the genome, and the original IS can then be inserted into a new location (Figure 2C). There is a preference for transposition in cut-and-paste insertion sequences when the DNA is hemimethylated (cited in Skipper et al 2013). After the replication fork

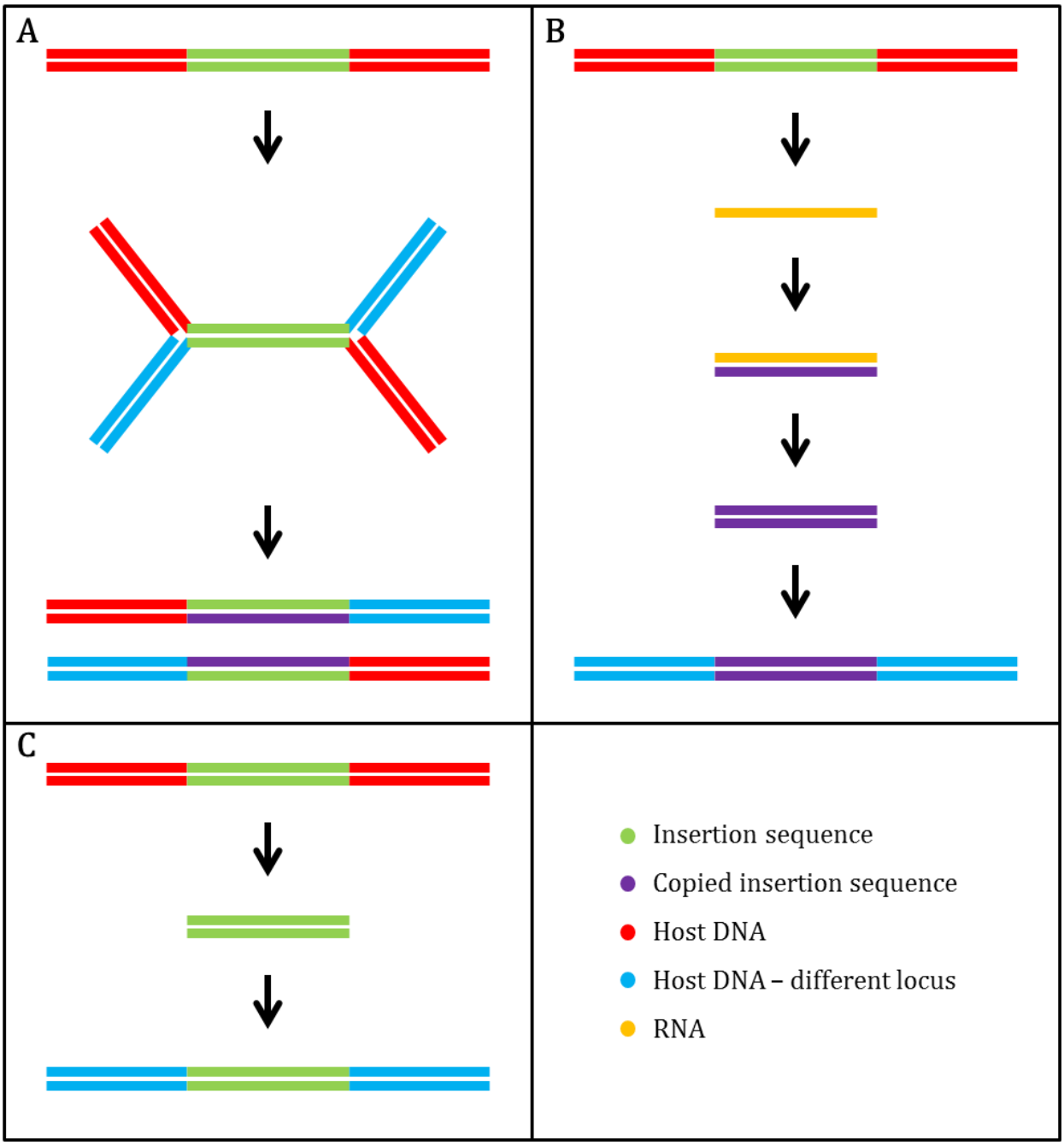

Figure 2: Mechanisms of Transposition
A) Copy-paste semi-conservative replication. B) Copy-paste reverse transcription. C) Cut-and-paste transposition (adapted from Curcio and Derbyshire 2003).

passes, and there are now two copies of the IS on two replicated DNA strands, one of the ISs can be transposed so that the two copies are on the same strand, or it can be inserted

before the replication fork, so one replicated strand has two copies, and the other replicated strand has one copy but in a different locus (Figure 3) (cited in Skipper et al 2013).

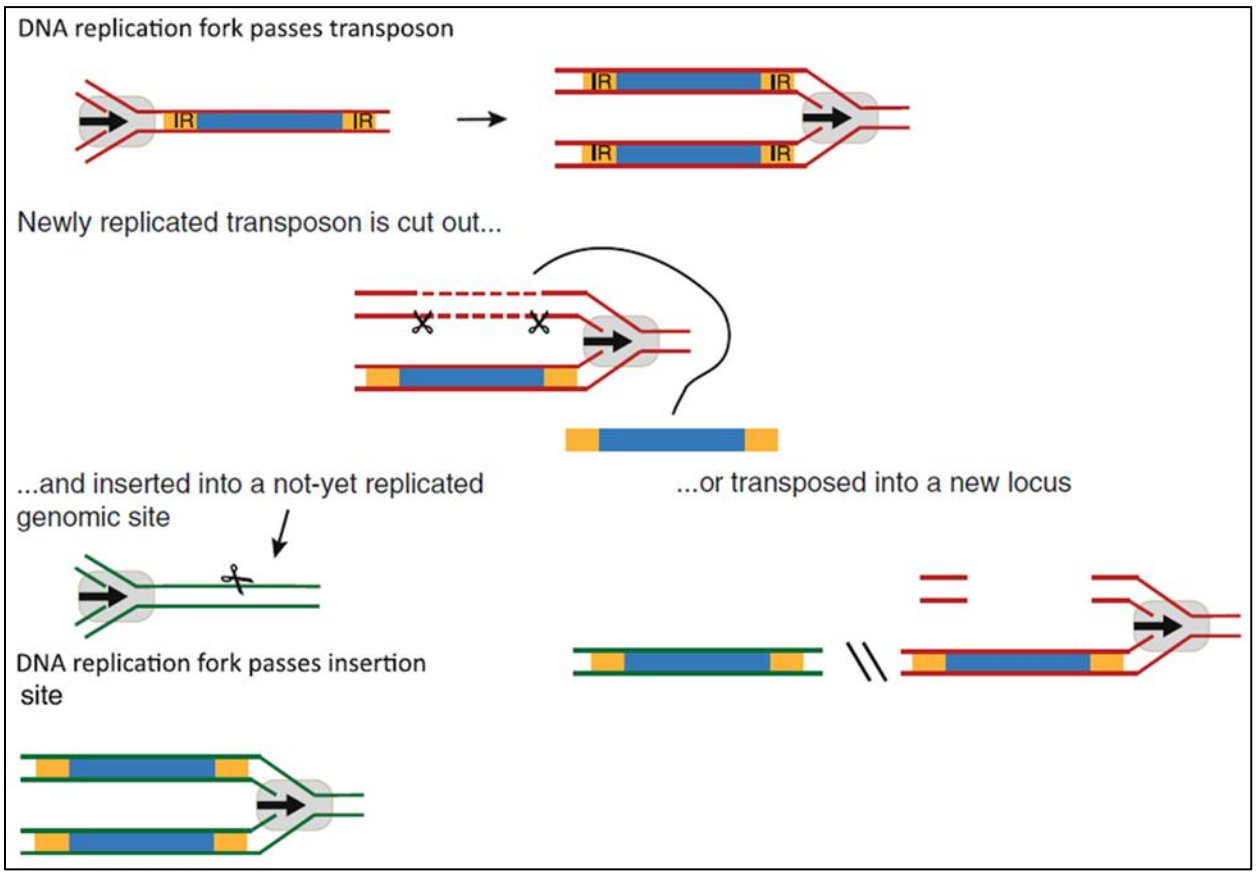

Figure 3: Replication of a Cut-and-Paste Insertion Sequence
After replication, one copy of the IS can be transposed into a new locus, resulting in two copies on one DNA strand and either one or zero copies on the sister strand.

Lateral Gene Transfer

Lateral gene transfer is the process by which genes can be passed from one organism or individual to another non-offspring individual. There are three mechanisms found in prokaryotes. Conjugation is the transfer of a plasmid or DNA directly from one bacterium to another. Usually this is done through a conjugation pilus. Transduction is when bacterial

genes are carried by a lysogenic phage that infects a new host, transferring those genes. Transformation is the uptake and incorporation of environmental DNA by competent bacteria. While transposases can move DNA around inside a cell, lateral transfer is dependent on the insertion sequences being transferred from one cell to another cell via any of these methods. If the insertion sequence is inserted into a prophage or a plasmid, for example, it can then be passed to another bacterium.

Transposons

In addition to the insertion sequences themselves, transposases can also cut out and transfer other genes. If an organism has two copies of an insertion sequence, the transposase may cut at the end of each insertion sequence rather that at each end of one copy. In doing so, any DNA and genes located between each IS are excised from the host DNA and can be moved to a new location. This gene or group of genes flanked by insertion sequences is known as a transposon. Transposons can transfer genes from one host to another through any form of lateral gene transfer. One of the best characterized is the Tn10 transposon, found in many pathogenic bacteria, which contains genes for tetracycline resistance and is flanked by two IS10s, IS10L and IS10R (cited in Mahillon and Chandler 1998).

IS10 group of Insertion Sequences

IS10s are a group of insertion sequences within the IS4 family. Like other members of the IS4 family, IS10s have one ORF coding for a transposase with a DDE (aspartate-aspartate-glutamate) active site domain. The insertion sequence is excised and inserted using a cut-and-paste method by making a break in the 3' end between the inverted repeat and the direct repeat. The –OH group of the 3' then attacks the 5' end of the complementary strand, creating a hairpin and a clean break between the insertion sequence and host DNA (Kennedy et al. 1998). This leaves the host DNA with two direct repeats where there was no repeat initially, and the IS10 is inserted into a different location in the genome.

Translational Regulation

Although there is a preference for certain sequences for insertion, IS10s can insert anywhere into a genome. This is hazardous for the host genome, because insertion into an essential gene would be lethal. IS10s are therefore subject to tight regulation. To regulate translation, an anti-sense sRNA is located on the complementary strand of the IS. The anti-sense RNA is able to bind the transposase mRNA at the ribosome binding site to prevent translation (cited in Ross et al. 2013 and Ellis et al. 2015). When there are more copies of an IS10 within the genome of an organism, there are more copies of the anti-sense RNA as well, and there is an increased inhibition of translation and a decrease in the amount of transposase (cited in Ellis et al. 2015). This can prevent a large number of copies of IS10s from accumulating in the bacterial genome.

If an IS10 is inserted into the host genome directly after a host promoter, production of transposase mRNA can be upregulated if the promoter belongs to a frequently transcribed gene. Translation in said mRNAs can be regulated through their secondary structures. The host promoter will transcribe an mRNA that is longer than the mRNA transcribed using the IS promoter itself (cited in Mahillon and Chandler 1998). This longer mRNA is capable of forming stable secondary structures that sequester the start of translation in the mRNA (cited in Mahillon and Chandler 1998). This prevents the translation of the transposase, and there is a lesser chance that an IS10 found in between genes or in a non-essential gene could be inserted into an essential gene.

Bioluminescence and IS451

Bioluminescence is a process by which organisms produce light. In *Vibrio*, bioluminescence is regulated through quorum sensing. As the numbers of *Vibrio* increase, the *lux* operon is turned up, and the genes *luxCDABEGH* are transcribed and translated. The proteins LuxA and LuxB together produce luciferase, which is essential for producing light. However, bioluminescence is not an essential process in bacteria. While some bioluminescent bacteria may be selected for based on their light production, such as *Aliivibrio fischeri* in the Hawaiian Bobtail Squid, the process itself is not necessary to the lifecycle of the bacterium. Therefore, disruption of the *luxA* gene of BCB451 simply knocks out light production by producing a non-functional LuxA, and does not impede its ability to grow.

Discovery of IS451

In 2005, Dr. Liz O'Grady was examining isolates of *Vibrio* from Boca Ciega Bay (O'Grady and Wimpee 2008). She was looking for naturally dark mutants of bioluminescent *Vibrio* species by selection on Thiosulfate-citrate-bile salts-sucrose (TCBS) agar, screening for *luxA*, and then looking for light production by the isolates. She found three *Vibrio harveyi* isolates that produced no light. PCR amplification of the *lux* genes in these isolates revealed that one of them, BCB451, had a *luxA* gene that was at least 1kb larger than normal. The *luxA* gene was sequenced, and approximately in the middle of *luxA*, there was an insertion sequence. A phylogenetic analysis showed that it was an insertion sequence in the IS10 family but was only 79% similar to the closest member (O'Grady and Wimpee 2008). It was renamed IS451.

IS451 is only in non-essential genes

Arie Brenner digested BCB451 genomic DNA with BamHI, EcoRI, KpnI and PstI and used a Southern Blot for IS451 to find the approximate number of copies of IS451 in BCB451 (Brenner MS Thesis 2011). There were approximately 10-12 copies found in the genome. She then looked at the BCB451 plasmid library that was made by Dr. O'Grady. Using primer walking, Arie Brenner was able to obtain the sequence of each clone that contained IS451 and determined that there were 11 different clone sequences in addition to the one copy of IS451 in *luxA*. She created a map from the sequence data, showing that the only gene that was disrupted by IS451 was *luxA*, and all the other copies in the genome were only adjacent to various domains, putative domains, and ORFs. In the inverted repeats that

flank the sequence, there is a 19 base pair insertion that was initially thought to disrupt the removal or insertion of the sequence, but sequencing of the additional copies showed that it was in every copy of IS451 in BCB451 (Brenner MS Thesis 2011).

Thesis Statement

- Is IS451 found in other *Vibrio* besides BCB451 and how commonly is it found?
- If IS451 is found in other *Vibrio*, how similar is the sequence and how many copies are found in each isolate?
- How closely related are *Vibrio* that contain IS451?

Materials and Methods

Media

2X Artificial Seawater (2X ASW) consists of 58.44g NaCl, 10.15g $MgCl_2$, 6g $MgSO_4$ anhydrous, and 1.49g KCl per liter. Seawater Complete (SWC) consists of 375mL 2X ASW, 5g tryptone, 3g yeast extract, and 3mL glycerol per liter. Lysogeny Broth (LB) consists of 10g tryptone, 5g yeast, and 10g NaCl per liter.

DNA Isolation

Cultures were grown overnight at 25°C in SWC and shaken at 200rpm. DNA was isolated using a CTAB/phenol/chloroform extraction as outlined in Current Protocols (Ausubel 1988).

Polymerase Chain Reaction

Primers were designed using sequence data from BCB451 plasmid library(Tables 1 and 2). Each reaction consisted of 10 µL 2X GoTaq Green master mix, 7 µL H_2O, 1 µL forward primer, 1 µL reverse primer, and 1 µL genomic DNA (approximately 10-100 µg). PCR program contained one cycle of 94°C for 5:00, thirty cycles of 94°C for 0:30, 46-56°C for 0:30, and 72°C for 0:30, and one long extension of 72°C for 5:00.

Nucleotide Sequencing

Sequences were obtained via Sanger sequencing at the University of Chicago and assembled using NCBI BLAST and Geneious version 6.1.6 (Madden 2002 and Kearse et al., 2012). Sequences were aligned using MUSCLE on the European Molecular Biology Laboratory-European Bioinformatics Institute website (Edgar 2004 and Li et al. 2015).

Table 1: BCB451 primers for the flanking regions of IS451

Primer Name	Clone	Primer Sequence	Direction
3flnk1	3	GTCAGATGCAATCATCTCGG	Reverse
3flnk2	3	CGCAATACCTATGCCGCCGC	Forward
7flnk1	7	AAAAGCTTTACCCAGACGCC	Reverse
7flnk2	7	GTAGGCACCGGAAACCACCG	Forward
13flnk1	13	GATTCATTAATGCAGCTGGC	Reverse
13flnk2	13	GCACAGTATCAAACCTGCCG	Forward
13flnk3	13	AGTGAGCTGATACCGCTCGC	Reverse
16flnk1	16	TCTTCACCATTACATAGCGC	Reverse
18flnk1	18	AGAGAGCAACAAATGGTGCC	Reverse
18flnk2	18	TGGGAAGTAAACTTCGCCCC	Forward
22flnk1	22	CTTTTCGCCCAACTTGTGGC	Forward
22flnk2	22	TTCTGGATTGTCGATAGCGG	Reverse
22flnk3	22	GAAGTTTTCACTCTGCAGGC	Forward
32flnk1	32	TTTTATCGGTCTACGCTGGC	Reverse
43flnk1	43	TTACTCATCTACGCAGGCAC	Reverse
48flnk1	48	GTATCTAAGCGCCGTCTGCC	Forward
53flnk1	53	ATAACGGTGGCCTGACTCCG	Forward
57flnk1	57	TGAAAGCGCATCAGACAGCG	Forward

Table 2: Internal primers found in IS451

Primer Name	Location	Primer Sequence	Direction
IR 1F1	IS451 5' IR	CTGAAGAATCCCCTAATGA	Forward
IS451F1	IS451	AATCATTAAGTTAAGGTAGATGC	Forward
IS1F	IS451	AATCATTAAGTTAAGGTAGATGCAC	Forward
IS451_F1	IS451	TCCCTTTATCAATTCTGCCC	Forward
IS451_F2	IS451	AGCTTAATGTTAGCCTGCCG	Forward
ISint5	IS451	TTGTCCTCGCTTGGTTGGGC	Reverse
ISint5R	IS451	GCCCAACCAAGCGAGGACAA	Forward
IS451 F-C	IS451	ACATTAAGCGCATGGATCGA	Forward
ISint4	IS451	CGGCGAGACGCTCTTTATGC	Reverse
ISint4R	IS451	GCATAAAGAGCGTCTCGCCG	Forward
IS451 F-B	IS451	AGCGTCTCGCCGTTTATCGA	Forward
IS451 F-A	IS451	TGTTTGTTCAGGAAATCCAA	Forward
IS451_F3	IS451	GTTCAGGAAATCCAATGCCC	Forward
IS451_F4	IS451	AGCATAAACGATTGATGGCG	Forward
ISint3	IS451	TGAAGCGCAATGGATGCACG	Reverse
IS451 midF	IS451	TCATCGTCACCGATGCTGGC	Forward
IS451 midR	IS451	GCTCAGAGCCAACCTCGGCG	Reverse
ISint2R	IS451	CAAGGTTCCTTGGTTGCGGC	Reverse
IS451 R-A	IS451	GTTCGTTGCTAGAACCCAAG	Reverse
ISint2	IS451	GCCGCAACCAAGGAACCTTG	Forward
IS451 R-C	IS451	AAGGTTCGTTGCTAGAACCC	Reverse
IS451_R1	IS451	CATAAAGCCTGACTAGCTGC	Reverse
IS451 R-B	IS451	ACTCTTTAAATCACGGAACG	Reverse
ISint1	IS451	AGCCCTAATCCATAAGCAGG	Reverse
IS451_R2	IS451	TATGTCGAAACGCTCTGGGC	Reverse
IS451_R3	IS451	AGCCTTGCTTTTGAGCGTGC	Reverse
IS451_R4	IS451	CACATTTCTTGTTCGCACGG	Reverse
IS451_R5	IS451	AGCTCTTGAGTAGTGATGCG	Reverse
IS1232R	IS451	GTTGCTGGGTAAGTTGGGTAC	Reverse
IRI R3	IS451 3' IR	ATCACCCAACGCATAGCCATTTTGG	Reverse
IRI R2	IS451 3' IR	ATCACCCAACGCATAGCCA	Reverse
IRI R4	IS451 3' IR	TAAATCACCCAACGCATAGCC	Reverse
IRI R1	IS451 3' IR	CTCATAAATCACCCAACGC	Reverse
IR 1F2	IS451 3' IR	CTGAAGGATCCCCTCATAA	Reverse

RNA Secondary Structure

RNAFold webserver run by the Institute for Theoretical Chemistry at the University of Vienna was used to generate minimum free energy secondary structures of RNA (Gruber et al. 2008).

Colony Hybridizations

A 0.46 mm nylon transfer membrane was placed on SWC plates and isolates were plated and grown overnight at 25°C. Colonies were lysed by placing the membrane in 0.75 M NaOH, 1M Tris HCl (pH 7.4), 0.5M Tris HCl (pH 7.4) with 1.5M NaCl, and finally 5X SSC with a two minute drying step between each solution. DNA was bound to the membrane by UV crosslinking. Membranes were hybridized with ^{32}P-labelled IS451 probe (produced using primers IS1F and IS1232R) and exposed to X-ray film with an intensifying screen overnight at -80°C and developed.

Restriction Digests

Restriction digests were performed with the restriction enzymes BamHI, EcoRI, KpnI and PstI using a mixture of 2µL restriction enzyme, 2µL 10X Buffer and 16µL genomic DNA (approximately 100-1000 µg). Digestion was incubated for 1 hour at 37°C.

Southern Blot Analysis

Southern blots were performed by standard procedures (Sambrook et al. 1989). Briefly, digested genomic DNA was run on a 75% agarose gel at 25 Volts overnight. DNA from the gel was blotted onto nylon film, prehybridized, and hybridized using a ^{32}P-labelled IS451 probe (produced using primers IS1F and IS1232R).

Phylogenetic Analysis

Trees were made by using the Molecular Evolutionary Genetics Analysis software version 7 (Kumar et al. 2015). Sequences were entered into MEGA7 and aligned using MUSCLE. Trees were made by the Maximum Likelihood method, and evolutionary history models were chosen based on the lowest number of parameters and criteria to keep variance low (Kumar et al. 2015). A General Time Reversible model with discrete Gamma distribution used to model evolutionary rate differences among sites (Nei and Kumar 2000). A rate variation model allowed for some sites to be evolutionarily invariable. 500 bootstrap replicates were used to represent the evolutionary history of the taxa analyzed (Felsenstein 1985).

Results

Verification of IS451 copies in BCB451 genomic DNA

In the previous work on IS451, there were twelve different clones from the plasmid library of BCB451 that contained IS451 (Brenner MS Thesis 2011). Cloning using partial restriction digests can lead to inaccuracies and cloning artifacts, however. It is possible for two distant pieces of the genome to ligate together or for portions of DNA to be missing in a sequence. It was therefore essential to verify that the sequences in each of the clones of BCB451 are also found in genomic DNA. PCR was used to amplify the IS451 in BCB451, and in all cases, the flanking regions of IS451 in the clones were also found to be flanking IS451 in the genome of BCB451 (Figure 4). There were some inaccuracies in the clone sequence data that were corrected by looking at genomic DNA. A missing sequence segment of 17 base pairs from the copy of IS451 in Clone 7 was found to be consistent with the other copies of IS451 in BCB451. Clone 22 was missing an 806 base pair part of the flanking region, but the correct sequence was obtained from the genomic DNA.

The previous Southern Blot of BCB451 indicated that there were at least 10 copies of IS451 in BCB451 (Brenner MS Thesis 2011). Since there were twelve individual clones, it was possible that multiple restriction fragments were the same size and appeared as a single band in the Southern Blot. Another possibility is that several of the truncated clones might be the left and right side of the same copy of IS451. Clones 16, 32/37, and 43/54 were missing the 5' end of IS451, and clones 22, 48, 53, and 57/58 were missing the 3' end of IS451 (Figure 4). Primers from the flanking regions of the truncated clones were paired to examine whether any of the IS451 in those clones might be part of the same IS451 copy. No

paired primers were able to amplify IS451 from genomic DNA; therefore, each clone most likely represents a single copy IS451 in BCB451.

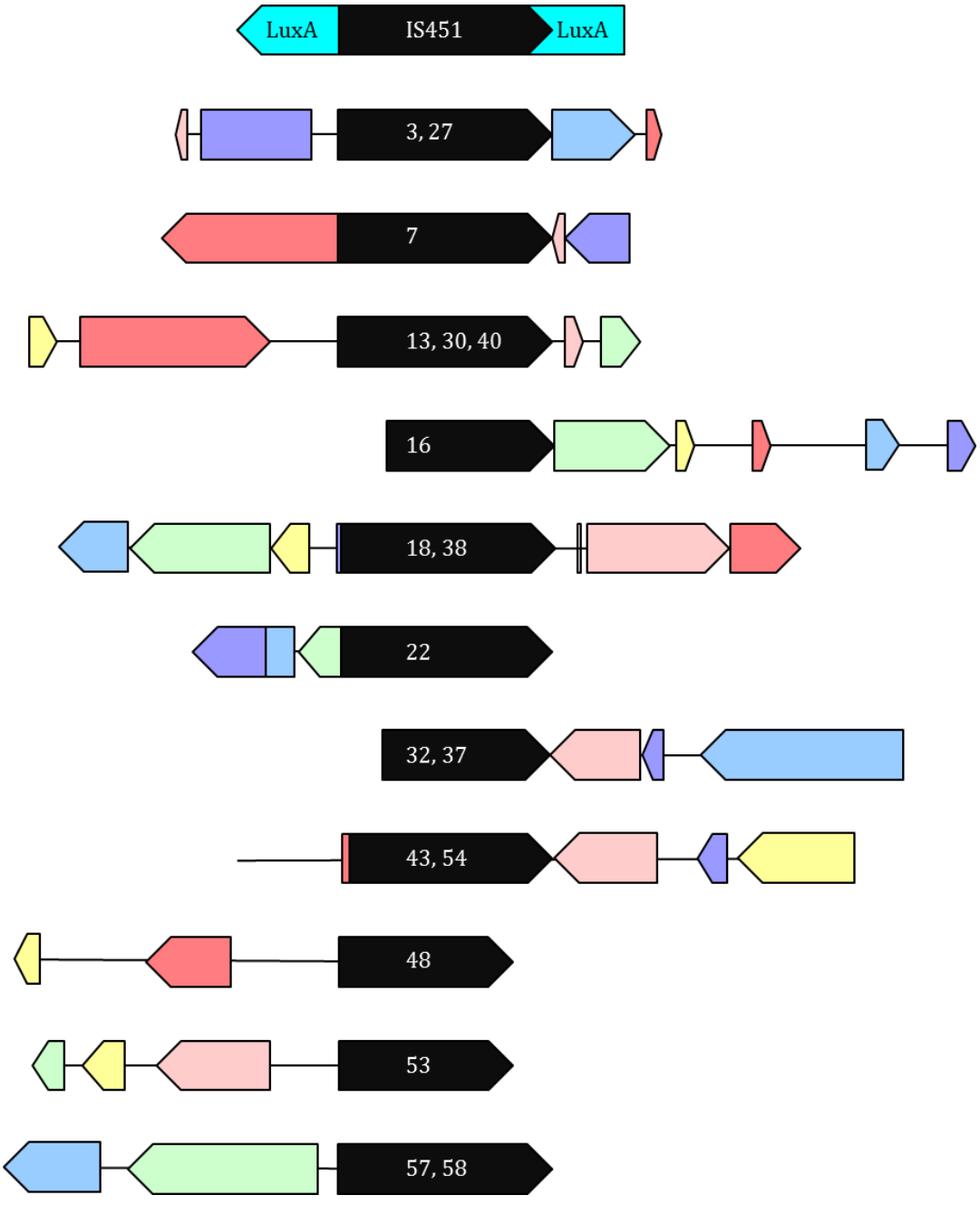

Figure 4: Aligned Map of Unique Clones
IS451 is depicted in black, with the clone numbers in white inside of each IS451. Colored arrows denote different ORFs in each clone (similarly colored ORFs are not related).

Predicting RNA Secondary Structure of IS451

There are a large number of copies of IS451 in BCB451. The majority of genomes with an IS4 (which includes IS10) contain only one, with the next largest groups having 2, 3, or 6-8 IS4 sequences per genome (De Palmenaer et al. 2008). Seeing 12 IS4s in one genome is unusual, since more copies of an insertion sequence will result in higher levels of antisense RNA that will repress translation of transposases, keeping transposition and copy numbers low (Ross et al. 2013). However, IS451 is adjacent to several ORFS, and is located in one nonessential gene, *luxA*. It is possible that IS451 could follow a host promoter belonging to one of those ORFS, which would produce a long mRNA and prevent translation through the RNA secondary structure. If the secondary structure does not block the ribosome binding site or the start site of translation, it could lead to an upregulation of the transposase and explain the large number of copies. This would not occur in the copy of IS451 in *luxA*, because it is inserted backwards in the gene, and only the reverse complement would be transcribed by the lux operon promoter.

The RNAfold webserver created by the University of Vienna was used to predict the secondary structure of the IS451 mRNA (Gruber et al. 2008). Bases 51-61 in the mRNA transcribed from an external promoter have an approximately 100% binding probability with a segment of RNA 104 bp in the sequence that contains the start codon. When that sequence is removed, binding probability to block the start AUG drops to around 25% (Figure 5). This predicts that IS451 translation could be regulated through mRNA length.

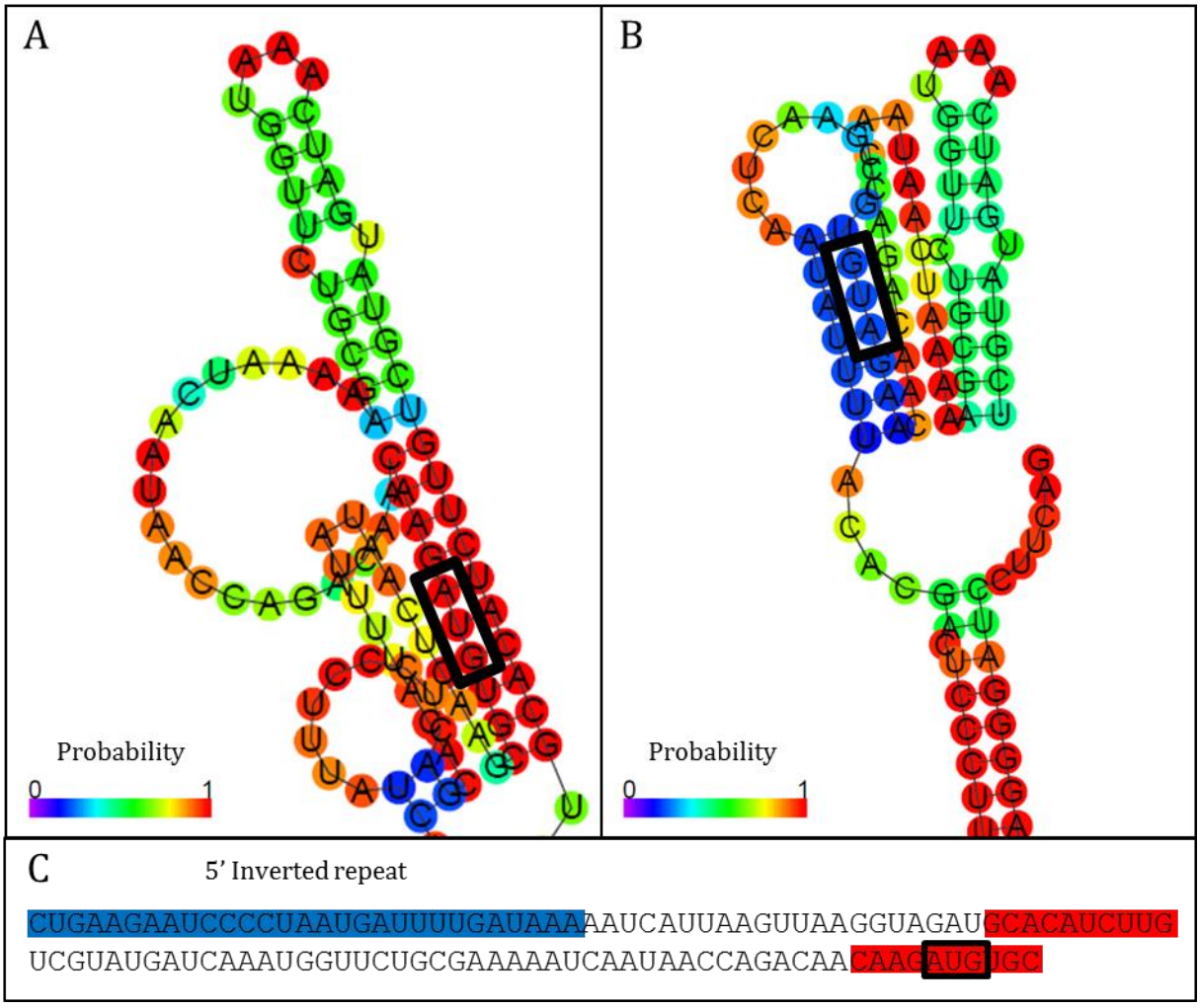

Figure 5: IS451 RNA Secondary Structure
A) The secondary structure in mRNA produced with a host promoter has high probability binding that blocks the start of translation (black box). B) The mRNA produced from an IS promoter has low probability binding to the start of translation (black box). C) The high affinity binding sites (highlighted in red) in the 5' end of an mRNA made from a host promoter.

Screening of isolates from Boca Ciega Bay for IS451

The multiple copies of IS451 in BCB451 demonstrate that it is capable of moving around within a single genome. What was not known is whether IS451 would be found in other genomes. IS451 almost certainly did not evolve in BCB451. Every copy has the same

sequence with no point mutations or other signs of progression from an ancestral insertion sequence to IS451 (Brenner MS Thesis 2011). Therefore, IS451 was probably passed from a different bacterium to BCB451 and could possibly be found in different isolates.

Although it is unknown if IS451 is endemic to Boca Ciega Bay, transfer of an insertion sequence would require close contact between two bacteria. Since IS451 was found in a *Vibrio* from Boca Ciega Bay, it was logical to start looking within *Vibrio* isolates from the same region. Screening isolated organisms was preferable to metagenomic analysis. Metagenomics would have identified whether IS451 was located in the environment, but could not show what organism it is found in or how rare it is.

Since insertion sequences are important in lateral gene transfer and evolution, discovering how rare IS451 is in the community of Boca Ciega Bay could shed light on rates of gene transfer in *Vibrio* in the area. Even though IS451 does not appear to be part of a transposon, the transfer of insertion sequences relies on the same mechanisms of transfer as any non-transposase genes. If IS451 were common, it could indicate a large amount of gene transfer amongst *Vibrio* in the Boca Ciega Bay area.

1,107 isolates collected from Boca Ciega Bay from 2004-2012 were probed for IS451. From this, three potential positives were obtained (Figure 6). However, colony hybridization is less specific than a Southern Blot, and has a higher rate of producing false positives. A Southern Blot for IS451 was therefore performed on the potential positives and revealed two isolates, BCB226 and T2-8, that either contain IS451 or an IS451-like IS10.

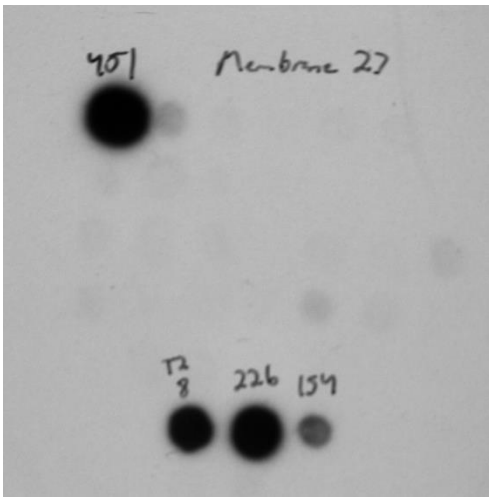

Figure 6: Colony Hybridization of BCB Isolates
A probe for IS451 indicates three isolates that potentially contain IS451. BCB451 is used as a positive control.

Examining the number of copies in other strains with IS451 could show whether 12 copies is a standard copy number for IS451 in a *Vibrio* genome. A Southern Blot of BCB226 and T2-8 using the restriction enzymes PstI, EcoRI, BamHI, and KpnI indicated that BCB226 contains approximately 14-15 copies of IS451 or an IS451-like IS10, and T2-8 has at least one (Figure 7). These are estimates, as the Southern Blot for BCB451 predicted 10 copies of IS451 and there are 12 copies. BCB226 or T2-8 could have even more copies than predicted by the Southern Blot and any of those potential copies could be a different IS10 than IS451. The large copy number in BCB226 is consistent with BCB451 and may indicate that either IS451 naturally produces a greater amount of transposase than some other IS10s or that both isolates have a mutation that enables greater transcription or translation in insertion sequences.

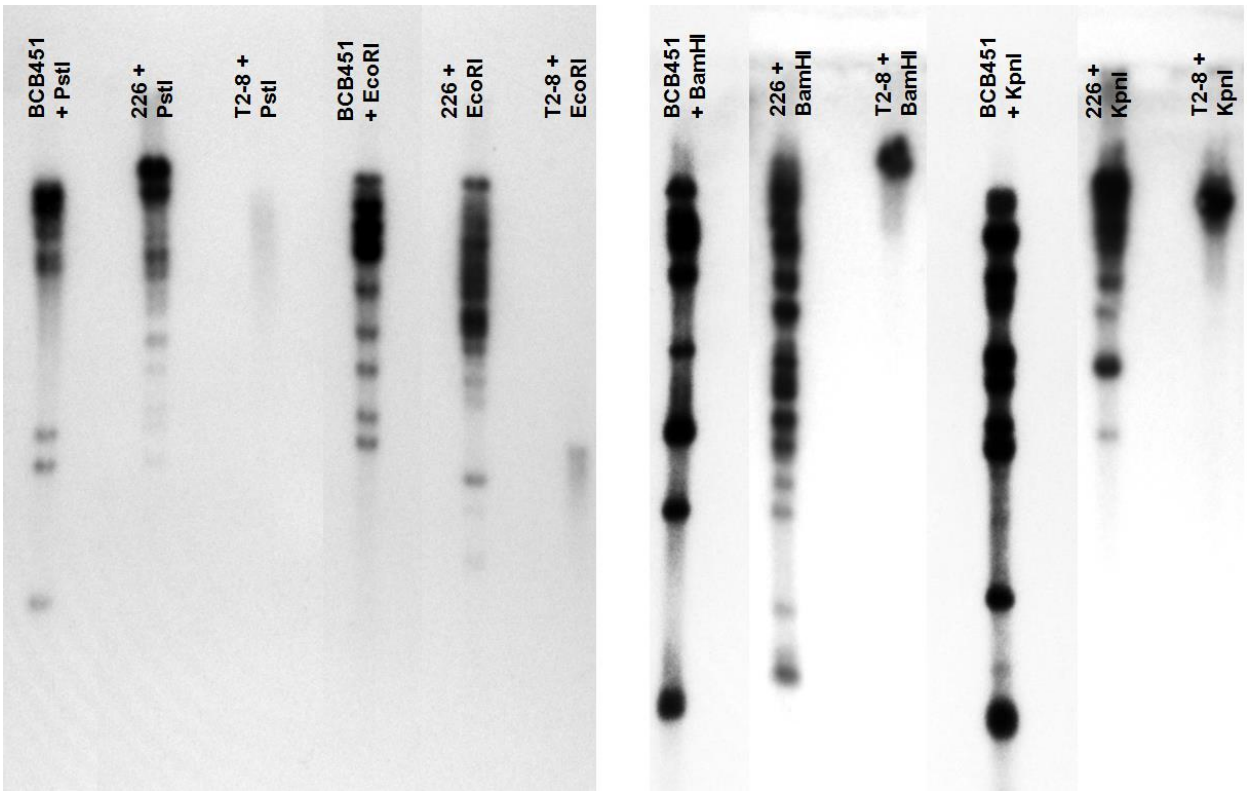

Figure 7: Southern Blot of Positive Isolates
Positive isolates were digested with restriction enzymes and probed for IS451, revealing multiple copies in 226 and at least one copy in T2-8. BCB451 was used as a positive control.

IS451 in BCB226

There is a 19 bp insertion in the 3' inverted repeat of IS451 in BCB451. When initially found in the *luxA* copy of IS451, it was thought that it might inhibit transposition of IS451 (O'Grady and Wimpee 2008). However, each full length and 5' truncated clone contained the insertion and therefore the insertion does not appear to be any negative effect on the ability of IS451 to move within the genome. It also presents a way to look at the transmission of IS451. If the insertion is missing in other IS451 positive genomes, it could show that the insertion was added either within BCB451 before movement and copying of

the IS, or that it was inserted in an intermediate between two IS451 positive bacteria. Alternatively, if the insertion is there, it could imply a close contact between BCB451 and another host, with one of the two bacteria transferring the IS to the other, either directly or through an intermediate.

Primers from both inverted repeats and primers that spanned various parts of the insertion were used to see how similar the sequence of IS451 in BCB226 is to the sequence of IS451 in BCB451 and to see if the insertion in the inverted repeat in IS451 could be found in BCB226. The sequences are 99% identical, but the 5' inverted repeat could not be sequenced using BCB451-IS451 primers. The primer was able to amplify a sequence of the correct length that can align with IS451 in BCB226, but it also amplified another sequence that made the sequence messy. The 5' inverted repeat is similar enough to bind the primer, but it is unknown if it is the same sequence or not (Figure 8A). In the 3' inverted repeat, out of the 19 bp that make up the insertion, 17 are the same as those in IS451 in BCB451. The other two bases were unable to be read in the sequence (Figure 8B).

Figure 8: Inverted Repeats in IS451 in BCB451 and BCB226
The inverted repeat is highlighted in blue and the insertion is highlighted in pink. A) The 5' IR in BCB226 was unable to be amplified using primers. B) The insertion and the left part of the 3' inverted repeat are found in IS451 in BCB226.

Characterization of Additional Insertion Sequences found in Vibrio

Like host genes that can be amplified using consensus primers, closely related ISs can have enough sequence similarity that primers designed for one IS can amplify another. It is also possible for a probe of one IS to hybridize to a similar IS. This was seen while screening Boca Ciega Bay *Vibrio* for IS451. During the colony hybridization, BCB154 seemed to be positive for IS451, but a Southern Blot revealed that it was not. The false positive was likely due to another IS10 in the isolate that was able to hybridize to the IS451 probe. Similarly, primers designed for IS451 in order to amplify it in BCB226 and T2-8 instead returned sequences for three new unique insertion sequences: IS226-A and IS226-B in BCB226, and IST2-8 in T2-8.

These three new additional insertion sequences are similar enough to IS451 be picked up by IS451 primers, but are different from each other and from IS451. IS451 is 72% similar to IS226-A, 73% similar to IS226-B, and 84% similar to IST2-8. IS226-A is 75% similar to IS226-B and 71% similar to IST2-8. IS226-B and IST2-8 are 71% similar (Table 3).

Table 3: Sequence Similarity of Boca Ciega Bay Vibrio Insertion Sequences

	IS226-A	IS226-B	IST2-8
IS451	72%	73%	84%
IS226-A		75%	71%
IS226-B			71%

IS226-A

IS226-A was amplified from BCB226 by using primers IS451_F1 and IS451_R2 (Table 2). IS226-A is only 88% identical to its most closely related IS10 sequence. IS226-A has

multiple stop codons after the start of translation in IS451 and the known functional transposase IS10R, leading to a 90 aa truncation in the N-terminus of the protein. It also contains a stop codon before the stop in IS451 and IS10 R, which truncates the C-terminus by 113 aa. Overall, the protein is predicted to be 199 aa compared to the 402 aa that make up IS451 and IS10R. There are still conserved aspartate and glutamate residues in the amino acid sequence, however, so the transposase may still have a functional active site (Figure 9).

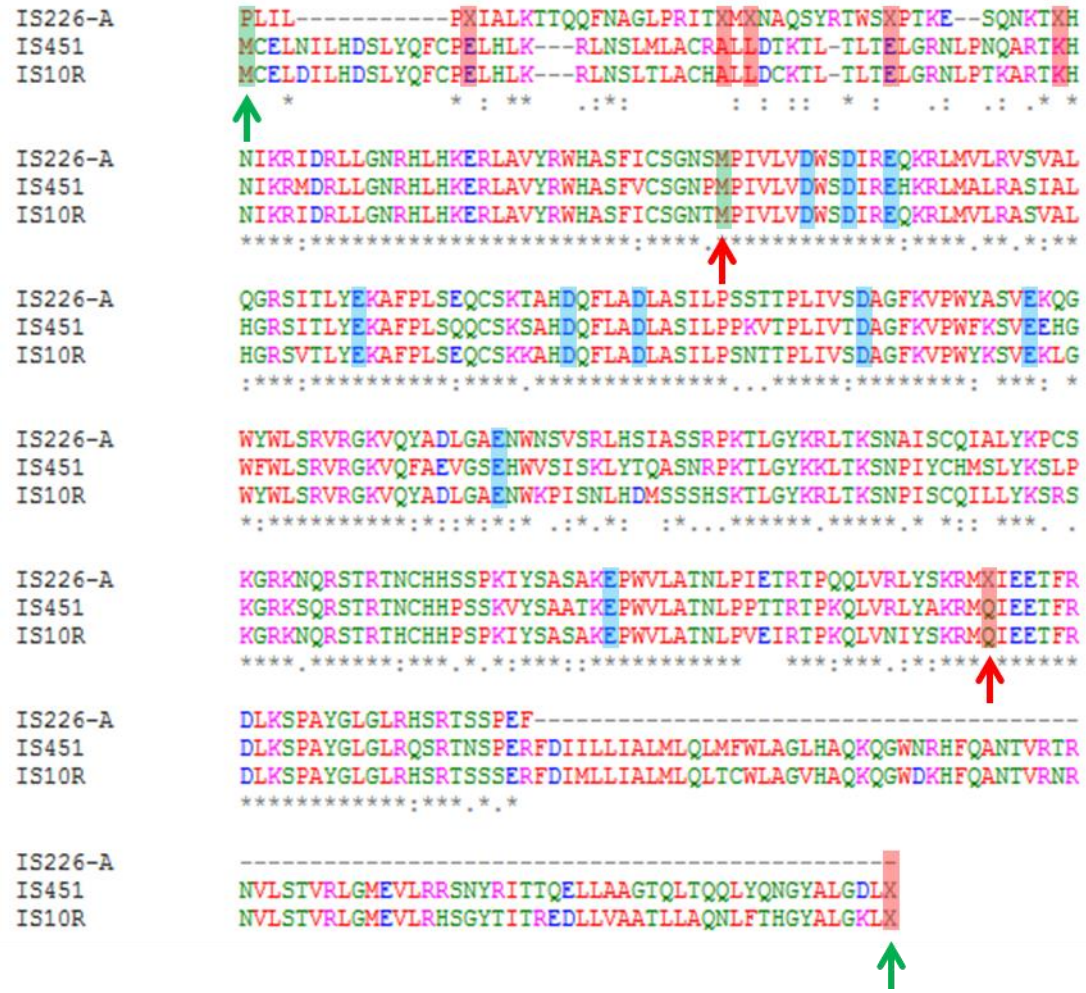

Figure 9: Amino Acid Sequence of IS226-A

Green highlights are the start of translation, blue highlights are glutamate and aspartate in the same location in each sequence, and red highlights are the end of translation. IS226-A is truncated (between red arrows) when compared to IS451 and IS10R (between green arrows).

IS226-B

IS226-B was amplified from BCB226 by using primers IS451F1, IS1F, IS451_F1, IS451_F2, or ISint4R paired with ISint1 (Table 2). IS226-B is 79% similar to its closest related IS10 sequence. There are at least two copies of IS226-B in BCB 226, and likely three or more. In all sequences of IS226-B, there is a base substitution at nucleotide 274 when counting from the ATG at the start of translation. This results in an amino acid substitution from a proline to a lysine or vice versa. There are at least two copies in the genome, one with a C, and one with a G, but there is likely to be at least one more copy of IS226-B with a C in the genome of BCB226 because the peaks for the two nucleotides in the chromatogram are uneven (Figure 10).

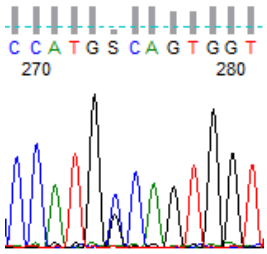

Figure 10: Chromatogram of IS226-B
Nucleotide 274 shows a double peak, with the peak for C being larger than the peak for G.

IST2-8

IST2-8 was amplified from T2-8 by using primer IS451 midF paired with IRI R1, IRI R2, IRI R3, or IRI R4 (Table 2). Of the three new insertion sequences, it has the greatest sequence similarity to IS451, with 84%. The reverse primers all spanned the insertion in the IR of IS451. The insertion could not be sequenced itself, as the primer on the right part of the IR did not work for IST2-8, but the primers were able to bind and amplify the left part of the IR, so it is possible that either the insertion or a similar insertion is also found in IST2-8 (Figure 11).

Figure 11: 3' Inverted Repeat of IST2-8
The inverted repeat is highlighted in blue and the insertion in the IR is highlighted in pink.

The IS451 specific primers used in T2-8 did not amplify the sequence of IS451 in T2-8, despite the positive result on the Southern Blot (Figure 7). Since the sequence of IST2-8 is so similar to IS451, the band in the Southern is probably IST2-8. T2-8 likely has at least one copy of IST2-8, but does not contain IS451.

Phylogeny of IS451, IS226-A, IS226-B and IST2-8

IS451, IS226-A, IS226-B and IST2-8 have relatively similar sequences, but it is not known how closely related they are. Depending on the diversity of the IS10 group of insertion sequences, they could share common ancestors, or they could be found on completely different branches of a tree. If IS10s all have close sequence similarity, these insertion sequences could be far apart evolutionarily. If, instead, IS10s are far more diverse, the Boca Ciega Bay Vibrio insertion sequences could form their own subgroup. IS10 sequences from the ISFinder database were used to assemble a phylogenetic tree (Siguier et al. 2006).

IS451, IS226-A, IS226-B, and IST2-8 are closely related to IS10A, as well as IS10L and IS10R, the two insertion sequences that make up the Tn10 transposon (Figure 12). IS226-A and IS226-B are more closely related to each other than to IS451, showing more sequence similarity. IS451 and IST2-8 are closely related, and share a common ancestor.

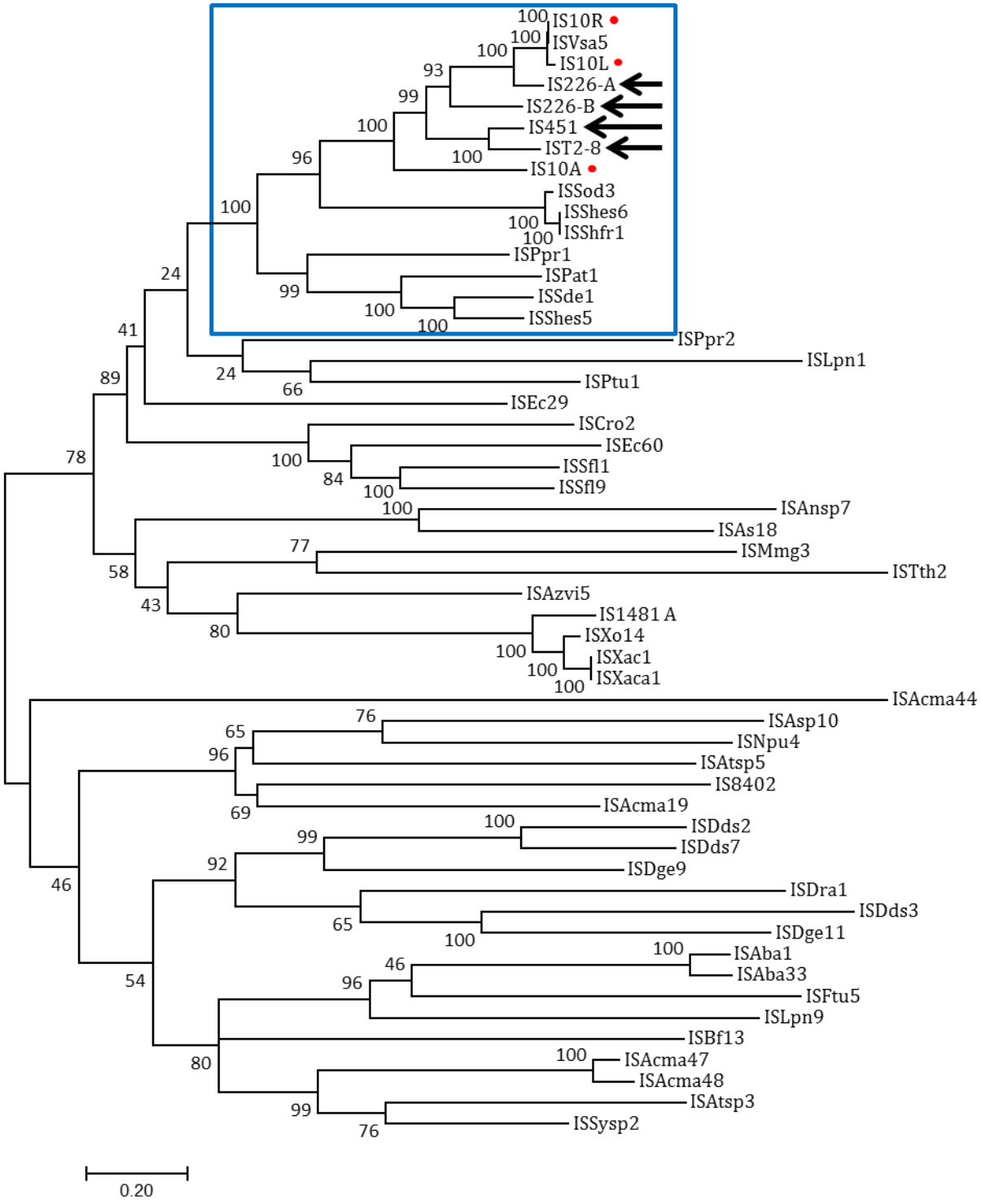

Figure 12: IS10 Phylogenetic Tree

IS451, IS226-A, IS226-B and IST2-8 are all indicated by black arrows. All the insertion sequences in the blue box were originally discovered in marine bacteria, except for IS10A, IS10L, and IS10R (red dots).

Characterization of BCB226 and T2-8

IS451 is found in two isolates, BCB451 and BCB226. BCB451 is a strain of *V. harveyi*, and BCB226 is an unknown species in the *Vibrio* genus. Determining the species of BCB226 would determine whether IS451 has species specificity. If BCB226 is a *V. harveyi*, then IS451 may preferentially be transferred between *V. harveyi*. If BCB226 is a different species, not only is there no species specificity, but it would suggest that the two different species had relatively close contact in order to transfer IS451. T2-8 is an unknown *Vibrio* species as well, yet contains an insertion sequence (IST2-8) that displays common ancestry with IS451. It could be that such closely-related insertion sequences could be found only in *V. harveyi*. To look for similarities amongst the IS451-positive isolates, BCB226 and T2-8 were grown overnight to look for signs of light production. Neither one produced any visible light. Either they were non-luminous Vibrio, or, like BCB451, they were naturally-occurring dark mutants.

Vibrio housekeeping genes (*ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, *topA*) were amplified from IS451, BCB226 and T2-8 genomic DNA (Table 4). Sequence analysis of the housekeeping genes indicated that BCB226 is most closely related to *Vibrio rotiferianus* and T2-8 is most closely related to *Vibrio mediterranei* (Table 5). *V. rotiferianus* is found within the *V. harveyi* clade of *Vibrio*, including *V. campbellii*, *V. harveyi*, *V. parahaemolyticus*, *V. natriegens*, *V. alginolyticus*, and *V. mttili* (Sawabe et al. 2008). *Vibrio mediterranei* is in a separate *Vibrio* clade (Sawabe et al. 2008).

Table 4: Housekeeping Genes in *Vibrio*

Housekeeping Gene	Function
ftsZ	Bacterial cytoskeleton component
gapA	Glyceraldehyde phosphate dehydrogenase
gyrB	DNA gyrase subunit B
mreB	Bacterial cytoskeleton component
pyrH	Aspartate carbamoyltransferase catalytic subunit
recA	DNA repair
rpoA	RNA polymerase alpha subunit
topA	Topoisomerase

Table 5: BLAST Results of *Vibrio* Housekeeping Genes

		First BLAST Result		Second BLAST Result		Third BLAST Result	
Isolate	Gene	Alignment	Identity	Alignment	Identity	Alignment	Identity
226	*ftsZ*	*Vibrio harveyi*	96%	*Vibrio* rotiferianus	96%	*Vibrio* sp.	93%
226	*gapA*	*Vibrio rotiferianus*	99%	*Vibrio* crosai	98%	*Aliivibrio finisterrensis*	97%
226	*gyrB*	*Vibrio rotiferianus*	98%	*Vibrio* sp.	92%	*Vibrio harveyi*	92%
226	*mreB*	*Vibrio* sp.	91%	*Vibrio* sp.	91%	*Vibrio* sp.	91%
226	*pyrH*	*Vibrio* sp.	99%	*Vibrio rotiferianus*	96%	*Vibrio rotiferianus*	96%
226	*recA*	*V. parahaemolyticus*	99%	*V. parahaemolyticus*	99%	*Vibrio harveyi*	96%
226	*rpoA*	*Vibrio rotiferianus*	99%	*Vibrio rotiferianus*	99%	*Vibrio campbellii*	99%
226	*topA*	*Vibrio harveyi*	95%	*Vibrio harveyi*	95%	*Vibrio rotiferianus*	94%
T2-8	*ftsZ*	*Vibrio mediterranei*	98%	*Vibrio mediterranei*	98%	*Vibrio* sp.	99%
T2-8	*gapA*	*Vibrio mediterranei*	99%	*Vibrio mediterranei*	99%	*Vibrio* sp.	99%
T2-8	*gyrB*	*Vibrio mediterranei*	95%	*Vibrio mediterranei*	99%	*Vibrio mediterranei*	99%
T2-8	*mreB*	*Vibrio mediterranei*	99%	*Vibrio mediterranei*	99%	*Vibrio mediterranei*	99%
T2-8	*pyrH*	*Vibrio mediterranei*	99%	*Vibrio mediterranei*	99%	*Vibrio mediterranei*	99%
T2-8	*recA*	*Vibrio mediterranei*	93%	*Vibrio mediterranei*	93%	*Vibrio shilonii*	93%
T2-8	*rpoA*	*Vibrio mediterranei*	99%	*Vibrio* sp.	99%	*Vibrio shiloni*	99%
T2-8	*topA*	*Vibrio mediterranei*	95%	*Vibrio mediterranei*	95%	*Vibrio* sp.	94%

The housekeeping genes *recA*, *pyrH*, *rpoA*, *ftsZ*, *mreB*, and *topA* from BCB451, BCB226, and T2-8 were concatenated alongside housekeeping genes from other representatives of the *Vibrio* genus to generate a phylogenetic tree (Figure 13). BCB451 and BCB226 are within the *V. harveyi* clade. T2-8 is most closely related to the *V. splendidus* clade.

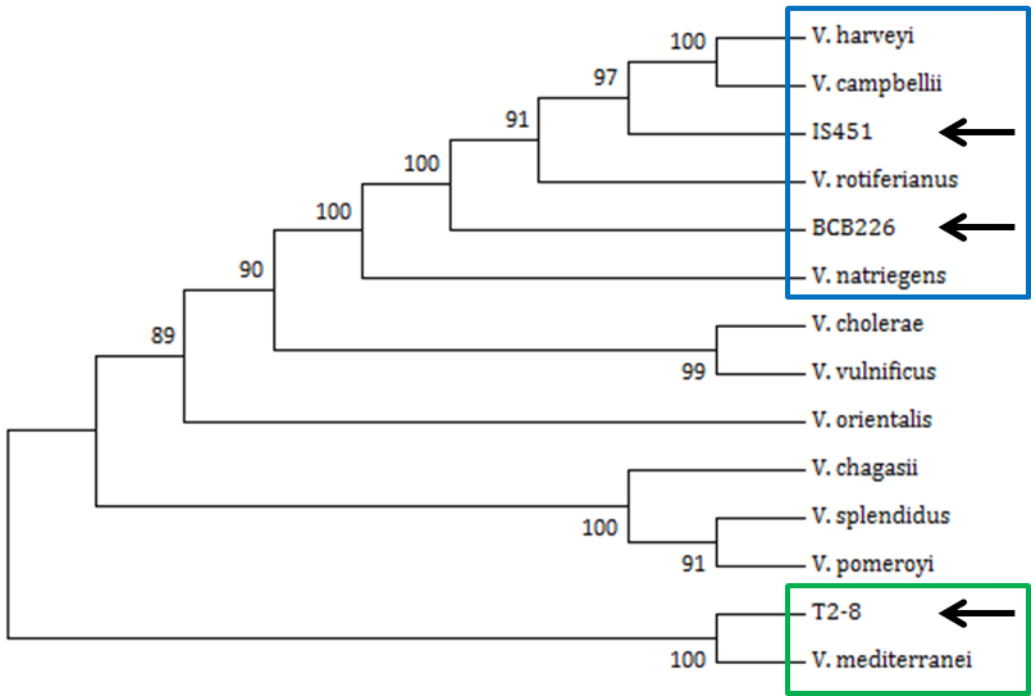

Figure 13: *Vibrio* Phylogenetic Tree
BCB451, BCB226 and T2-8 are indicated by black arrows. The *V. harveyi* clade is outlined in blue and the *V. mediterranei* clade is outlined in green.

Cloning of the BCB226 genome

BCB226 is viable with at least 14-15 copies of IS451-like IS10s in its genome, but any one of those copies could be either in intergenic space or in nonessential genes. It is also unknown whether all of the copies predicted by the Southern Blot are IS451, or if they include IS226-A and/or IS226-B. Cloning was performed in order to determine where IS451 resides in the genome of BCB225.

A partial digest of BCB226 DNA was cloned into Xl10 competent *E. coli*, but the colony hybridization yielded false positives, with every colony being hybridized to the probe for IS451. A second cloning and colony hybridization in Top10 cells revealed once again that

all the colonies came back as false positives. This was comparable to the results of the Southern Blot to look for IS451 in *Vibrio* isolates. T2-8 was a false positive and contains IST2-8, which is similar enough to IS451 to hybridize to the IS451 probe. The XL10 and Top10 competent *E. coli* cells used during cloning probably contain one or more IS10s that are similar to IS451.

The same issue was encountered when a PCR method was attempted instead of using a radiolabelled probe. Colonies of transformed XL10 were selected at random for a colony PCR using primers for IS451. The same bands were in each clone, suggesting that the primers were picking up IS10s within the *E. coli*.

Discussion

High number of copies of IS451

BCB451 has at least 12 copies of IS451 in its genome. Since the majority of sequenced Proteobacteria contain fewer than 8 copies of IS4s in their genomes, this is unusual (De Palmenaer et al. 2008). BCB226 also has a large number of copies, with approximately 14-15 IS451-like IS10s in its genome. These copies could all be IS451, but if the probe for IS451 is able to hybridize with IS226-A or IS226-B, as it was able to hybridize with IST2-8 and another IS10 in *E. coli*, then there may be fewer than 11 copies of IS451. There would be at least one copy of IS451, one copy of IS226-A, and at least 2 copies of IS226-B. There are two possible explanations for why both BCB451 and BCB226 have such large numbers of IS10s. One possibility is that IS451 is naturally able to produce greater quantities of transposase. The other is that both BCB451 and BCB226 have mutations that increase the amount of transposition.

There are different strengths of promoters in IS10s. Of the two IS10s that make up the Tn10 transposon, only IS10R is functional. IS10L has only an 18 bp difference from IS10R, but 7 of those differences are in the regulatory region, and no transposase can be produced from IS10L. IS10R, despite producing a functional transposase, has a weaker promoter than the promoter for the antisense sRNA, and few transposases are produced (cited in Ross et al. 2010). If IS451 could bind RNA polymerase more strongly, more transposase could be produced. There could also be a defect in the antisense sRNA in IS451. The antisense sRNA that blocks translation must necessarily be present, since it is complementary to the strand

that codes for the transposase, but if the promoter is weak, then the amount of antisense sRNA that is transcribed could be small.

There are mechanisms in the host genome for normal cellular functions and regulation that can regulate insertion sequences as well. A component of the antisense sRNA regulation is the host protein Hfq. It can block translation of the transposase on its own, but it can also bind the antisense sRNA and reconfigure it from a double-stranded secondary structure to a partially single-stranded formation that can more easily bind to the transposase mRNA (Ellis et al. 2015). If there is a defect in Hfq in either BCB451 or BCB226, that could explain the large number of copies of IS451-like IS10s in their genomes. The two bacteria could also have a defect in DNA methylation. IS10 transposases have a preference for unmethylated or hemimethylated DNA for excision and insertion (cited in Mahillon 1998). If either BCB451 or BCB226 has a defect in the *dam* gene, which is the gene responsible for methylation, then there would be more transposition of the IS10s in their genomes.

Of the two possibilities, a naturally high rate of transposition in IS451 seems more likely. IS451 is very rare in *Vibrio* in the Boca Ciega Bay area. Out of 1,107 isolates, it was found in only two, BCB226 and BCB451 (the isolate in which it was originally found). Considering the number of IS451-positive isolates and that an Hfq or *dam* defect would not be advantageous for the initial transfer of IS451 into the genomes of those positive isolates, the chances that IS451 would only be found in mutants seems small. What is far more likely is that IS451 is very rare in Boca Ciega Bay *Vibrio*, and because it naturally produces more transposase, there are more copies in the two Vibrio in which it happened to be transmitted to.

Transmission of IS451

Since the IS451s found in BCB451 and BCB226 have 100% sequence identity over 96% of the copies in BCB451, the order of transmission cannot be determined. The lack of mutations between the copies in the two bacteria suggests either a direct transmission or transmission with few intermediates. With whole genome sequencing, it might be possible to estimate the transmission order. For example, if IS451 in BCB451 is adjacent to an identical gene that is found in BCB226, then it is possible that a transposon was transferred directly between the two. If said gene is in a different location in the genome as compared to a non-BCB451 *V. harveyi*, as well, then it could be possible that BCB451 received the IS from BCB226. If the transfer event was just the IS with no additional genes, or a plasmid with the IS; however, it would be impossible to determine the order of transmission.

If additional isolates with IS451 could be found, transmission order might be able to be determined. Additional IS451 genes with slight sequence variability could, in principle, allow the determination of an ancestral sequence, thus allowing some estimation of the order of transmission to other species. In *Vibrio* in the Boca Ciega Bay region, however, IS451 is extremely rare. Whereas some IS will be found in multiple genera, others are found in similar species, so finding IS451 could be possible in other genera, but it is most likely to be found in other *Vibrio* species (Siguier et al. 2006). IS451 has no specificity at the species level, being found in *V. harveyi*, and a relative of *V. rotiferianus*. There is also no preference for bioluminescent organisms, as *V. rotiferianus* does not produce any light.

Another possibility is that IS451 is rare in Boca Ciega Bay because it was introduced to the region from another area. Since sampling was done through plating seawater, and these

organisms are transients in seawater that prefer to be in the gut of an organism, they could have been transferred to the Boca Ciega Bay area by hitchhiking in a shrimp or a fish. This could also provide an environment for the transfer of IS451 between BCB451 and BCB226. *V. harveyi* is a pathogen of shrimp and *V. rotiferianus* is found in rotifers, so inhabiting the gut of the same marine organism could have brought the two bacteria into close enough proximity to transfer IS451 (Gomez-Gil et al. 2003).

Diversity of IS10s

The IS10 group of insertion sequences is more diverse than previously thought. IS451, IS226-A, IS226-B, and IST2-8 are all closely related, and help fill in the IS10 subgroup of IS10 that contains IS10R and IS10L from Tn10. Most of the insertion sequences that were most closely related to those found in the Boca Ciega Bay isolates were discovered in various marine bacteria (Siguier et al. 2006). This suggests a great amount of lateral gene transfer among marine bacteria, as well as evolution of insertion sequences within them. Because the majority of the subgroup were found in marine bacteria, it is also possible that the ancestor sequences of IS10A, IS10L, and IS10R were transferred from marine bacteria before they were passed to the pathogens (such as *Pseudomonas aeruginosa* and *Salmonella typhimurium*) in which they are now found (Siguier et al. 2006).

Future directions

The next logical step for confirming the number of copies of IS451, IS226-A, IS226-B and IST2-8 in BCB451, BCB226 and T2-8 is whole genome sequencing. Cloning and PCR for IS451 proved to be inefficient, as IS10s have similar enough sequences, and are so ubiquitous, that hybridizing probes and primers will attach to other IS10 sequences in both genomic DNA as well as the DNA of the competent *E. coli*.

References

Aziz RK, Breitbart M, Edwards RM. 2010. Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Research. 38(13):4207-4217.

Brenner A. MS Thesis 2001. A new subfamily of IS10 insertion sequences discovered in bioluminescent bacteria.

Curcio MJ, Derbyshire KM. 2003. The outs and ins of transposition: from mu to kangaroo. Nature Reviews Molecular Cell Biology. 4(11):865-877.

De Palmenaer D, Siguier P, Mahillon J. 2008. IS4 family goes genomic. BMC Evolutionary Biology. 8(18).

Edgar R. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics 5:113.

Ellis MJ, Trussler RS, Haniford DB. 2015. Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation. Molecular Microbiology. 96(3):633-650.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Gomez-Gil B, Thompson FL, Thompson CC, Swings J. 2003. Vibrio rotiferianus sp. nov., isolated from cultures of the rotifer Brachionus plicatilis. Int J Syst Evol Microbiology. 53(1):239-243.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA Websuite. Nucleic Acids Research. 36: W70-W74. http://rna.tbi.univie.ac.at/

Kearse, M, Moir, R, Wilson, A, Stones-Havas, S, Cheung, M, Sturrock, S, Buxton, S, Cooper, A, Markowitz, S, Duran, C, Thierer, T, Ashton, B, Mentjies, P, and Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12): 1647-1649.

Kennedy AK, Guhathakurta a, Kleckner N, Haniford DB. 1998. Tn10 transposition via a DNA hairpin intermediate. Cell. 95(1):125-134.

Kumar S, Stecher G, and Tamura K. (2015). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution (submitted).

Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Research. 43(W1):W580-4.

Madden T. The BLAST Sequence Analysis Tool. 2002. In: McEntyre J, Ostell J, editors. The NCBI Handbook. National Center for Biotechnology Information. Chapter 16. http://www.ncbi.nlm.nih.gov/books/NBK21097/

Mahillon J, Chandler M. 1998. Insertion Sequences. Microbiology and Molecular Biology Reviews. 62(3):725-774.

Nei M. and Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

O'Grady EA, Wimpee CF. 2008. Mutations in the lux operon of natural dark mutants in the genus Vibrio. Applied and Environmental Microbiology. 74(1): 61-66.

Ross JA, Ellis MJ, Hossain S, Haniford DB. 2013. Hfq restructures RNA-In and RNA-Out and facilitates antisense pairing in the Tn10/IS10 system. RNA. 19(5):670-684.

Sawabe T, Kita-Tsukamoto, K, Thompson F. 2007. Inferring the evolutionary history of Vibrios by means of multilocus sequence analysis. Journal of Bacteriology. 189(21):7932-7936.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34: D32-D36. http://www-is.biotoul.fr

Skipper KA, Andersen PR, Sharma N, Mikkelsen JG. 2013. DNA transposon-based gene vehicles – scenes from an evolutionary drive. Journal of Biomedical Science. 20(92).

Appendix A: Alignment of IS451 in BCB451 and BCB226

```
BCB226    ----------------------------------------------------ATCTTG
BCB451    CTGAAGAATCCCCTAATGATTTTGATAAAAATCATTAAGTTAAGGTAGATGCACATCTTG
                                                              ******

BCB226    TCGTATGATCAAATGGTTCTGCGAAAAATCAATAACCAGACAACAAGATGTGCGAACTCA
BCB451    TCGTATGATCAAATGGTTCTGCGAAAAATCAATAACCAGACAACAAGATGTGCGAACTCA
          ************************************************************

BCB226    ATATTTTACACGACTCCCTTTATCAATTCTGCCCTGAACTGCACTTAAAGCGGCTTAATA
BCB451    ATATTTTACACGACTCCCTTTATCAATTCTGCCCTGAACTGCACTTAAAGCGGCTTAATA
          ************************************************************

BCB226    GCTTAATGTTAGCCTGCCGTGCTTTACTCGATACCAAAACGTTGACACTCACTGAACTTG
BCB451    GCTTAATGTTAGCCTGCCGTGCTTTACTCGATACCAAAACGTTGACACTCACTGAACTTG
          ************************************************************

BCB226    GCCGCAATTTGCCCAACCAAGCGAGGACAAAGCACAACATTAAGCGCATGGATCGATTAC
BCB451    GCCGCAATTTGCCCAACCAAGCGAGGACAAAGCACAACATTAAGCGCATGGATCGATTAC
          ************************************************************

BCB226    TTGGCAACCGTCACCTGCATAAAGAGCGTCTCGCCGTTTATCGATGGCATGCTAGTTTTG
BCB451    TTGGCAACCGTCACCTGCATAAAGAGCGTCTCGCCGTTTATCGATGGCATGCTAGTTTTG
          ************************************************************

BCB226    TTTGTTCAGGAAATCCAATGCCCATAGTCTTAGTCGATTGGTCTGATATTCGTGAGCATA
BCB451    TTTGTTCAGGAAATCCAATGCCCATAGTCTTAGTCGATTGGTCTGATATTCGTGAGCATA
          ************************************************************

BCB226    AACGATTGATGGCGTTACGTGCATCCATTGCGCTTCACGGTCGCTCTATTACCCTCTATG
BCB451    AACGATTGATGGCGTTACGTGCATCCATTGCGCTTCACGGTCGCTCTATTACCCTCTATG
          ************************************************************

BCB226    AGAAAGCTTTTCCTCTATCTCAGCAATGCTCTAAGTCGGCACATGATCAGTTTTTAGCGG
BCB451    AGAAAGCTTTTCCTCTATCTCAGCAATGCTCTAAGTCGGCACATGATCAGTTTTTAGCGG
          ************************************************************

BCB226    ATTTGGCTAGCATCTTACCGCCTAAGGTCACGCCACTCATCGTCACCGATGCTGGCTTTA
BCB451    ATTTGGCTAGCATCTTACCGCCTAAGGTCACGCCACTCATCGTCACCGATGCTGGCTTTA
          ************************************************************

BCB226    AGGTTCCTTGGTTCAAATCCGTTGAAGAGCATGGTTGGTTTTGGTTAAGTCGAGTCCGTG
BCB451    AGGTTCCTTGGTTCAAATCCGTTGAAGAGCATGGTTGGTTTTGGTTAAGTCGAGTCCGTG
          ************************************************************

BCB226    GCAAGGTCCAGTTCGCCGAGGTTGGCTCTGAGCACTGGGTCTCGATAAGCAAGTTGTACA
BCB451    GCAAGGTCCAGTTCGCCGAGGTTGGCTCTGAGCACTGGGTCTCGATAAGCAAGTTGTACA
          ************************************************************
```

```
BCB226    CTCAAGCGTCCAATCGACCTAAAACTCTAGGTTACAAAAAGCTCACAAAAAGTAATCCAA
BCB451    CTCAAGCGTCCAATCGACCTAAAACTCTAGGTTACAAAAAGCTCACAAAAAGTAATCCAA
          ************************************************************

BCB226    TCTACTGCCATATGTCGCTGTACAAGTCGTTACCCAAAGGAAGAAAGAGCCAACGTTCGA
BCB451    TCTACTGCCATATGTCGCTGTACAAGTCGTTACCCAAAGGAAGAAAGAGCCAACGTTCGA
          ************************************************************

BCB226    CAAGAACCAACTGCCATCATCCATCATCTAAAGTGTACTCTGCCGCAACCAAGGAACCTT
BCB451    CAAGAACCAACTGCCATCATCCATCATCTAAAGTGTACTCTGCCGCAACCAAGGAACCTT
          ************************************************************

BCB226    GGGTTCTAGCAACGAACCTTCCTCCAACGACTCGAACTCCAAAGCAGCTAGTCAGGCTTT
BCB451    GGGTTCTAGCAACGAACCTTCCTCCAACGACTCGAACTCCAAAGCAGCTAGTCAGGCTTT
          ************************************************************

BCB226    ATGCTAAACGCATGCAAATTGAAGAAACGTTCCGTGATTTAAAGAGTCCTGCTTATGGAT
BCB451    ATGCTAAACGCATGCAAATTGAAGAAACGTTCCGTGATTTAAAGAGTCCTGCTTATGGAT
          ************************************************************

BCB226    TAGGGCTAAGACAAAGTCGCACTAACAGCCCAGAGCGTTTCGACATAATTTTACTGATCG
BCB451    TAGGGCTAAGACAAAGTCGCACTAACAGCCCAGAGCGTTTCGACATAATTTTACTGATCG
          ************************************************************

BCB226    CACTCATGCTACAACTCATGTTCTGGTTAGCAGGTTTGCACGCTCAAAAGCAAGGCTGGA
BCB451    CACTCATGCTACAACTCATGTTCTGGTTAGCAGGTTTGCACGCTCAAAAGCAAGGCTGGA
          ************************************************************

BCB226    ACAGGCATTTCCAAGCGAATACCGTGCGAACAAGAAATGTGTTATCAACCGTTCGGTTGG
BCB451    ACAGGCATTTCCAAGCGAATACCGTGCGAACAAGAAATGTGTTATCAACCGTTCGGTTGG
          ************************************************************

BCB226    GTATGGAAGTTCTACGAAGGTCAAACTATCGCATCACTACTCAAGAGCTACTGGCCGCGG
BCB451    GTATGGAAGTTCTACGAAGGTCAAACTATCGCATCACTACTCAAGAGCTACTGGCCGCGG
          ************************************************************

BCB226    GTACCCAACTTACCCAGCAACTTTACCAAAATGGCTATGCNTTGGGNGAT----------
BCB451    GTACCCAACTTACCCAGCAACTTTACCAAAATGGCTATGCGTTGGGTGATTTATGAGGGG
          ****************************************** ***** ***

BCB226    ---------
BCB451    ATCCTTCAG
```

Appendix B: Alignment of Insertion Sequences found within Boca Ciega Bay *Vibrio*

Alignment of IS451, IS226-A, IS226-B and IST2-8

```
IS451    CTGAAGAATCCCCTAATGATTTTGATAAAAATCATTAAGTTAAGGTAGATGCACATCTTG
IST2-8   ------------------------------------------------------------
IS226-A  ------------------------------------------------------------
IS226-B  ------------------------------------------------------ATCTTG

IS451    TCGTATGATCAAATGGTTCTGCGAAAAATCAATAACCAGACAACAAGATGTGCGAACTCA
IST2-8   ------------------------------------------------------------
IS226-A  ------------------------------------------------------------
IS226-B  TCATGTGATCAAATGGTTTCGTCAAAAAACAACAACCAAACAACAAGATGTGCGAACTCA

IS451    ATATTTTACACGACTCCCTTTATCAATTCTGCCCTGAACTGCACTTAAAGCGGCTTAATA
IST2-8   ------------------------------------------------------------
IS226-A  ----------------------------------TGAATTGCACTTAAAACGACTCAACA
IS226-B  ATATCTTACATGACTCTCTCTACCAATTCTGCCCTGAACTGCACTTAAAGCGACTTAATA

IS451    GCTTAATGTTAGCCTGCCGTGCTTTACTCGATACCAAAACGTTGACACTCACTGAACTTG
IST2-8   ------------------------------------------------------------
IS226-A  GTTTAACGCTGGCTTGCCACGCATTACTTGAATGTAAAACGCTCAATCTTACCGAACTTG
IS226-B  GCTTAATGCTAGCTTGCCGAGCATTGCTTGATAGCAAAACCCTTACGCTCACCGAACTTG

IS451    GCCGCAATTTGCCCAACCAAGCGAGGACAAAGCACAACATT-----AAGCGCATGGATCG
IST2-8   ------------------------------------------------------------
IS226-A  GTCGTAACCTACCAAAGAAAGCCAGAACAAAACATAACATAACATCAAACGAATCGACCG
IS226-B  GTCGCAACTTACCTTGCCAAGCTAGAACTAAGCACAATATA-----AAGCGGATGGATCG

IS451    ATTACTTGGCAACCGTCACCTGCATAAAGAGCGTCTCGCCGTTTATCGATGGCATGCTAG
IST2-8   ------------------------------------------------------------
IS226-A  ATTGTTAGGTAATCGTCACTTACACAAAGAGCGACTCGCTGTATACCGCTGGCATGCTAG
IS226-B  TCTTTTAGGTAACCACCACCTGCATCAAGAGAGACTCGCCGTTTATCGCTGGCATGCTAG

IS451    TTTTGTTTGTTCAGGAAATCCAATGCCCATAGTCTTAGTCGATTGGTCTGATATTCGTGA
IST2-8   ------------------------------------------------------------
IS226-A  CTTTATCTGTTCGGGCAATTCGATGCCCATTGTTCTTGTTGATTGGTCTGATATTCGTGA
IS226-B  CTTTATCTGTGCGGGTAACCCCATGSCAGTGGTTCTCGTCGATTGGTCTGATATCCGTGA

IS451    GCATAAACGATTGATGGCGTTACGTGCATCCATTGCGCTTCACGGTCGCTCTATTACCCT
```

```
IST2-8   ------------------------------------------------------------
IS226-A GCAAAAACGGCTTATGGTATTGCGAGTTTCAGTGGCGCTACAGGGGCGTTCTATTACTCT
IS226-B ACACAAGAGACTTATGGTGCTCCGAGCATCAGTGGCGCTACAAGGGCGTTCGGTCACTTT


IS451    CTATGAGAAAGCTTTTCCTCTATCTCAGCAATGCTCTAAGTCGGCACATGATCAGTTTTT
IST2-8   ------------------------------------------------------------
IS226-A TTATGAGAAAGCGTTTCCGCTTTCAGAGCAATGTTCAAAGACAGCTCATGACCAATTTCT
IS226-B ATATGAAAAGGCCTTTCCACTTTCAAAGCAGTGTTCTAAGTCAGCTCATGACCAGTTTTT


IS451    AGCGGATTTGGCTAGCATCTTACCGCCTAAGGTCACGCCACTCATCGTCACCGATGCTGG
IST2-8   ------------------------------------------------------------
IS226-A AGCCGACCTCGCAAGCATTCTACCGAGTAGCACTACACCGCTCATTGTCAGTGATGCAGG
IS226-B ATCTGACCTGGCAAGCATTTTACCTTCTTCTGTGACCCCGCTCATTATCAGTGATGCTGG


IS451    CTTTAAGGTTCCTTGGTTCAAATCCGTTGAAGAGCATGGTTGGTTTTGGTTAAGTCGAGT
IST2-8   -------------------------GTCGAGCA-CATGGCTGGTTTTGGTTAAGTCGAGT
IS226-A CTTTAAAGTGCCATGGTATGCATCTGTTGAAAAGCAGGGTTGGTACTGGTTAAGCCGAGT
IS226-B ATTTAAAGTGCCATGGTACAAGTCTGTTGAGGAGCATGGTTGGTACTGGTTAAGTCGAGT
                                  ** **  * ** ** **** ******** *****

IS451    CCGTGGCAAGGTCCAGTTCGCCGAGGTTGGCTCTGAGCACTGGGTCTCGATAAGCAAGTT
IST2-8   ACGTGGCAAGGTTCAGTTTGCAGAGCTTGGCTCTGAGCACTGGAGTCCAATCAGTAAGTT
IS226-A AAGAGGGAAAGTTCAATATGCAGACTTAGGTGCTGAAAACTGGAACTCTGTCAGTAGGTT
IS226-B AAGAGGTAAAATCCAATTTGCTGAACTCGGTGCTGAAAACTGGCAACCTGTCAGTAGCTT
          * ** **  * ** *  ** **  * **  **** *****    *  * ** *  **

IS451    GTACACTCAAGCGTCCAATCGACCTAAAACTCTAGGTTACAAAAAGCTCACAAAAAGTAA
IST2-8   GCACGCTCAGGCATCCAGTAGACCTAAAACATTAGGCTACAAAAAACTCACCAAAAGCAA
IS226-A ACATAGCATTGCCTCAAGTCGTCCAAAGACTTTAGGTTATAAGAGGCTGACTAAAAGTAA
IS226-B ACATAGCAAAGCCTCTAGCAGAGCTAAAAGTCTTGGTTATCAAAAACTGACCCAAAGCAA
          *       ** ** *   *  * ** *   * ** **  * *  ** **  **** **

IS451    TCCAATCTACTGCCATATGTCGCTGTACAAGTCGTTACCCAAAGGAAGAAAGAGCCAACG
IST2-8   TCCAATCCATTGCCACATGGCACTCTATAAGTCGTTACCCAAAGGAAGAAAAAACCAGCG
IS226-A TGCAATCTCATGTCAAATTGCACTGTATAAACCTTGCTCTAAAGGTCGAAAAAATCAGCG
IS226-B TGCCATCAACTGCCAAATAGCGCTGTATAGAGCATTGCCTAAAGGCCGCAAGAACCAGCG
        * * ***   ** ** **  * ** ** *   * *   * *****  * ** *  ** **

IS451    TTCGACAAGAACCAACTGCCATCATCCATCATCTAAAGTGTACTCTGCCGCAACCAAGGA
IST2-8   TTCGACGCGAACCAACTGTCATCATCCATCATCGAAAATATACTCTGCCGCGACCAAGGA
IS226-A TTCGACAAGGACTAATTGTCATCATTCGTCACCTAAAATTTACTCTGCGTCAGCAAAAGA
IS226-B TTCTACCCGAACAAACTGTCACCACCCATCGCCCAAGGTGTACTCTGACTCAGCAAAGGA
        *** **  * ** ** ** ** **  * **  * **  * *******   *  * ** **

IS451    ACCTTGGGTTCTAGCAACGAACCTTCCTCCAACGACTCGAACTCCAAAGCAGCTAGTCAG
IST2-8   GCCTTGGGTGCTGGCGACGAATCTTCCTCCATCAACCCGAACACCAAAGCAGCTGGTCAA
IS226-A ACCGTGGGTTCTGGCGACTAACTTACCCATTGAAACTCGAACCCCCCAGCAACTTGTTAG
IS226-B GCCATGGGTATTGGCAACTAACTTACCAACGGCGGCTCGTAGTCCTAAGCAGTTGGTGAG
         ** *****  * ** ** **  * **        * ** *  **  ****  * ** *
```

```
IS451    GCTTTATGCTAAACGCATGCAAATTGAAGAAACGTTCCGTGATTTAAAGAGTCCTGCTTA
IST2-8   TCTTTATGCTAAACGCATGCAAATTGAAGAAACTTTCCGTGACTTAAAAAGTCCAGCTTA
IS226-A  GCTCTACTCGAAGCGTATGTAGATTGAAGAAACATTCCGAGACTTGAAAAGCCCTGCGTA
IS226-B  GCTTTACTCAAAGCGCATGCAGATAGAGGAAACCTTCCGCGACTTAAAGAGT--------
          ** **  * ** ** *** * ** ** ***** ***** ** ** ** **

IS451    TGGATTAGGGCTAAGACAAAGTCGCACTAACAGCCCAGAGCGTTTCGACATAATTTTACT
IST2-8   CGGGCTCGGTCTCAGACAAAGCAGAACTCACTCCCCAGAGCGCTTCGACATTATTTTACT
IS226-A  CGGATTAGGGTTGCGTCATAGCCGAACGAGCA----------------------------
IS226-B  ------------------------------------------------------------

IS451    GATCGCACTCATGCTACAACTCATGTTCTGGTTAGCAGGTTTGCACGCTCAAAAGCAAGG
IST2-8   GATCGCACTCATGTTACAACTCATGTTTTGGTTAGCAGGTTTGCACGCTCAAAAGCAAGG
IS226-A  ------------------------------------------------------------
IS226-B  ------------------------------------------------------------

IS451    CTGGAACAGGCATTTCCAAGCGAATACCGTGCGAACAAGAAATGTGTTATCAACCGTTCG
IST2-8   CTGGGACAGACACTTCCAGGCGAATACCGTGCGAATAAGGAGTGTATTATCAACCGTTCG
IS226-A  ------------------------------------------------------------
IS226-B  ------------------------------------------------------------

IS451    GTTGGGTATGGAAGTTCTACGAAGGTCAAACTATCGCATCACTACTCAAGAGCTACTGGC
IST2-8   CTTAGGTATGGAAGTTCTACGAAGATCAGACTACGAAATCACAACGTTAGAACTCCTAGA
IS226-A  ------------------------------------------------------------
IS226-B  ------------------------------------------------------------

IS451    CGCGGGTACCCAACTTACCCAGCAACTTTACCAAAATGGCTATGCGTTGGGTGATTTATG
IST2-8   AGCAGGGGCAGAGCTTACCCGGCAACTTCACCAATAT----------------------
IS226-A  ------------------------------------------------------------
IS226-B  ------------------------------------------------------------

IS451    AGGGGATCCTTCAG
IST2-8   --------------
IS226-A  --------------
IS226-B  --------------
```

IS451 Aligned with IS226-A

```
IS451     TGAACTGCACTTAAAGCGGCTTAATAGCTTAATGTTAGCCTGCCGTGCTTTACTCGATAC
IS226-A   TGAATTGCACTTAAAACGACTCAACAGTTTAACGCTGGCTTGCCACGCATTACTTGAATG
          **** ********** ** ** ** ** **** * * ** ****  ** ***** **

IS451     CAAAACGTTGACACTCACTGAACTTGGCCGCAATTTGCCCAACCAAGCGAGGACAA----
IS226-A   TAAAACGCTCAATCTTACCGAACTTGGTCGTAACCTACCAAAGAAAGCCAGAACAAAACA
          ****** * *  ** ** ******** ** **  * ** **  **** ** ****

IS451     -AGCACAACATTAAGCGCATGGATCGATTACTTGGCAACCGTCACCTGCATAAAGAGCGT
IS226-A   TAACATAACATCAAACGAATCGACCGATTGTTAGGTAATCGTCACTTACACAAAGAGCGA
          * ** ***** ** ** ** ** *****  * ** ** ****** * ** ********

IS451     CTCGCCGTTTATCGATGGCATGCTAGTTTTGTTTGTTCAGGAAATCCAATGCCCATAGTC
IS226-A   CTCGCTGTATACCGCTGGCATGCTAGCTTTATCTGTTCGGGCAATTCGATGCCCATTGTT
          ***** ** ** ** ********** *** * ***** ** *** * ******** **

IS451     TTAGTCGATTGGTCTGATATTCGTGAGCATAAACGATTGATGGCGTTACGTGCATCCATT
IS226-A   CTTGTTGATTGGTCTGATATTCGTGAGCAAAAACGGCTTATGGTATTGCGAGTTTCAGTG
          * ** ********************** *****  * ****  ** ** *  **  *

IS451     GCGCTTCACGGTCGCTCTATTACCCTCTATGAGAAAGCTTTTCCTCTATCTCAGCAATGC
IS226-A   GCGCTACAGGGGCGTTCTATTACTCTTTATGAGAAAGCGTTTCCGCTTTCAGAGCAATGT
          ***** ** ** ** ******** ** ********** ***** ** **  *******

IS451     TCTAAGTCGGCACATGATCAGTTTTTAGCGGATTTGGCTAGCATCTTACCGCCTAAGGTC
IS226-A   TCAAAGACAGCTCATGACCAATTTCTAGCCGACCTCGCAAGCATTCTACCGAGTAGCACT
          ** *** * ** ***** ** *** **** **  * ** *****  *****  **

IS451     ACGCCACTCATCGTCACCGATGCTGGCTTTAAGGTTCCTTGGTTCAAATCCGTTGAAGAG
IS226-A   ACACCGCTCATTGTCAGTGATGCAGGCTTTAAAGTGCCATGGTATGCATCTGTTGAAAAG
          ** ** ***** ****  ***** ******** ** ** ****    *** ****** **

IS451     CATGGTTGGTTTTGGTTAAGTCGAGTCCGTGGCAAGGTCCAGTTCGCCGAGGTTGGCTCT
IS226-A   CAGGGTTGGTACTGGTTAAGCCGAGTAAGAGGGAAAGTTCAATATGCAGACTTAGGTGCT
          ** *******  ******** *****  * ** ** ** ** *  ** **  * **  **

IS451     GAGCACTGGGTCTCGATAAGCAAGTTGTACACTCAAGCGTCCAATCGACCTAAAACTCTA
IS226-A   GAAAACTGGAACTCTGTCAGTAGGTTACATAGCATTGCCTCAAGTCGTCCAAAGACTTTA
          **  *****  ***  * ** * ***  * *    ** ** * *** ** ** *** **

IS451     GGTTACAAAAAGCTCACAAAAAGTAATCCAATCTACTGCCATATGTCGCTGTACAAGTCG
IS226-A   GGTTATAAGAGGCTGACTAAAAGTAATGCAATCTCATGTCAAATTGCACTGTATAAACCT
          ***** ** * *** ** ********* ******  ** ** **  * ***** **  *

IS451     TTACCCAAAGGAAGAAAGAGCCAACGTTCGACAAGAACCAACTGCCATCATCCATCATCT
IS226-A   TGCTCTAAAGGTCGAAAAAATCAGCGTTCGACAAGGACTAATTGTCATCATTCGTCACCT
          *   * *****  **** *  ** ********** ** ** ** ****** * *** **

IS451     AAAGTGTACTCTGCCGCAACCAAGGAACCTTGGGTTCTAGCAACGAACCTTCCTCCAACG
```

```
IS226-A   AAAATTTACTCTGCGTCAGCAAAAGAACCGTGGGTTCTGGCGACTAACTTACCCATTGAA
          *** * ********  ** * ** ***** ******** ** ** *** * **

IS451     ACTCGAACTCCAAAGCAGCTAGTCAGGCTTTATGCTAAACGCATGCAAATTGAAGAAACG
IS226-A   ACTCGAACCCCCCCAGCAACTTGTTAGGCTCTACTCGAAGCGTATGTAGATTGAAGAAACA
          ******** **  **** ** ** ***** **  * ** ** *** * **********

IS451     TTCCGTGATTTAAAGAGTCCTGCTTATGGATTAGGGCTAAGACAAAGTCGCACTAACA
IS226-A   TTCCGAGACTTGAAAAGCCCTGCGTACGGATTAGGGTTGCGTCATAGCCGAACGAGCA
          ***** ** ** ** ** ***** ** ********* *  * ** ** ** ** * **
```

IS451 aligned with IS226-B

```
IS451     ATCTTGTCGTATGATCAAATGGTTCTGCGAAAAATCAATAACCAGACAACAAGATGTGCG
IS226-B   ATCTTGTCATGTGATCAAATGGTTTCGTCAAAAAACAACAACCAAACAACAAGATGTGCG
          ******** * *************  *  ***** *** ***** ***************

IS451     AACTCAATATTTTACACGACTCCCTTTATCAATTCTGCCCTGAACTGCACTTAAAGCGGC
IS226-B   AACTCAATATCTTACATGACTCTCTCTACCAATTCTGCCCTGAACTGCACTTAAAGCGAC
          ********** ***** ***** ** ** ****************************** *

IS451     TTAATAGCTTAATGTTAGCCTGCCGTGCTTTACTCGATACCAAAACGTTGACACTCACTG
IS226-B   TTAATAGCTTAATGCTAGCTTGCCGAGCATTGCTTGATAGCAAAACCCTTACGCTCACCG
          ************** **** ***** ** ** ** **** ******  * ** ***** *

IS451     AACTTGGCCGCAATTTGCCCAACCAAGCGAGGACAAAGCACAACATTAAGCGCATGGATC
IS226-B   AACTTGGTCGCAACTTACCTTGCCAAGCTAGAACTAAGCACAATATAAAGCGGATGGATC
          ******* ***** ** **   ****** ** ** ******** ** ***** *******

IS451     GATTACTTGGCAACCGTCACCTGCATAAAGAGCGTCTCGCCGTTTATCGATGGCATGCTA
IS226-B   GTCTTTTAGGTAACCACCACCTGCATCAAGAGAGACTCGCCGTTTATCGCTGGCATGCTA
          *  *  * ** ****  ********* ***** * ************** **********

IS451     GTTTTGTTTGTTCAGGAAATCCAATGCCCATAGTCTTAGTCGATTGGTCTGATATTCGTG
IS226-B   GCTTTATCTGTGCGGGTAACCCCATGSCAGTGGTTCTCGTCGATTGGTCTGATATCCGTG
          * *** * *** * ** ** ** *** *  * **  * ***************** ****

IS451     AGCATAAACGATTGATGGCGTTACGTGCATCCATTGCGCTTCACGGTCGCTCTATTACCC
IS226-B   AACACAAGAGACTTATGGTGCTCCGAGCATCAGTGGCGCTACAAGGGCGTTCGGTCACTT
          * ** **  ** * **** * * ** *****  * ***** ** ** ** **  * **

IS451     TCTATGAGAAAGCTTTTCCTCTATCTCAGCAATGCTCTAAGTCGGCACATGATCAGTTTT
IS226-B   TATATGAAAAGGCCTTTCCACTTTCAAAGCAGTGTTCTAAGTCAGCTCATGACCAGTTTT
          * ***** ** ** ***** ** **  **** ** ******** ** ***** *******

IS451     TAGCGGATTTGGCTAGCATCTTACCGCCTAAGGTCACGCCACTCATCGTCACCGATGCTG
IS226-B   TATCTGACCTGGCAAGCATTTTACCTTCTTCTGTGACCCCGCTCATTATCAGTGATGCTG
          ** * **  **** ***** *****  **   ** ** ** *****  ***  *******
```

```
IS451     GCTTTAAGGTTCCTTGGTTCAAATCCGTTGAAGAGCATGGTTGGTTTTGGTTAAGTCGAG
IS226-B   GATTTAAAGTGCCATGGTACAAGTCTGTTGAGGAGCATGGTTGGTACTGGTTAAGTCGAG
          * ***** ** ** **** *** ** ***** *************  *************

IS451     TCCGTGGCAAGGTCCAGTTCGCCGAGGTTGGCTCTGAGCACTGGGTCTCGATAAGCAAGT
IS226-B   TAAGAGGTAAAATCCAATTTGCTGAACTCGGTGCTGAAAACTGGCAACCTGTCAGTAGCT
          *  * ** **  **** ** ** **  * **  ****  *****    *  * ** *  *

IS451     TGTACACTCAAGCGTCCAATCGACCTAAAACTCTAGGTTACAAAAAGCTCACAAAAAGTA
IS226-B   TACATAGCAAAGCCTCTAGCAGAGCTAAAAGTCTTGGTTATCAAAAACTGACCCAAAGCA
          *  * *   **** ** *   ** ****** *** *****  **** ** **  **** *

IS451     ATCCAATCTACTGCCATATGTCGCTGTACAAGTCGTTACCCAAAGGAAGAAAGAGCCAAC
IS226-B   ATGCCATCAACTGCCAAATAGCGCTGTATAGAGCATTGCCTAAAGGCCGCAAGAACCAGC
          ** * *** ******* **  ******* *   * ** ** *****  * **** *** *

IS451     GTTCGACAAGAACCAACTGCCATCATCCATCATCTAAAGTGTACTCTGCCGCAACCAAGG
IS226-B   GTTCTACCCGAACAAACTGTCACCACCCATCGCCCAAGGTGTACTCTGACTCAGCAAAGG
          **** **  **** ***** ** ** *****  * ** ********** * ** * ****

IS451     AACCTTGGGTTCTAGCAACGAACCTTCCTCCAACGACTCGAACTCCAAAGCAGCTAGTCA
IS226-B   AGCCATGGGTATTGGCAACTAACTTACCAACGGCGGCTCGTAGTCCTAAGCAGTTGGTGA
          * ** *****  * ***** *** * **  *  ** **** * *** ****** * ** *

IS451     GGCTTTATGCTAAACGCATGCAAATTGAAGAAACGTTCCGTGATTTAAAGAGT
IS226-B   GGCTTTACTCAAAGCGCATGCAGATAGAGGAAACCTTCCGCGACTTAAAGAGT
          *******  * ** ******** ** ** ***** ***** ** *********
```

IS451 aligned with IST2-8

```
IS451     GTTGAAGAGCATGGTTGGTTTTGGTTAAGTCGAGTCCGTGGCAAGGTCCAGTTCGCCGAG
IST2-8    GTCG-AGCACATGGCTGGTTTTGGTTAAGTCGAGTACGTGGCAAGGTTCAGTTTGCAGAG
          ** * **  ***** ******************** ********** ***** ** ***

IS451     GTTGGCTCTGAGCACTGGGTCTCGATAAGCAAGTTGTACACTCAAGCGTCCAATCGACCT
IST2-8    CTTGGCTCTGAGCACTGGAGTCCAATCAGTAAGTTGCACGCTCAGGCATCCAGTAGACCT
          *****************    * ** ** ****** ** **** ** **** * *****

IS451     AAAACTCTAGGTTACAAAAAGCTCACAAAAAGTAATCCAATCTACTGCCATATGTCGCTG
IST2-8    AAAACATTAGGCTACAAAAAACTCACCAAAAGCAATCCAATCCATTGCCACATGGCACTC
          *****  **** ******** ***** ***** ********* * ***** *** * **

IS451     TACAAGTCGTTACCCAAAGGAAGAAAGAGCCAACGTTCGACAAGAACCAACTGCCATCAT
IST2-8    TATAAGTCGTTACCCAAAGGAAGAAAAAACCAGCGTTCGACGCGAACCAACTGTCATCAT
          ** ********************** * *** ********  ********** ******

IS451     CCATCATCTAAAGTGTACTCTGCCGCAACCAAGGAACCTTGGGTTCTAGCAACGAACCTT
```

```
IST2-8    CCATCATCGAAAATATACTCTGCCGCGACCAAGGAGCCTTGGGTGCTGGCGACGAATCTT
          ******** *** * ********** ******** ******** ** ** ***** ***

IS451     CCTCCAACGACTCGAACTCCAAAGCAGCTAGTCAGGCTTTATGCTAAACGCATGCAAATT
IST2-8    CCTCCATCAACCCGAACACCAAAGCAGCTGGTCAATCTTTATGCTAAACGCATGCAAATT
          ****** * ** ***** ********** ****  ************************

IS451     GAAGAAACGTTCCGTGATTTAAAGAGTCCTGCTTATGGATTAGGGCTAAGACAAAGTCGC
IST2-8    GAAGAAACTTTCCGTGACTTAAAAAGTCCAGCTTACGGGCTCGGTCTCAGACAAAGCAGA
          ******** ******** ***** ***** ***** **  * ** ** ********  *

IS451     ACTAACAGCCCAGAGCGTTTCGACATAATTTTACTGATCGCACTCATGCTACAACTCATG
IST2-8    ACTCACTCCCCAGAGCGCTTCGACATTATTTTACTGATCGCACTCATGTTACAACTCATG
          *** **  ********* ******** ********************** ***********

IS451     TTCTGGTTAGCAGGTTTGCACGCTCAAAAGCAAGGCTGGAACAGGCATTTCCAAGCGAAT
IST2-8    TTTTGGTTAGCAGGTTTGCACGCTCAAAAGCAAGGCTGGGACAGACACTTCCAGGCGAAT
          ** ************************************ **** ** ***** ******

IS451     ACCGTGCGAACAAGAAATGTGTTATCAACCGTTCGGTTGGGTATGGAAGTTCTACGAAGG
IST2-8    ACCGTGCGAATAAGGAGTGTATTATCAACCGTTCGCTTAGGTATGGAAGTTCTACGAAGA
          ********** *** * *** ************** ** ********************

IS451     TCAAACTATCGCATCACTACTCAAGAGCTACTGGCCGCGGGTACCCAACTTACCCAGCAA
IST2-8    TCAGACTACGAAATCACAACGTTAGAACTCCTAGAAGCAGGGGCAGAGCTTACCCGGCAA
          *** ****    ***** **   *** ** ** *  ** **  *  * ******* ****

IS451     TACCCAGCAACTTTACCAAAATGGCTATGCGTTGGGTGATTTATGAGGGGATCCTTCAG
IST2-8    TACCCGGCAACTTCACCAATAT----------------------------------
          ***** ******* ***** **
```

IS226-A aligned with IS226-B

```
IS226-A   TGAATTGCACTTAAAACGACTCAACAGTTTAACGCTGGCTTGCCACGCATTACTTGAATG
IS226-B   TGAACTGCACTTAAAGCGACTTAATAGCTTAATGCTAGCTTGCCGAGCATTGCTTGATAG
          **** ********** ***** ** ** **** *** *******  ***** *****  *

IS226-A   TAAAACGCTCAATCTTACCGAACTTGGTCGTAACCTACCAAAGAAAGCCAGAACAAAACA
IS226-B   CAAAACCCTTACGCTCACCGAACTTGGTCGCAACTTACCTTGCCAAGCTAGAACTAAGCA
          ***** ** *  ** ************** *** ****     **** ***** ** **

IS226-A   TAACATAACATCAAACGAATCGACCGATTGTTAGGTAATCGTCACTTACACAAAGAGCGA
IS226-B   CAATATA-----AAGCGGATGGATCGTCTTTTAGGTAACCACCACCTGCATCAAGAGAGA
          ** ***     ** ** ** ** **  * ******** *  *** * **  ***** **
```

```
IS226-A   CTCGCTGTATACCGCTGGCATGCTAGCTTTATCTGTTCGGGCAATTCGATGCCCATTGTT
IS226-B   CTCGCCGTTTATCGCTGGCATGCTAGCTTTATCTGTGCGGGTAACCCCATGSCAGTGGTT
          ***** ** ** ************************ **** **  * *** *  * ***

IS226-A   CTTGTTGATTGGTCTGATATTCGTGAGCAAAAACGGCTTATGGTATTGCGAGTTTCAGTG
IS226-B   CTCGTCGATTGGTCTGATATCCGTGAACACAAGAGACTTATGGTGCTCCGAGCATCAGTG
          ** ** ************** ***** ** **  * ********  * ****  ******

IS226-A   GCGCTACAGGGGCGTTCTATTACTCTTTATGAGAAAGCGTTTCCGCTTTCAGAGCAATGT
IS226-B   GCGCTACAAGGGCGTTCGGTCACTTTATATGAAAAGGCCTTTCCACTTTCAAAGCAGTGT
          ******** ********  * *** * ***** ** ** ***** ****** **** ***

IS226-A   TCAAAGACAGCTCATGACCAATTTCTAGCCGACCTCGCAAGCATTCTACCGAGTAGCACT
IS226-B   TCTAAGTCAGCTCATGACCAGTTTTTATCTGACCTGGCAAGCATTTTACCTTCTTCTGTG
          ** *** ************* *** ** * ***** ********* ****   *

IS226-A   ACACCGCTCATTGTCAGTGATGCAGGCTTTAAAGTGCCATGGTATGCATCTGTTGAAAAG
IS226-B   ACCCCGCTCATTATCAGTGATGCTGGATTTAAAGTGCCATGGTACAAGTCTGTTGAGGAG
          ** ********* ********** ** ****************    ********  **

IS226-A   CAGGGTTGGTACTGGTTAAGCCGAGTAAGAGGGAAAGTTCAATATGCAGACTTAGGTGCT
IS226-B   CATGGTTGGTACTGGTTAAGTCGAGTAAGAGGTAAAATCCAATTTGCTGAACTCGGTGCT
          ** ***************** *********** *** * **** *** **  * ******

IS226-A   GAAAACTGGAACTCTGTCAGTAGGTTACATAGCATTGCCTCAAGTCGTCCAAAGACTTTA
IS226-B   GAAAACTGGCAACCTGTCAGTAGCTTACATAGCAAAGCCTCTAGCAGAGCTAAAAGTCTT
          ********* *  ********** **********  ***** **  *  * ** * * *

IS226-A   GGTTATAAGAGGCTGACTAAAAGTAATGCAATCTCATGTCAAATTGCACTGTATAAACCT
IS226-B   GGTTATCAAAAACTGACCCAAAGCAATGCCATCAACTGCCAAATAGCGCTGTATAGAGCA
          ****** * *  *****  **** ***** ***   ** ***** ** ******* * *

IS226-A   TGCTCTAAAGGTCGAAAAAATCAGCGTTCGACAAGGACTAATTGTCATCATTCGTCACCT
IS226-B   TTGCCTAAAGGCCGCAAGAACCAGCGTTCTACCCGAACAAACTGTCACCACCCATCGCCC
          *   ******* ** ** ** ******** **  * ** ** ***** **  * ** **

IS226-A   AAAATTTACTCTGCGTCAGCAAAAGAACCGTGGGTTCTGGCGACTAACTTACCCATTGAA
IS226-B   AAGGTGTACTCTGACTCAGCAAAGGAGCCATGGGTATTGGCAACTAACTTACCAACGGCG
          **  * *******  ******** ** ** *****  **** ********** *  *

IS226-A   ACTCGAACCCCCCAGCAACTTGTTAGGCTCTACTCGAAGCGTATGTAGATTGAAGAAACA
IS226-B   GCTCGTAGTCCTAAGCAGTTGGTGAGGCTTTACTCAAAGCGCATGCAGATAGAGGAAACC
          **** *  **  ****  * ** ***** ***** ***** *** **** ** *****

IS226-A   TTCCGAGACTTGAAAAGC
IS226-B   TTCCGCGACTTAAAGAGT
          ***** ***** ** **
```

IS226-A aligned with IST2-8

```
IS226-A   GTTGAAAAGCAGGGTTGGTACTGGTTAAGCCGAGTAAGAGGGAAAGTTCAATATGCAGAC
IST2-8    GTCGAGCA-CATGGCTGGTTTTGGTTAAGTCGAGTACGTGGCAAGGTTCAGTTTGCAGAG
          ** **  * ** ** **** ******** ****** * ** ** ***** * ******

IS226-A   TTAGGTGCTGAAAACTGGAACTCTGTCAGTAGGTTACATAGCATTGCCTCAAGTCGTCCA
IST2-8    CTTGGCTCTGAGCACTGGAGTCCAATCAGTAAGTTGCACGCTCAGGCATCCAGTAGACCT
           * **  **** ****** * ****** *** **       ** ** *** * **

IS226-A   AAGACTTTAGGTTATAAGAGGCTGACTAAAAGTAATGCAATCTCATGTCAAATTGCACTG
IST2-8    AAAACATTAGGCTACAAAAAACTCACCAAAAGCAATCCAATCCATTGCCACATGGCACTC
          ** ** ***** ** ** *  ** ** ***** *** *****   ** ** ** *****

IS226-A   TATAAACCTTGCTCTAAAGGTCGAAAAAATCAGCGTTCGACAAGGACTAATTGTCATCAT
IST2-8    TATAAGTCGTTACCCAAAGGAAGAAAAAACCAGCGTTCGACGCGAACCAACTGTCATCAT
          *****  * *   * *****  ******* **********  * ** ** *********

IS226-A   TCGTCACCTAAAATTTACTCTGCGTCAGCAAAAGAACCGTGGGTTCTGGCGACTAACTTA
IST2-8    CCATCATCGAAAATATACTCTGCCGCGACCAAGGAGCCTTGGGTGCTGGCGACGAATCTT
           * *** * ***** ********  *  * ** ** ** ***** ******** **  *

IS226-A   CCCATTGAAACTCGAACCCCCCAGCAACTTGTTAGGCTCTACTCGAAGCGTATGTAGATT
IST2-8    CCTCCATCAACCCGAACACCAAAGCAGCTGGTCAATCTTTATGCTAAACGCATGCAAATT
          **      *** ***** **  **** ** ** *  ** **  * ** ** *** * ***

IS226-A   GAAGAAACATTCCGAGACTTGAAAAGCCCTGCGTACGGATTAGGGTTGCGTCATAGCCGA
IST2-8    GAAGAAACTTTCCGTGACTTAAAAAGTCCAGCTTACGGGCTCGGTCTCAGACAAAGCAGA
          ******** ***** ***** ***** ** ** *****  * **  *  * ** *** **

IS226-A   AC----------GAGCA
IST2-8    ACTCACTCCCCAGAGCG
          **          ****
```

IS226-B aligned with IST2-8

```
IS226-B   GTTGAGGAGCATGGTTGGTACTGGTTAAGTCGAGTAAGAGGTAAAATCCAATTTGCTGAA
IST2-8    GTCGAGCA-CATGGCTGGTTTTGGTTAAGTCGAGTACGTGGCAAGGTTCAGTTTGCAGAG
          ** *** * ***** ****  ************** * ** **  * ** ***** **

IS226-B   CTCGGTGCTGAAAACTGGCAACCTGTCAGTAGCTTACATAGCAAAGCCTCTAGCAGAGCT
IST2-8    CTTGGCTCTGAGCACTGGAGTCCAATCAGTAAGTTGCACGCTCAGGCATCCAGTAGACCT
          ** **  ****  *****   **  ******  ** **     * ** ** ** *** **

IS226-B   AAAAGTCTTGGTTATCAAAAACTGACCCAAAGCAATGCCATCAACTGCCAAATAGCGCTG
IST2-8    AAAACATTAGGCTACAAAAAACTCACCAAAAGCAATCCAATCCATTGCCACATGGCACTC
          ****   * ** **  ******* *** ******** * *** * ***** ** ** **

IS226-B   TATAGAGCATTGCCTAAAGGCCGCAAGAACCAGCGTTCTACCCGAACAAACTGTCACCAC
```

```
IST2-8    TATAAGTCGTTACCCAAAGGAAGAAAAAACCAGCGTTCGACGCGAACCAACTGTCATCAT
          ****   * ** ** *****  * ** ********** ** ***** ******** **

IS226-B   CCATCGCCCAAGGTGTACTCTGACTCAGCAAAGGAGCCATGGGTATTGGCAACTAACTTA
IST2-8    CCATCATCGAAAATATACTCTGCCGCGACCAAGGAGCCTTGGGTGCTGGCGACGAATCTT
          *****  * **  * ******* * *  * ******** *****  **** ** **  *

IS226-B   CCAACGGCGGCTCGTAGTCCTAAGCAGTTGGTGAGGCTTTACTCAAAGCGCATGCAGATA
IST2-8    CCTCCATCAACCCGAACACCAAAGCAGCTGGTCAATCTTTATGCTAAACGCATGCAAATT
          **  *  *  * ** *  ** ****** **** *  *****  * ** ******** **

IS226-B   GAGGAAACCTTCCGCGACTTAAAGAGT
IST2-8    GAAGAAACTTTCCGTGACTTAAAAAGT
          ** ***** ***** ******** ***
```