UNIVERSITY OF CALIFORNIA, SAN DIEGO

Insights in Two Plasmid Families from the SIO Pier Metagenome

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Science

in

Biology

by

Elizabeth Anne Latham

Committee in charge:

Brian Palenik, Chair
Eric Allen, Co-chair
Bianca Brahamsha
Susan Golden

2010

The thesis of Elizabeth Anne Latham is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____
Co-chair

____
Chair

University of California, San Diego
2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My deepest thanks to my advisor Brian Palenik. Thanks for keeping me around. I would also like to give a special thanks to my committee Biance Branhamsa, Eric Allen, and Susan Golden. I am sincerely grateful for their support and guidance in all things science and non-science.

Thanks to the whole Palenik Lab: Emy Daniels, Rhona Stuart, Todd Johnson and to Chris Dupont for getting me started and harding my skin.

Without my family and friends this thesis would not be possible. Thank you for putting up with me.

ABSTRACT OF THE THESIS

Insights in Two Plasmid Families from the SIO Pier Metagenome

by

Elizabeth Anne Latham

Master of Science in Biology

University of California, San Diego, 2010

Professor Brian Palenik, Chair
Professor Eric Allen, Co-chair

From coastal California seawater, a complex and diverse environment, the marine cyanobacteria of the genus *Synechococcus* were enriched by flow cytometry-based sorting and the population metagenome was analyzed with 454 sequencing technology. Interestingly, at least three distinct mobile DNA elements not found in model *Synechococcus* strain genomes were detected in the assembled contigs. These contigs were confirmed to be plasmids. p27638e, p31635e, and p31454e, were sequenced and described. p27638e, p31635e, and p314543e are 1425, 1558, and 1479 kb in length, respectively. Further analysis of natural samples has shed light on their diversity over time. A recombinant plasmid, pTOPO27638 was constructed for sequencing and development of a shuttle vector capable of transforming both *Escherichia coli* and *Synechococcus* CC9311. Sixteen presumptive ORFs were identified which may code small peptides. The deduced amino acid sequence of the largest ORFs were significantly similar to that of a putative plasmid replication, *rep*A. Comparison of the replication proteins from p27638e, p31635e and p31454e to other plasmid replication proteins show that it may replicate via rolling circle replication. p27638e and p31635e were found to be positively correlated with marine *Synechococcus* over time in the environment.

Introduction

The unicellular cyanobacteria *Synechococcus* are an ecologically important component of the phytoplankton community in the temperate to tropical oceans (J. Waterbury, Watson, Valois, & Franks, 1986) (J. B. Waterbury, Watson, Guillard, & Brand, 1979). Because *Synechococcus* are one of the most abundant primary producers and are so widespread in the marine environment, they are an important nutrient cycler and as such are responsible for an estimated 20-40% of carbon fixation {Li, 1994}. As expected, *Synechococcus* have become one of the most studied cyanobacterial genera. However, they still hold surprises.

Like the vast majority of prokaryotes, cyanobacteria are known to contain a diverse range of plasmids, which vary in their size, copy number and function (Bose & Carmichael, 1990; Felkner & Barnum, 1988; Miyake, Nagai, Shirai, Kurane, & Asada, 1999; Rebiere, Castets, Houmard, & Demarsac, 1986; Tominaga, Ashida, Sawa, & Ochiai, 1992; Xu & McFadden, 1997). The complete sequences of about 40 cyanobacterial plasmids have been deposited into the Genbank database as of December 2009, however most of these are associated with freshwater, thermophillic or halophilic cyanobacteria including *Anabaena, Microcystis, Leptolyngbya, Synechocystis, Nostoc*, and freshwater *Synechococcus*. Plasmids may be the main means for acquisition of new genetic material or horizontal gene transfer - the mobilization and expression of genetic material across different organisms, proven to be a widespread natural occurrence (Koonin, Makarova, & Aravind, 2001). Despite their potential use in genetic engineering and the role they may play in horizontal gene transfer and evolution, cyanobacterial plasmids have not been studied in detail in the marine environment (Thomas & Nielsen, 2005). There is an especially large void of information when marine cyanobacterial plasmids are considered despite their potential importance. There is evidence to suggest a high incidence of plasmids in marine bacterial communities (P. A. Sobecky, 1999; Patricia A. Sobecky, Mincer, Chang, Toukdarian, & Helinski, 1998; Zhang, Wang, Leung, & Gu, 2007). However, as of yet none have been found to be associated with marine *Synechococcus* despite its widespread distribution and abundance. The initial goal of this study was to show that plasmids occur within the marine cyanobacterial populations, in particular *Synechococcus*.

Putative plasmids were detected in a marine cyanobacterial metagenome population from Oct 10, 2006 from the Scripps Institution of Oceanography (SIO) pier, Ca via 454. Within the metagenome 15 large contigs were discovered that did not represent material from a *Synechococcus* chromosome. When circularized, these contigs showed open reading frames similar to those of plasmid replication proteins (Palenik, Ren, Tai, & Paulsen, 2009). Based on nucleotide identity, these plasmids were broken up into three families. Two of these families were pursued in this study. Contigs 27638 and 31635 were considered one family, while contig 31454 was another. Contig 27638 was constructed of 142 sequencing reads, 31454 of 152 reads, and 31635 of 209 reads. When circularized, a discrete plasmid replication protein was revealed for these two families that resembled previously described plasmids of freshwater cyanobacteria. These contigs were used to design primers that would amplify the entire plasmid, which was then sequenced. Here, we report the complete DNA sequence of plasmid p31454e, p27638e, and p31635e, all of which were amplified from environmental DNA (e is used to denote a plasmid from environmental DNA). Sequence analysis of these plasmids was conducted to determine replication protein and other possible open reading frames.

Many small plasmids, such as the plasmids of this study, replicate by the rolling circle mechanism. This was originally observed in single-stranded DNA bacteriophages of Escherichia coli (Baas & Jansz, 1988). This replication mechanism is more strongly associated with gram positive bacteria, although it is also found in gram negative bacteria and archaea (Holmes, Pfeifer, & Dyallsmith, 1995; Yasukawa, Hase, Sakai, & Masamune, 1991). Rolling circle replication may even be more ubiquitous across hosts than previously thought. It is also observed in animal parvoviruses and mitochondrial DNA in plants (Backert, Meissner, & Borner, 1997; Berns, 1990). Due to its prevalent nature, it would not be surprising to also find rolling circle plasmids associated with marine cyanobacteria. Three structural elements are required for a plasmid to be considered rolling circle replication: i) the replication gene (rep) and its corresponding regulatory elements, ii) double stranded origin (dso), and iii) the single stranded origin (sso). The rolling circle plasmids, or rep 1 superfamily, replicate via a two-step process. In the first step (leading strand synthesis), DNA synthesis is initiated by the replication initiation protein, which recognizes and introduces a nick at the plus origin of replication on the supercoiled DNA. This happens at the double-strand origin. The plus strand is replicated first using host encoded protein, while the second strand is being displaced. The second step is the synthesis of the lagging or second strand. The second strand is initiated by a different region, the single strand origin sso, after the first by the same host machinery (del Solar, Giraldo, Ruiz-Echevarria, Espinosa, & Diaz-Orejas, 1998). The intermediate single strand DNA, which is created due to the uncoupling of the synthesis of the first and second strand, is a defining property of all rolling circle plasmids (del Solar et al., 1998).

The next step was to show that these plasmids were able to replicate in *Synechococcus*. The strains used in this study, *Synechococcus* CC9311 and CC9902, were chosen because they are the dominant coastal strains (Palenik et al., 2009), while *Synechococcus* WH8102 is prevalent in an open ocean environment (J. B. Waterbury, Willey, Franks, Valois, & Watson, 1985). We found plasmids 27638 and 31454 may be able to replicate in *Synechococcus* CC9311 and E. coli. Another objective of this study was to construct an effective vector and protocol for the transfer of DNA fragments in to *Synechococcus* CC9311 and E. coli. To ensure its stable replication in *Synechococcus* cells, a vector would typically have an autonomous replicon. Often plasmids isolated from their host are used as shuttle vectors.

Often plasmids confer an advantage to their host by encoding a favorable trait while favoring their own maintenance. This includes, but is not limited to, antibiotic and heavy metal resistance, toxin production, gas vacuolation, and involvement in adaptation to nutrient deficiencies (Bose & Carmichael, 1990; Chen, Holtman, Magnuson, Youderian, & Golden, 2008; Lau, Sapienza, & Doolittle, 1980). For many plasmids, no obvious function can be found as of yet; they are known as cryptic plasmids. A third possibility exists, that some plasmids may be parasitic i.e. the plasmids do not confer an advantage to their host but only encode traits that ensure their own survival at the cost of their host (Koonin et al., 2001). This study attempts to provide insight into the function of these plasmid families within their host.

The complete genome of the strains used in this study, *Synechococcus* CC9311, CC9902, and WH8102, are available (Palenik et al., 2003; Palenik et al., 2009). The plasmids detected in the metagenomic analysis were not identified in pure culture isolates from which the complete genomes originated (Palenik et al., 2003). This calls into question whether cultured isolates are an accurate reflection of microbes in a natural context. It appears that the genomic variability of natural populations may be larger than cultured isolates. This is reflected in the idea of a pangenome, which is the full set of genes in a given species. A single genome will not represent the pan genome, which may be orders of magnitude larger (Medini, Donati, Tettelin, Masignani, & Rappuoli, 2005). Relatively little is known about plasmid diversity within a single plasmid or family, even less is known about how this could potentially vary with time. To further flush out any missing diversity associated with *Synechococcus*, the two-plasmid families were cloned and sequenced. The replication region and the area surrounding it, known as the cargo, were analyzed at two time points. We decided to look at diversity during May, a time of *Synechococcus* abundance and October when cell numbers are lower. This revealed a surprising level of diversity within each plasmid group. Six cyanobacterial metagenomes have been completed which revealed other contigs that may be plasmids that would fall under these two families (unpublished data). These contigs were included in the analysis of the cargo and repA diversity, but primers were not designed to amplify these other contigs.

In summary, this study investigates the diversity, characteristics, and abundance of two plasmid families detected in the assembled contigs from a marine cyanobacterial population metagenome from the SIO pier, CA. Nucleotide sequences of these plasmids were obtained to study their function and replication means. In the cargo region, we discovered many putative open reading frames, which may indicate the presence of genes. The replication region was examined for the defining characteristics of rolling circle plasmids. Conjugation and electroporation were used to confirm the ability of p27638e and p31454e to replicate in *Synechococcus*. Quantitative PCR (qPCR) was used to assess the abundance of plasmid families 27638 and 31454 at the SIO pier. Temporal variation was also observed.

Material and Methods

Bacterial Strains and Cultivation

Three isolated strains of marine *Synechococcus* were used in this study, CC9311, CC9902, and WH8102. *Synechococcus* sp. CC9902 was isolated from water samples from SIO, San Diego, California. *Synechococcus* sp WH8102 was isolated from the Sargasso Sea (J. B. Waterbury et al., 1985). *Synechococcus* sp CC9311 was originally isolated from the California Coast (Toledo & Palenik, 1997). Samples were cultivated in SN medium (J. Waterbury & Willey, 1988) prepared with seawater from SIO Pier. Cultures were incubated without shaking at 25 °C with constant light intensity (20 microeinsteins m^{-2} s^{1}). SN agar medium was prepared by using purified agarose (J. Waterbury & Willey, 1988) or Seaplaque agarose (FMC). Salts, trace metals, and vitamins were filter sterilized before being added to sterile seawater and agar. Escherichia coli strains were grown in Luria-Bertani (LB) medium broth cultures or by LB agar plating (Maniatis, Fritsch, & Sambrook, 1982). E. coli strains hosting the plasmids were grown overnight at 37 degrees C by standard methodology. The antibiotics chloramphenical (10 µg/ml), kanamycin (50µg/ml), and ampicillin (100µg/ml) were added, when required, to solid and liquid culture medium for the selection and maintenance of plasmids under investigation in E. coli. Kanamycin at 20 µg/ml was used when needed for plasmid maintenance and selection in *Synechococcus*. Plasmid purification kits, used with E. coli, were from Qiagen (Valencia, CA, USA)

Synechococcus was plated in seawater agarose pour plates (Brahamsha, 1996).

For pour plating to obtain single colonies, 200ul of *Synechococcus* sp. strain CC9311 were added to 40 ml of SN medium with 0.3% agar and poured immediately into a Falcon plastic petri dish with cells embedded in the plating medium. The same procedure was done with CC9902 and WH8102. To prepare the plating medium, SN medium was supplemented with 0.3% SeaPlaque and autoclaved to dissolve and sterilize. The liquid agarose solution was then cooled, at which time *Synechococcus* cells and antibiotics were added at appropriate concentrations directly to the liquid agarose SN medium. The agar concentration for plates to be spread was 0.6%. All plates were incubated at 25C at a low light intensity (10 microeinsteins m^{-2} s^{-1}) for 24 hours then moved to a higher light intensity (20 microeinsteins m^{-2} s^{-1}).

PCR conditions and cloning conditions

Basic PCR and the variant touchdown PCR were used. PCR reagents were purchased from Invitrogen (Carlsbad, CA, USA) or Promega (Madison, WI, USA). A total of 40 PCR cycles were run under the following conditions summarized in Table 1. The PCR amplifications were performed with Mastercycler gradient thermocycler, Eppendorf (Hauppauge, NY). Amplified products were analyzed with 1.0% agarose gels, stained with ethidium bromide, and photographed on a UV transilluminator. Single bands were seen in gels of PCR reactions with described primers. To create clone libraries, PCR products were then transformed into pCR4- TOPO vector using Sequencing TOPO® TA Cloning® Kit for Sequencing into TOP 10 chemically competent cells (Invitrogen). Random colonies were picked and grown under manufacturers instructions.

Preparation and cloning of Plasmid DNA

Primers (Table 1) were used to amplify the entire plasmids 27638, 31638, and 31454 from environmental DNA extracted on Oct 10, 2006. The gel purified PCR products were cloned into the ligation site in the commercially available TOPO vector pCR4. The resulting plasmids were designated pTOPO31454e, pTOPO27638e, and pTOPO31635e. This construct was used directly for electroporation.

The recombinant plasmids, pMUT31454e and pMUT27638e, were prepared by ligating EcoR1-digested linear pTOPO31454e and pTOPO27638e to EcoR1 digested linear pMUT100. Plasmid pMUT100 was derived from pBR322 with the 1.23 kb kanamycin cassette from pUK4K cloned into the PstI site as the source of kanamycin resistance marker (Brahamsha, 1996). Plasmids were obtained from 4 ml of an overnight cell culture and transformed into E. coli MC 1060 (pRK24, pRL153) for conjugation.

The constructs were confirmed by restriction analysis and sequencing using primers which flank the EcoR1 sites of pCR4 and pMUT100.

Primer design

Three sets of primers were designed for each plasmid family, one set to target the replication gene, another the area surrounding the repA or cargo, and lastly essentially the whole plasmid.

Oligonucleotide primers were designed based on the contigs 27638 and 31635 as well as contigs 11531 and 31454 which were constructed from 454 sequence reads from the metagenome analysis from the SIO pier on Oct 10, 2006. Primers were designed to amplify regions of the contigs that were not perfectly conserved, so degeneracy was included in the middle of the primer. Primers that would amplify the complete plasmid were designed to exactly match the region of contigs 27638, 31635, 31454, and 11513. A region outside of the repA gene and putative genes was chosen, so hypothetically no plasmid functions would be interrupted.

Primers were synthesized by Integrated DNA technologies Inc. (Coralville, Iowa). Table 1, Fig. 2, 3, and 4 summarizes plasmid primer sequences, product size, primer position and references for respective sequences. For 27638 repA primers were used for qPCR, however for family 31454 new primers were designed to amplify a smaller region.

Cargo and repA Library

Genomic DNA from environmental samples used to construct cargo and repA libraries was extracted by Tai (V Tai & Palenik, 2009). Clone libraries were constructed from environmental DNA extracted on Oct 10, 2006, May 17, 2007, and May 15, 2008. The DNA from May 15, 2008 was collected at a 20m depth.

Environmental clone libraries consisting of the repA region and libraries of the area surrounding the repA gene from the environment DNA for families 27638 and 31454 were obtained. Nucleotide sequencing of DNA was carried out by SeqXcel, San Diego, USA. Sequences in the database exhibiting significant similarity to regions of p27638e and p31454e at the amino acid or nucleotide level (BLASTn and BLASTp $p<10^{-5}$) were aligned (Altschul, Gish, Miller, Myers, & Lipman, 1990).

Sequences were aligned using the CLC Genomics Workbench v3 software (Aarhus, Denmark). Phylogenetic relationships were estimated from nucleotide and predicted protein sequences. Neighbor joining trees were obtained using CLC Genomics. Putative similarities to other known proteins were investigated using NCBI-BLAST(Altschul et al., 1990). Rolling circle analysis, investigating direct and inverted repeats, was also performed by CLC genomics. Putative open reading frames were determined by the online tool NEB cutter V2.0 (Vincze, Posfai, & Roberts, 2003). Promoter motifs were searched for with CLC genomics. Potential ribosomal binding sites (RBS) were examined manually in regions upstream of the transcription start site. E. coli Shine Dalgarno sequence or the sequence complementary to the 3’ end of the 16s rRNA of *Synechococcus* CC9311, CC9902, and WH8102 were used (Bryant, 1994).

Electrotransformation of marine *Synechococcus*

The pTOPO31454e and pTOPO27638e were introduced into the cyanobacteria CC9311 via electroporation by a modification of the procedure described by S. Mazard (personal communication). The transposon bearing plasmid pRL27 was used as a positive control (Larsen, Wilson, Guss, & Metcalf, 2002). Electrocompetent cells were prepared from 10ml of cultures at optical density$_{750}$ between 0.1 to 0.2. Cells were to be prepared fresh and kept on ice until use. Electroporation was performed with approximately 200 ug of plasmid DNA and 20ul of the concentrated cells (2-3.10^{8} cells per electroporation) in 0.5 M sucrose with all media and antibiotics removed via centrifugation and washing with 0.5M sucrose.

Plasmids were introduced into the *Synechococcus* strains by using the following conditions: Capacitor, 25µF; Resistance, 200Ω; voltage, 2.5kV; pulse length, 4.5 to 4.8ms and cuvette gap, 0.2mm using electroporator (BTX, ECM630). Immediately after the voltage pulse, the cells were suspended in SN. Cells were grown in very low light conditions (10 microeinsteins m^{-2} s^{1}) for a day before being moved to a slightly higher light level (20 microeinsteins m^{-2} s^{1}) . After another day cells were moved to normal light levels before being plated into SN plus antibiotics as previously described (20 µg/ml Kanamycin). Single colonies appeared after two weeks, they were then picked and transferred to liquid SN media with antibiotics (20 µg/ml Kanamycin).

Conjugation from E. coli to marine *Synechococcus*

E. coli MC1061 was mated with *Synechococcus* CC9311, CC9902 and WH8102 with the helper plasmids pRK24, pRL528. Conjugation was done according to the method described by Brahamsha (Brahamsha, 1996). Antibiotic resistant colonies started appearing after 2-3 weeks.

CC9311 cultures that were transformed with the Tn5-containing plasmid pRL27 using the same conjugation protocol as the transfer of pMUT27638e or pTOPO27638e. Following conjugation, CC9311 cultures mated with the E. coli donor strain BW20767 containing pRL27 grew under kanamycin the selection (McCarren & Brahamsha, 2005). CC9311 electrotransformed with pRL27 directly also formed colonies under kanamycin selection. There was no growth of CC9311 electrotransformed with no DNA or CC9311 mated with E. coli that have no plasmid to conjugate.

Isolation of Pure *Synechococcus* cultures after Conjugation or Electroporation

Colonies were excised using a sterile pipette tip and transferred back to liquid medium containing 20 ug ml^-1 kanamycin. Putative conjugants were repour plated several times to eliminate contaminants and better isolate colonies. When a visible pellet has grown, samples were checked for the presence of contaminants, by spotting 50 ul on an LB plate (with no antibiotics) and incubating this at room temperature.

Synechococcus plasmid isolation

Plasmid DNA was isolated from transformed *Synechococcus* CC9311 by a modification of the standard bacterial alkaline lysis protocol (Maniatis et al., 1982) (McCarren & Brahamsha, 2005). PCR was performed using primers specific to the rpoC1 gene of CC9311 (V Tai & Palenik, 2009), in addition to pCR4 TOPO vector specific Kanamycin cassette primers or pBR322 specific primers for pMUT100 constructs, and plasmid specific primers. Plasmids recovered from *Synechococcus* 9311 were back-transformed into E. coli Top 10 chemically competent cells and electrocompotent DH5 alpha to confirm presence of plasmids constructs within *Synechococcus* and to indicate if pMUT31454e and pMUT27638e are efficient shuttle vectors. EcoR1 restriction pattern and sequencing were used to confirm.

Abundance of plasmid families using qPCR

The incidence and relative abundance of p31454e and p27638e in nature microbial communities were determined by real time quantitative PCR. The PCR primers amplified plasmid repA fragment from DNA samples from the environment. qPCR, quantitative PCR, was performed on genomic DNA from environmental samples extracted by Tai (V. Tai, Ren, Paulsen, & Palenik, 2008)(Tai, Ren, Paulsen, & Palenik, 2008)

Plasmid DNA of repA clones from environmental libraries ware used as the standards. Serial, 10 fold dilutions of the plasmids DNA were used as standard curves for the repA gene. All qPCR reactions were performed in duplicate in a MX3000P qPCR Thermocycler (Stratagene) according to manufacturer's instructions. At the end of each reaction, the threshold cycle (Ct) value was acquired and analyzed. Standard curves of Ct against the logarithm of the molecular numbers of plasmid DNA per reaction were plotted. The Ct value obtained for an unknown sample was then used to extrapolate the molecular number of plasmid DNA present based on the amount of DNA added and the length of the template assuming that the average weight of a base pair is 650 daltons and Avogadro's number: 6.022x10^{23} (number of copies = (amount * 6.022x10^{23}) / (length * 1x10^{9} * 650))(V Tai & Palenik, 2009)(Tai & Palenik, 2009). The experiment was performed in duplicate and average results are reported.

The results from the repA primer set using samples from the SIO pier time series were compared to results obtained from a previous study using flow cytometry(V Tai & Palenik, 2009). Plasmids numbers were matched analyzed against *Synechococcus* abundance, Pearson correlation coefficient was calculated using Excel.

qPCR was also used to assess abundance or copy numbers of plasmid p27638e per cell inrelative to *Synechococcus* uCC9311 under selection in culture after electrotransformation or conjugation. The repA primers were used along with CC9311 specific primers (Forward rpoC1 F-I 5’ TGA AAG GGA TYC CCA GTT ATG T 3’ and reverse rpoC1 R-I 5’ CCC TTA CTI CCA GCA ATC TC 3’), standards for rpoC1 clade I were derived from Tai (V Tai & Palenik, 2009).

Results

1. The complete sequence of the plasmid p27638e and p31454e

Contigs assembled from the SIO metagenome were shown to be circular genomic elements. PCR was performed on environmental DNA from Oct 10, 2006 and May 17, 2007 using primers near the ends of the contigs that would amplify nearly the whole plasmid. The pcr product showed a single band via gel electrophoresis indicating a circular piece of DNA (Fig. 1). If these contigs were part of a larger genome more than one band would have been expected. PCR products were cloned into E. coli and sequenced.

Whole nucleotide sequences were blasted against NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis) databases (http://camera.calit2.net). There were no significantly similar sequences found outside of SIO cyanobacterial metagenomes. As previously mentioned, SIO pier metagenomes revealed that other assembled contigs may have plasmids that fall within these two families as summarized in appendix 1, however, primers were not designed to incorporate other potential plasmids. When considered on a larger scale, p31454e and contig11513 fall within one group, while p27638e and p31635e are included in another group of all the plasmid like contigs from all of the SIO pier metagenomes (Ma, Yingfei, personal communication).

1.1 p31454e

The whole sequence of circular plasmid p31454e was obtained. The total length of the circular plasmid is 1479 contains a repA gene of 978 nucleotides and molecular weight predicted to be 477.2kDa. The overall G+C content of 31454 is 55.2%. This is within the expected range for marine unicellular *Synechococcus* CC9311, which has a chromosomal G+C content of 52.5%. Fig. 2 shows a schematic of the p31454e DNA sequence which reveals several major structural features, including a total of 5 putative open reading frames, ORFs, of sizes 33, 45, 52, 56, and 61, which are named ORFs a-e respectively, which were identified as putative coding regions (Table 3). Contig31454, 1479 bp, is 96% identical to p31454e. The related putative plasmid contig 11531 is 85% identical on the nucleotide level. However, the repA region is identical at the nucleotide level, so the variation appears in the cargo region. Contig 11531 and p31454e share two ORFs. Primers were designed to include both of these plasmids for further analysis of repA and cargo diversity in the environment.

1.2 p27638e

p27638e is 1425 bp in length and contains a different repA gene of 858 nucleotides and predicted size of 460 kDA along with 4 potential small open reading frames coding for putative proteins of 28, 30, 35, and 28 amino acids, named A-D respectively (Fig. 3). p27638e has a G + C content of 55.4% which is similar to *Synechococcus* CC9311 and p31454e. p27638e is 98% identical to contig27638, 1423bp.

Within the same family is p31635e (Fig. 4), which contains a similar repA gene, is 1558 bp, as compared to contigs31635 of 1553 bp in length and 98 % identical. p31635e is 63% identical to p27638e at the nucleotide level; primers were designed to amplify both plasmids with the exception of the PCR whole plasmid amplification primers. p31635e contains a similar repA protein to p27638e, but the two plasmids only have one potential ORF in common that being D from p27638e and H from p31635e.. p31635e has 6 putative ORFs, sizes 88, 45, 57, 28, 81, and 31 (E-J respectively) (Table 2).

2. p27638e replication in *Synechococcus* CC9311

Next we wanted to see if these plasmids could replicate in marine *Synechococcus* CC9311, the dominant coastal strain. To do this a plasmid was constructed, pMUT26638e, that combined the putative environmental plasmid, p27638e, and pMUT100 containing a kanamycin resistance cassette that cannot self-replicate in *Synechococcus*. After this, conjugation and electrotransformation were applied. CC9311 cultures electrotransformed with pMUT27638e grew under kanamycin selection in liquid culture, indicating that conjugation with E. coli was successful enabled by pMUT27638e. Plating confirmed that E. coli was not a contaminant in CC9311 cultures. The same results were discovered when pTOPO27638e was transformed into CC9311 by electroporation.

Plasmid DNA was extracted from CC9311 via alkaline lysis (Brahamsha, 1996). PCR was performed using the primers specific to rpoC clade 1 (CC9311), TOPO pCR4 kanamycin cassette, 27638 repA family, whole 27638, and cargo primers. These products were amplified and strong single bands were obtained (Fig. 5). To confirm plasmid presence and due to low copy number, plasmids from transformed strain

CC9311 were recovered from back transformed E. coli.

Electrotransformation and conjugation with other plasmid constructs, including pMUT31454e, pMUT31635e, pTOPO31454e, and pTOPO31635e, with CC9311 have also been performed, however E. coli contaminants have yet to be eliminated, so conclusive results cannot be drawn. Preliminary results indicate that plasmids families 27638 and 31454 can also replicate be expressed in *Synechococcus* CC9902, however these two are not pure cultures. To date, we have not been able to show replication any of the plasmid constructs in *Synechococcus* WH8102.

3. Analysis of the repA gene

3.1 Plasmid family 31454

To assess the diversity of the plasmids families, environmental clone libraries were produced for the repA region. The sequences were also analyzed for characteristics of rolling circle plasmids. The repA proteins of rolling circle plasmids typically consists of approximately 300 amino acids(Delsolar, Moscoso, & Espinosa, 1993). The repA gene of p31454e is 327 amino acids. All environmental clones matched p31454e. The repA gene was not detected in environmental samples from May 17, 2007 or May 15, 2008.

Searches using Genbank and CAMERA revealed that the repA protein showed homology with replication proteins encoded by plasmids with known replication initiator protein (rep). repA 31454 was found to have the putative conserved domain for the Rep 1 rolling circle superfamily. A phylogenic analysis suggests that the repA genes of the family 31454 are not more similar to other cyanobacterial plasmids (*Nostoc*, *Synechocystis* and *Cylindrosperms*) than other gram-negative bacteria (*Zymomona*s sp and *Shigella*) or gram-positive bacteria (*Streptococcus thermos* and *Bacillus subtillus*) based on their grouping (Fig. 6). The start codon for the repA genes of the 27638 and 31454 family is ATG, which was also found to be the case with all of the plasmids and contigs with significant homology. ATG may be the only initiator for plasmids encoding repA protein of this group(Spiers & Bergquist, 1992).

Detailed analysis of the region upstream of p31454e family repA showed that it possessed characteristics of plasmid replication origins and found several features typical of the rolling circle type replicons, dso (del Solar et al., 1998). Comparison of the nucleotide sequences of the origins (dso) of pTA1015 (*B. subtilis*), pZOM1 (*Z. mobilis*), pBS512 (*S. boydii*), pCA2.4 *Synechocystis* PCC6803, plasmid of *Nostoc* and plasmid of *Streptococcus thermos* shows significant homology (Fig. 7). This stretch of homology is unique to the origin sequences, in other words there are no areas of sequence outside of the repA gene besides this region that show this level of similarity between p31454e and other plasmids. We also expected to find interons, AT – rich region containing sequence repeats, in the upstream region of the repA, but these were not found.

The amino acid sequence of repA p31454e was deduced from the nucleotide sequences. The protein consists of 326 amino acids. The amino acids sequence homology between rep p31454e and similar proteins was examined by alignment and phylogeny (Fig. 6 and 8). The results show that the repA p31454e has significant homology to proteins encoded by the pKYM, pCA2.4, pUC110, and rep genes from *Nostoc* sp and *Cylindrosperms*. pKYM , was isolated from *Shigella sonnei*, a gram negative bacterium, and has the most similarity to p31454e . It has been suggested that pKYM belongs to the pUB110 plasmid family (Yasukawa et al., 1991). The metagenome contig11531 was not included in the analysis because a complete repA gene was not found. There are several hundred base pairs near the beginning of the repA that were expected to be there, but are not.

The p31454e rep like protein shares two regions that have been characterized in other plasmids: region 1 (HFH) is thought to be the metal binding domain of the rep proteins and region 2 is the active site for DNA to protein binding leading to replication (Fig. 8). The greatest similarity was obtained with the replication proteins of plasmids from the *Synechocystis* sp, *Zymomonas mobili*s, pKYM, *Shigella sonnei*, *Clindrospermum* sp. A.1345, *Nostoc* sp, *Psychrobacter* sp. PRwf-1, *Streptococcus thermophilu*s, pUH1, pUB110, and Paenibacillus popilliae.

In the rep proteins tThere is conservation of the amino acid sequence around the Tyr within the active sitein the rep proteins within the active site. The tyrosine residues in Rep proteins of plasmids are believed to be important for their replication, in that it is thought to act as an active center to bind DNA strands (Fig. 8___). The plasmids included in the analysis replicate by a rolling-circle mechanism where the tyrosines of some repA proteins accept the 5’ end of the cleaved DNA at the replication origin. This suggests that plasmid p31454e replicates by a rolling circle mechanism and that the Tyr residue plays an important part of its function.

The dso consensus of the pC194/pUB110 group of plasmids was found in all of the 31454 family sequence at 152 – 167 nt (del Solar et al., 1998)(17). The sso are generally imperfect palindromes up to 300 bp long with the possible ability to form hairpin secondary structures. p31454e has a sequence TAGCGG which is homologous to a general consensus sequence of six nucleotides which may act as the site of initiation for the second or lagging strand of DNA or sso (17) (at nucleotides 660-665) (del Solar et al., 1998). Another sequence similar to the sequence complementary to the 3' end of the RBS of *Synechecoccus* CC9311, CC9902, and WH8102 was found downstream of the coding region. The palindrome GCGATCGC is over represented in DNA sequences of many cyanobacterial strains, and . Tthere are two sites found in p31454e and 1 in p27638e (Chen et al., 2008).

3.2 Plasmid family 27638

The rep gene for the other plasmid family was found to be shorter, with contig27638 at 259 and contig31635 having 248 amino acids. Which This is smaller thaen plasmid family 31454 and the typical repA gene (Khan, 2005). The putative repA protein of p27638e showed significant similarity with the replication protein of a plasmid endogenous to *Leptolyngbya boryana (*YP001687742.1). The p27638e shared one well-conserved region with a high degree of homology to the plasmid replication protein of *Leptolyngbya boryana*. This area contains the tyrosine residue (marked with a box in Ffig. 9) involved in the DNA-protein binding for initiation of replication (del Solar et al., 1998)(17). Other smaller areas of conservation have no known functional characteristics assigned to them as of yet. The palindrome CGACCCTAGGGTCG was found before the start of the repA protein of p27638e but not p31635e, this might be the sso for p27638e.

The 27638 plasmid family repA gene does show some changes in diversity with time in environmental sampleswith time. The clones that matched most closely with p31635e were only detected during May 07, while clones found to match p27638e were found in May 07 and Oct 06. Three clones, Oct 06 4, Oct 06 7 and May 07 8 form a distinct cluster separate from p27638e and p31635e (Fig. 10). These clones do not match any known contigs detected in the cyanobacterial metagenome (Appx 2). This may suggest that there is even more plasmid diversity thaen previously detected.

4. Analysis of the cargo region and its putative open reading frames

A schematic diagram of the p27638e, p31635e, p31454e gene organization is shown in Fig. 2, 3, and 4, respectively. p27638e has four ORFs and p31635e had six ORFs that correspond to more than 20 amino acids while p31454e has five. Other small sized ORFs (less than 20 amino acids) were neglected. Based on the positions and number of ORFs, it may be that the majority of the plasmid's genome may be coding. ORF D from p27638e and ORF H from p31635e are 55% identical based on their alignment (Fig. 11), all other ORFs are different from each other (Fig. 11).

As it stands very little information is available about the genes encoded by cyanobacterial plasmids. To identify what kind of proteins are encoded by the ORFS of plasmid family 27638 and 31454, homology searches for the deduced amino acid sequences in GenBank and CAMERA data bases were performed. Comparisons revealed that none of the conserved proteins show any significant homology with known proteins. Surprisingly, the proteins of the plasmids showed no similarities to the ORFs encoded by the plasmids for which they showed significant homology in the repA gene. Nor do the putative ORFs found have any conserved domains.

The alignments and phylogenies of 27638 and 31454 family plasmids were inferred using sequences of the cargo from DNA extracted on data points Oct 10, 2006, May 17, 2007, and May 15, 2008. Six unique clones were sequenced on the time point May 17, 2007 and five for Oct 10, 2006 for plasmid family 27638. Putative plasmids: contigs 52093, 44876, 58180, 49088, 58316, 50652, 55949, 46567, 00646, and 71237 were also included in the analysis (Appx 4). For plasmid family 31454, five clones were sequenced from May 17, 2007, four for May 15, 2008 and two for Oct 10, 2006. Related putative plasmids discovered in other cyanobacterial metagenome included contigs: 01198, 52534, 11531, and 45314 were also evaluated against the clone library (Appx 5).

4.1 Cargo of plasmid family 31454

The neighboring joining tree for the nucleotide sequence of the cargo region of the 31454 plasmid family revealed changes in plasmid diversity with time (Fig. 12). The clones from May 17, 07 group together and are distinct from May 15, 2008 and Oct 10, 2006. It should also be noted that the clones from a depth of 20m are different from surface clones observed. Since the same date was not analyzed at surface depth a more detailed analysis cannot be done. As with the repA gene, 31454 plasmid family is more conserved than the 27638 plasmid family in the cargo region.

Two putative ORFsS (Table 3) fromwith the 31454 plasmid family were identified using ATG as the potential start codon, two ORFs withfor GTG, and one for TTG as start codons (this sentence doesn't really make sense, re-word). . The nucleotide range, putative initiaation codon, number of codons, and sequences found are shown in Table 33 for 31454 plasmid family. None of the predicted ORFs shared any significant similarity based on protein BLAST to any proteins in the NCBI database.

4.2 Cargo of plasmid family 27638

The resulting neighbor joining phylogenies reveal distinct changes in plasmid diversity with time for the plasmid family 27638 (Fig. 13). The clones of 27638 plasmid family form two different phylogenetic clusters. Clones from the data point May 07 group most strongly with p27635 while sequences from Oct 10 2006 group with p31635e. The data point May 15, 2008 at a depth of 20m was also probed for this plasmid family, however none were found.

Using ATG, TTG, and GTG as start codons:, two, five, and three putative ORFs respectivelyS were identifited for the 27638 plasmid family. Table 2 shows the nucleotide range, putative initiation codon, number of codons, and sequences found. The majority of potential ORFs use start codons other than ATG for the plasmids under study. This is interesting because the majorityit's the tendency of cyanobacterial genes to initiate with ATG. Protein BLAST revealed no similarity to these potential ORFs to proteins in the NCBI database.

5. Abundance of plasmid families using qPCR over time

The relative copy number and occurrence of the repA gene for plasmid families 27638 and 31454 was determined by real time qPCR. A ten fold dilution of the total DNA (0.001-10ng ul^{-1}) from cloned plasmid DNA for 31454 and 27638 was used to make the standard curve. The standard curves obtained from a ten fold dilution of the total DNA (0.001-10ng ul^{-1}) from cloned plasmid DNA for 31454 and 27638 were linear ($R^2>0.99$) in the range analysis and the slopes were 3.494 and 3.178 for 31454 and 27638 plasmid family respectively.

p27638e was found at highest abundance when during *Synechococcus* blooms (Fig. 14). Peaks occur annually on March, May, and June. p27638e was discovered to significantly correlate with the total *Synechococcus* population off the SIO pier during the time course examined (Fig. 15 $p<0.01$, $R^2 = 0.7299$) . p27638e was also evaluated against in Clade II WH8102, clade I CC9311, and clade IV CC9902, however similar levels of positive correlate were seen with individual clades as with total numbers. p31454e was not positively correlated with total *Synechococcus* abundance or any clade (Fig. 15 $p>0.05$, $R^2 = 0.07007$).

In the environment the ratio of *Synechococcus* to p27638e was found on average over time to be 87:1 ($\underline{+}$ 89 SD) to total *Synechococcus*. In culture under selection, *Synechococcus* CC9311 to pTOPO27638 was 1:34.

Conclusions and Future Direction

This paper represents a systematic effort to reveal the existence of plasmids indigenous to the marine *Synechococcus* population. We also looked at the diversity of plasmids, abundance over a time course, and ability to replicate in *Synechococcus* sp. CC9311, CC9902, and WH8102 of the two plasmid families 27638 and 31454 found in a marine cyanobacterial metagenome.

Two families of plasmids isolated from the SIO metagenome have been described. Three of the cyanobacterial plasmids were completely sequenced. The sequences did not show any similarity to any other known sequence in the NCBI or CAMERA database. This hints that the plasmid population discovered here may be unique to this area. Both families were found to code replication proteins and one of these falls into the rolling circle replication group of plasmids, plasmid family 31454. However, plasmid family 27638 does not significantly match any known replication initiation protein groups, although its closest match is also rolling circle. The smallest plasmid found to be associated with cyanobacterium was that of *Leptolyngbya boryana* plasmid pPBS1, which is 1.5 kb. The largest plasmid was 400kb, belonging to *Nostoc* sp. PCC7120. So the size of p27638e, 1553 bp, and p31454e, 1479 bp, is not an anomaly for this group of organisms. The similar G+C content of the host *Synechococcus* CC9311 and p31454e and p27638e suggests that the plasmids have resided in this host for a significant amount of time or that they are indigenous. The G+C content of plasmids relative to their host may be a useful tool in cyanobacterial classification along with their replication compatibility.

The repA gene encoded by p31454e, isolated from the SIO pier metagenome, shows significant similarity with rep proteins from other gram-negative and -positive bacteria at the protein level, but are not similar in nucleotide sequences with one exception. All rolling circle plasmids are divided loosely into four groups: pT181, pC194/pUB110, pLS1/pE194, and pSN2, based on rep protein, dso, and control of replication(Khan, 2005) (Novick, 1989). Based on known characteristics, plasmid family 31454 would most likely fall into the pC194/pUB110 group, while plasmid family 27638 remains unique. The tyrosine residue in the Rep proteins of plasmids that replicate by rolling circle mechanism is thought to be important for their role. Each protein contained a typical active center sequence, KYNNK, where K is lysine, Y is tyrosine, and N is the amino acid residue (Ilyina & Koonin, 1992). The tyrosine accepts the 5’ end of the single strand break introduced by these proteins. The importance of this Tyr residue would need to be confirmed by mutagenesis. The consensus nucleotide sequence of the origins for this family and many others is 5’ – CTTANNNNGATANNT- 3’. The nick site sequence is found at the double-stranded origin. CTGATA is the conserved motif found in p31454e, pUB110, and pC194. Single stranded origins of replication vary much more; this is hypothesized to be because they must interact with proteins on the host which are of course very diverse (Seery, Nolan, Sharp, & Devine, 1993). In the plasmid family 27638, the dso CTGATA can be detected in addition to the binding site KYXXY. If p31454e and p27638e do in fact replicate by a rolling circle mechanism, a specific single stranded circular DNA (ssDNA) would be found as an intermediate (Riele, Michel, & Ehrlich, 1986). The presence of this DNA would support the hypothesis that the plasmids replicate by a rolling circle mechanism (Maniatis et al., 1982) (Sambrook & Russell, 2001)..

Plasmid families 31454 and 27638 have inverted repeats, a dso, and sso, however, the direct repeats, which are characteristic of rolling circle plasmids, were not found. This suggests that these small cryptic plasmids replicate via rolling circle replication mode, but this replicon may be a new class of rolling circle replication. Also expected were genes involved in maintaining the plasmids stable inheritance in the host, which is typical of many low copy number plasmids (Chen et al., 2008). One would expect a site-specific recombinase, a partitioning system, or possibly a toxin-antitoxin cassette. However, related putative plasmids have been found to have all these domains based on metagenomic data (Ma, Yingfei, personal communication).

It is interesting to note that there was less homology to known cyanobacterial plasmids than expected. Similarities appear to be independent of gram-positive or gram-negative bacteria, since likeness is found with both pKYM, endogenous to gram-negative bacteria, and pUB100, from gram positive. Cyanobacterial plasmids have been found to be quite divergent from each other, in addition to other plasmids derived from both gram positive and gram negative bacteria (Seery et al., 1993). Relatedness with plasmids does not always correlate with relatedness of host strains in that genetically unrelated strains have been found to harbor similar plasmids (Lau et al., 1980). The similarity of p31454e to the other more widespread strain plasmids suggests horizontal gene transfer of the repA region might have happened not very long ago. However, as of now little is known about how possible interspecific transmission comes about.

There is a great deal of variation and flexibility in the area outside of the cargo region of the plasmid. As previously mentioned, many plasmids contain genes in these regions that are not only beneficial to their own propagation, but also to their hosts. Although no proteins were functionally identified, it is still possible that these genes are essential for plasmid maintenance and the host’s ability to flourish. It is unclear what would cause the plasmid sequences to change over time and often these types of genes are located on transposons, which may account for much of the variation (Couturier, Bex, Bergquist, & Maas, 1988). This may indicate that there is very little selective pressure on this region, bolstering the theory that perhaps these plasmids are not beneficial to their host or coding a protein. There is no significant sequence similarity between 31454 and 27638 plasmid family cargo region and other plasmid’s cargo region that they shared homology within the repA gene. This may be a reflection of each host’s individual needs, as this will promote a unique strategy for plasmid maintenance and the persistence in their host based on their environment. In that sense, the changes observed in cargo composition may also be a reflection of variation in the *Synechococcus* strains. In order to tease out this potential relationship further, qPCR could be performed. More specific primers could be developed that differentiate the different plasmids in each family then compare to changes in *Synechococcus* strains over time. Regardless, it seems appropriate that further classification of environmental plasmids would be done based on areas of less variation such as the replication region.

As typical of DNA sequence analysis by computer program, many small ORFs were predicted to be present but that may not be functional. To test the likelihood of these ORFs actually coding, several recombinant plasmids could be constructed that would interrupt some ORFs while keeping others intact. Then test their stability in

Synechococcus. This could also potentially reveal if any ORFs confer beneficial traits to their host. In addition, this will help with the potential use as a shuttle vector and identify all the regions required for plasmid maintenance, in that the plasmid can be whittled down to their minimum size.

The primary preliminary contribution of this paper is the beginning of the underpinning of a shuttle vector that would allow the transfer of DNA to *Synechococcus* CC9311 and E. coli. Conditions using a conjugation system and electrotransformation with pMUT27638e and pTOPO27638e respectively give kanamycin resistance and expression of inserted DNA. Pour plating methods then allow the isolation of individual colonies, which can then be grown in larger liquid volumes for additional examination. These methods illustrated have been repeated multiple times, however the final result of the modified *Synechococcus* does take over a month to obtain. It should be noted that preliminary results indicate that both plasmid families can express in *Synechococcus* CC9311 and CC9902, both of which are dominant coastal strains indigenous to the area the plasmids were originally found, while WH8102 is an open ocean strain. Perhaps, a different collection of plasmids is found in this differing environment.

Marine microbes are becoming more important for biotechnological purposes such as biofuels, so optimizing shuttle vectors will become even more essential for ongoing research. Since plasmids from the marine environment are not closely related to well-characterized plasmids, shuttle vectors derived from the host species may have the greatest likelihood of success (Dahlberg, Linberg, Torsvik, & Hermansson, 1997). Testing stability and compatibility of these two cryptic plasmids will also be important in vector use. Strains should be tested under selecting and non-selecting conditions to see how long the plasmids last in the population.

Little is known about the mechanisms that control the copy number of cyanobacterial plasmids including p27638e and p31454e, nor why in cultured isolates the plasmids would be lost completely. Replication of some plasmids that replicate via rolling circle is regulated by the replication protein and copy number is dictated by both synthesis and inactivation of that protein (Novick, 1989). However, in cyanobacterium *Synechococcus* sp. PCC7002, salinity may determine the abundance of some endogenous plasmids by affecting the amount of replication protein (Yano, Kawata, & Kojima, 1995) . It appears that growth medium may be important for the persistence of plasmids within host populations in some situations. Perhaps, culture conditions are not conducive for p27638e, p31454e, or other plasmids to continue within isolates of *Synechococcus*.

qPCR shows the changes in plasmid families 27638 and 31454 relative to *Synechococcus* over a year period. The plasmid family 27638 was positively correlated with total *Synechococcus* abundance. A strong correlation between p31454e and *Synechococcus* may be lacking due to its broad host range. Plasmid family 31454 was found to be similar to pC194 and pUB110 both of which have been isolated from widely divergent species of bacteria (Seery et al., 1993) (Scheerabramowitz, Gryczan, & Dubnau, 1981). Since the correlation of p27638e does not differ greatly between total *Synechococcus* and individual clades this suggests that the 27638 plasmid family may be compatible with many strains of *Synechococcus*. Preliminary work conjugating pMUT27638e into CC9902 support this conclusion. Plasmids are rarely placed into a larger ecological context, but being subject to the rules of natural selection, it makes sense that there may be top-down or bottom-up environmental effects. It would be interesting to further tease out the relationship between plasmids and the marine environment. Perhaps *Synechococcus* cell numbers determine plasmid abundance, or maybe the plasmids determine the bacterial populations. By causing much of the movement of genes in the marine bacterial populations, plasmids may have a large and direct effect on the ecology as a whole(Patricia A. Sobecky, 2002).

In conclusion, we put forward that two small cryptic plasmids, p27638e and p31454e are potentially endogenous to marine *Synechococcus*. All of the plasmids under study potentially encode a rep protein and may replicate by a rolling circle mechanism. The purpose of this thesis is to understand the genetic diversity and abundance of plasmid families 27638 and 31454 at different time points, in addition to their ability to replicate within marine *Synechococcus*.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). BASIC LOCAL ALIGNMENT SEARCH TOOL. Journal of Molecular Biology, 215(3), 403-410.

Baas, P. D., & Jansz, H. S. (1988). SINGLE-STRANDED-DNA PHAGE ORIGINS. Current Topics in Microbiology and Immunology, 136, 31-70.

Backert, S., Meissner, K., & Borner, T. (1997). Unique features of the mitochondrial rolling circle-plasmid mp1 from the higher plant Chenopodium album (L.). Nucleic Acids Research, 25(3), 582-589.

Berns, K. I. (1990). PARVOVIRUS REPLICATION. Microbiological Reviews, 54(3), 316-329.

Bose, S. G., & Carmichael, W. W. (1990). PLASMID DISTRIBUTION AMONG UNICELLULAR AND FILAMENTOUS TOXIC CYANOBACTERIA. Journal of Applied Phycology, 2(2), 131-136.

Brahamsha, B. (1996). A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. Applied and Environmental Microbiology, 62(5), 1747-1751.

Bryant, D. (1994). The molecular biology of cyanobacteria: Kluwer Academic Pub.

Chen, Y., Holtman, C. K., Magnuson, R. D., Youderian, P. A., & Golden, S. S. (2008). The complete sequence and functional analysis of pANL, the large plasmid of the unicellular freshwater cyanobacterium *Synechococcus* elongatus PCC 7942. Plasmid, 59(3), 176-192.

Couturier, M., Bex, F., Bergquist, P. L., & Maas, W. K. (1988). IDENTIFICATION AND CLASSIFICATION OF BACTERIAL PLASMIDS. Microbiological Reviews, 52(3), 375-395.

Dahlberg, C., Linberg, C., Torsvik, V. L., & Hermansson, M. (1997). Conjugative plasmids isolated from bacteria in marine environments show various degrees of homology to each other and are not closely related to well-characterized plasmids. Applied and Environmental Microbiology, 63(12), 4692-4697.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., & Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. Microbiology and Molecular Biology Reviews, 62(2), 434-+.

del Solar, G., Moscoso, M., & Espinosa, M. (1993). ROLLING CIRCLE-REPLICATING PLASMIDS FROM GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA - A WALL FALLS. Molecular Microbiology, 8(5), 789-796.

Felkner, R. H., & Barnum, S. R. (1988). PLASMID CONTENT AND HOMOLOGY OF 16 STRAINS OF FILAMENTOUS, NONHETEROCYSTOUS CYANOBACTERIA. Current Microbiology, 17(1), 37-41.

Holmes, M. L., Pfeifer, F., & Dyallsmith, M. L. (1995). ANALYSIS OF THE HALOBACTERIAL PLASMID PHK2 MINIMAL REPLICON. Gene, 153(1), 117-121. Retrieved Feb, from Note database.

Ilyina, T. V., & Koonin, E. V. (1992). CONSERVED SEQUENCE MOTIFS IN THE INITIATOR PROTEINS FOR ROLLING CIRCLE DNA-REPLICATION ENCODED BY DIVERSE REPLICONS FROM EUBACTERIA, EUKARYOTES AND ARCHAEBACTERIA. Nucleic Acids Research, 20(13), 3279-3285. Retrieved Jul, from Article database.

Khan, S. A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. Plasmid, 53(2), 126-136.

Koonin, E. V., Makarova, K. S., & Aravind, L. (2001). Horizontal gene transfer in prokaryotes: Quantification and classification. Annual Review of Microbiology, 55, 709-742.

Larsen, R. A., Wilson, M. M., Guss, A. M., & Metcalf, W. W. (2002). Genetic analysis of pigment biosynthesis in Xanthobacter autotrophicus Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. Archives of Microbiology, 178(3), 193-201.

Lau, R. H., Sapienza, C., & Doolittle, W. F. (1980). CYANOBACTERIAL PLASMIDS - THEIR WIDESPREAD OCCURRENCE, AND THE EXISTENCE OF REGIONS OF HOMOLOGY BETWEEN PLASMIDS IN THE SAME AND DIFFERENT SPECIES. Molecular & General Genetics, 178(1), 203-211.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982). MOLECULAR CLONING A LABORATORY MANUAL.

McCarren, J., & Brahamsha, B. (2005). Transposon mutagenesis in a marine *synechococcus* strain: Isolation of swimming motility mutants. Journal of Bacteriology, 187(13), 4457-4462.

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. Current Opinion in Genetics & Development, 15(6), 589-594. Retrieved Dec, from

Miyake, M., Nagai, H., Shirai, M., Kurane, R., & Asada, Y. (1999). A high-copy-number plasmid capable of replication in thermophilic cyanobacteria. Applied Biochemistry and Biotechnology, 77-9, 267-275.

Novick, R. P. (1989). STAPHYLOCOCCAL PLASMIDS AND THEIR REPLICATION. Annual Review of Microbiology, 43, 537-565.

Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., et al. (2003). The genome of a motile marine *Synechococcus*. Nature, 424(6952), 1037-1042.

Palenik, B., Ren, Q., Tai, V., & Paulsen, I. T. (2009). Coastal Synechococcus metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. Environmental Microbiology, 11(2), 349-359.

Rebiere, M. C., Castets, A. M., Houmard, J., & Demarsac, N. T. (1986). PLASMID DISTRIBUTION AMONG UNICELLULAR AND FILAMENTOUS CYANOBACTERIA - OCCURRENCE OF LARGE AND MEGA-PLASMIDS. Fems Microbiology Letters, 37(3), 269-275.

Riele, H. T., Michel, B., & Ehrlich, S. D. (1986). SINGLE-STRANDED PLASMID DNA IN BACILLUS-SUBTILIS AND STAPHYLOCOCCUS-AUREUS. Proceedings of the National Academy of Sciences of the United States of America, 83(8), 2541-2545.

Sambrook, J., & Russell, D. W. (2001). Molecular cloning: A laboratory manual. Molecular cloning: A laboratory manual.

Scheerabramowitz, J., Gryczan, T. J., & Dubnau, D. (1981). ORIGIN AND MODE OF REPLICATION OF PLASMIDS PE194 AND PUB110. Plasmid, 6(1), 67-77. Article database.

Seery, L., Nolan, N., Sharp, P., & Devine, K. (1993). Comparative analysis of the pC194 group of rolling circle plasmids. Plasmid, 30(3), 185.

Sobecky, P. A. (1999). Plasmid ecology of marine sediment microbial communities. Hydrobiologia, 401(0), 9-18. Retrieved May 1, from

Sobecky, P. A. (2002). Approaches to investigating the ecology of plasmids in marine bacterial communities. Plasmid, 48(3), 213-221. Retrieved November, from

Sobecky, P. A., Mincer, T. J., Chang, M. C., Toukdarian, A., & Helinski, D. R. (1998). Isolation of broad-host-range replicons from marine sediment bacteria. Applied and Environmental Microbiology, 64(8), 2822-2830. Retrieved Aug., from

Spiers, A. J., & Bergquist, P. L. (1992). Expression and regulation of the RepA protein of the RepFIB replicon from plasmid P307. Journal of Bacteriology, 174(23), 7533-7541.

Tai, V., & Palenik, B. (2009). Temporal variation of Synechococcus clades at a coastal Pacific Ocean monitoring site. The ISME Journal, 3(8), 903-915.

Tai, V., Ren, Q., Paulsen, I. T., & Palenik, B. (2008). Dominance of *Synechococcus* clades I and IV during a Coastal Marine Time-Series. Abstracts of the General Meeting of the American Society for Microbiology, 108, 430. Retrieved 2008, from Meeting database.

Thomas, C., & Nielsen, K. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature reviews microbiology, 3(9), 711-721.

Toledo, G., & Palenik, B. (1997). Synechococcus diversity in the California current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. Applied and Environmental Microbiology, 63(11), 4298-4303. Retrieved Nov., from

Tominaga, H., Ashida, H., Sawa, Y., & Ochiai, H. (1992). FUNCTION-ANALYSIS OF CYANOBACTERIAL PLASMIDS - PPF1 (PHORMIDIUM-FOVEORARUM) AND PMA1(MICROCYSTIS-AERUGINOSA). Photosynthesis Research, 34(1), 181-181.

Vincze, T., Posfai, J., & Roberts, R. J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Research, 31(13), 3688-3691.

Waterbury, J., Watson, S., Valois, F., & Franks, D. (1986). Biological and ecological characterization of the marine unicellular cyanobacterium Synechococcus. Photosynthetic picoplankton, 71–120.

Waterbury, J., & Willey, J. (1988). Isolation and growth of marine planktonic cyanobacteria.

Waterbury, J. B., Watson, S. W., Guillard, R. R. L., & Brand, L. E. (1979). WIDESPREAD OCCURRENCE OF A UNICELLULAR, MARINE, PLANKTONIC, CYANOBACTERIUM. Nature, 277(5694), 293-294.

Waterbury, J. B., Willey, J. M., Franks, D. G., Valois, F. W., & Watson, S. W. (1985). A CYANOBACTERIUM CAPABLE OF SWIMMING MOTILITY. Science (Washington D C), 230(4721), 74-76.

Xu, W. D., & McFadden, B. A. (1997). Sequence analysis of plasmid pCC5.2 from cyanobacterium Synechocystis PCC 6803 that replicates by a rolling circle mechanism. Plasmid, 37(2), 95-104. Article database.

Yano, S., Kawata, Y., & Kojima, H. (1995). Salinity-dependent copy number change of endogenous plasmids in Synechococcus sp. strain PCC 7002. Current Microbiology, 31(6), 357-360.

Yasukawa, H., Hase, T., Sakai, A., & Masamune, Y. (1991). ROLLING-CIRCLE REPLICATION OF THE PLASMID PKYM ISOLATED FROM A GRAM-NEGATIVE BACTERIUM. Proceedings of the National Academy of Sciences of the United States of America, 88(22), 10282-10286.

Zhang, R., Wang, Y., Leung, P. C., & Gu, J.-D. (2007). pVC, a small cryptic plasmid from the environmental isolate of Vibrio cholerae MP-1. Journal of Microbiology, 45(3), 193-198. Retrieved Jun, from

Figures and Tables

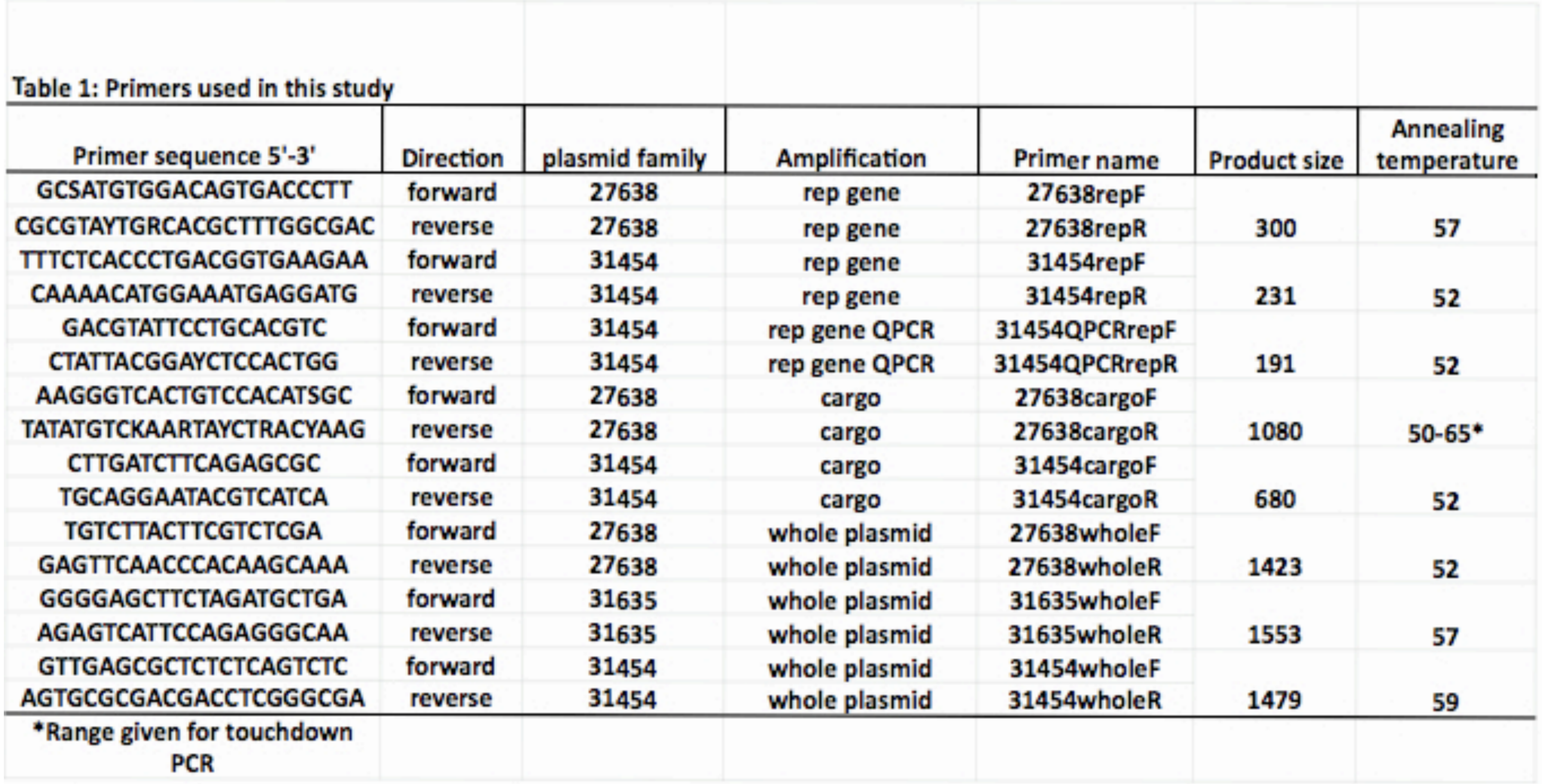

Table 1: Primers used in this study

Primer sequence 5'-3'	Direction	plasmid family	Amplification	Primer name	Product size	Annealing temperature
GCSATGTGGACAGTGACCCTT	forward	27638	rep gene	27638repF	300	57
CGCGTAYTGRCACGCTTTGGCGAC	reverse	27638	rep gene	27638repR		
TTTCTCACCCTGACGGTGAAGAA	forward	31454	rep gene	31454repF	231	52
CAAAACATGGAAATGAGGATG	reverse	31454	rep gene	31454repR		
GACGTATTCCTGCACGTC	forward	31454	rep gene QPCR	31454QPCRrepF	191	52
CTATTACGGAYCTCCACTGG	reverse	31454	rep gene QPCR	31454QPCRrepR		
AAGGGTCACTGTCCACATSGC	forward	27638	cargo	27638cargoF	1080	50-65*
TATATGTCKAARTAYCTRACYAAG	reverse	27638	cargo	27638cargoR		
CTTGATCTTCAGAGCGC	forward	31454	cargo	31454cargoF	680	52
TGCAGGAATACGTCATCA	reverse	31454	cargo	31454cargoR		
TGTCTTACTTCGTCTCGA	forward	27638	whole plasmid	27638wholeF	1423	52
GAGTTCAACCCACAAGCAAA	reverse	27638	whole plasmid	27638wholeR		
GGGGAGCTTCTAGATGCTGA	forward	31635	whole plasmid	31635wholeF	1553	57
AGAGTCATTCCAGAGGGCAA	reverse	31635	whole plasmid	31635wholeR		
GTTGAGCGCTCTCTCAGTCTC	forward	31454	whole plasmid	31454wholeF	1479	59
AGTGCGCGACGACCTCGGGCGA	reverse	31454	whole plasmid	31454wholeR		

*Range given for touchdown PCR

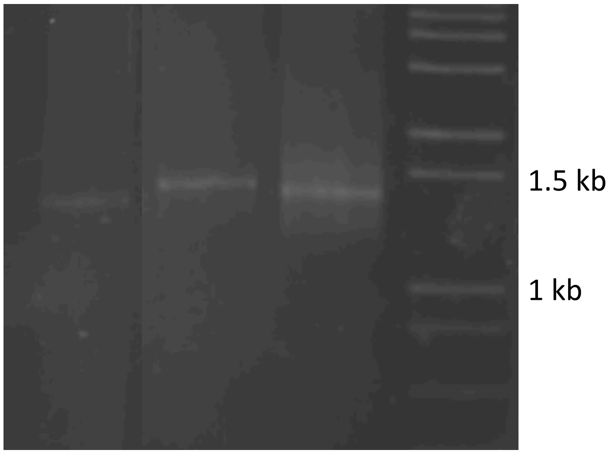

Figure 1: 0.8% gel of PCR amplification of whole plasmids. Lane 1 p31454E, lane 2 p27638E, Lane 3 p31635E.

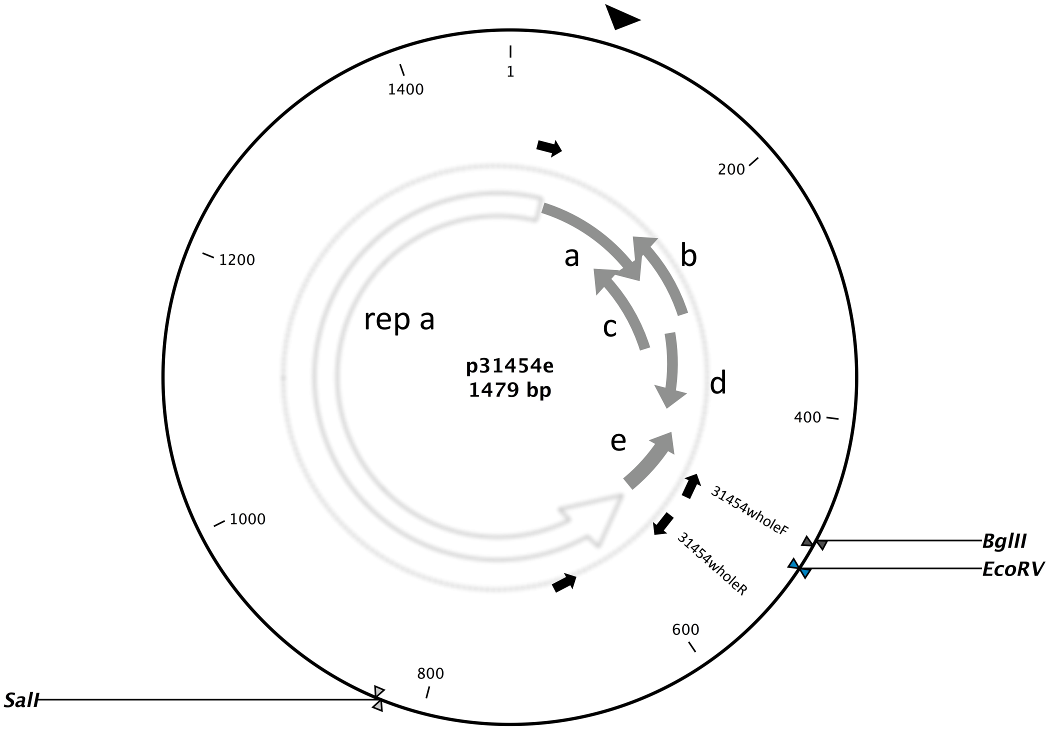

Figure 2: Physical map of plasmid 31454. Gene position and direction of transcription are indicated by arrows. Hollow arrows indicate ORFS. Solid arrows are primer representation. Unique restriction sites are given. Triangle indicates double stranded origin nick site (dso)

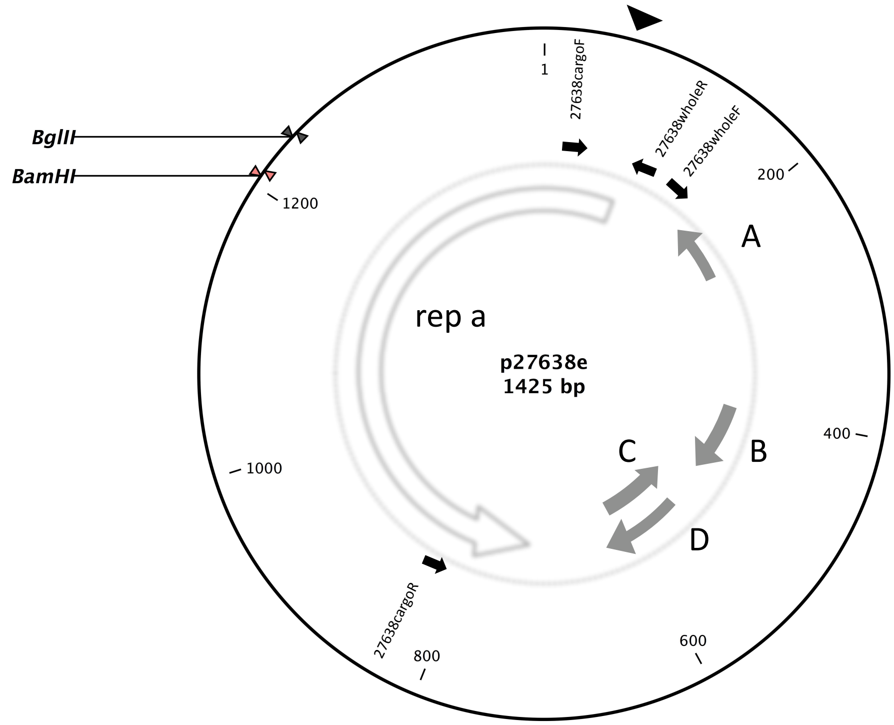

Figure 3: Physical map of plasmid 27638. Gene position and direction of transcription are indicated by arrows. Hollow arrows indicate rep a gene, grey arrows putative ORFs. Solid arrows are primer representation. Unique restriction sites are given. Triangle indicates double stranded origin nick site (dso)

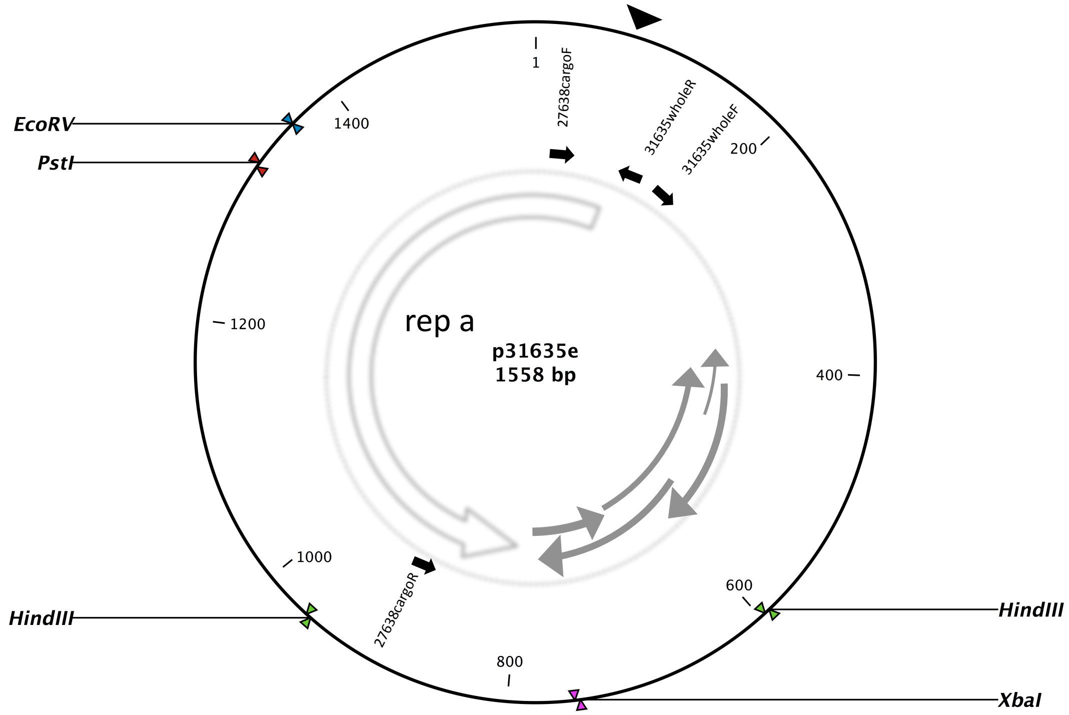

Figure 4: Physical map of plasmid 31635. Gene position and direction of transcription are indicated by arrows. Hollow arrows indicate rep a gene, grey arrows putative ORFs. Solid arrows are primer representation. Unique restriction sites are given. Triangle indicates double stranded origin nick site (dso)

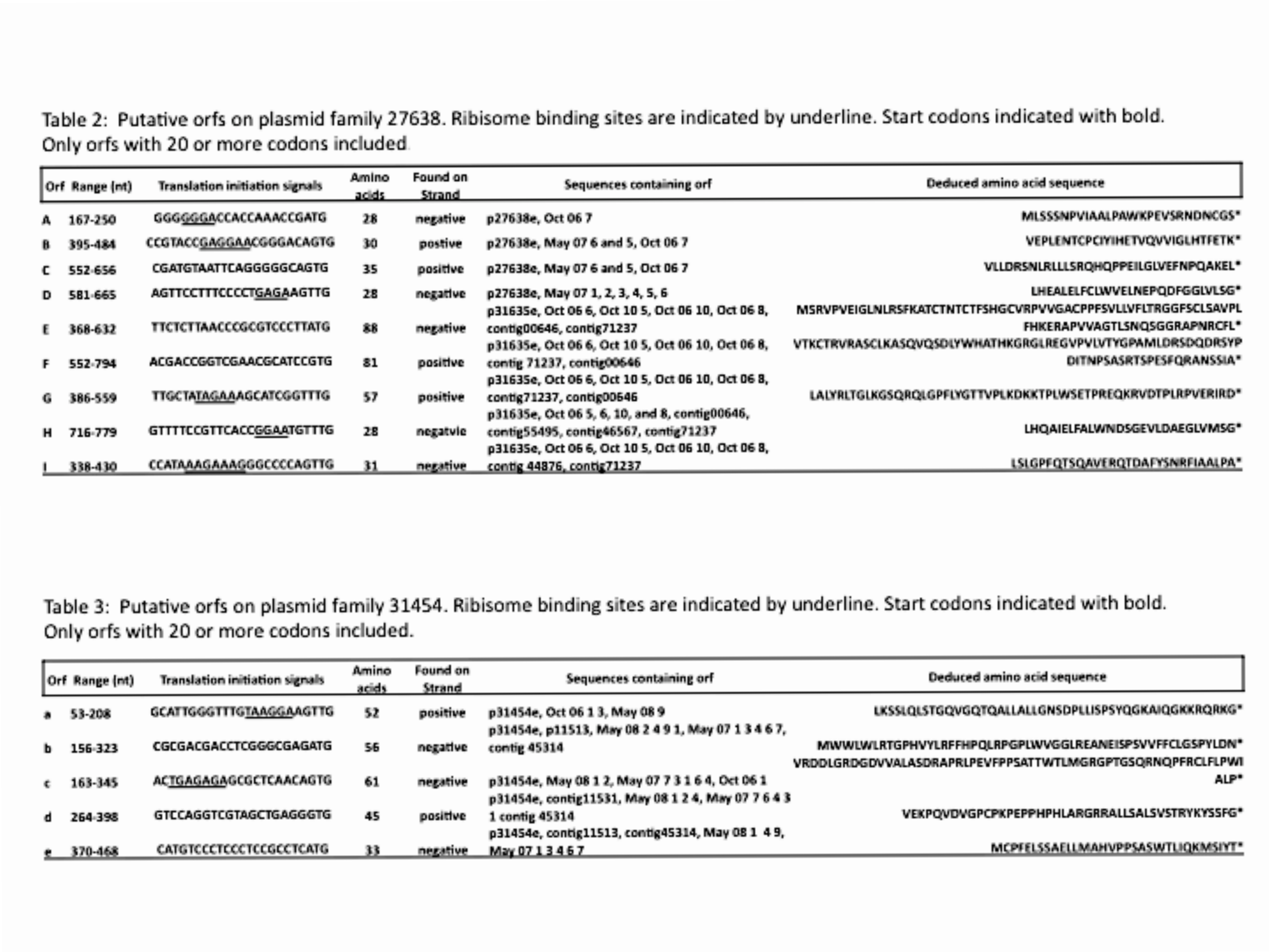

Table 2: Putative orfs on plasmid family 27638. Ribisome binding sites are indicated by underline. Start codons indicated with bold. Only orfs with 20 or more codons included.

Orf	Range (nt)	Translation initiation signals	Amino acids	Found on Strand	Sequences containing orf	Deduced amino acid sequence
A	167-250	GGGGGGACCACCAAACCGATG	28	negative	p27638e, Oct 06 7	MLSSSNPVIAALPAWKPEVSRNDNCGS*
B	395-484	CCGTACCGAGGAACGGGACAGTG	30	postive	p27638e, May 07 6 and 5, Oct 06 7	VEPLENTCPCIYIHETVQVVIGLHTFETK*
C	552-656	CGATGTAATTCAGGGGGCAGTG	35	positive	p27638e, May 07 6 and 5, Oct 06 7	VLLDRSNLRLLLSRQHQPPEILGLVEFNPQAKEL*
D	581-665	AGTTCCTTTCCCCTGAGAAGTTG	28	negative	p27638e, May 07 1, 2, 3, 4, 5, 6	LHEALELFCLWVELNEPQDFGGLVLSG*
E	368-632	TTCTCTTAACCCGCGTCCCTTATG	88	negative	p31635e, Oct 06 6, Oct 10 5, Oct 06 10, Oct 06 8, contig00646, contig71237	MSRVPVEIGLNLRSFKATCTNTCTFSHGCVRPVVGACPPFSVLLVFLTRGGFSCLSAVPL FHKERAPVVAGTLSNQSGGRAPNRCFL*
F	552-794	ACGACCGGTCGAACGCATCCGTG	81	positive	p31635e, Oct 06 6, Oct 10 5, Oct 06 10, Oct 06 8, contig 71237, contig00646	VTKCTRVRASCLKASQVQSDLYWHATHKGRGLREGVPVLVTYGPAMLDRSDQDRSYP DITNPSASRTSPESFQRANSSIA*
G	386-559	TTGCTATAGAAAGCATCGGTTTG	57	positive	p31635e, Oct 06 6, Oct 10 5, Oct 06 10, Oct 06 8, contig71237, contig00646	LALYRLTGLKGSQRQLGPFLYGTTVPLKDKKTPLWSETPREQKRVDTPLRPVERIRD*
H	716-779	GTTTTCCGTTCACCGGAATGTTTG	28	negatvie	p31635e, Oct 06 5, 6, 10, and 8, contig00646, contig55495, contig46567, contig71237	LHQAIELFALWNDSGEVLDAEGLVMSG*
I	338-430	CCATAAAGAAAGGGCCCCAGTTG	31	negative	p31635e, Oct 06 6, Oct 10 5, Oct 06 10, Oct 06 8, contig 44876, contig71237	LSLGPFQTSQAVERQTDAFYSNRFIAALPA*

Table 3: Putative orfs on plasmid family 31454. Ribisome binding sites are indicated by underline. Start codons indicated with bold. Only orfs with 20 or more codons included.

Orf	Range (nt)	Translation initiation signals	Amino acids	Found on Strand	Sequences containing orf	Deduced amino acid sequence
a	53-208	GCATTGGGTTTGTAAGGAAGTTG	52	positive	p31454e, Oct 06 1 3, May 08 9	LKSSLQLSTGQVGQTQALLALLGNSDPLLISPSYQGKAIQGKKRQRKG*
b	156-323	CGCGACGACCTCGGGCGAGATG	56	negative	p31454e, p11513, May 08 2 4 9 1, May 07 1 3 4 6 7, contig 45314	MWWLWLRTGPHVYLRFFHPQLRPGPLWVGGLREANEISPSVVFFCLGSPYLDN*
c	163-345	ACTGAGAGAGCGCTCAACAGTG	61	negative	p31454e, May 08 1 2, May 07 7 3 1 6 4, Oct 06 1	VRDDLGRDGDVVALASDRAPRLPEVFPPSATTWTLMGRGPTGSQRNQPFRCLFLPWI ALP*
d	264-398	GTCCAGGTCGTAGCTGAGGGTG	45	positive	p31454e, contig11531, May 08 1 2 4, May 07 7 6 4 3 1 contig 45314	VEKPQVDVGPCPKPEPPHPHLARGRRALLSALSVSTRYKYSSFG*
e	370-468	CATGTCCCTCCCTCCGCCTCATG	33	negative	p31454e, contig11513, contig45314, May 08 1 4 9, May 07 1 3 4 6 7	MCPFELSSAELLMAHVPPSASWTLIQKMSIYT*

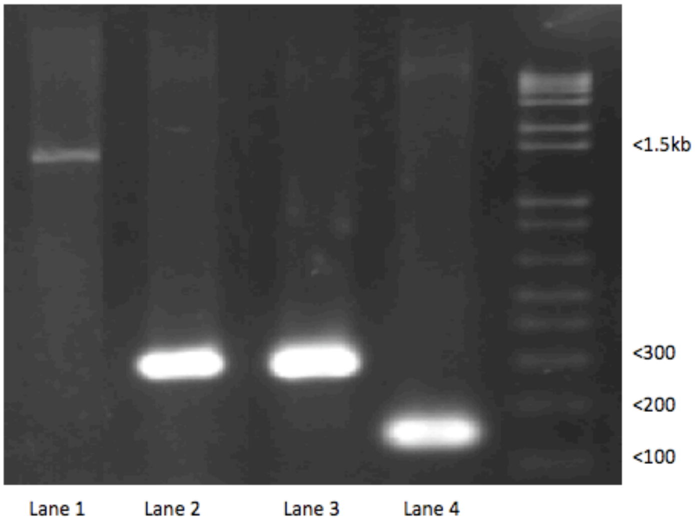

Figure 5: 1.0% gel with plasmid pTOPO27638 extracted from *Synechococcus* CC9311. Lane 1 amplification of whole plasmid, Lane 2 *rep*A gene, Lane 3 *rpo*C1 CC9311 specific, and Lane 4 Topo specific Kan primers

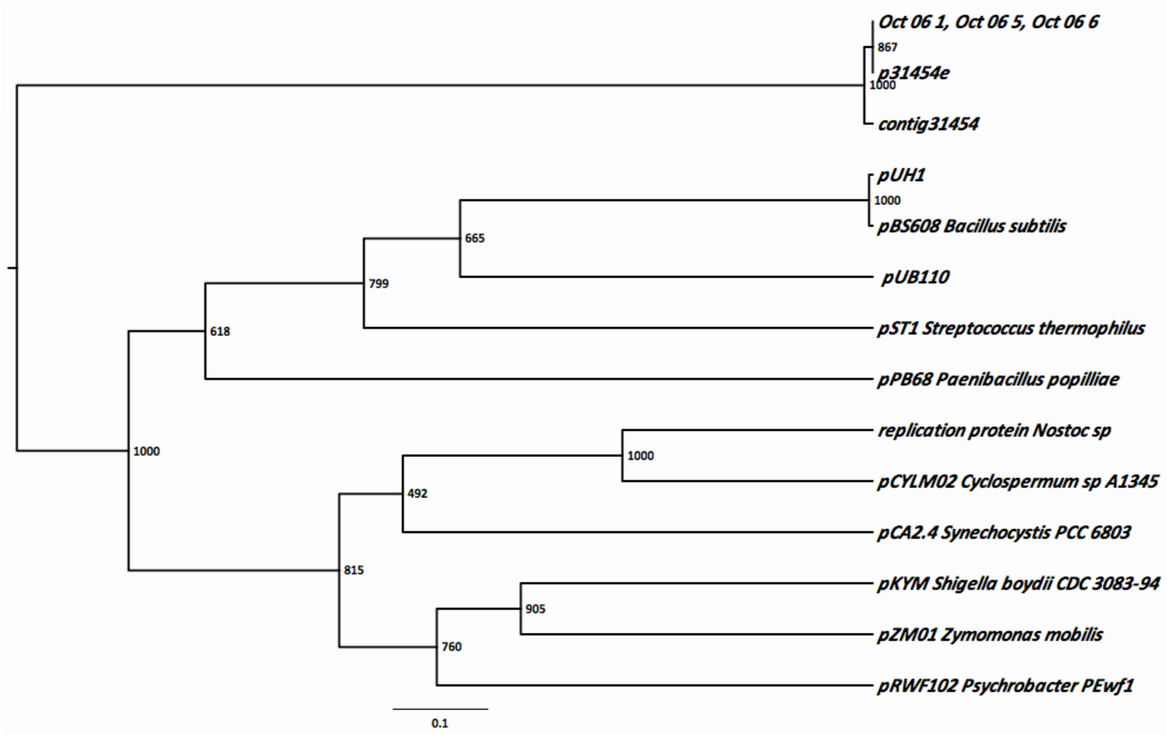
Oct 06 1, Oct 06 5, Oct 06 6
867
p31454e
1000
contig31454
pUH1
1000
pBS608 Bacillus subtilis
665
pUB110
799
pST1 Streptococcus thermophilus
618
pPB68 Paenibacillus popilliae
1000
replication protein Nostoc sp
1000
pCYLM02 Cyclospermum sp A1345
492
pCA2.4 Synechocystis PCC 6803
815
pKYM Shigella boydii CDC 3083-94
905
pZM01 Zymomonas mobilis
760
pRWF102 Psychrobacter PEwf1
0.1

Figure 6: Neighbor-joining phylogeny of the 31454 plasmid family based on amino acid sequence analysis of *rep*A region from a Scripps Institution of Oceanography pier sample on Oct 10, 2006 and various gram negative bacteria (highest blast hits). The sequence information was obtained from the following Accession number:AAA25513.1 (*Nostoc* sp), YP_001965480.1 (*Cyclospermum* sp A1345), replication protein [*Shigella boydii* CDC 3083-94], AAA02970.1 [*Synechocystis* sp.], NP_045297.1[*Zymomonas mobilis*], CAB43206.1 [*Streptococcus thermophilus*], AAA99151.1 [Bacillus subtilis], YP_001274383 (*Psychrobacter*),_AAB36954.1 (*Paenibacillus*) . Bootstrap values (out of 1000) are shown adjacent to branch notes. The length of the horizontal branches, based on pairwise analysis, correspond to evolutionary distances and the scale bars show the number of substitutions per site.

```
A T A C T T A A G G - G A T A A C T   pZM01
A T A C T T A A G G - G A T A A A T   pKYM
C T T C T T A T C T T G A T A C T T   pC194
T T T C T T A T C T T G A T A C T T   pUB100
C T A C T T A T A C A G A T A A T T   replication-associated protein Nostoc sp.
A C C C T T A C C A - G A T A A G G   pCA2.4
C G C C T T A C C T T G A T A A C T   p31454 contig 11531

T C C C T T A - - - - G A T A G A T   replication-associated protein L. boryana
T G - C T T G G T T A G A T A C T T   p27638e
C T - G T T G G T T C G A T A T C T   p31635e
```

Figure 7: Alignment of the putative double stranded origin nick site (5’ 157-175 3’). Sequences within the plus origins of replication region of plasmids: pZM01,pKYM, pC194, pUB100, replication-associated protein *Nostoc* sp. and pCA2.4.

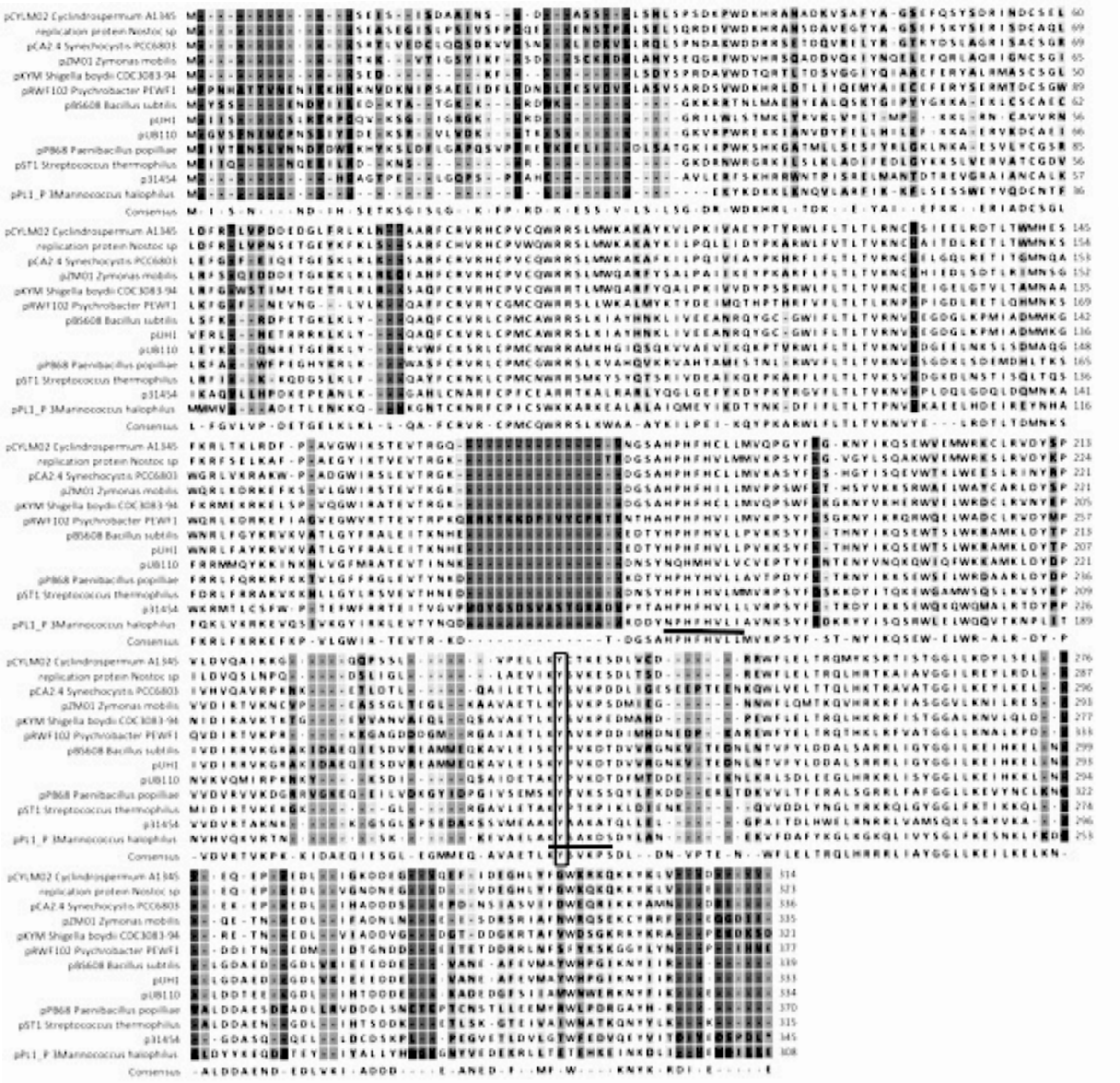

Figure 8: Comparison of the amino acid sequence of p31454 potential *rep*A protein with that of relative plasmid replication proteins. The conserved Tyr residue involved in the linkage sites to the DNA when nicking occurs at the origin for rolling circle replication is marked with a box and the active sites are underlined (del Solar, 1998) (Ilyina, 1992)

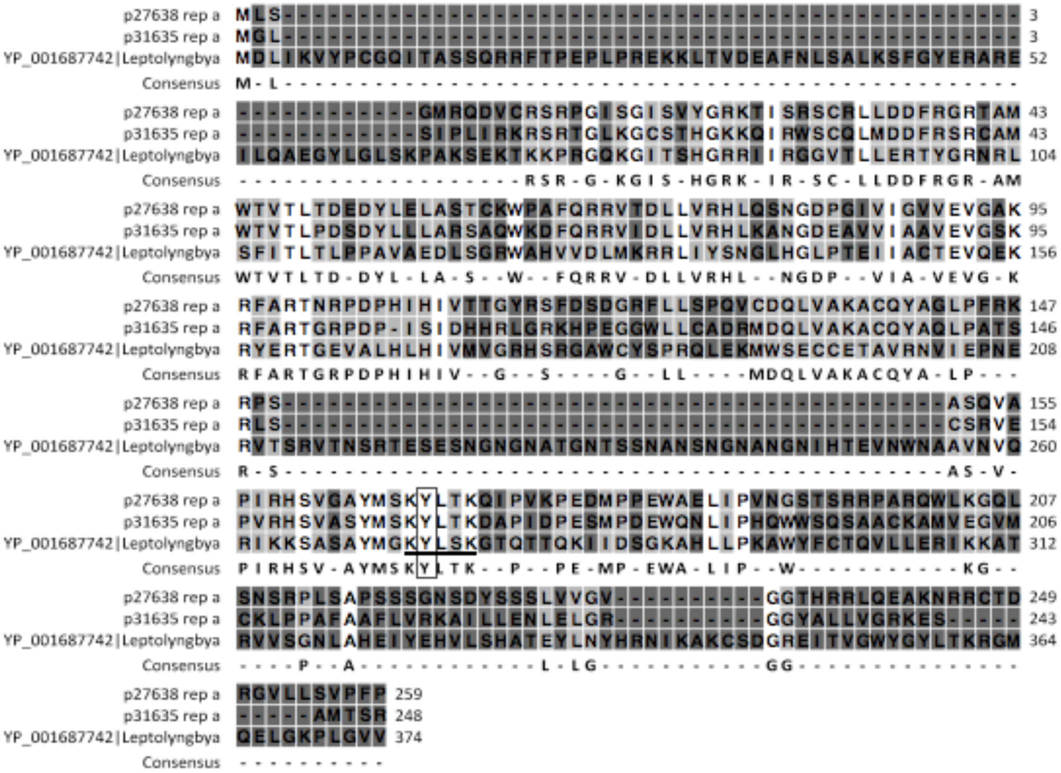
p27638 rep a MLS--- 3
p31635 rep a MGL--- 3
YP_001687742|Leptolyngbya MDLIKVYPCGQITASSQRRFTPEPLPREKKLTVDEAFNLSALKSFGYERARE 52
Consensus M-L---
p27638 rep a ------------GMRQDVCRSRPGISGISVYGRKTISRSCRLLDDFRGRTAM 43
p31635 rep a ------------SIPLIRKRSRTGLKGCSTHGKKQIRWSCQLMDDFRSRCAM 43
YP_001687742|Leptolyngbya ILQAEGYLGLSKPAKSEKTKKPRGQKGITSHGRRIIRGGVTLLERTYGRNRL 104
Consensus -------------------RSR-G-KGIS-HGRK-IR-SC-LLDDFRGR-AM
p27638 rep a WTVTLTDEDYLELASTCKWPAFQRRVTDLLVRHLQSNGDPGIVIGVVEVGAK 95
p31635 rep a WTVTLPDSDYLLLARSAQWKDFQRRVIDLLVRHLKANGDEAVVIAAVEVGSK 95
YP_001687742|Leptolyngbya SFITLTLPPAVAEDLSGRWAHVVDLMKRRLIYSNGLHGLPTEIIACTEVQEK 156
Consensus WTVTLTD-DYL-LA-S--W--FQRRV-DLLVRHL--NGDP--VIA-VEVG-K
p27638 rep a RFARTNRPDPHIHIVTTGYRSFDSDGRFLLSPQVCDQLVAKACQYAGLPFRK 147
p31635 rep a RFARTGRPDP-ISIDHHRLGRKHPEGGWLLCADRMDQLVAKACQYAQLPATS 146
YP_001687742|Leptolyngbya RYERTGEVALHLHIVMVGRHSRGAWCYSPRQLEKMWSECCETAVRNVIEPNE 208
Consensus RFARTGRPDPHIHIV--G--S----G--LL----MDQLVAKACQYA-LP---
p27638 rep a RPS--ASQVA 155
p31635 rep a RLS--CSRVE 154
YP_001687742|Leptolyngbya RVTSRVTNSRTESESNGNGNATGNTSSNANSNGNANGNIHTEVNWNAAVNVQ 260
Consensus R-S--AS-V-
p27638 rep a PIRHSVGAYMSKYLTKQIPVKPEDMPPEWAELIPVNGSTSRRPARQWLKGQL 207
p31635 rep a PVRHSVASYMSKYLTKDAPIDPESMPDEWQNLIPHQWWSQSAACKAMVEGVM 206
YP_001687742|Leptolyngbya RIKKSASAYMGKYLSKGTQTTQKIIDSGKAHLLPKAWYFCTQVLLERIKKAT 312
Consensus PIRHSV-AYMSKYLTK--P--PE-MP-EWA-LIP--W-----------KG--
p27638 rep a SNSRPLSAPSSSGNSDYSSSLVVGV----------GGTHRRLQEAKNRRCTD 249
p31635 rep a CKLPPAFAAFLVRKAILLENLELGR----------GGYALLVGRKES----- 243
YP_001687742|Leptolyngbya RVVSGNLAHEIYEHVLSHATEYLNYHRNIKAKCSDGREITVGWYGYLTKRGM 364
Consensus ----P--A------------L-LG-----------GG---------------
p27638 rep a RGVLLSVPFP 259
p31635 rep a -----AMTSR 248
YP_001687742|Leptolyngbya QELGKPLGVV 374
Consensus ----------

Figure 9: Comparison of the amino acid sequence of the p27638 and p31635 potential rep a protein with that of *Leptolyngbya*. The conserved Tyr residue involved in the linkage sites to the DNA when nicking occurs at the origin for rolling circle replication is marked with a box and the active site is underlined del Solar, 1998) (Ilyina, 1992)

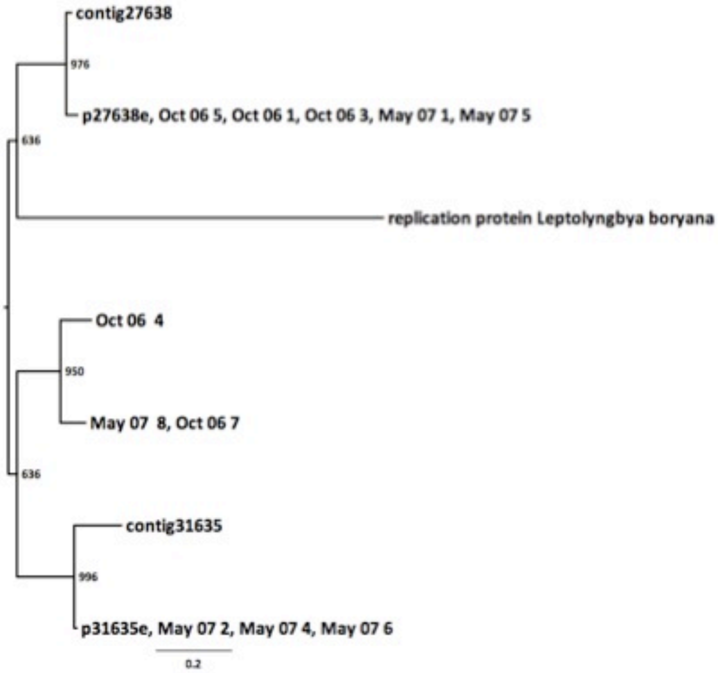
contig27638
976
p27638e, Oct 06 5, Oct 06 1, Oct 06 3, May 07 1, May 07 5
636
replication protein Leptolyngbya boryana
Oct 06 4
950
May 07 8, Oct 06 7
636
contig31635
996
p31635e, May 07 2, May 07 4, May 07 6
0.2

Figure 10: Neighbor-joining phylogeny of 27638 and 31635 based on amino acid sequence analysis of *rep*A region from a Scripps Institution of Oceanography pier sample on Oct 10, 2006 and May 15, 200 . and plasmid replication gene of cyanobacterium *Leptolyngba boryana*. The sequence information was obtained from the following accession number: YP 001687742. Bootstrap values (out of 1000) are shown adjacent to branch notes. The length of the horizontal branches, based on pairwise analysis, correspond to evolutionary distances and the scale bars show the number of substitutions per site.

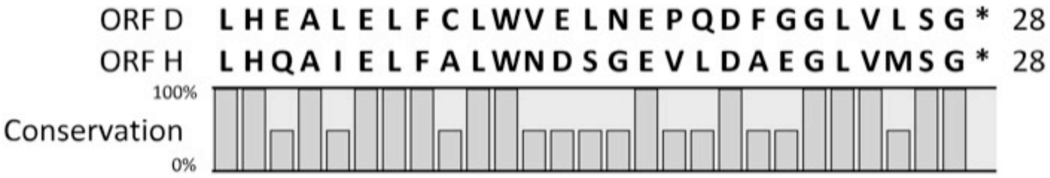
ORF D LHEALELFCLWVELNEPQDFGGLVLSG* 28
ORF H LHQAIELFALWNDSGEVLDAEGLVMSG* 28
100%
Conservation
0%

Figure 11: Alignment of proteins from plasmid family 27638 that are significantly similar

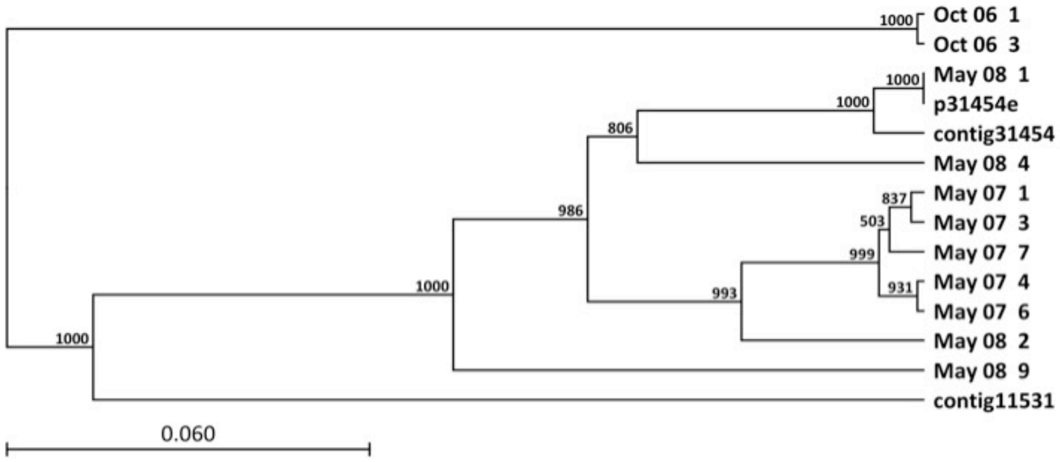

Figure 12: Neighbor-joining phylogeny of the 31454 plasmid family from the environment based on analysis of trimmed nucleotide sequences surrounding the rep a gene or cargo region from a Scripps Institution of Oceanography pier sample on Oct 10, 2006, May 17, 2007 and May 15, 2008 at 20m depth . Bootstrap values (out of 1000) are shown adjacent to branch notes. The length of the horizontal branches correspond to evolutionary distances and the scale bars show the number of substitutions per site.

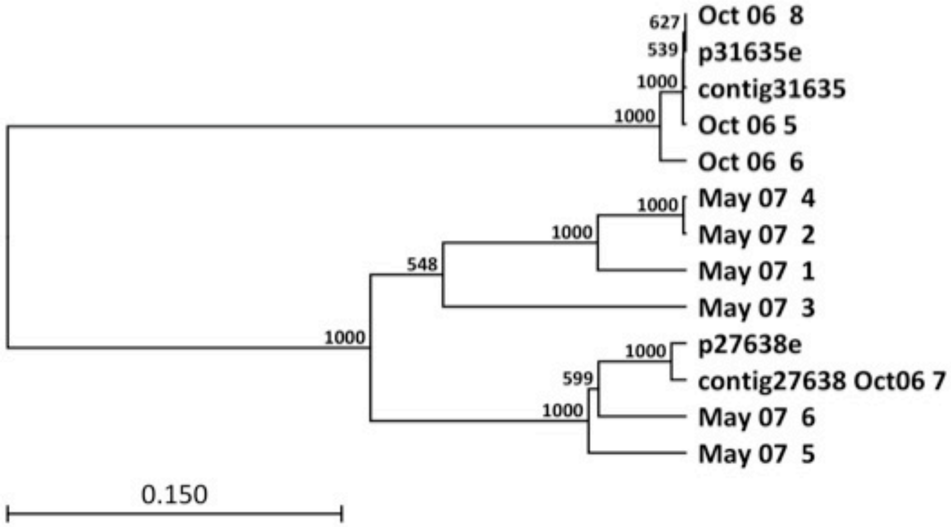

Figure 13: Neighbor-joining phylogeny of the 27638 plasmid family based on analysis of trimmed nucleotide sequences surrounding the rep a gene or cargo region from a Scripps Institution of Oceanography pier sample on Oct 10, 2006 and May 15, 200 . Bootstrap values (out of 1000) are shown adjacent to branch notes. The length of the horizontal branches correspond to evolutionary distances and the scale bars show the number of substitutions per site.

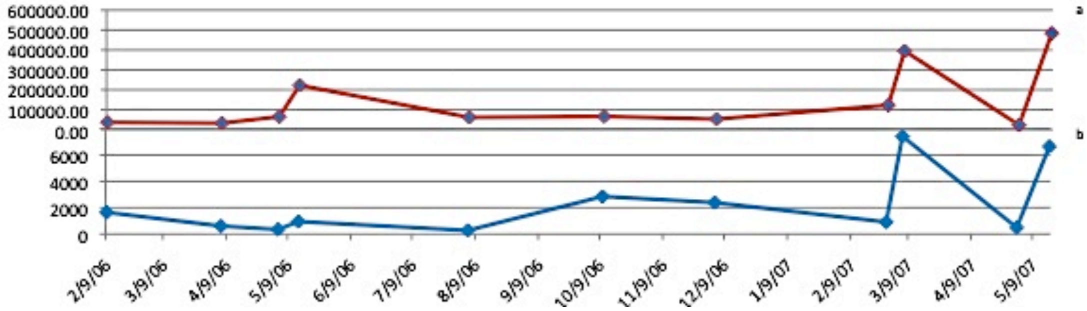

Figure 14: (a) abundance of *Synechococcus* from Scripps Institution of Oceanography (SIO pier surface samples (b) abundance of plasmids 27638 and 31635.

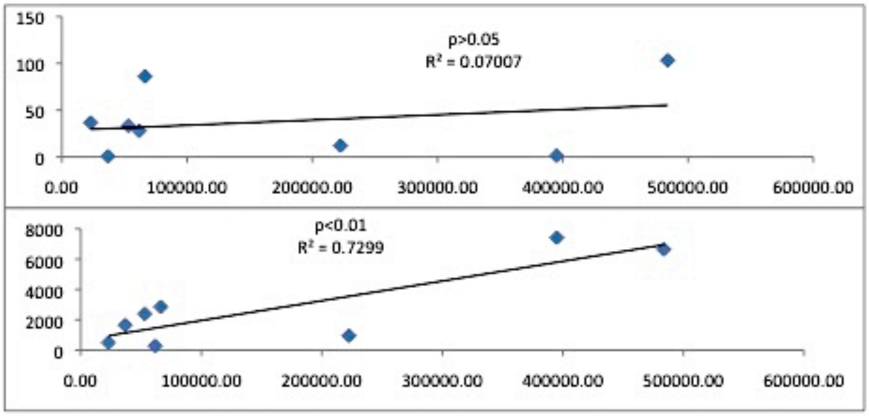

Figure 15: Scatter plot of total *Synechococcus* abundances versus plasmid based on environmental samples assayed by quantitative PCR. Results from correlation provided. (a) 31454 versus total *Synechococcus*. (b) 27638 and 31635 versus total *Synechococcus*. Best fit line is shown. Results from correlation analysis: correlation cofficient and associated p value.

Appendix

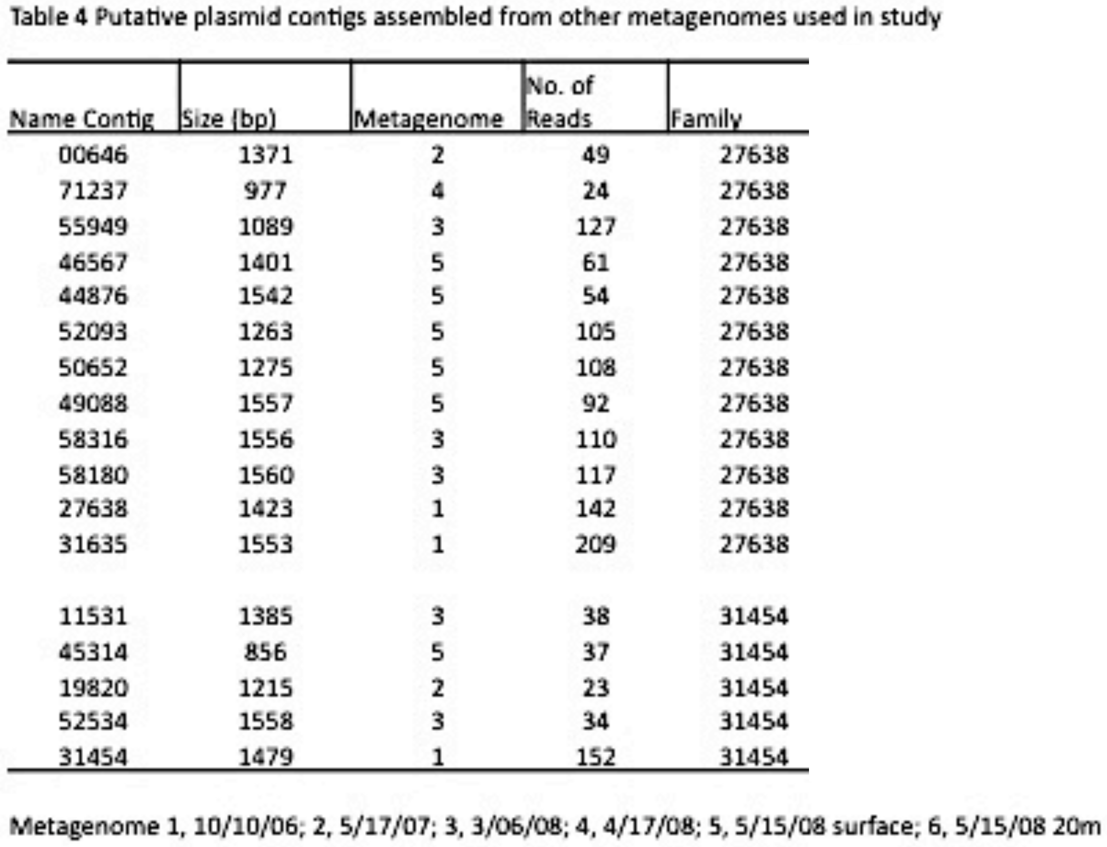

Table 4 Putative plasmid contigs assembled from other metagenomes used in study

Name Contig	Size (bp)	Metagenome	No. of Reads	Family
00646	1371	2	49	27638
71237	977	4	24	27638
55949	1089	3	127	27638
46567	1401	5	61	27638
44876	1542	5	54	27638
52093	1263	5	105	27638
50652	1275	5	108	27638
49088	1557	5	92	27638
58316	1556	3	110	27638
58180	1560	3	117	27638
27638	1423	1	142	27638
31635	1553	1	209	27638
11531	1385	3	38	31454
45314	856	5	37	31454
19820	1215	2	23	31454
52534	1558	3	34	31454
31454	1479	1	152	31454

Metagenome 1, 10/10/06; 2, 5/17/07; 3, 3/06/08; 4, 4/17/08; 5, 5/15/08 surface; 6, 5/15/08 20m

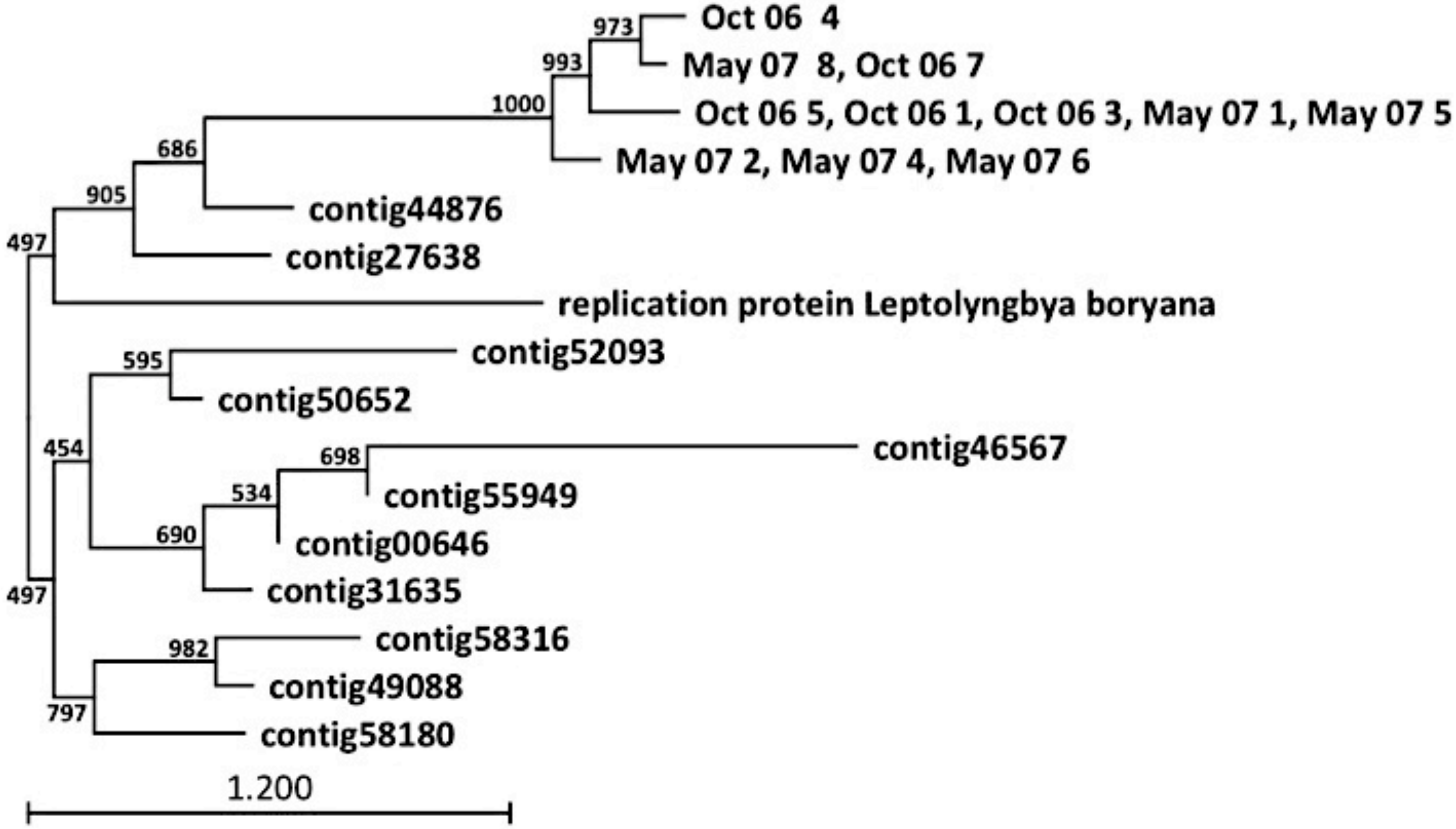

Figure 16: Un-rooted neighbor-joining phylogeny of plasmid family 27638 based on amino acid sequence analysis of *rep*A region from a Scripps Institution of Oceanography pier sample on Oct 10, 2006 and May 15, 200 and plasmid replication gene of cyanobacterium *Leptolyngba*. The sequence information was obtained from the following accession number: YP 001687742. Bootstrap values (out of 1000) are shown adjacent to branch notes. The length of the horizontal branches correspond to evolutionary distances and the scale bars show the number of substitutions per site.

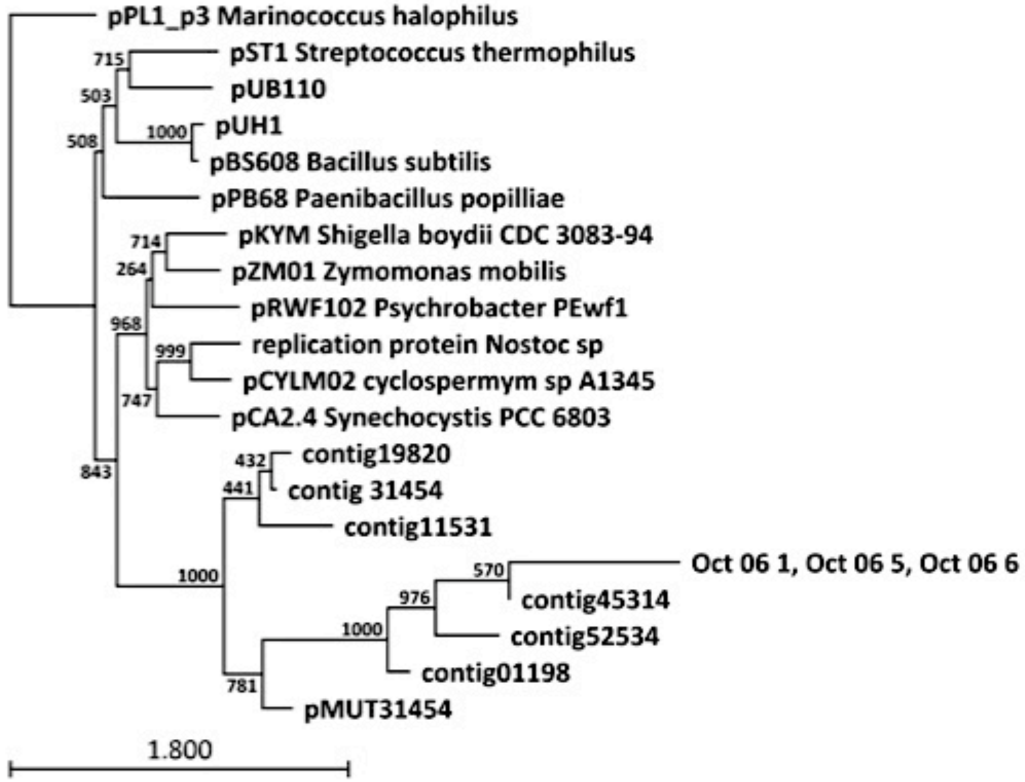
pPL1_p3 Marinococcus halophilus
715
pST1 Streptococcus thermophilus
503
pUB110
1000
pUH1
508
pBS608 Bacillus subtilis
pPB68 Paenibacillus popilliae
714
pKYM Shigella boydii CDC 3083-94
264
pZM01 Zymomonas mobilis
pRWF102 Psychrobacter PEwf1
968
999
replication protein Nostoc sp
pCYLM02 cyclospermym sp A1345
747
pCA2.4 Synechocystis PCC 6803
432
contig19820
843
441
contig 31454
contig11531
1000
570
Oct 06 1, Oct 06 5, Oct 06 6
976
contig45314
1000
contig52534
contig01198
781
pMUT31454
1.800

Figure 17: Neighbor-joining phylogeny of the 31454 plasmid family based amino acid sequence analysis of rep a region from a Scripps Institution of Oceanography pier sample on Oct 10, 2006 and various gram negative bacteria (highest blast hits). The sequence information was obtained from the following Accession number:AAA25513.1 (*Nostoc* sp), YP_001965480.1 (*Cyclindrospermum* sp A1345), replication protein *[Shigella boydii* CDC 3083-94], AAA02970.1 [*Synechocystis* sp.], NP_045297.1[*Zymomonas mobilis*],
CAB43206.1 [*Streptococcus thermophilus*], AAA99151.1 [B*acillus subtilis]*
Bootstrap values (out of 1000) are shown adjacent to branch notes. The length of the horizontal branches correspond to evolutionary distances and the scale bars show the number of substitutions per site.