Investigation of a Novel *Mollicute*-like Organism Inhabiting the Human Gastrointestinal Tract

A thesis
submitted in partial fulfillment
of the requirements for the degree
of
Master of Science in Biological Sciences
at
The University of Waikato
by
ANDREW SHANE CARE

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

The University of Waikato
2009

Abstract

The microflora inhabiting the human gastrointestinal tract can be considered an essential 'metabolic organ', in a symbiotic relationship with its host. Due to the low cultivability and inappropriate sampling methodology the microflora is poorly explored and ill-defined.

Preliminary, molecular-based research at the University of Waikato revealed the presence of 16S rRNA gene sequences originating from novel *Mollicute*-like species inhabiting the human GI tract. A ~830bp 'consensus' sequence representing these novel *Mollicute*-like sequences was classified within the *Mollicute* Genus *Anaeroplasma* the type species of which is *Anaeroplasma abactoclasticum*. It also displayed near exact matches with 16S rRNA sequences obtained from the human GI tract and matches of high similarity to those from the mouse GI tract in the NCBI database.

This thesis describes an attempt to design and create primers that would amplify and characterize full-length versions of these *Mollicute*-like sequences from samples obtained from the mucosal surface of the human gastrointestinal tract. Primers sets targeted extended 5' and 3' versions of these novel 'known' sequences and were designed from sequence matches found in the preliminary work and other related sequences from the NCBI database.

The attempt to amplify a full-length version of these novel *Mollicute*-like sequences was proven to be unsuccessful. No sequences were classified within the Genus *Anaeroplasma*, although 81% of amplicons from the 5' extending primer sets were classified within the same division as the *Mollicutes,* the *Firmicutes*, only 6% of the sequenced amplicons from the 3' extending primer set belonged to this division. Phylograms containing these 'relevant' sequences and the 'consensus' sequence grouped the 'consensus' sequence separately, indicating a lower relatedness than would have been seen if any of the amplicons contained the 'consensus' sequence.

Acknowledgements

Thank you to my supervisor David Musgrave for his knowledge and guidance throughout this entire project.

Also, many thanks to those in the 'Molecular Genetics' Laboratory at The University of Waikato, for creating such a friendly and helpful work environment:

Ray Cursons

Dick Wilkins

Raewyn Towers

Olivia Patty

Amanda Bean

Ali Ruthe

and Gregory Jacobson

Thank you to friends and family who have encouraged me to further my education and after five years of university, encourage me to find employment.

Table of Contents

Index of Figures

Index of Tables

Chapter 1: Literature Review

1.1 Introduction

1.1.1 The Human Gastrointestinal Tract

Humans have evolved a complex symbiotic relationship with the trillion or more (10^{13} to 10^{14}) anaerobic micro-organisms inhabiting the distal regions of their gastrointestinal (G.I) tract (Gill et al., 2006). Collectively known as the gut 'microflora', these microbes endow their human hosts with fundamental genetic and metabolic properties that they have not had to evolve themselves, contributing significantly to their health and well-being (Backhed, Ley, Sonnenburg, Peterson, & Gordon, 2005; Turnbaugh et al., 2007). Some of these properties include: the catabolism of polysaccharides, the provision of essential nutrients (e.g. vitamin K), protection against epithelial cell damage and pathogenic infection, immunity, regulation of host fat storage and protection against the development of G.I. cancer (Eckburg et al., 2005). These properties taken into account along with the shear number of micro-organisms, it is of no surprise that many consider the gut's microbiota to be a 'metabolic organ' (Turnbaugh et al., 2007), located within a 'Superorganism' (Gill et al., 2006). 'Superorganism' is a term used throughout the relevant literature in reference to the human body, describing it as a composite of human and microbial cells, hence, the human genetic landscape is seen as a collaboration of the human genome and the genomes of the resident microflora (collectively termed the 'microbiome')(Gill et al., 2006).

1.1.2 Studying the microflora

It is widely accepted that up to 99.8% of the microbes present in many environments are not readily culturable (Streit & Schmitz, 2004), and with the G.I tract playing host to a plethora of microenvironments, it is understandable that its microflora is poorly studied and ill-defined to date. With the current limitations of laborious and costly culture based techniques and continuing advancements and utilisation of DNA-based

molecular research, headway has been made in determining the genetic composition of the microflora, described as the 'microbiome'.

1.1.3 Studying the microbiome

The microbiome is estimated to contain ≥100 times as many genes as our 2.85-billion bp human genome (Gill et al., 2006), determining its composition is an enormous task, one which has being addressed via 'metagenomics'.

'Metagenomics' (synonymous with Environmental Genomics, Ecogenomics or Community Genomics) is a term that encompasses various approaches taken in the direct genomic analysis of the microbial communities inhabiting an environmental sample with the aim to access their genomic potentials, while bypassing the isolation and cultivation of individual species.(Naomi, 2006). At present this is done by isolating DNA from an environmental sample, cloning the DNA into a vector, transforming the clones into a host bacterium, and screening the transformants that result (Figure 1.1-A). The clones can then be screened for phylogenetic markers for instance 16SrRNA or rpoB, or other conserved genes by hybridisation or multiplex PCR, or expression of specific traits, for example antibiotic production, or they can be sequenced randomly (Handelsman, 2004).

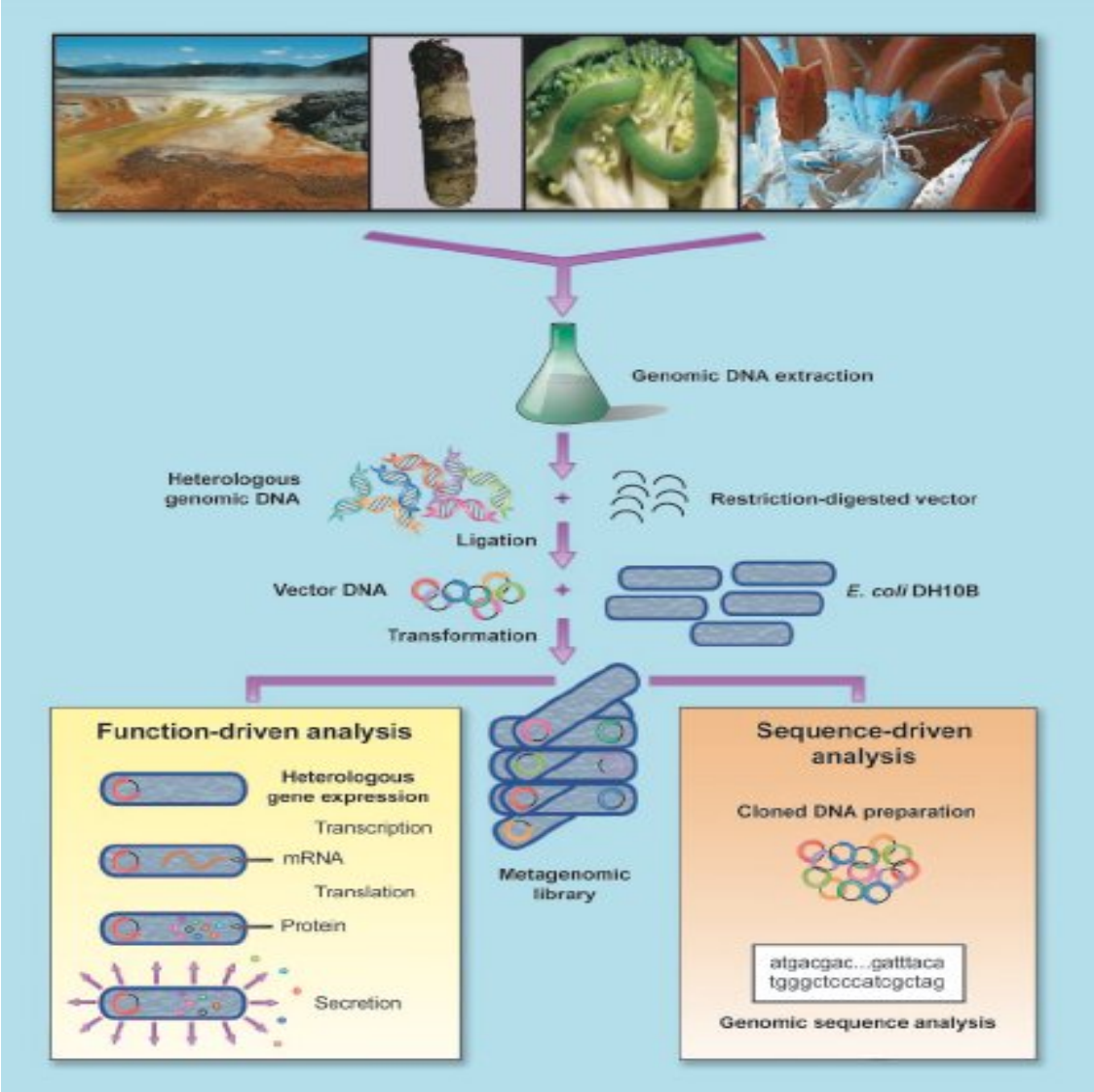


Figure 1.1-A: Flowchart summarising the 'Metagenomic Process' (Handelsman, 2004).

1.1.3.1 Metagenomic Libraries

Whole-genome shotgun sequencing and assembly has been applied to diverse microbial communities including the microbiome. This approach has been validated in work by (Venter et al., 2004), and works on the premise that sequences belonging to abundant species will be soundly represented in random shotgun sequencing data, whereas species of lower abundance will be represented to a lesser extent (Gill et al., 2006).

Libraries of whole-genome shotgun sequences retrieved from two adult human fecal samples were assembled and analysed by (Gill et al., 2006), in an attempt to analyse the human distal gut microbiome. In this study ~78 million bp of DNA sequence was obtained, of which 33,753,108bp belonged to 17,668 contigs which were assembled into 14,572 scaffolds. The remaining 45,078,063bp of DNA couldn't be assembled into contigs; this is thought to be related to the depth of sequence coverage and the low relative abundance of the organisms from which the sequences were derived. A total of 50,164 open reading frames (ORFs) were predicted in the data set, with 19,866 being unique; 589 contigs were assigned to Archaea and 13,130 to Bacteria. All other contigs didn't have matches with any known ORF's or gave ambiguous results. The ability of this metagenomic method to define an environment's diversity was shown in the comparison of shot-gun sequences to a sequenced genome. This was done with the lactic acid producing bacterium *Bifidobacterium longum*; 1965 reads, representing 1,617,706bp of relevant sequence was compared to, and covered ~0.7% of the *B. longum* genome. 52% of the reads had less than 95% identity when aligned with the genome sequence (Figure 1.1-B). These results give reason to believe that these reads did not originate from a single strain, but rather from multiple strains, hence giving insight into the diverse nature of the microbiome.

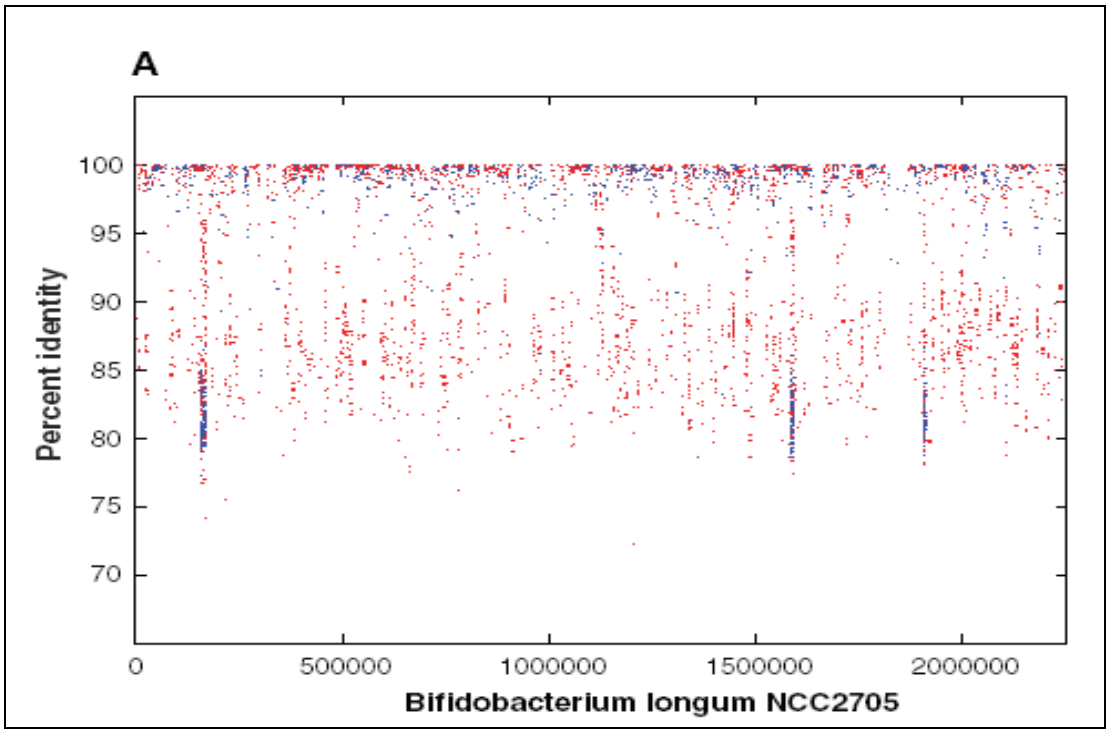


Figure 1.1-B: Displays the diversity within the microbiome, using one species and its multiple strains, all of which were found using shotgun sequencing methodology. Percentage identity plot of the sequence alignments between the relevant random reads obtained from human fecal samples and the completed genome of *Bifidobactrium longum*. The *x* axis represents the coordinate across the genome, and the *y* axis represents the percentage identity of the match (Gill et al., 2006).

1.1.3.2 Phylogenetic Determination using 16SrRNA

Ribosomal RNA's are part of the machinery required for DNA translation during protein synthesis. The ribosome of Prokaryotes is 70 Svedberg (S) units in size and is composed of two subunits, the large subunit 50S (contain the 5S rRNA and the 23S rRNA) and the small subunit 30S (contains 16S rRNA molecule). Due to the complexity of its structure and function of the rRNA molecules, ribosomal RNA genes are highly conserved and are employed in the measurement of evolutionary relationships among organisms. 16S rRNA has become the most widely used of the rRNAs in the assembly of Prokaryote phylogenies, this is attributable to the choice of variable regions where the sequence differs markedly within the 16S rRNA gene.

Variable regions are unique to a particular organism or organisms, therefore they allow the identification and phylogenetic positioning of bacteria (Blaut et al., 2002; Turnbaugh et al., 2007). 16S rDNA data is integral to the determination of the human gut microflora's composition, work of such nature was undertaken by (Eckburg et al., 2005).

(Eckburg et al., 2005) obtained human mucosal samples from the: caecum, ascending colon, transverse colon, descending colon, sigmoid colon, and the rectum of three healthy adult patients undergoing colonoscopy; fecal samples were collected 1 month later. Their aim was to characterise the adherent microbial populations and make comparisons between the subjects and mucosal sites using 16S rDNA sequences. 16S rDNA was amplified by polymerase chain reaction, for bacterial 16S rDNA amplification the primers used were: Bact-8F (5'-AGAGTTTGATCCTGGCTCAG-3') and Bact-1391R (5'-GACGGGCGGTGTGTRCA-3'). All 21 samples yielded bacterial products; these were cloned, bi-directionally sequenced and subjected to numerical ecology approaches.

11,831 near-full length, non-chimeric, bacterial 16S rDNA sequences underwent phylogenetic analysis; with the parameter of 99% minimum similarity within a phylotype, 395 bacterial phylotypes were determined (Figure 1.1-C). 244 (62%) of these were novel phylotypes and 80% had not been cultivated at the time of publication (2005). The majority of organisms belonged to the two bacterial phyla of the *Firmicutes* and *Bacteriodetes*. The *Firmicutes* phylum was represented by 301 phylotypes, 191 of them were novel and the bulk of them (95%) belonged to the Class *Clostridia*, while 4.5% were *Mollicutes*, 13 of the 20 *Mollicute* phylotypes found were novel. Interestingly, there was a large variation between *Bacteriodetes* phylotypes between subjects.

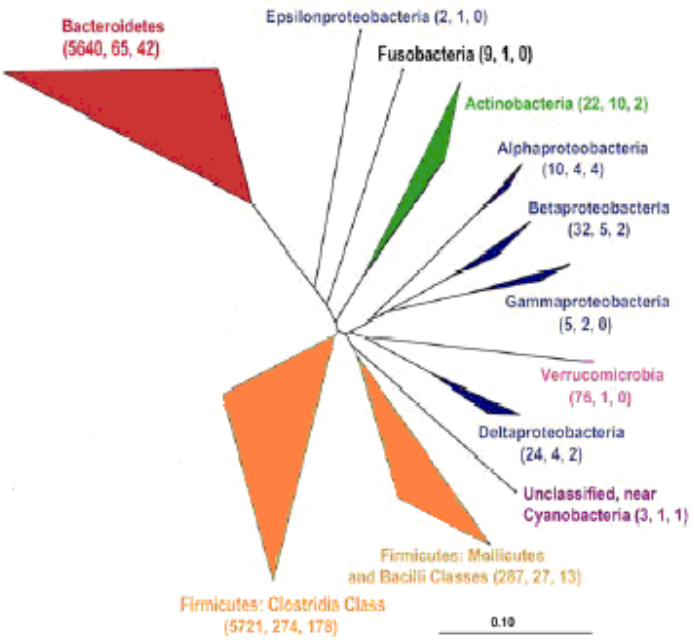


Figure 1.1-C: Phylogenetic tree based on a combined human intestinal 16S rDNA sequence data set. The label for each clade includes, in order, the total number of recovered sequences, phylotypes, and novel phylotypes (in parentheses). 20 *Mollicute* phylotypes were found, 13 of which were novel. The angle where each triangle joins the tree represents the relative abundance of sequences, and the lengths of the two adjacent sides indicate the range of branching depths within that clade (Eckburg et al., 2005).

1.1.3.3 Defining the Human Gut Microflora from Fecal Samples

Much of the early data used in the assistance of defining the human gut microflora has been derived from fecal samples, yet there has been little exploration into the possible differences that might occur between fecal and mucosal communities, between sites and between subjects. (Eckburg et al., 2005) took relative abundance and genetic divergence of sequences within communities and applied the statistical method of double principal coordinate analysis (DPCoA) to access the aforementioned differences. The findings displayed significant interpersonal differences, differences between mucosal and fecal compositions, and fewer differences between intrapersonal sampling sites (Figure 1.1-D).

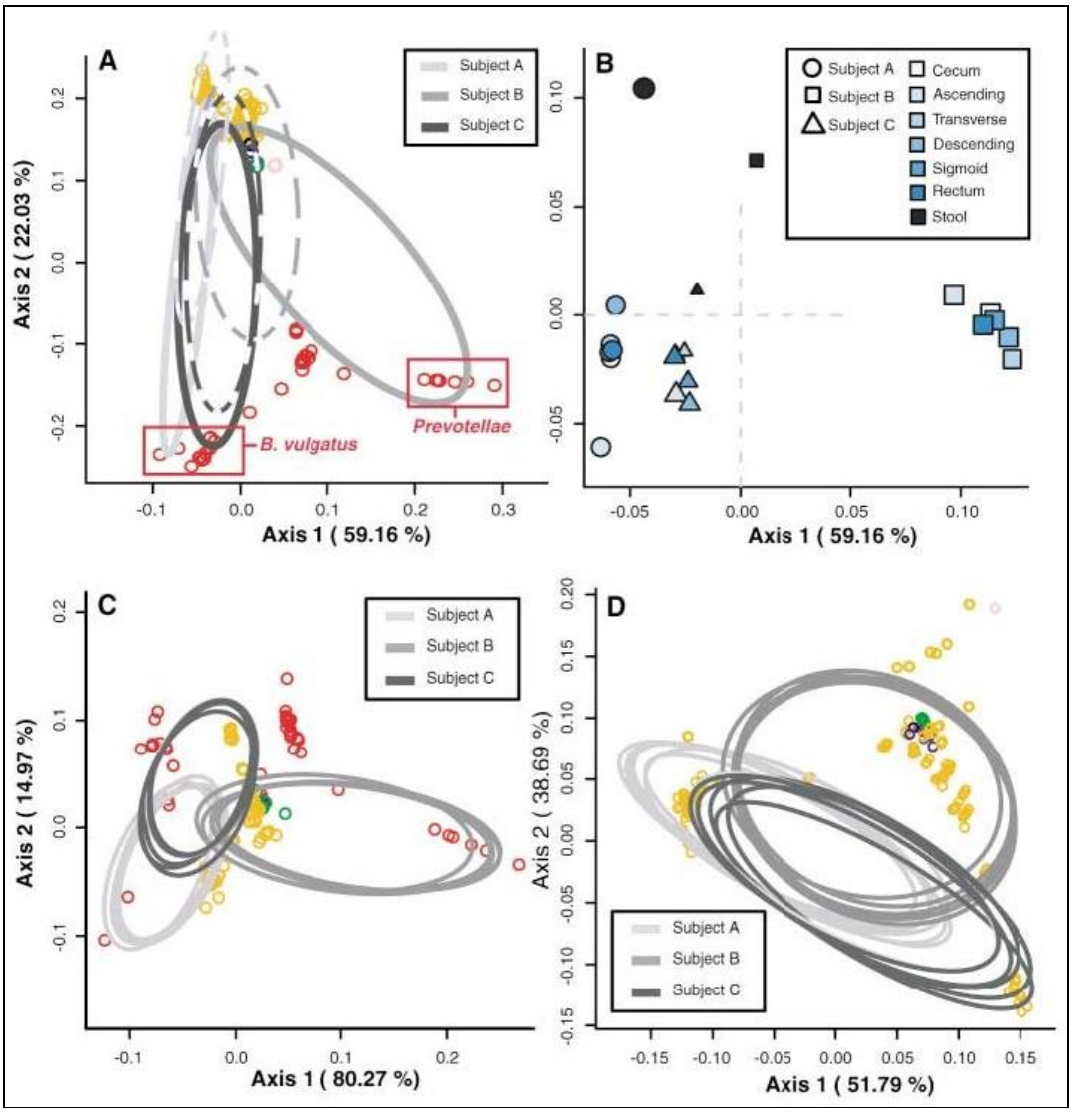


Figure 1.1-D: DPCoA for (A) colonic mucosa (solid lines) and stool (dashed lines), (C) colonic mucosal sites alone, and (D) mucosal sites excluding *Bacteroidetes* phylotypes. Phylotypes are represented as open circles, colored according to phylum. Phylotype points are positioned in multidimensional space according to the square root of the distances between them. Ellipses indicate the distribution of phylotypes per sample site, except in (A), where all mucosal sites are represented by one ellipse. Percentages shown along the axes represent the proportion of total Rao dissimilarity diversity coefficient (which accounts for both phylotype abundance and dissimilarity), captured by that axis. (A) Is the best possible two-dimensional representation of the Rao dissimilarities between all samples. (B) Is an enlarged view of (A), depicting the centroids of each site-specific ellipse. Subject ellipse distributions remain distinct after stool phylotypes (C) and Bacteroidetes phylotypes (D) are excluded from the analysis, as so many phylotypes were already found to vary between subjects (Eckburg et al., 2005).

1.1.4 Determining the Functionality of the Gut Microflora

Many functions of the microflora have been determined in the past, many of which have already been mentioned. Despite this, with such an incomplete knowledge of the composition of the microflora it is expected that much of the functionalities of the organisms that comprise it are yet to be determined. Consequently, this lack of understanding has motivated investigations into the microflora's functionality. The ascertainment of many of these potential functions could very well lend hand to various fields, particularly those of medicine and pharmacology.

1.1.4.1 The Human Gut Microflora and Obesity

The importance of understanding the functionality of the human gut microflora comes to light in work carried out by (Backhed et al., 2004; R. E. Ley et al., 2005; R. E. Ley, Turnbaugh, Klein, & Gordon, 2006).

(Eckburg et al., 2005; Gill et al., 2006) used the mouse obesity model, where, genetically obese *ob/ob* mice, lean *ob/+*, lean wild-type +/+ siblings and their lean *ob/+* mothers were raised on polysaccharide-rich diets. As the gut's microbial community was inherited from their mothers, it was an opportunity to measure the composition and any changes in that composition between the lean and obese mice. Molecular analyses of 16S rRNA sequences amplified from the caecum of the mice showed that the distal gut is dominated by the presence of two bacterial divisions, the *Bacteriodetes* (20-40%) and the *Firmicutes* (60-80%). These results were compared to 16S rRNA sequences amplified from the human colon using the same primers as with the mice, with significant similarities between them at this division level (Figure 1.1-E).

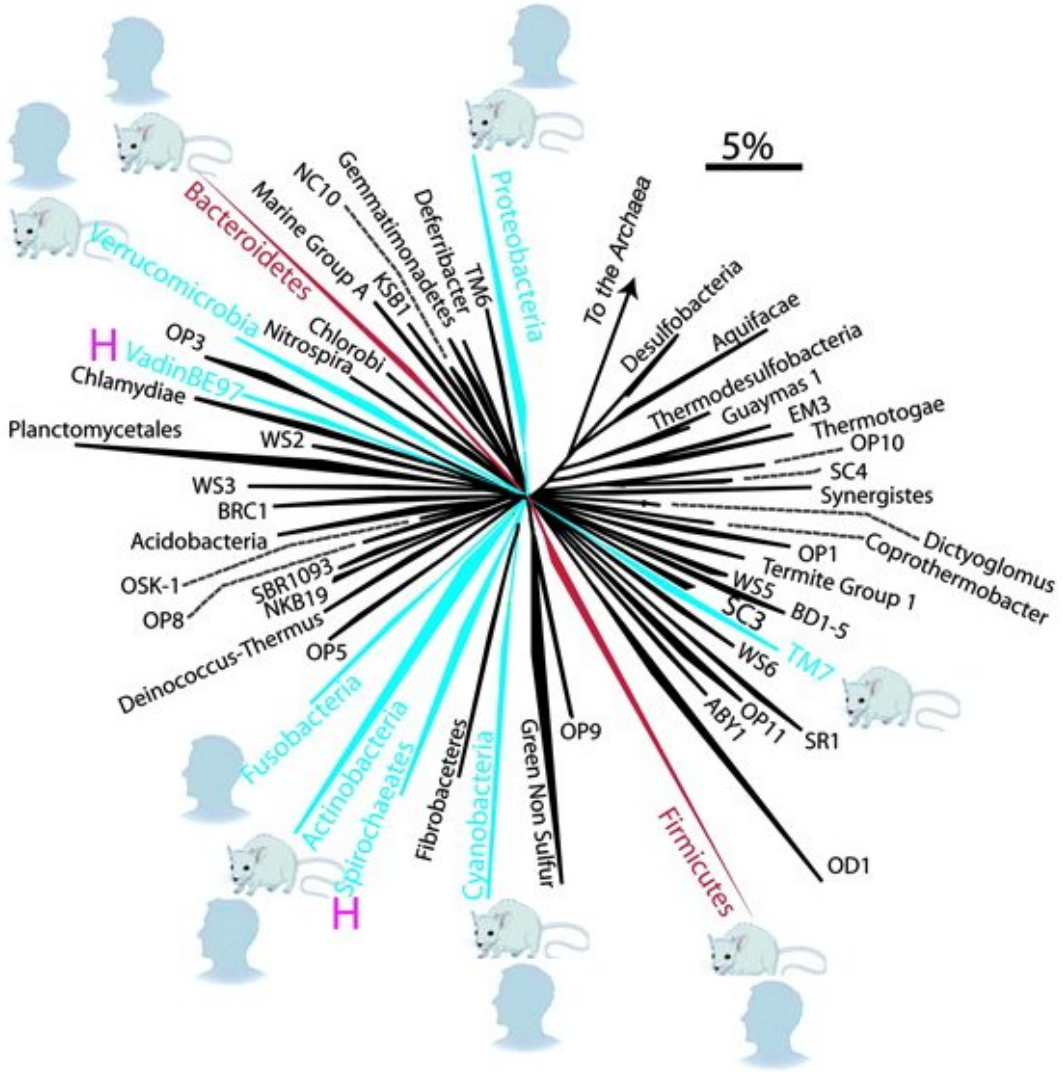


Figure 1.1-E: Phylogenetic tree displaying the bacterial Divisions. Those found in the mouse gut are represented with a mouse symbol, while those from the human gut are represented with a human-head symbol. Divisions labeled 'H' are those found in human fecal samples. Dominant Divisions from both Mice and Humans are labeled in red (*Bacteriodetes* and *Firmicutes*), rarer Divisions are blue, and undetected Divisions are black (R. E. Ley et al., 2005).

It has also been shown that a decrease in the *Bacteriodetes* to *Firmicutes* ratio leads to an increase in the efficiency with which food is converted and stored as fat by microbial manipulations of physiological and metabolic pathways (Backhed et al., 2005; R. E. Ley et al., 2005; R. E. Ley et al., 2006). Higher proportions of the *Firmicutes* division were found in the genetically obese mice (Figure 1.1-F).

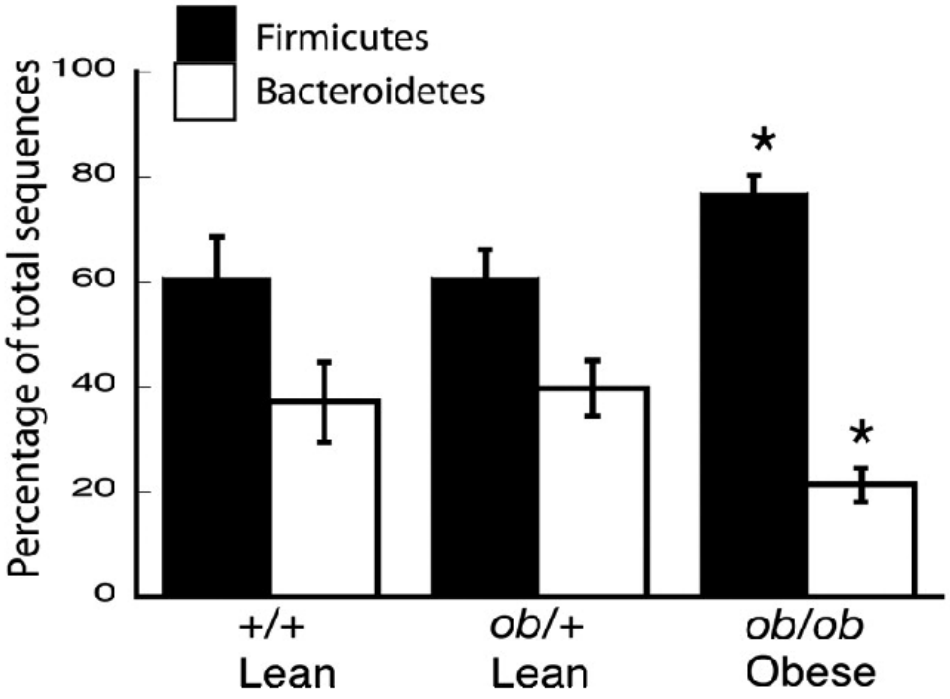


Figure 1.1-F: Graph depicting the percentage of sequences found in the caecal contents of each of the transgenic siblings from the mouse obesity model. The *ob/ob* obese siblings has a higher ratio of *Firmicute* sequences to *Bacteriodetes*, while it's lean (+/+, *ob*/+) have a markedly less obvious difference in ratio, and a higher amount of *Bacteriodetes* sequences than that of the obese individual. (R. E. Ley et al., 2005).

This microflora has been suggested to contribute to obesity by an increased efficiency in food conversion to fat, despite a lowered quantity of food intake and increased overall metabolic rate in comparison to their normal littermates (R. E. Ley et al., 2005). (P. J. Turnbaugh et al., 2006) transplanted the microbiota from the caecum of obese (*ob/ob*) or lean (*ob*/+) mice into adult germ-free mice. 16S rRNA sequence analysis confirmed the obese *ob/ob* donor's microbiota had a higher proportion of *Firmicutes* to *Bacteriodetes* than the lean ob/+ donor, concordant with the aforementioned studies. Figure 1.1-G, displays the comparison between the ceacal contents of the recipients. Figure 1.1-G (a), there was higher amount of short-chain fatty acids present in the *ob/ob* recipients implying a greater conversion of carbohydrate to fats; (b) there was also less energy in the feces of the *ob/ob* recipients implying a greater efficiency harvesting energy from food than the *ob*/+ recipients; (c) recipients of the transplanted *ob/ob* microbiota gained significantly more fat (%) than the recipient of the *ob*/+ microbiota.

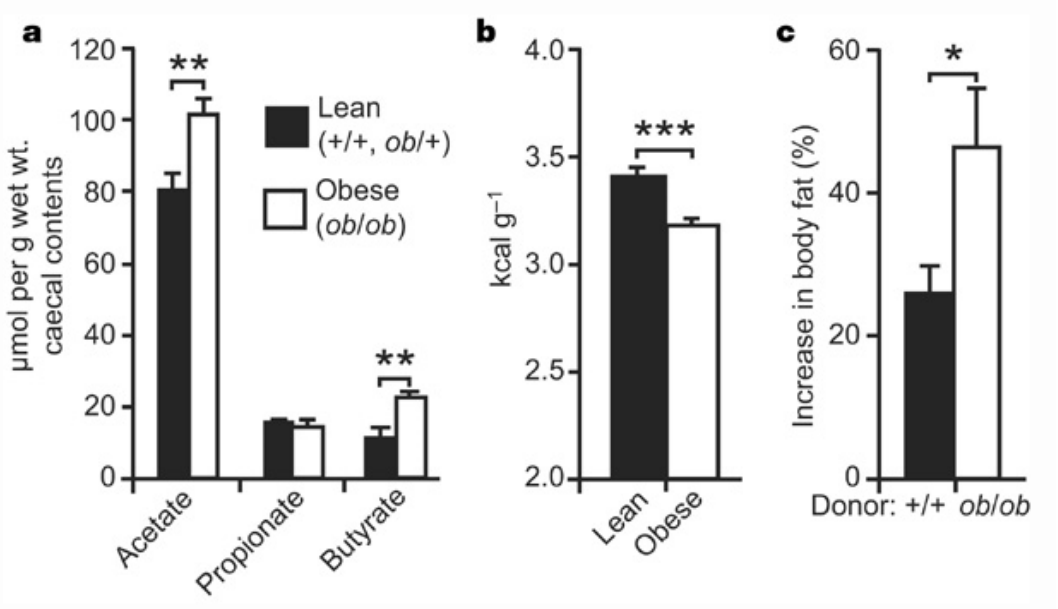


Figure 1.1-G: (a) Gas- chromatography mass-spectrometry quantification of short-chain fatty acids found in the caecal content of both lean (+/+, *ob*/+) and obese (*ob/ob*) mice. **(b)** Bomb calorimetry of the gross energy content ($kcalg^{-1}$) of feces from both lean and obese mice. **(c)** Increase of body fat (%) between lean and obese mice. (P. Turnbaugh et al., 2006)

A decrease in *Firmicutes* and an increase in *Bacteriodetes* in the human G.I tract has been correlated with weight loss in obese individuals when undergoing calorie restricted diet regimes (R. E. Ley et al., 2006). In work by (R. E. Ley et al., 2006)12 obese individuals were restricted to either fat-restricted (FAT-R) or carbohydrate-restricted (CARB-R), low calorie diets. 16S rRNA data was collected over the course of 1 year from fecal samples as means of keeping track of any changes in the composition of their gut microbiota. Figure 1.1-H is taken from this work, (a) displays the consistency of bacterial lineages within people over time, with intrapersonal differences being significantly less than interpersonal differences, this information is complemented above in Figure 1.1-D (Eckburg et al., 2005). (b) Shows the slight decrease of *Firmicutes* and steady increase in *Bacteriodetes* sequences found in samples over time. This increase in *Bacteriodetes* also correlated with weight loss, as can be seen in (c).

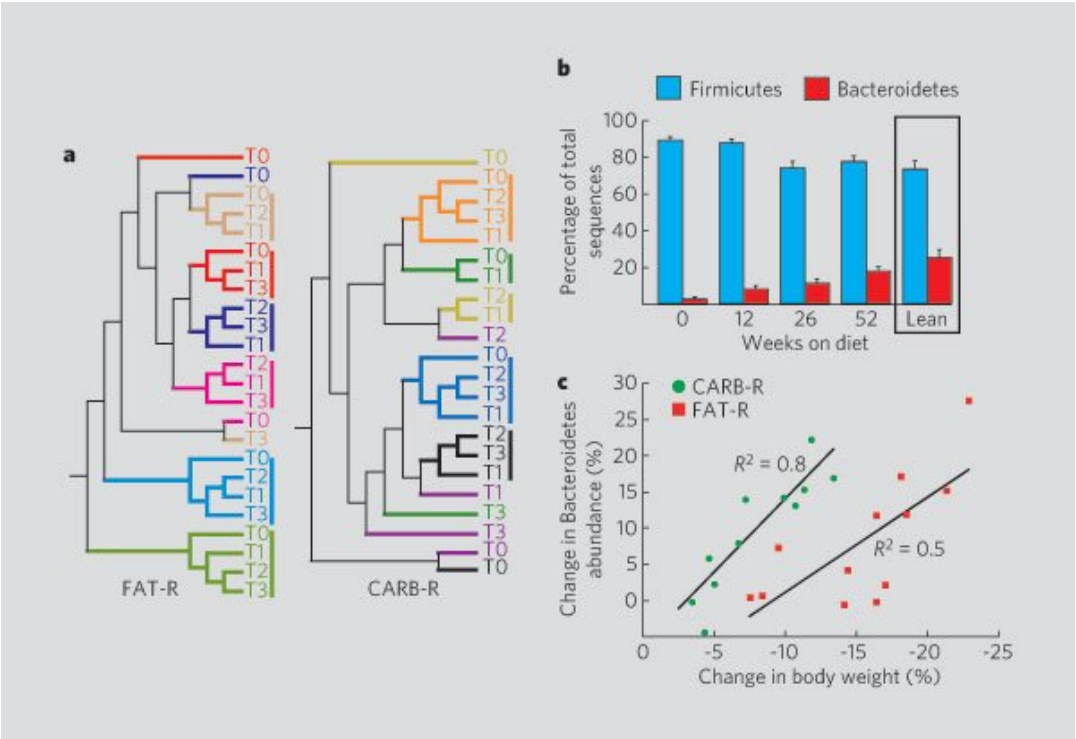


Figure 1.1-H: (a) 16S rRNA sequences taken from fecal contents of two obese human subject groups over time (T0, baseline; T1, 12 weeks; T2, 26 weeks; T3, 52 weeks), one group had a fat-restricted (FAT-R) diet, while the other had a carbohydrate restricted diet (CARB-R) both diets were calorie restricted. **(b)** Relative abundance of *Firmicute* to *Bacteriodetes* sequences averaged from all samples taken over time. **(c)** Change in relative abundance of *Bacteriodetes* in subjects with weight loss above a threshold of: 2% for CARB-R diets and 6% for FAT-R diets (R. E. Ley et al., 2006).

With the increasing importance of finding ways to combat obesity in the western world, this research could be seen as an opportunity to further understand the microflora component of this diverse syndrome. They also reinforce the view that manipulation of the gut's microflora can contribute significantly to human health.

1.2 Preliminary Analysis

1.2.1 Discovery of a Novel Mollicute-like Sequence

Culture independent, molecular based research undertaken at the Institut Pasteur and The University of Waikato, has revealed the presence of a novel *Mollicute* inhabiting the human gastrointestinal tract and suggested to be associated with the mucosal surfaces of the Esophagus, Stomach, Colon and Ileum. The research was initiated by the procurement of three healthy gastrointestinal mucosal tissue samples from each of nine elderly male patients undergoing gastrointestinal surgery for the treatment of various cancers; these were accompanied by further DNA samples, which were gained from the fecal matter of two healthy French teenagers. These samples were supplied by Francoise Rieu-Lesme (Rieu-Lesme, Delbes, & Sollelis, 2005).

(Kubs & Musgrave, 2007) extracted total DNA from both the tissue and fecal samples was amplified using the primer set: 109F (5' ACGGCTTSAGTAAACRCGTRG 3') and 927R (5' CCR KYC AAT TCC TTT AAG TTT C 3'), which resulted in the amplification of a ~830bp region of the bacterial 16S rRNA gene. These amplicons were cloned into plasmids then sequenced or sequenced directly from their PCR products, using M13F/M13R or 109F/927R respectively. Sequencing results were edited in Chromas prior to analysis using BLAST from the NCBI website and Classifier from the RDPII website.

A number of polymorphisms were also identified in the novel *Mollicute*-like sequences, which suggested that not only had a new genus being identified but closely related stains had also being identified (Table 1.2-A). Phylogenetic clustering of the *Mollicute*-like sequences also suggested interpersonal diversity, with sequences from individuals grouping together in Figure 1.2-A.This detail has been found in other studies that have explored the interpersonal differences in the human gut microflora (Figure 1.1-D) (P. J. Turnbaugh et al., 2006).

Table 1.2-A: Potential polymorphisms from all "full-length" sequences. "ConSeq"= 'Consensus' sequence. "Position"= the position of the nucleotide difference, numbering corresponds with the JalView alignment. B= no nucleotide. "."= same nucleotide as the consensus sequence (Kubs & Musgrave, 2007).

	ConSeq	T	C	A	B	G	A	A	D	A	A	A	A	G	D	A	T	A	T	G	A	A	A	G	C	A	C	T	T	
	Position	25	68	86	3[illegible]	[illegible]0	124	[illegible]40	1[illegible]4	182	192	20[illegible]	238	259	261	27[illegible]	299	312	317	346	375	394	425	449	456	483	4[illegible]	493	7[illegible]		
Sample ID																														
RL[illegible]-1			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Polymorphism1
RL[illegible]-5			B	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	Polymorphism2
RL[illegible]-11			.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	Polymorphism3
RL2-49			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Polymorphism1
RL2-65			.	T	.	.	B	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	Polymorphism4
RL2-67			.	.	.	.	.	.	.	T	.	.	G	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Polymorphism5
DM1-39			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	G	.	.	.	.	.	.	.	Polymorphism3
DM1-41			.	.	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	.	Polymorphism7
DM1-43			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Polymorphism1
DM1-47			B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Polymorphism8
DME.2					.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.		Polymorphism9
DME.3					.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.		Polymorphism10
DME.6					.	.	.	G	.	.	.	.	.	A	.	.	.	G	.	.	G	.	.	.	.	.	.		Polymorphism11
DME.9					G	B	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C		Polymorphism12
DME.15					.	.	.	G	.	.	.	.	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.		Polymorphism13
DME.17					.	.	.	G	.	.	G	.	.	.	.	.	.	G	.	.	.	.	A	.	.	.	.		Polymorphism14
DME.20					.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.		Polymorphism15
DME.24					.	.	.	G	.	.	A	G	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.		Polymorphism16

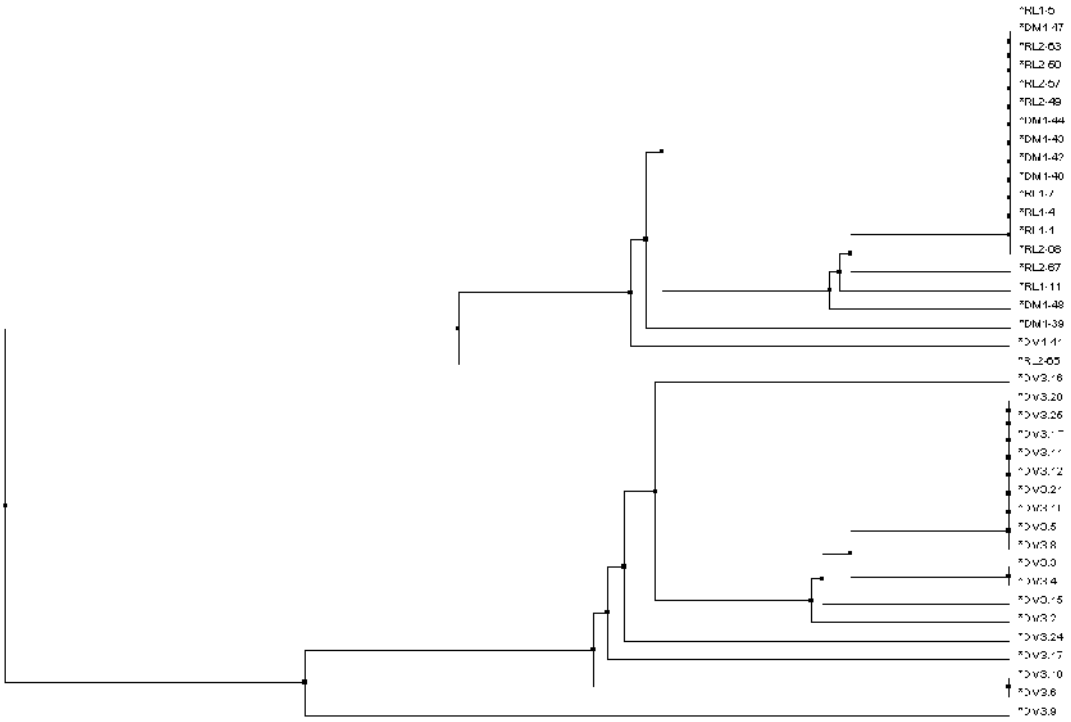


Figure 1.2-A: The interpersonal diversity of the *Mollicute*-like sequences is displayed in this phylogenetic tree. The tree was calculated from all of the "full-length" *Mollicute*-like polymorphisms using JalView % identity .(Kubs & Musgrave, 2007).

BLAST submission for all full length sequences, and most partial length sequences, gave 'best hits' with a 16S rRNA gene from an uncultured bacterium clone RL386_aao86c09 or an uncultured bacterium clone RL386_aao86h04 16S ribosomal RNA gene, partial sequence (P. J. Turnbaugh et al., 2006), this sequence was then Classified within the *Mollicute* Genus *Anaeroplasma* (94% confidence), the type species of which is *Anaeroplasma abactoclasticum* an anaerobic *Mollicute* found in the caecal contents of some mammalian species (S. Razin, Yogev, & Naot, 1998).

A consensus sequence 'DM1 Consensus' was constructed from the *Mollicute*-like sequences taken from one individual (DM1) and was used as a representative for all of the relevant *Mollicute*-like sequences. This was credible due to the fact the DM1 consensus sequence was virtually identical to the total consensus sequence from (Kubs & Musgrave, 2007).

1.2.1.1 The Novel Mollicute-like Sequence and Obesity Studies

The DM1 consensus sequence was submitted to the non-redundant BLAST database in October 2007, its 'best hit' was with the uncultured bacterium clone RL386_aao86c09, a sequence submitted from the work (R. E. Ley et al., 2006). A phylogenetic tree constructed from the BLAST results and utilising a neighbor joining algorithm within the NCBI BLAST results webpage is displayed in Figure 1.2-B. It shows the DM1 consensus sequence (Iclj7127, highlighted in yellow) is most closely related to the sequences obtained from the human (R. E. Ley et al., 2006) and is similar to sequences obtained from the mouse gastrointestinal microflora (R. E. Ley et al., 2005). The closest named organisms belong to the Genus *Anaeroplasma* (highlighted in blue).

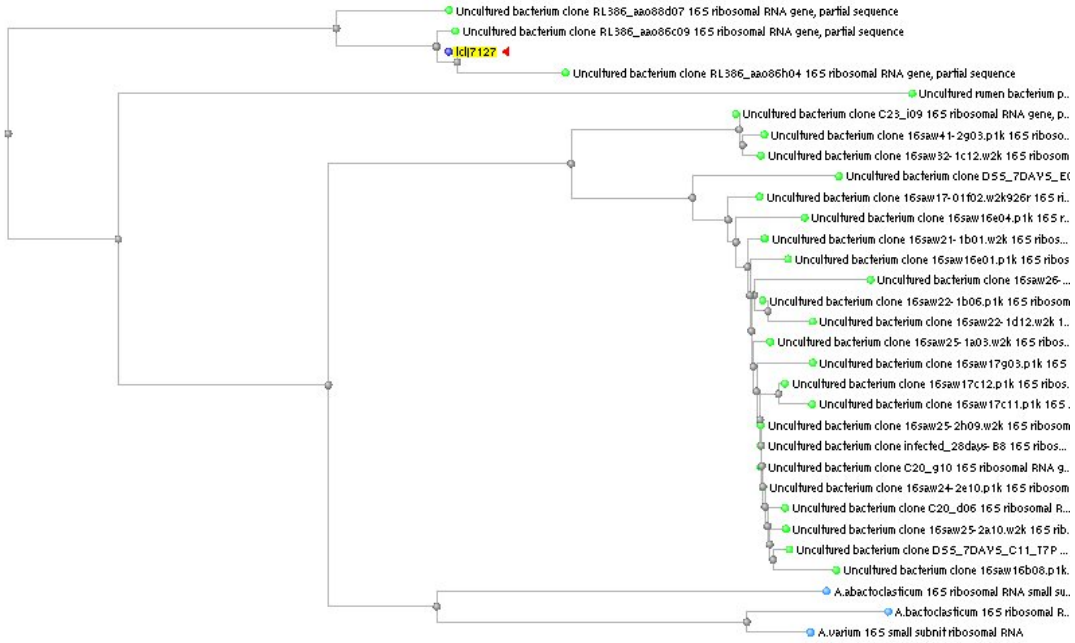


Figure 1.2-B: Neighborhood joining, phylogenetic sub-tree constructed from BLAST 'hits' with the submitted DM1 'consensus' sequence. As this sequence is virtually identical to the 'consensus' sequence from (Kubs & Musgrave, 2007) report, due to the novel nature of the DM1 'consensus' sequence it is of no surprise that this tree places the it in its own branch

Figure 1.2-C: Neighbourhood joining, phylogenetic tree constructed from BLAST 'hits', which had a maximum sequence difference of 0.05 with the submitted DM1 'consensus' sequence (Iclj30579, highlighted in yellow). This tree shows the DM1 'consensus' sequence having high homology with uncultured bacterial clones isolated from the human GI tract.

These results are interesting as these 'hits' come from the aforementioned studies that suggest that obesity is associated with the functionality of the gut microflora (see 1.1.4.1). The ratio of *Firmicutes* to *Bacteriodetes* has being found to correlate to the efficiency in which energy is harvested from ingested food and are converted to fat, with a higher proportion of *Firmicutes* increasing this efficiency (Backhed et al., 2005; R. E. Ley et al., 2005; R. E. Ley et al., 2006). As the novel anaerobic *Mollicute*-like species identified by (Kubs & Musgrave, 2007) belongs to the bacterial division of the *Firmicutes*, and were present in all of the tissue samples and fecal samples tested, it is not only possible that they constitute an important component of the microbial flora of the human gastrointestinal tract, but could also have functionality associated with obesity.

1.3 Mollicutes

A noteworthy observation is that *Mollicutes* or '*Mycoplasmas*' are generally pathogenic in nature, it is therefore interesting that these preliminary studies have revealed the presence of these organisms with the human gastrointestinal tract and suggest that they are there in large quantities (Kubs & Musgrave, 2007). Below a summary is given concerning the characteristics of the *Mollicutes* and the methods employed in identifying them.

1.3.1 Overview

The smallest known self-replicating cellular organisms, distinguished from other bacteria by their lack of cell walls, the Genus of *Mycoplasma* (*myces*, a fungus; *plasma*, a form, in Latin), has had its title, by and large, accepted as the epithet for all the microorganisms belonging to its own bacterial class: the *Mollicutes (mollis,* soft*; cutis,* skin*,* in Latin). Hence this 'trivial' term is interchanged with '*Mollicutes*' throughout the literature in which it is present (S. Razin et al., 1998).

Prevalent throughout the natural world, *Mollicutes* have been established as parasitic or commensal organisms residing within plants, arthropods, fish, reptiles, mammals

and humans (S. Razin, 1992). Attributable to this noted parasitic nature it is of no surprise that many have being implicated or identified as the causative pathogenic agents of numerous diseases. *Mollicutes* are considered to be both host and tissue specific, this fastidious nature is mirrored in the difficulties arising during isolation and culturing procedures.

Exhibited by their requirements of complex media in order to grow, *Mollicutes* are dependant on their hosts for the provision of essential nutrients. This dependency is directly related to the absence of genes necessary for functional and structural biochemical pathways and as a result *Mollicutes* possess limited metabolic abilities and small genomes. Their known genomes range in size from 577Kbp (*Mycoplasma genitalium*) to 2220Kbp (*Spiroplasma ixodetis*) and it would seem as a general rule that phylogenetically 'earlier' *Mollicutes* have larger genomes than those that come 'later', this is compatible with the view that *Mollicutes* have undergone 'reductive evolution' (Bove, 1993; S. Razin et al., 1998; Woese, 1987).

Mollicutes are suggested to have regressively evolved (i.e. reduction in genome size) from the low G+C, Gram positive, *Clostridium-Streptococcus-Lactobacillus* phylogenetic branch of the eubacteria (Pettersson, Tully, Bolske, & Johansson, 2001). Akin to their ancestors, they possess circular DNA genomes with low G+C content, which usually ranges from 24-33 mol% G+C (S. Razin et al., 1998). Over time, phenotypes displaying decreased numbers of rRNA operons and tRNA genes, the aforementioned lack of a cell wall, size, fastidious growth, and limited metabolic activities began to arise and are considered results of reductive evolution (Bove, 1993). Owing to this, *Mollicutes* have become of great interest with respect to the 'minimal cell' concept (Maniloff, 1996), of which 'the goal has been to prove the dogma of the completeness of molecular biology, that is, that the logic of life is finite, relatively simple and subject to full exploration whereby the minimal amount of essential genes required for life can be determined' (S. Razin, 1997).

Consistent with their low G+C and therefore high A+T biased genomic evolution; synonymous codons containing A+T have being favoured in the 'later' *Mollicutes*.

Such A+T selective pressures are seen in the reassignment of the UGA 'stop' codon to a 'tryptophan' codon (traditionally UGG) (S. Razin et al., 1998).

The number of required genes are further decreased through the lack of cell walls and the absence of intracytoplasmic membranes, which coincidently leaves *Mollicutes* having only one type of membrane, the plasma membrane (Rottem & Kahane, 1993). Located on the cell surface, lipoproteins constitute over two-thirds of the plasma membrane, a proportion that is considerably higher that those seen in other bacteria. Being antigenic in nature and found to be highly variable, it is of no revelation that these lipoproteins play an integral role in colonising and maintaining the host-*Mollicute* relationship (Sirand-Pugnet et al., 2007). The other third of the plasma membrane is composed of lipids, but despite this, most *Mollicutes* are partially or totally incapable of carrying out the necessary fatty acid synthesis. This has resulted in the reliance upon their host organisms for the supply of required lipids and cholesterol (a unique component of some *Mollicutes*) in order to facilitate their own growth (S. Razin et al., 1998).

The dependency on other organisms in-situ by *Mollicutes* are typified in-vitro by the complex media that are called upon when utilising culture based techniques for *Mollicute* identification. Fortunately, it would seem that such time-consuming burdens have been relieved somewhat by the application of molecular techniques. But despite this progress, it is generally accepted, that only the 'tip of the iceberg' has being revealed in respect to the number of *Mollicutes* identified and the actual number present in the environment.

Table 1.3-A: Properties Distinguishing Mollicutes from Other Bacteria (adapted from (Shmuel Razin, 1995)).

Properties Distinguishing Mollicutes from Other Bacteria		
Property	**Mollicutes**	**Other Bacteria**
Cell wall	Absent	Present
Plasma membrane	Cholesterol present in most species	Cholesterol absent
Genome size	577-2220 Kbps	1450- >6000Kbp
G + C content of genome	23-41 mol%	25-75 mol%
No. of rRNA operons	1-3	1-10
5S rRNA length	104-113 nucleotides	>114 nucleotides
No. of tRNA genes	30 (*M. capricolum*) 33 (*M. pneumoniae*)	51 (*B. subtilis*) 78 (*E. coli*)
UGA codon usage	Tryptophan codon in *Mycoplasma,Ureaplasma, Spiroplasma,* and *Mesoplasma*	Stop codon
RNA polymerase	Resistant to rifampicin	Rifampicin sensitive

1.3.2 Taxonomy and Phylogeny

'Prior to the era of rapid and inexpensive sequencing, the outlook for developing a comprehensive and usable taxonomy of the prokaryotes was bleak' (Lilburn & Garrity, 2004). Although numerous *Mollicute* species have had their entire genomes sequenced, we are some distance from being able to form what could be considered, highly accurate taxonomic and phylogenetic descriptions of this Class, or even the Domain of Bacteria itself. Taking this into account, most *Mollicute* taxonomy has had to be constructed using a combination of phenotypic characteristics and phylogenetic information from fractional genomic sequences, primarily those from ribosomal genes (S. Razin et al., 1998).

1.3.2.1 Phylogenetic Markers

Their presence and consistent functionality throughout the three Domains of life (*Bacteria, Archaea* and *Eucarya* makes the complex and modular protein

synthesising machinery, ideal candidates for phylogenetic analysis. The prokaryotic ribosomal RNA's (5S, 16S and 23S) contain highly conserved regions of sequence, for this reason, comparing these sequences by alignment and determining the degree of similarity between them, enable evolutionary relationships to be established. Higher similarities are indicative of closer evolutionary relationships and vice versa. Variable regions within these large ribosomal gene sequences are employed as evolutionary chronometers, measuring the evolutionary distance between two organisms (Madigan, Martinko, & Parker, 2003).

16S rRNA is the most widely used of the ribosomal phylogenetic markers; 16S rDNA sequence analysis of the prokaryotes has shown that the *Mollicutes* are taxonomically positioned within the bacterial Division of the *Firmicutes.* It has also revealed that the *Mollicutes'* class has revealed over 200 species placed into the following Genera: *Mycoplasma, Eperythrozoon, Haemobrtonella, Ureaplasma, Entomoplasma, Mesoplasma, Spiroplasma, Acholeplasma, Phytoplasma Anaeroplasma* and *Asteroleplasma*; all of which belong to the Orders: *Mycoplasmatales, Entomoplasmatales, Acholeplasmatales*, and *Anaeroplasmatales* (Garrity, Bell, & Lilburn, 2004). The presence of species of uncertain taxonomic positioning has eventuated in the creation of the additional Order: *Incertae sedisi* which contains the proposed Genera: *Erysiperothrix, Bulleidia, Holdemania* and *Solobacteria* (Garrity et al., 2004) (see Table 1.2-A.).

16S rDNA sequence data is a prerequisite when describing a new species for the '*Mollicutes* Taxonomy Committee', naturally, new insights and developments in molecular techniques has invariably led to the preference of complementary data accompanying that of the 16S rDNA sequence (S. Razin et al., 1998). Investigated complementary rDNA data, such as sequencing and targeting of 5S (Woese, Maniloff, & Zablen, 1980), tRNA (Stakenborg et al., 2005) and the 16S-23S rRNA intergenic transcribed spacer (ITS) regions (Volokhov et al., 2006; H. Wang, Kong, Jelfs, James, & Gilbert, 2004) have proved to be practical in the detection and characterisation of many *Mollicutes* (Volokhov et al., 2006). With access to these potential targets, methods such as microarray and reverse line blot hybridisations and

differing polymerase chain reaction (PCR) analytical techniques, have successfully established them as useable distinguishing phylogenetic markers.

Other complementary markers that are non-rDNA in origin that have been applied to date include the elongation factor EF-Tu (*tuf*) gene and the heat shock protein (*hsp70*) gene, not unlike the previously discussed rDNA genes, these genes display regions of conserved sequence and hence are suitable candidates for constructing phylogenetic relationships (S. Razin et al., 1998).

1.3.2.2 Phenotypic Markers

As emphasised earlier, isolation of *Mollicutes* is difficult, time-consuming and labour intensive, if possible at all. In spite of these major limiting factors, there are a number of phenotypic markers present that aid in the classification of isolated members of the *Mollicute* class. Some markers are taken from a proteomic approach, i.e. study of functional proteins (i.e. enzymes), which ultimately influence the cell's nutritional requirements (e.g. cholesterol, arginine, urea) and the eventuating metabolic waste products. The variable antigenic membrane surface of the *Mollicutes* provides a unique structure for serological based identification, the membrane also harbors polar lipid and fatty acids which have their own profiles (S. Razin et al., 1998; Worliczek, Kampfer, Rosengarten, Tindall, & Busse, 2007).

1.4 Hypothesis and Aims

In light of the acquired data from the preliminary study carried out by (Kubs & Musgrave, 2007), we hypothesise that these novel *Firmicutes* are members of a new *Mollicute* Genus related to *Anaeroplasma*, are widely distributed throughout the human G.I tract and are a significant component of the total microbiota.

Our aim is to take a molecular biological approach and use the newly deposited human sequences (R. E. Ley et al., 2005; R. E. Ley et al., 2006) and others that also relate to the DM1 consensus sequence to design and create primers that would amplify and characterise full-length versions of these *Mollicute*-like sequences from our human gastrointestinal samples.

With the increasing importance of finding ways to combat obesity in the western world, this research could be seen as an opportunity to further understand the microflora component of this diverse syndrome. Furthermore, research into the roles of *Mollicute* species in both the micro-organism G.I population dynamics and human health could be gained, assisting ultimately in finding medical solutions to the disease and ill health that they are responsible for, or suspected of causing in humans.

Chapter 2: Expanding the 109F/927R Amplicon

2.1 Introduction

A number of methods that are common place in molecular biology, such as sequencing and PCR require 'primers' in order to function. Primers are generally short, chemically synthesized single stranded oligonucleotides, with an approximate length of twenty bases. They are designed to hybridise to a region of target DNA, which is then copied by DNA polymerase. Our aim was to expand the length of the ~830bp 'known' 16S rRNA gene sequences (using the DM1 Consensus sequence as a representative) determined in preliminary studies using the primers: 109F/927R, primers were designed and tested with this objective in mind.

2.2 Materials and Methods

2.2.1 Primer Design

2.2.1.1 Ribosomal Database Project 10

Relevant and related sequences were obtained from the Ribosomal Database Project 10 (RDP) website (http://rdp.cme.msu.edu/). Multiple 'out-group' sequences were also selected, these 'out-groups' were not related to the relevant sequences and were used to assist in determining unique regions of sequence belonging to all of the relevant sequences. After these regions were determined the 'out-groups' were removed to increase the ease with which the alignment could be read and the primers designed.

2.2.1.2 ClustalW2

Sequences obtained from the RDP website were collectively submitted into the ClustalW2 sequence alignment program located on the European Bioinformatics

Institute (EBI) website (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The alignment result was printed and manually checked, regions of unique sequence present in the relevant sequences and absent in 'out-group' sequences were subsequently highlighted and considered as potential primer designing sites.

2.2.1.3 Forward Primer Designs.

Potential forward primers ran 5'-3' within the target sequences. Potential forward primers design sites were narrowed down and selected with respect to: their position in the sequences (5' end targeted), the number of unique nucleotides present within close proximity of one another (primers are generally ≤20bp) and the number of unique nucleotides present at the 3' end of the potential primer region, to promote primer specificity and stability. All potential forward primers were tested for viability using Primer3.

2.2.1.4 Reverse Primer Designs

Potential reverse primers were reverse and complementary to their potential binding sites within the target sequence. Potential reverse primer designing sites were narrowed down and selected with respect to: their position in the sequences (3' end targeted), the number of unique nucleotides present within close proximity of one another (primers are generally ≤20bp) and the number of unique nucleotides present at the 5' end of the potential primer target region, to promote primer specificity and stability. All potential reverse primers were tested for viability using Primer3.

2.2.2 Theoretical Testing of Designed Primers

2.2.2.1 Primer3

The viability and the compatibility of potential primers with existing or other potential primers was determined using the Primer3 Input 0.4.0 program located at http://frodo.wi.mit.edu/. This program required the input of a relevant target

sequence; primer/s were then submitted, and analysed with the target sequence. If the primer/s were viable they were ordered from Sigma-Aldrich® (http://www.sigmaaldrich.com/life-science/custom-oligos/custom-dna.html) for use in subsequent PCR reactions.

2.2.2.2 BLAST

In order to deduce whether primers would be able to target the desired 16S rRNA sequences, they were submitted to the BLAST tool located on the National Center for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The parameters for BLAST were: Nucleotide BLAST; Non-redundant Database (Nucleotide Collection (nr/nt); Megablast (optimised for highly similar sequences).

2.2.2.3 Classifier

In order to deduce whether primers would be able to target the desired 16S rRNA sequences, they were submitted to the Classifier tool located on the Ribosomal Database Project (RDP) website (http://rdp.cme.msu.edu/classifier/classifier.jsp). The classification algorithm that aligned 16S rRNA sequences was designed by (Q. Wang, Garrity, Tiedje, & Cole, 2007)

2.2.3 Practical Testing of Designed Primers

2.2.3.1 DNA Extraction and Isolation

2.2.3.1.A Samples

Longitudinal samples representing a transect of the human gastrointestinal tract, were obtained from patients undergoing surgery of such nature at the Hopital Europeen Georges Pompidou, Paris, France, courtesy of Dr Benjamin

Wyplosz. These were later processed by Dr David Musgrave either at The Institut Pasteur Paris, France or The University of Waikato, Hamilton, New Zealand.

2.2.3.1.B DNA Extraction and Isolation

Total DNA was extracted and isolated from samples for use in subsequent experiments using an adapted Benzyl alcohol-guanidine hydrochloride organic extraction method (Fredricks & Relman, 1998).

Cells were lysed on the surface of the sample by the addition and mixing of equal quantities of lysis buffer (1mL of buffer to each 1g of sample) (5.0M guanidine hydrochloride, 100mM Tris pH 8.0 in Milli-Q® (M.Q.), deionised water (>18 MΩ resistance)) with the sample. After 5mins shaking in a cold room (4°C), 0.4mL of MQ water was added to the tube, followed by 0.8mL of 99% Benzyl Alcohol, after which the tube was centrifuged at 7000g for 5mins. 0.4mLs of the resulting aqueous supernatant was removed by pipette and placed into a new tube, where the DNA was precipitated by the addition of 1/9 volume of 3.0M sodium acetate pH5.2 and an equal volume of isopropanol and incubated at room temperature for 5mins; the tube was then centrifuged at 16,000g for 15mins, the supernatant removed, and the DNA pellet air dried for 2h at room temperature. The pellet was re-suspended in 200µL of TE 10/0.1, the DNA pellet then had an equal amount of Phenol added to denature and remove any proteins that were present. The tube and its contents were mixed for 10mins, placed on ice for another 10mins and spun at 5000g for 2mins. If the aqueous phase was cloudy the phenol wash was repeated until clear.

The supernatant was transferred to a new tube with an equal volume of phenol/chloroform/isoamyl alcohol (P/C/IAA) solution (25:24:1); the phenol denatured any proteins; the chloroform denatured proteins and also stabilised the organic/aqueous boundary while the isoamyl alcohol assisted in separating

the two phases to decrease any foaming. After addition of the (P/C/IAA) solution mixing was carried out, the tube was then placed on ice and spun, and the aqueous phase removed and transferred to a new tube. The (P/C/IAA) treatment was then repeated.

Upon completion of the second (P/C/IAA) treatment the aqueous phase was treated solely with 200µL of chloroform, and again, mixed, placed on ice, spun, with the aqueous phase removed and transferred to a new tube. The DNA was again precipitated out of the solution by addition of 1/9 volume of 3.0M sodium acetate pH5.2 and an equal volume of isopropanol, incubated, spun and air dried prior to re-suspension in 0.1mL TE 10/0.1. All processed samples were labeled and stored at ≤ -20°C.

2.2.3.2 Polymerase Chain Reaction

The designed primers were tested practically using the Polymerase Chain Reaction (PCR). PCR is a fast, efficient and economically viable method employed in the amplification of specific regions of DNA that has become fundamental in molecular biology. It utilises the specificity of primers designed to select a target sequence and the heat-stable properties of some DNA polymerases (e.g. *Taq* polymerase, isolated from *Thermus aquaticus*) to replicate these sequences at the high temperatures required for PCR (Hartwell et al., 2004).

2.2.3.2.A Primers

Dry primer stocks were ordered and received from Sigma-Aldrich®. They were re-suspended in DNA-free Tris-EDTA (10/1) to have a final concentration of 1mM. This concentration was attained by the addition of an equivalent number of µL of TE (10/1) to the number of nanomoles received for each primer; these values were stated in the delivery's supplementary 'Oligonucleotide Data Sheet'. From these 1mM stocks, 100µM and 10µM solutions were made by re-suspension in DNA free TE (10/0.1) and water

respectively. All primers were stored at ≤ -20°C, with the 10μM solutions used in all subsequent PCR reactions, at final concentrations of 0.4μM.

2.2.3.2.B dNTPs

While placed on ice, 20μL was removed from 100mM stocks of each of the four dNTPs. These were then mixed carefully in an Eppendorf tube and followed by the addition of 120uL 10mM Tris buffer at pH8, giving a total volume of 200μL at a concentration of 10mM. 20μL aliquots were placed into PCR tubes, which were labeled and placed into a larger 15mL falcon tube, which was labeled and stored at ≤ -20°C for use in subsequent PCR reactions, at final concentrations of 40μM.

2.2.3.2.C DNA Free Water

Milli-Q®, deionised water (>18 MΩ resistance) was filtered through an Advantec Cellulose Acetate 0.20μM Syringe Filter Unit into sterile Petri dishes, which were then exposed to U.V light for 1hr, before undergoing a final filtration, before being aliquoted into labeled 1.5mL tubes stored at ≤ -20°C.

2.2.3.2.D DNA Free TE Solutions

Autoclaved solutions of Tris pH 8 (1.0M) and EDTA (0.5M) were mixed with Milli-Q®, deionised water (>18 MΩ resistance) to produce two solutions with final compositions of TE 10/1.0 and 10/0.1. These were then filtered, exposed to U.V light and filtered again before being aliquoted into labeled 1.5mL tubes and stored at ≤ -20°C as done with the DNA free water.

2.2.3.2.E 10X PCR Reaction Buffer

Invitrogen™ 10X PCR Reaction Buffer was stored at ≤ -20°C and used in all subsequent PCR reactions to give a final concentration of 1X.

2.2.3.2.F Magnesium Chloride

Invitrogen™ $MgCl_2$ 50mM was stored at ≤ -20°C and used in all subsequent PCR reactions at a final concentration of 2.5mM.

2.2.3.2.G Taq Polymerase

Invitrogen™ Platinum Taq DNA Polymerase 5U/µL was stored at ≤ -20°C and used in all subsequent PCR reactions at a final concentration of 0.5U/25µl.

100µL of PCR reaction mix (required for four 25µL reactions) contained:

10µL of 10X PCR reaction buffer	(FC 1X)
5µL of 50mM MgCl2	(FC 2.5mM)
0.4µL TAQ 5U/µl	(FC 0.5U/25µL)
4µL of each 10µM primer	(FC 0.4µM)
0.4µL 10mM dNTPs	(FC 40µM)

80.2 µL of sterile DNA free water to bring volume up to100ul

2.2.3.2.H PCR protocols

Except for gradient PCRs, all PCRs were run in the MJ Research Inc™ PTC-100™ Peltier Thermal Cycler using the custom designed protocol: DM55LONG.

DM55LONG:

95°C	3min	
95°C	30sec	} 30 Cycles
55°C	30sec	
72°C	1min	
72°C	7min	

Thermal gradient PCRs were run in the MJ Research Inc™ PTC-200™ Peltier Thermal Cycler using the custom designed protocol: DMGRAD.

DMGRAD:

95°C	3min	
95°C	30sec	} 30 Cycles
50-65°C	30sec	
72°C	1min 30sec	
72°C	7min	

2.2.3.3 Gel Electrophoresis

Gel electrophoresis is a method used to separate, purify and identify fragments of DNA depending upon their individual size, weight and chemical composition. Separation of these fragments is achieved with the application of an electrical current across a cross-linking, porous gel matrix. One cross linker, agarose, is capable of separating a broad range of DNA fragments, with lower concentrations (e.g. 0.1-0.2% w/v) creating larger pores in the gel and resolving larger fragments and vice-versa.

Separation of DNA fragments in this work was done so following the 'Agarose gel electrophoresis' method outlined in (Sambrook & Russell, 2001).

2.2.3.3.A Gel Preparation

All gels were cast using 'Ultrapure™' agarose powder.

Tris-borate-EDTA (TBE) gels

TBE gels were made using 10x Tris-borate-EDTA (TBE) pH8.0 buffer, 6.3mM ethidium bromide and double distilled water. Final composition of the TBE gels were 1.0% agarose, 1x TBE (90mM Tris–borate, 2mM EDTA) pH8.0 buffer and 0.03mM ethidium bromide.

Tris-acetate-EDTA (TAE) gels

TAE gels were made using: 50x TAE pH8.0 buffer, 6.3mM ethidium bromide and double distilled water. Final composition of the TAE gels were 0.9% agarose, 1x TAE (40mM Tris-acetate, 1mM EDTA) pH8.0 buffer and 0.03mM ethidium bromide.

2.2.3.3.B Sample Preparation

DNA that was to undergo electrophoresis from extractions, PCRs, precipitations, had 1.3x or 6x Gel-loading buffer concentrations (0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll in TE 10/1; method modified from (Sambrook & Russell, 2001), to replace H20 with TE 10/1) added to and mixed with, giving a final buffer concentration of 1x.

2.2.3.3.C Electrophoresis

Electrophoresis of gels was carried out using an Owl Lightning Volt (Model OSP-250L) power supply. Small gels were electrophoresed in an Owl Separation System (Model: B1A) at 100V (constant voltage) for 1hr, while larger gels were electrophoresed in a Gibco/BRL Gel Electrophoresis Apparatus (Model: Horizon 11.4) at 110V (constant voltage) for 1hr. All gel results were illuminated with ultraviolet light (U.V), light and photographed; photos were then stored digitally using Scion Image and printed for further analysis.

2.3 Results

The desired amplicons for the two primer sets 1AF/927R or 927R NEW and 109F or 109F NEW/1391R are pictured in Figure 2.3-A.

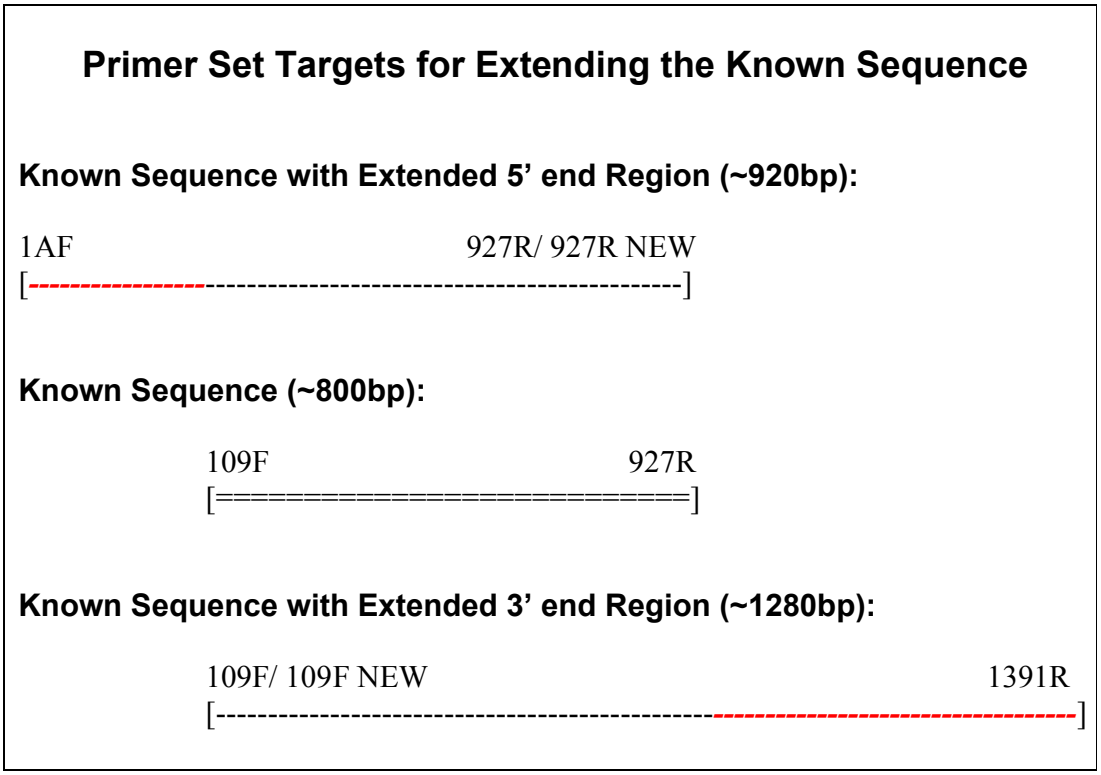


Figure 2.3-A: Depiction of primer set targets and their theoretical or known amplicons.

2.3.1 1 Anaeroplasma Forward (1AF), 927 Reverse (927R) and 927R NEW Primer Design

With the intention of extending the 5' end region of the known 16S sequence, 1AF was designed as a forward primer to target sequences which belonged to members of the bacterial genus: *Anaeroplasma*. This primer was used in conjunction with the specific 927R reverse primer, which was used in preliminary analysis to amplify the *Mollicute*-like 'known' sequence, and the modified version of 927R, 927R NEW which was designed to be more specific to sequences similar to the DM1 consensus sequence.

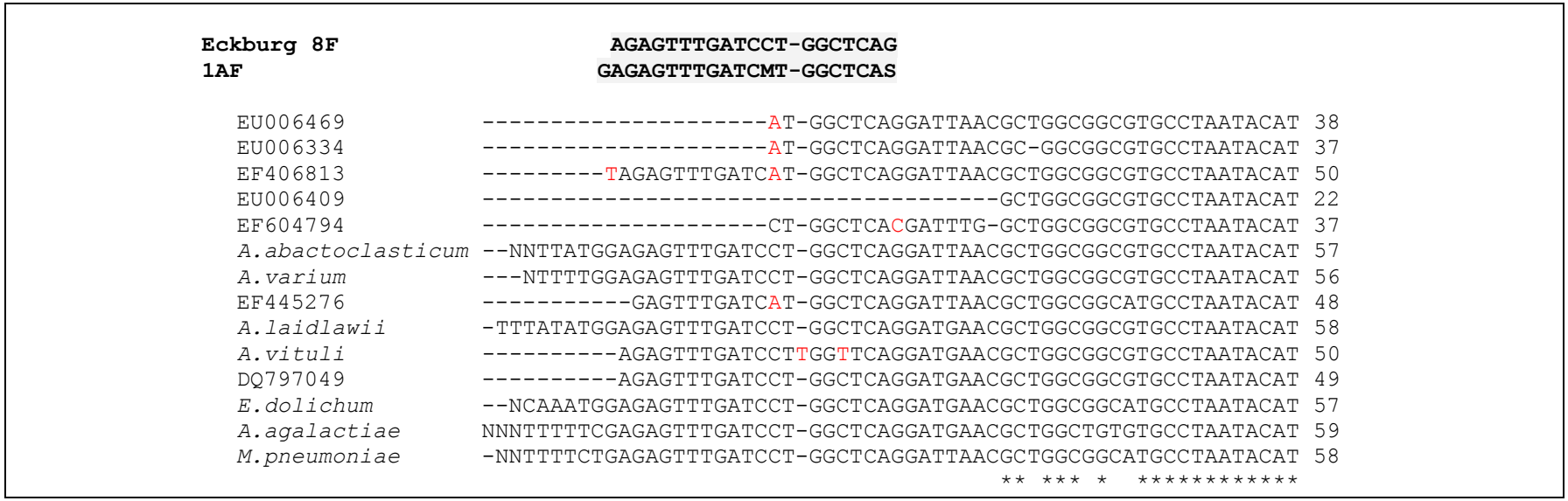
2.3.2 1391R Reverse, 109 Forward (109F) and 109F NEW Primer Design

With the intention of extending the 3' end region of the known 16S sequence, 1391R was a reverse primer designed for this purpose. This primer was used in conjunction with the specific 109F forward primer, which was used in preliminary analysis to amplify the *Mollicute*-like known sequence, and the modified version of 109F, 109F NEW, which was designed to be more specific to sequences similar to the DM1 consensus sequence.

2.3.3 Sequences Obtained from the Ribosomal Database Project 10 and NCBI Databases and Aligned with ClustalW2

Below Figures 2.3B-2.3E display ClustalW2 alignments of the 16S rRNA gene sequences belonging to species of the *Anaeroplasma* genus, these were downloaded from the RDP10 website. The other sequences in the alignment belong to uncultured species with similar sequences to that of the DM1 consensus sequence, all of which were found in the NCBI database. The sequence differences are highlighted in red and the relevant primers positioned above their target region within the alignment.

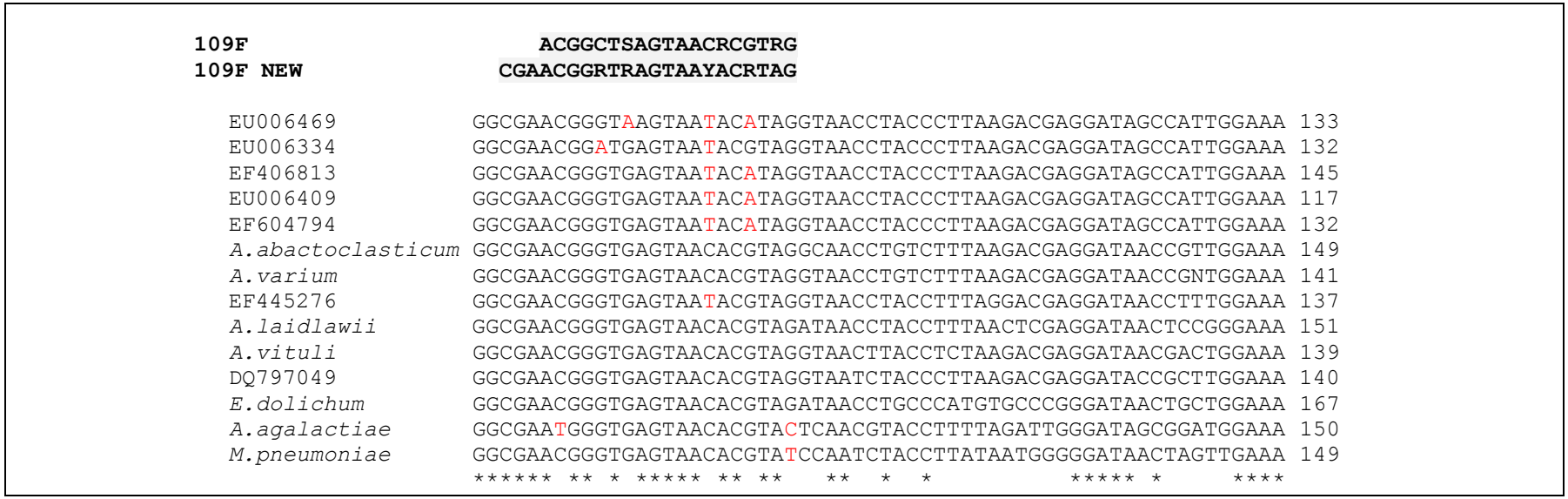
Figure 2.3-B: Region of ClustalW2 result used for the design of the forward primer 1AF.



```
Eckburg 8F                      AGAGTTTGATCCT-GGCTCAG
1AF                            GAGAGTTTGATCMT-GGCTCAS

EU006469          ----------------------AT-GGCTCAGGATTAACGCTGGCGGCGTGCCTAATACAT 38
EU006334          ----------------------AT-GGCTCAGGATTAACGC-GGCGGCGTGCCTAATACAT 37
EF406813          ---------TAGAGTTTGATCAT-GGCTCAGGATTAACGCTGGCGGCGTGCCTAATACAT 50
EU006409          ----------------------------------------GCTGGCGGCGTGCCTAATACAT 22
EF604794          ----------------------CT-GGCTCACGATTTG-GCTGGCGGCGTGCCTAATACAT 37
A.abactoclasticum --NNTTATGGAGAGTTTGATCCT-GGCTCAGGATTAACGCTGGCGGCGTGCCTAATACAT 57
A.varium          ---NTTTTGGAGAGTTTGATCCT-GGCTCAGGATTAACGCTGGCGGCGTGCCTAATACAT 56
EF445276          -----------GAGTTTGATCAT-GGCTCAGGATTAACGCTGGCGGCATGCCTAATACAT 48
A.laidlawii       -TTTATATGGAGAGTTTGATCCT-GGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT 58
A.vituli          ----------AGAGTTTGATCCTTGGTTCAGGATGAACGCTGGCGGCGTGCCTAATACAT 50
DQ797049          ----------AGAGTTTGATCCT-GGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT 49
E.dolichum        --NCAAATGGAGAGTTTGATCCT-GGCTCAGGATGAACGCTGGCGGCATGCCTAATACAT 57
A.agalactiae      NNNTTTTTCGAGAGTTTGATCCT-GGCTCAGGATGAACGCTGGCTGTGTGCCTAATACAT 59
M.pneumoniae      -NNTTTTCTGAGAGTTTGATCCT-GGCTCAGGATTAACGCTGGCGGCATGCCTAATACAT 58
                                                    ** *** *  ************
```

Figure 2.3-C: Region of ClustalW2 result used for the design of the forward primer 109F/109F NEW



```
109F                        ACGGCTSAGTAACRCGTRG
109F NEW                  CGAACGGRTRAGTAAYACRTAG

   EU006469               GGCGAACGGGTAAGTAATACATAGGTAACCTACCCTTAAGACGAGGATAGCCATTGGAAA 133
   EU006334               GGCGAACGGATGAGTAATACGTAGGTAACCTACCCTTAAGACGAGGATAGCCATTGGAAA 132
   EF406813               GGCGAACGGGTGAGTAATACATAGGTAACCTACCCTTAAGACGAGGATAGCCATTGGAAA 145
   EU006409               GGCGAACGGGTGAGTAATACATAGGTAACCTACCCTTAAGACGAGGATAGCCATTGGAAA 117
   EF604794               GGCGAACGGGTGAGTAATACATAGGTAACCTACCCTTAAGACGAGGATAGCCATTGGAAA 132
   A.abactoclasticum      GGCGAACGGGTGAGTAACACGTAGGCAACCTGTCTTTAAGACGAGGATAACCGTTGGAAA 149
   A.varium               GGCGAACGGGTGAGTAACACGTAGGTAACCTGTCTTTAAGACGAGGATAACCGNTGGAAA 141
   EF445276               GGCGAACGGGTGAGTAATACGTAGGTAACCTACCTTTAGGACGAGGATAACCTTTGGAAA 137
   A.laidlawii            GGCGAACGGGTGAGTAACACGTAGATAACCTACCTTTAACTCGAGGATAACTCCGGGAAA 151
   A.vituli               GGCGAACGGGTGAGTAACACGTAGGTAACTTACCTCTAAGACGAGGATAACGACTGGAAA 139
   DQ797049               GGCGAACGGGTGAGTAACACGTAGGTAATCTACCCTTAAGACGAGGATACCGCTTGGAAA 140
   E.dolichum             GGCGAACGGGTGAGTAACACGTAGATAACCTGCCCATGTGCCCGGGATAACTGCTGGAAA 167
   A.agalactiae           GGCGAATGGGTGAGTAACACGTACTCAACGTACCTTTTAGATTGGGATAGCGGATGGAAA 150
   M.pneumoniae           GGCGAACGGGTGAGTAACACGTATCCAATCTACCTTATAATGGGGGATAACTAGTTGAAA 149
                          ****** ** * ***** ** **   **  *  *          ***** *     ****
```

Figure 2.3-D: Region of ClustalW2 result used for the design of the Reverse Primers 927R/927.

```
927R Reverse Complement:                    --GAAACTTAAA-GGAATTGACGGG
927R NEW Reverse Complement:                -TMAAACTWAAA-GGAYRTGACGGG

  EU006469          CCTGAGTAGTACGTACGCAAGTATGNAACTAAAA-GGAATGACGGGGAGCCCGCACAAGC 886
  EU006334          CCTGAGTAGTTCGTCCGCAAGTATTAAACTTAAA-GGAACTGACGGGAGTTCGCACAAGC 885
  EF406813          CCTGAGTAGTACGTACGCAAGTATGAAACTTAAA-GGAATTGACGGGAGCCCGCACAAGC 899
  EU006409          CCTGAGTAGTACGTACGCAAGTATGAAACTTAAA-GGAATTGACGGGAGCCCGCACAAGC 871
  EF604794          CCTGAGTAGTACGTACGCAAGTATGAAACTTAAA-GGAATCGACGGGAGCCCGCACGAGC 886
  A.abactoclasticum CCTGAGTAGTACGTACGCAAGTATGAAACTTAAA-GGAATTGACGGGAGCCNGCACAAGC 909
  A.varium          CCTGAGTAGTACGTACGCAAGTATGAAACTTAAA-GGAATTGACGGGAGCCNGCACAAGC 897
  EF445276          CCTGAGTAATACGTACGCAAGTATGAAACTTAAA-GGAGTTGACGGGAGCCCGCACAAGC 889
  A.laidlawii       CCTGAGTAGTACGTACGCAAGTATGAAACTCAAA-GGAATTGACGGGACCCCGCACAAGC 906
  A.vituli          CCTGAGTAGTACGTACGCAAGTATGAAACTCAAA-GGAATTGACGGGACCCCGCACAAGC 918
  DQ797049          CCTGAGTAGTACGTACGCAAGTATGAAACTTAAA-GGAATTGACGGGATCCCGCACAAGC 894
  E.dolichum        CCTGGGGAGTATGCACGCAAGTGTGAAACTCAAA-GGAATTGACGGGNGCCNGCACAAGC 944
  A.agalactiae      CCTGAGTAGTACGTTCGCAAGAATAAAACTTAAA-GGAATTGACGGGGATNCGCACAAGC 929
  M.pneumoniae      CCTGGGTAGTACATTCGCAAGAATGAAACTCAAACGGAATTGACGGGGACCCGCACAAGT 936
                    **** * * *     ******  *  **** *** ***      ***     **** **
```

Figure 2.3-E: Region of ClustalW2 result used for the design of the Reverse Primer 1391R.

```
1391R:                                                GTTTGACGGGCGGTGTGTA
1391R Reverse Complement:                             TACACACCGCCCGTCAAAC

EU006469                TGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTTTG 1364
EU006334                TGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTTTG 1363
EF406813                TGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTTTG 1377
EU006409                TGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTTTG 1349
EF604794                TGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGACAAACCATGAGAGTTTG 1364
A.abactoclasticum       TGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTTTG 1387
A.varium                TGTCGCGGTGAATACGTTCCCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTTTG 1372
EF445276                TGCCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGAGAGTCTG 1364
A.laidlawii             TGTTGCGGTGAATACGTTCTCGGGGTTTGTACACACCGCCCGTCAAACCACGAAAGTGGG 1385
A.vituli                TGCTGCGGTGAATACGTTCTCGGGGTTTGTACACACCGCCCGTCAAACCACGAAAGTTGA 1396
DQ797049                TGCCGCGGTGAATACGTTCCCGGGTCTTGCACTCACCGCCCGTCAA-------------- 1357
E.dolichum              TGCTGCGGTGAATACGTTCTCGGGCNTTGTACACACCGCCCGTCAAACCATGGGAGTCAG 1422
A.agalactiae            CGCTACGGTGAATACGTTCTCGGGTCTTGTACACACCGCCCGTCAAACCNTGGGAGCTGG 1407
M.pneumoniae            TGTCGCGGTGAATACGTTCTCGGGTCTTGTACACACCGCCCGTCAAACTATGAAAGCTGG 1399
                          *   ************** ****  *** ** ********* ***
```

2.3.4 *Blast Results for Submitted Primers*

Figure 2.3-F: BLAST result for the 1AF primer.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
FJ210678.1	Pseudomonas aeruginosa strain Ls4 16S ribosomal RNA gene, partial seque	42.1	78.3	100%	0.016	100%
EU862567.1	Pseudomonas sp. ABc22 16S ribosomal RNA gene, partial sequence	42.1	82.3	100%	0.016	100%
EU862560.1	Pseudomonas sp. ABc2 16S ribosomal RNA gene, partial sequence	42.1	82.3	100%	0.016	100%
EU016674.1	Uncultured marine bacterium HF4000APKG3108 fosmid sequence	42.1	42.1	100%	0.016	100%
EU016564.1	Uncultured marine microorganism HF4000_005D21 fosmid sequence	42.1	42.1	100%	0.016	100%
EU795204.1	Uncultured bacterium HF0770_37D02 genomic sequence	42.1	42.1	100%	0.016	100%
EU795180.1	Uncultured bacterium HF0070_01M22 genomic sequence	42.1	42.1	100%	0.016	100%
EU795178.1	Uncultured bacterium HF0010_09O16 genomic sequence	42.1	42.1	100%	0.016	100%
EU795120.1	Uncultured Rickettsiales bacterium HF4000_[384]005D21 genomic sequence	42.1	42.1	100%	0.016	100%
EU686599.1	Uncultured bacterium AD347-E12 genomic sequence	42.1	42.1	100%	0.016	100%
EU686641.1	Uncultured bacterium KM4-1-D1 genomic sequence	42.1	42.1	100%	0.016	100%
EU734937.1	Uncultured bacterium clone 028C19 16S ribosomal RNA gene, partial seque	42.1	74.3	100%	0.016	100%
EU669566.1	Uncultured Selenomonas sp. clone OCN091 16S ribosomal RNA gene, partia	42.1	42.1	100%	0.016	100%
EU669565.1	Uncultured Selenomonas sp. clone OCH033 16S ribosomal RNA gene, partia	42.1	42.1	100%	0.016	100%
EU554440.1	Escherichia coli strain Ag14 16S ribosomal RNA gene, partial sequence	42.1	42.1	100%	0.016	100%
EU373384.1	Staphylococcus epidermidis strain HNR19 16S ribosomal RNA gene, partial s	42.1	42.1	100%	0.016	100%
EU373368.1	Staphylococcus epidermidis strain TPL02 16S ribosomal RNA gene, partial s	42.1	42.1	100%	0.016	100%
EU373364.1	Staphylococcus epidermidis strain HNL22 16S ribosomal RNA gene, partial s	42.1	42.1	100%	0.016	100%
EU410957.1	Candidatus Pelagibacter ubique clone fosmid 01-003783, complete sequenc	42.1	42.1	100%	0.016	100%
EF583689.2	Methylobacterium radiotolerans strain IMBG187 16S ribosomal RNA gene, p	42.1	42.1	100%	0.016	100%

Figure 2.3-G: BLAST result for 109F primer.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
EU713915.2	Uncultured archaeon clone 40H-0S-16 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.17	100%
FJ264522.1	Uncultured archaeon clone Mn3b-A83 16S ribosomal RNA gene, partial s	38.2	38.2	100%	0.17	100%
EU924222.1	Uncultured archaeon clone LHC4_L2_E12 16S ribosomal RNA gene, parti	38.2	38.2	100%	0.17	100%
AB370920.1	Firmicutes bacterium HF6 gene for 16S ribosomal RNA, partial sequence	38.2	38.2	100%	0.17	100%
EU635914.1	Uncultured archaeon clone SSE_L4_C01 16S ribosomal RNA gene, partia	38.2	38.2	100%	0.17	100%
EU616791.1	Uncultured eukaryote clone ALAS78 16S ribosomal RNA gene, partial se	38.2	38.2	100%	0.17	100%
EU616779.1	Uncultured eukaryote clone ALAS26 16S ribosomal RNA gene, partial se	38.2	38.2	100%	0.17	100%
EU559693.1	Uncultured archaeon clone B1-60 16S ribosomal RNA gene, partial sequ	38.2	38.2	100%	0.17	100%
AB424702.1	Uncultured archaeon gene for 16S rRNA, partial sequence, clone: Fhm4A	38.2	38.2	100%	0.17	100%
AB424689.1	Uncultured archaeon gene for 16S rRNA, partial sequence, clone: Fhm2A	38.2	38.2	100%	0.17	100%
EU377247.1	Uncultured marine archaeon clone Au-Fg10-Arch47 16S ribosomal RNA g	38.2	38.2	100%	0.17	100%
EF203613.1	Uncultured archaeon clone 5A026 16S ribosomal RNA gene, partial sequ	38.2	38.2	100%	0.17	100%
EU284664.1	Uncultured archaeon clone IP1SAT-G9 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.17	100%
EU284663.1	Uncultured archaeon clone IP1SAT-G12 16S ribosomal RNA gene, partia	38.2	38.2	100%	0.17	100%
EU284650.1	Uncultured archaeon clone IP1SAT-B11 16S ribosomal RNA gene, partia	38.2	38.2	100%	0.17	100%
EU284648.1	Uncultured archaeon clone IP1SAT-A9 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.17	100%
EU284608.1	Uncultured archaeon clone COSAS-E4 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.17	100%
EU284604.1	Uncultured archaeon clone COSAS-C7 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.17	100%
EU280218.1	Uncultured archaeon clone HQSAT_10E9 16S ribosomal RNA gene, parti	38.2	38.2	100%	0.17	100%
EF591444.1	Uncultured crenarchaeote clone EJ1-33 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.17	100%

Figure 2.3-H: BLAST result for primer 927R

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
FJ445390.1	Jujube witches'-broom phytoplasma strain Jwb-Henan 16S ribosomal RN	44.1	44.1	100%	0.005	100%
FJ445389.1	Jujube witches'-broom phytoplasma strain WJwb-Henan 16S ribosomal R	44.1	44.1	100%	0.005	100%
FJ436792.1	Candidatus Phytoplasma ulmi 16S ribosomal RNA gene, partial sequence	44.1	44.1	100%	0.005	100%
FM878637.1	Mycoplasma sp. 38104 partial 16S rRNA gene, strain 38104	44.1	44.1	100%	0.005	100%
FJ429364.1	Rose phytoplasma Gorakhpur 16S ribosomal RNA gene, partial sequenc	44.1	44.1	100%	0.005	100%
FJ369857.1	Uncultured bacterium clone TS55_a01f02 16S ribosomal RNA gene, part	44.1	44.1	100%	0.005	100%
FJ366778.1	Uncultured bacterium clone TS27_a03g03 16S ribosomal RNA gene, part	44.1	44.1	100%	0.005	100%
FJ366559.1	Uncultured bacterium clone TS27_a01b08 16S ribosomal RNA gene, part	44.1	44.1	100%	0.005	100%
FJ366119.1	Uncultured bacterium clone TS25_a02h08 16S ribosomal RNA gene, part	44.1	44.1	100%	0.005	100%
FJ365009.1	Uncultured bacterium clone TS1_a01g07 16S ribosomal RNA gene, parti	44.1	44.1	100%	0.005	100%
EU910037.1	Uncultured eukaryote clone T6P2AeG01 18S ribosomal RNA gene, partia	44.1	44.1	100%	0.005	100%
EU266074.1	Brazilian Huanglongbing disease-associated phytoplasma 16S ribosomal	44.1	44.1	100%	0.005	100%
FM874799.1	Uncultured bacterium partial 16S rRNA gene, clone SA03F02	44.1	44.1	100%	0.005	100%
AM911412.1	Uncultured bacterium partial 16S rRNA, clone D11_CW02_full	44.1	44.1	100%	0.005	100%
FJ430017.1	Brevibacillus sp. C292 16S ribosomal RNA gene, partial sequence	44.1	44.1	100%	0.005	100%
FJ427297.1	Iranian potato phytoplasma strain Ch4 16S ribosomal RNA gene, partial	44.1	44.1	100%	0.005	100%
FJ427296.1	Candidatus Phytoplasma trifolii strain H3 16S ribosomal RNA gene, parti	44.1	44.1	100%	0.005	100%
FJ427295.1	Candidatus Phytoplasma trifolii strain A6 16S ribosomal RNA gene, parti	44.1	44.1	100%	0.005	100%
FJ409235.1	Candidatus Phytoplasma pini isolate TsJ 16S ribosomal RNA gene, partia	44.1	44.1	100%	0.005	100%
FJ409234.1	Candidatus Phytoplasma pini isolate URAN 16S ribosomal RNA gene, par	44.1	44.1	100%	0.005	100%

Figure 2.3-I: BLAST result for the primer 1391R

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
FJ263621.1	Paulownia witches'-broom phytoplasma strain PaWB-Beijing subgroup 16Sr1	38.2	38.2	100%	0.16	100%
FM200179.1	Candidatus phytoplasma fraxini partial 16S rRNA gene and partial 16S-23S	38.2	38.2	100%	0.16	100%
FM200178.1	Candidatus phytoplasma fraxini partial 16S rRNA gene and 16S-23S rRNA ir	38.2	38.2	100%	0.16	100%
FM214522.1	Uncultured bacterium partial 16S rRNA gene, clone Crozet_s_1041	38.2	38.2	100%	0.16	100%
FM214186.1	Uncultured bacterium partial 16S rRNA gene, clone Crozet_s_705	38.2	38.2	100%	0.16	100%
FM213774.1	Uncultured bacterium partial 16S rRNA gene, clone Crozet_s_293	38.2	38.2	100%	0.16	100%
EU184021.1	Candidatus Phytoplasma ulmi strain EYCZ1 16S ribosomal RNA gene, partia	38.2	38.2	100%	0.16	100%
EU244101.1	Uncultured bacterium clone sl1394 16S ribosomal RNA gene, partial sequen	38.2	38.2	100%	0.16	100%
EU244098.1	Uncultured bacterium clone sl1390 16S ribosomal RNA gene, partial sequen	38.2	38.2	100%	0.16	100%
EU220836.1	Spirobacillus cienkowskii 16S ribosomal RNA gene, partial sequence	38.2	38.2	100%	0.16	100%
EU005951.1	Uncultured marine bacterium clone KG_C11_100m75 16S ribosomal RNA ge	38.2	38.2	100%	0.16	100%
EU344590.1	Uncultured bacterium clone hoa45_76D10 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.16	100%
EU344550.1	Uncultured bacterium clone hoa45_36D05 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.16	100%
EU344547.1	Uncultured bacterium clone hoa46_30F04 16S ribosomal RNA gene, partial s	38.2	38.2	100%	0.16	100%
EU344523.1	Uncultured bacterium clone hoa45_87G11 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.16	100%
EU344491.1	Uncultured bacterium clone hoa33_75C10 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.16	100%
EU344472.1	Uncultured bacterium clone hoa32_50B07 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.16	100%
EU344443.1	Uncultured bacterium clone hoa31_91C12 16S ribosomal RNA gene, partial	38.2	38.2	100%	0.16	100%
EU344430.1	Uncultured bacterium clone hoa2_43h08 16S ribosomal RNA gene, partial se	38.2	38.2	100%	0.16	100%
EU344411.1	Uncultured bacterium clone hoa2_41a01 16S ribosomal RNA gene, partial se	38.2	38.2	100%	0.16	100%

2.3.5 Primer3 Results

The Figures shown below (Figures 2.3-J, 2.3-K) are the Primer3 results for the submitted primer sets: 1AF/927R and109F/1391R with the *Anaeroplasma abactoclasticum* 16S rRNA sequence used as a template. Any redundancy was removed from the primer with the appropriate nucleotide featured within the primer, as Primer3 did not recognize redundancies.

Figure 2.3-J: Primer3 result for the, 5' region extending primer set 1AF/927R.

Primer	Start	Length	tm (°C)	GC%	Comp.	3′ Comp.
1AF (LEFT)	8	21	59.03	52.38	7.00	5.00
927R (RIGHT)	896	22	59.04	40.91	5.00	0.00

```
SEQUENCE SIZE: 1453
INCLUDED REGION SIZE: 1453
PRODUCT SIZE: 889, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 1.00

    1 NNTTATGGAGAGTTTGATCCTGGCTCAGGATTAACGCTGGCGGCGTGCCTAATACATGCA
            >>>>>>>>>>>>>>>>>>>>
   61 AGTCGAATGGTAGTAGCAATACAAACCATGGCGAACGGGTGAGTAACACGTAGGCAACCT
  121 GTCTTTAAGACGAGGATAACCGTTGGAAACGACGGATAATACTGGATAGGACATCACAAA
  181 GGGCATCCTTAGATGTTTAAAGGTTATTATGCCACTTAGAGAGGGGCCTGCGGCGCATTA
  241 GCTAGTTGGTGAGATAACAGCCCACCAAGGCGATGATGCGTAGCCGGACTGAGAGGTTGA
  301 ACGGCCACAATGGAACTGAGAACGGTCCATACTCCTACGGGAGGCAGCAGTAGGGAATTT
  361 TCGGCAATGGGCGCAAGCCTGACCGAGCAACGCCGCGTGAACGAAGAAGTTATTCGTAAT
  421 GTAAAGTTCTTTTATCTGGGAAGAAAAAACAAATTGACTGTACCAGATGAATAAGCTCCG
  481 GCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGAGCNAGCGTTATCCGGAATTATT
  541 GGGCGTAAAGGGTGCGCAGATGGCACATTAAGTTTCTTGTAAAAATGCTGGGCTCAACCC
  601 AGTAGAGCAAGAGATACTGATGAGCTAGAGTACGAGAGAGGCAAGTGGAATTCCATGTGT
  661 AGCGGTAAAATGCGTAAATATATGGCAGAACACCAGTGGCGAAGGCGGCTTGCTAGCTCG
  721 ATACTGACATTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCC
  781 ACGCCCTAAACGATGAGTACTAAGTGCTGGGAGGAATCTCGGTGCTGAAGTTAACGCATT
  841 AAGTACTCCGCCTGAGTAGTACGTACGCAAGTATGAAACTTAAAGGAATTGACGGGAGCC
                                        <<<<<<<<<<<<<<<<<<<<
  901 NGCACAAGCGGTGGAGCATGTTGTTTAATTCGACGNTACGCGAAGAACCTTACCAGGTCT
  961 TGACATCCTCTGCGAAGCTATAGAGATATAGTGGAGGTTAGCAGAGAGACAGGTGGTGCA
 1021 TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
 1081 TATATCTAGTTACCATCATTCAGTTGGGGACTCTAGATAGACTGCCATTGATAAAATGGA
 1141 GGAAGGTGGGGATGACGTCAAATCATCATGCCCTATATGACCTGGGCTACAAACGTGCTA
 1201 CAATGGCTGAAACAAAGAGAAGCGAAGCGGTGACGTGGAGCTAACCTCATAAAAGCAGTC
 1261 TCAGTTCGGATTGAAGTCTGCAACTCGACTTCATGAAGTTGGAATCGCTAGTAATCGCGA
 1321 ATCAGAATGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGA
 1381 GAGTTTGTAATACCCGAAGCCGGTGGCCTAACCGCAAGGAAGGAGCCGTCTAAGGTAGGA
 1441 CAGATGATTGGGG

KEYS (in order of precedence):
>>>>>> left primer
<<<<<< right primer
```

Figure 2.3-K: Primer3 result for the, 3' region extending primer set 109F/1391R.

Primer	Start	Length	tm (°C)	GC%	Comp.	3′ Comp.
109F (LEFT)	95	19	53.52	52.63	7.00	2.00
1391R(RIGHT)	1375	19	62.93	57.89	3.00	2.00

```
SEQUENCE SIZE: 1453
INCLUDED REGION SIZE: 1453
PRODUCT SIZE: 1281, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 1.00

    1 NNTTATGGAGAGTTTGATCCTGGCTCAGGATTAACGCTGGCGGCGTGCCTAATACATGCA
   61 AGTCGAATGGTAGTAGCAATACAAACCATGGCGAACGGGTGAGTAACACGTAGGCAACCT
                                        >>>>>>>>>>>>>>>>>>>
  121 GTCTTTAAGACGAGGATAACCGTTGGAAACGACGGATAATACTGGATAGGACATCACAAA
  181 GGGCATCCTTAGATGTTTAAAGGTTATTATGCCACTTAGAGAGGGGCCTGCGGCGCATTA
  241 GCTAGTTGGTGAGATAACAGCCCACCAAGGCGATGATGCGTAGCCGGACTGAGAGGTTGA
  301 ACGGCCACAATGGAACTGAGAACGGTCCATACTCCTACGGGAGGCAGCAGTAGGGAATTT
  361 TCGGCAATGGGCGCAAGCCTGACCGAGCAACGCCGCGTGAACGAAGAAGTTATTCGTAAT
  421 GTAAAGTTCTTTTATCTGGGAAGAAAAAACAAATTGACTGTACCAGATGAATAAGCTCCG
  481 GCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGAGCNAGCGTTATCCGGAATTATT
  541 GGGCGTAAAGGGTGCGCAGATGGCACATTAAGTTTCTTGTAAAAATGCTGGGCTCAACCC
  601 AGTAGAGCAAGAGATACTGATGAGCTAGAGTACGAGAGAGGCAAGTGGAATTCCATGTGT
  661 AGCGGTAAAATGCGTAAATATATGGCAGAACACCAGTGGCGAAGGCGGCTTGCTAGCTCG
  721 ATACTGACATTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCC
  781 ACGCCCTAAACGATGAGTACTAAGTGCTGGGAGGAATCTCGGTGCTGAAGTTAACGCATT
  841 AAGTACTCCGCCTGAGTAGTACGTACGCAAGTATGAAACTTAAAGGAATTGACGGGAGCC
  901 NGCACAAGCGGTGGAGCATGTTGTTTAATTCGACGNTACGCGAAGAACCTTACCAGGTCT
  961 TGACATCCTCTGCGAAGCTATAGAGATATAGTGGAGGTTAGCAGAGAGACAGGTGGTGCA
 1021 TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
 1081 TATATCTAGTTACCATCATTCAGTTGGGGACTCTAGATAGACTGCCATTGATAAAATGGA
 1141 GGAAGGTGGGGATGACGTCAAATCATCATGCCCTATATGACCTGGGCTACAAACGTGCTA
 1201 CAATGGCTGAAACAAAGAGAAGCGAAGCGGTGACGTGGAGCTAACCTCATAAAAGCAGTC
 1261 TCAGTTCGGATTGAAGTCTGCAACTCGACTTCATGAAGTTGGAATCGCTAGTAATCGCGA
 1321 ATCAGAATGTCGCGGTGAATACGTTCTCGGGCTTTGTACACACCGCCCGTCAAACCATGA
                                          <<<<<<<<<<<<<<<<<<<
 1381 GAGTTTGTAATACCCGAAGCCGGTGGCCTAACCGCAAGGAAGGAGCCGTCTAAGGTAGGA
 1441 CAGATGATTGGGG

KEYS (in order of precedence):
>>>>>> left primer
<<<<<< right primer
```

2.3.6 PCR Results for Primer Sets

The Figure below (Figure 2.3-L) depicts the PCR results for all of the potential primer sets that were used in PCR reactions that contained template DNA from the human gastrointestinal mucosal surface sample DM20.

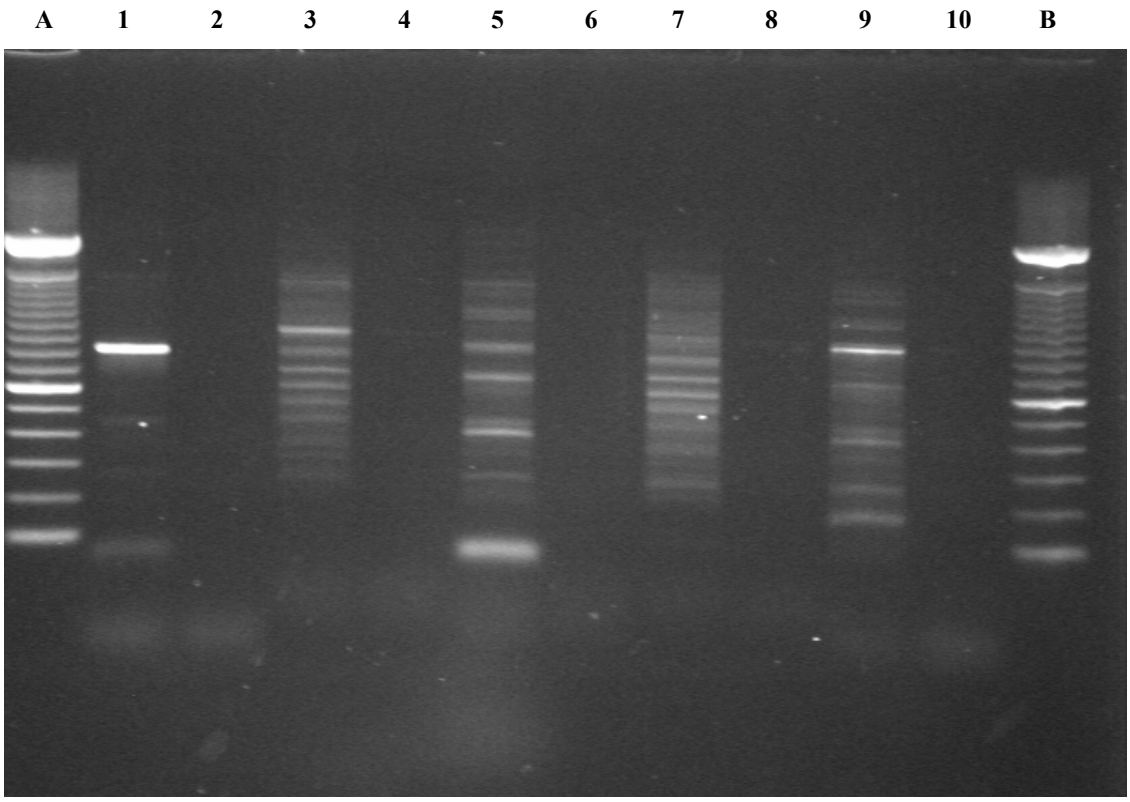


Figure 2.3-L: Small TBE gel depicting PCR results. A, 100kb ladder (bright band at 600bp); 1, Positive control: Mucosal surface sample DNA template DM20 (1:10 dilution) with the primer set 109F/927R; 2, Negative control: DNA free water with primer set 109F/927R; 3, 1AF/927R; 4, Negative control of 3; 5, 109F/1391R; 6, Negative control of 5; 7, 109F NEW/1391R NEW; 8, Negative control of 7; 9, 1AF/927R NEW; 10, Negative control of 9; B, 100kb ladder.

2.4 Discussion

2.4.1 Theoretical Assessment of Primers

16S rRNA sequences from organisms that were 'Best Hits' with DM1 in BLAST, along with sequences belonging to members of the Genus *Anaeroplasma* of which the 'Best Hit' Species *Anaeroplasma abactoclasticum* belongs, were aligned in ClustalW2. The ClustalW2 results (Figures: 2.3-B to 2.3-E) depict the intended primer target region within the submitted sequences, with any variations amongst them highlighted in red. Redundancy was applied to the primers (where necessary) to accommodate the aforementioned sequence variations, although, it can be observed that some of these were not taken into consideration, these were rational decisions, based on both the sequence integrity within and surrounding the target region, and the frequency of which any polymorphism appeared within the submitted sequences.

In order to establish whether the primers were potentially capable of targeting relevant 16S rRNA sequences, they were submitted to BLAST (Figures: 2.3-F to 2.3-I).

The 1AF primer sequence as shown in Figure 2.3-F, had 'Best Hits' with a 16S rRNA genes from a '*Pseudomonas aeruginosa* strain: Ls4' and other *Pseudomonas* species, it also had 'Hits' with both uncultured and well-known bacterium. These results indicated that 1AF was selective towards bacterial 16S rRNA sequences.

The 109F primer sequence as shown in Figure 2.3-G, had a 'Best Hit' with the 'Uncultured archaeon clone 40H-0S-16 16S ribosomal RNA gene, partial sequence' and 'Hits' with other uncultured archaeon 16S rRNA sequences. These results were not-surprising as the 109F primer was initially designed to amplify members of the *Crenarchaeota* from the human gut, yet as discussed in Chapter 1, the 109F/927R primer set amplified the *Anaeroplasma*-like 16S rRNA sequence that is the focus point of this thesis because of our inability to amplify crenarchaeal sequences from any of the human gut samples (Musgrave personal communication).

The 927R primer sequence as shown in Figure 2.3-H, had a 'Best Hit' with a 16S rRNA gene of a 'Jujube witches'-broom phytoplasma strain Jwb-Henan', it also had 'Hits' with a *Mycoplasma* species, and an uncultured bacterium sourced from human feces. As *Phytoplasmas* and *Mycoplasmas* belong to the *Mollicutes*, these results indicated 927R was selective towards *Mollicutes* as previous work had shown.

The 1391R primer sequence as shown in Figure 2.3-I, had a 'Best Hit' with a 16S rRNA gene of a 'Paulownia witches'-broom phytoplasma strain PaWB-Beijing subgroup', it also had 'Hits' with other Phytoplasmas and uncultured bacterium from environmental samples, including samples from the gut of the *Opisthocomus hoazin*, a leaf-eating bird found in South America (Godoy-Vitorino et al., 2008). These results displayed 1391R's selectivity towards not only bacteria, but also a bacterium found in the gut of other species, and also members of the *Mollicute* class.

The Primer3 results (Figures 2.3-J and 2.3-K) present the theoretical compatibility and viability of primer sets. Primer3 determines and takes into consideration a primer's: potential melting point (tm), Guanine/Cytosine content (GC %), self-complementarity's (Comp.) and 3' self-complementarity's (3' Comp.); and a primer set's: pair-complementarity's (Pair Any Compl.) and pair-3'complementarity's (Pair 3' Compl.).

The 1AF/927R primer set was intended to expand the 5' end region of the known sequence, Primer3 determined (Figure 2.3-J) that these primers were compatible with each other and the submitted target *Anaeroplasma abactoclasticum* sequence. Although the self-complementarity result for the 1AF primer (7.00) is relatively high, it was considered suitable on the grounds that the 8F primer which it was modeled from, is widely used as a universal bacterial primer.

The 109F/1391R primer set was intended to expand the 3' end region of the known sequence, Primer3 determined (Figure 2.3-K) that these primers were compatible with each other and the submitted target *Anaeroplasma abactoclasticum* sequence. Although the melting temperature (53.52°C) and self-complementarity (7.00) result for

the 109F primer is relatively high, it was considered suitable on the grounds that the 109F primer was used in preliminary work to amplify the amplicon of interest.

2.4.2 Practical Assessment of Primers

PCRs carried out with candidate primer sets (Figure 2.3-L) gave noteworthy results and were seen as promising steps towards extending the known amplicon at both its 5' and 3' ends.

1AF/927R: showed non specific amplification between ~300-1500bp, there was also a brighter band at ~950bp, this is promising as this is located where an amplicon containing the targeted known sequence with an extended 5' region would lie.

109F/1391R: showed a thick band at ~900kb, a position that the PCR amplicon for the human mucosal surface sample DM20 template and the primer set 109F/927R from preliminary work is expected to be located. This primer set also displayed non-specific bands between ~100bp and ~1500bp, bands seen at ~1200-1300bp were of interest, as this range is where the amplicon with an extended 3' region would be located within the gel.

109FNEW/1391R: showed a brighter band at ~900kb, a position that the PCR amplicon for the human mucosal surface sample DM20 template and the primer set 109F/927R from preliminary work is expected to be located. This primer set also displayed non-specific bands between ~200bp and ~1500bp; bands seen at ~1200-1300bp were of interest, as this range is where the amplicon with an extended 3' region would be located within the gel.

1AF/927R NEW: showed non specific amplification between ~200-1500bp, there was also a brighter band at ~950bp, this was promising as it is located where an amplicon containing the targeted known sequence with an extended front region would lie.

2.5 Conclusion

Merging the theoretical and practical results, it was deemed viable to determine the sequences of the non-specific PCR products and from there, relevant sequences would be determined, and it would be seen whether or not any of the sequences were extended versions of the known sequence.

It is also worth mentioning that much of the primer validation testing was hindered by contamination issues, these were eventually overcome by ordering new primers and refreshing all related reagents.

Chapter 3: Isolation and Sequencing of Amplicons

3.1 Introduction

Despite the absence of a clear-cut PCR product/s from which to work, due to the non-specific nature of the primers used, it was deemed viable to continue forward with the analysis of the potential products that were present, in an effort to extend the already known sequence found in preliminary studies. Potential products were selected, isolated, cloned, transformed and prepared for sequencing and analysis.

3.2 Materials and Methods

3.2.1 Isolating PCR products

PCR products had to be isolated before they could be used in cloning reactions, below is the approach taken.

3.2.1.1 Polymerase Chain Reaction

PCR reactions were carried out following the methodology outlined in 2.2.3.2.

3.2.1.2 Gel Electrophoresis

1μL of each PCR product had 11μL of 1X GLB added to it prior to being run on small, TBE, 1% agarose gels at 100V for 1hr, illuminated with UV light and photographed with SCION imager.

3.2.1.3 Bandstabs

Non-specific amplification can lead to the presence of multiple PCR products, in order to isolate and further amplify the PCR products of interest from these other products, band stabs were performed following the method developed by (Copper et al. 1992). PCR products were separated in gels, illuminated by UV light; the sought after band was then excised by stabbing an Eppendorf pipette tip perpendicularly into the gel, the excised PCR product was re-suspended in 40µL of DNA free water overnight at ≤ 4°C, 1µL of supernatant was then used as the template in PCR checks.

3.2.1.4 PEG/$MgCl_2$ PCR Product Precipitation

DNA was precipitated by the addition of the 25µL PCR product to 275µL of the 14.5% PEG/11mM $MgCl_2$ solution which then underwent mixing and incubation at room temperature prior to centrifugation at 12,000 rpm. The remaining PEG solution was removed from the precipitated DNA pellet by pipette. The pellet was then washed with 300µL, ice cold, 80% ethanol, followed by centrifugation at 12,000rpm for 5mins; removal of ethanol by pipette was followed by a further ethanol wash. After the final wash the DNA pellet was air dried and resuspended in 10uL TE 10/0.1 and stored until use in subsequent cloning protocols at ≤ -20°C.

3.2.2 Cloning PCR Products

PCR products were cloned using Topo TA Cloning®, ahead of being sequenced.

3.2.2.1 Topo TA Cloning®

Purified PCR products were inserted into a pCR® 2.1-TOPO plasmid vector (F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G) (Figure 3.2-A.) following the TOPO® Cloning Reaction outlined in the TOPO TA Cloning® Instruction Manual (Version R), Invitrogen ™ (Invitrogen, 2004).

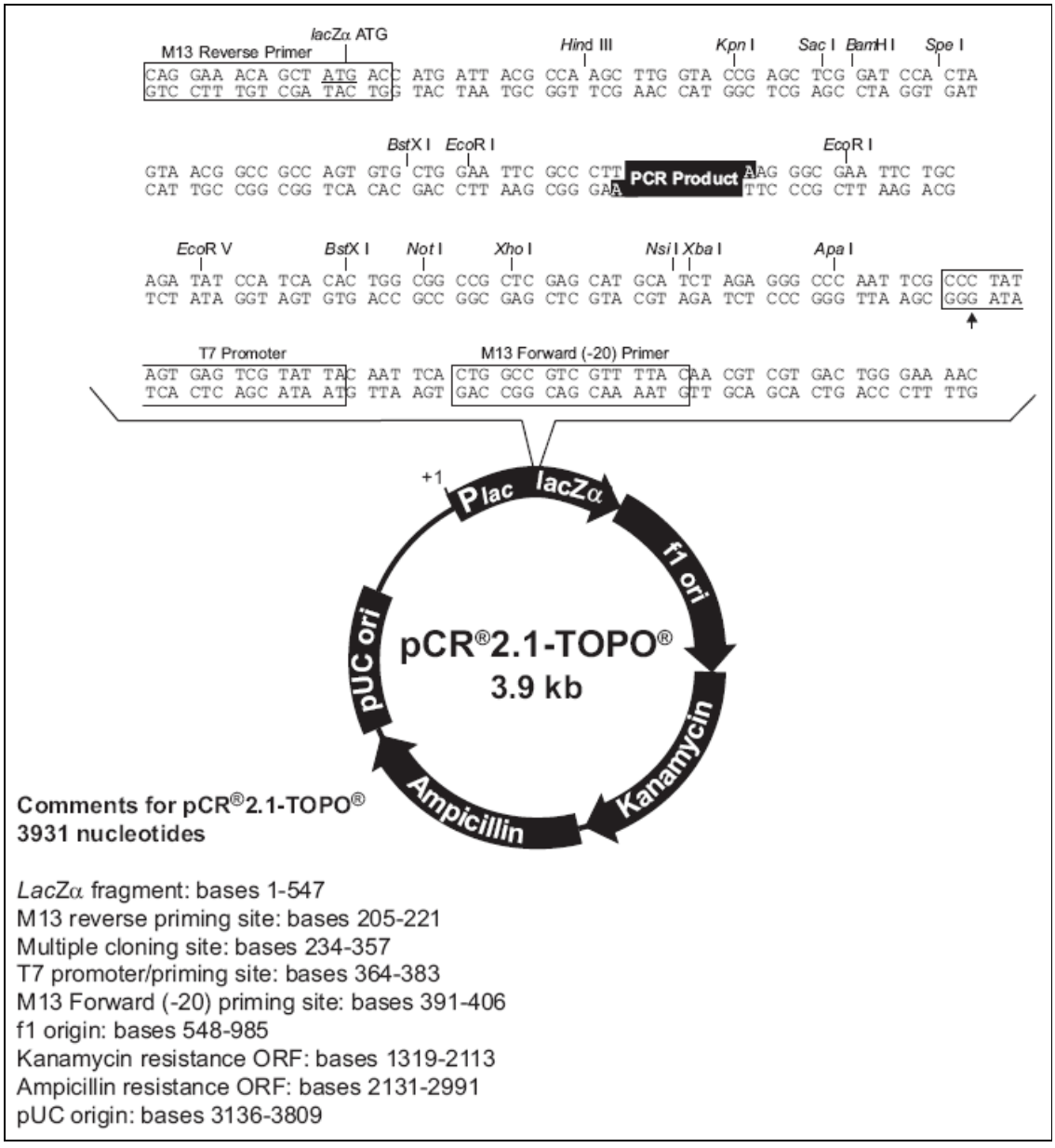


Figure 3.2-A: Map of pCR® 2.1-TOPO and the sequence surrounding the PCR product insertion site (Invitrogen, 2004)

3.2.2.2 Transformation of One Shot® TOP10 Competent Cells

The pCR® 2.1-TOPO plasmid was transformed into One Shot® TOP10 Competent Cells adhering to the method described in TOPO TA Cloning® Instruction Manual (Version R), Invitrogen ™ (Invitrogen, 2004). Transformants were spread and incubated overnight at 37°C on selective LB plates containing 50µg/mL kanamycin and 40µL of 40mg/mL X-gal in dimethylformamide. A stock solution of X-gal was made earlier using a method modified from (Sambrook & Russell, 2001). X-gal was dissolved in dimethylformamide to give a final concentration of 40mg/mL. The solution was stored in a glass tube, covered with aluminium foil, labelled and stored at ≤ -20°C.

3.2.2.3 Selecting Transformants

X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) is a galactoside substrate that is used to determine whether foreign DNA has been inserted into a plasmid vector. Recombinant plasmids will contain an inactivated *lacZ* gene which encodes the enzyme: β-galactosidase. This enzyme catabolises β-galactosides into galactose and 5-bromo-4-chloro-3-hydroxyindole, the former is used in essential metabolic pathways while the latter is oxidised to an insoluble blue product 5'-dibromo-4,4'-dichloro-indigo. Ergo, blue cell colonies are non-transformed $lacZ^+$, while white cell colonies are transformed $lacZ^-$ (see Figure 3.2-A). Here, 10 to 20 white colonies were selected at any one time and transferred to a reference LB/Kan plate and incubated overnight at 37°C, prior to screening.

3.2.3 Screening Clones for Plasmids Containing Relevant PCR Products

To determine whether or not a transformant's plasmid contains a PCR product of interest a 'screening method' was devised, this method is validated in Chapter 4. The method applied is presented, in full, below.

3.2.3.1 Direct Colony PCR

Direct Colony PCR is a quick and efficient screening method used to determine whether a colony contains a plasmid and an inserted PCR product, in this case clones were screened for PCR product inserts amplified by the primer combination: 109F/927R. The unpublished method required a colony to be placed into 200µL of DNA-free water, vortexed, spun down at 13K for 2mins. 1µl of the supernatant was added to a 25µL PCR reaction mix, containing the 109F/927R primers. The PCR was carried out using the protocol: DM55LONG. Products were then gel electrophoresed, illuminated by UV light and photographed; colonies that displayed a product ~900bp in size were streaked out for single colonies on LB/Kan plates.

3.2.4 Preparation of Plasmid DNA for Sequencing

Preceding sequencing, plasmids had to be isolated from cultures of each transformant and diluted to the correct concentration.

3.2.4.1 Plasmid DNA Preparation

The pCR® 2.1-TOPO plasmid (Invitrogen, 2004) was isolated from the selected transformants that were streaked out for single colonies. Universals containing 5mL of Terrific Broth containing Kanamycin at a final concentration of 50ug/mL and Glucose at a final concentration of 0.1%, were inoculated with cells from a single colony and incubated overnight at 37°C with vigorous aeration. The following day they were placed on ice for 10mins, 1.5mL quantities of culture were spun down sequentially two times into 1.5mL tubes and the cells finally resuspended in 200µL of cold GTE (50mM glucose, 25mM Tris HCl pH8.0, 10mM EDTA pH8.0). 0.2M Sodium Hydroxide/1% Sodium Dodecyl Sulfate was then added and mixed by inversion. 300µL 3M Potassium Acetate pH 5.2 was mixed in and a white clot formed. The tubes were placed on ice for 10mins before being spun down at 13K for

10mins. The supernatant was treated with 1µL of 20mg/ml RNAse (DNAse free) for 20mins in a 37°C water bath. 600µL Chloroform was mixed in by inversion to solubilize unwanted proteins, the tubes contents were then spun at 13K for 1min, this Chloroform step was repeated on the top aqueous phase. 600µL Isopropanol was then added to the top aqueous phase to precipitate DNA. Spun at 13K for 10mins, the Isopropanol was then removed by pipette and the DNA pellet washed with 500µL of ice cold Ethanol, this was spun at 13K for 2mins and the Ethanol removed by pipette. The pellet was air dried, re-suspended in 96µL of water. The DNA was once again precipitated by the addition and mixing of 24µL Sodium Chloride and 120 µL 13% PEG 8000, the tube was put on ice for 20mins then spun at 13K for 25mins at 4°C. The supernatant was removed and the DNA pellet washed with 400µL ice cold Ethanol. Once the Ethanol was removed by pipette and air drying, the pellet was resuspended in 20µL of TE10/0.1, and the tube stored at -20°C.

3.2.4.1.A Gel Electrophoresis Check

To evaluate the quality and establish an idea of the quantity of re-suspended plasmid DNA, solutions were run on large TAE, 0.9% agarose gels for 1hr, at 110V. These were illuminated by UV light and photographed using Scion imager.

3.2.4.1.B Determination of DNA concentration using a Nanodrop Spectrophotometer

To verify this information and give a numerical value to the concentration of plasmid DNA preparations, 2µL of re-suspended plasmid solutions were measured on a Nanodrop® ND-1000 Spectrophotometer, adhering to the protocol outlined in the Nanodrop V3.1.0 User's Manual (Nanodrop-Technologies, 2004).

3.2.4.1.C Dilution of Plasmid DNA

The Waikato DNA Sequencing Facility requires 12µL of plasmid DNA at a concentration of ~100ng/µL prior to sequencing. Dilution was achieved using the quantitative values (ng/µL) generated by the Nanodrop® ND-1000 Spectrophotometer (Nanodrop-Technologies, 2004).

3.3 Results

3.3.1 PCR Product Isolation

PCR products of interest were isolated prior to cloning by PCR amplification of the human mucosal surface sample DM20 using various primer sets. Bands of interest were excised from gels, by 'bandstab'. Excised bands were re-suspended before undergoing a further round of PCR, in order to check whether the correct band was removed and amplified, the remaining PCR mix was purified and precipitated using a PEG/$MgCl_2$ method with a small amount checked on a gel. The remaining concentrated precipitate was subsequently used in cloning, transformation and screened prior to sequencing.

3.3.1.1 5' End Extension

3.3.1.1.A 1AF/927R and 1AF/927R NEW Bandstabs

The figures shown below (Figures: 3.3-A and B) depict the results of the initial PCR carried out using the human mucosal surface sample DM20 template and the primer sets 1AF/927R and 1AF/927R NEW, and the succeeding 'bandstab' excisions of the relevant PCR products.

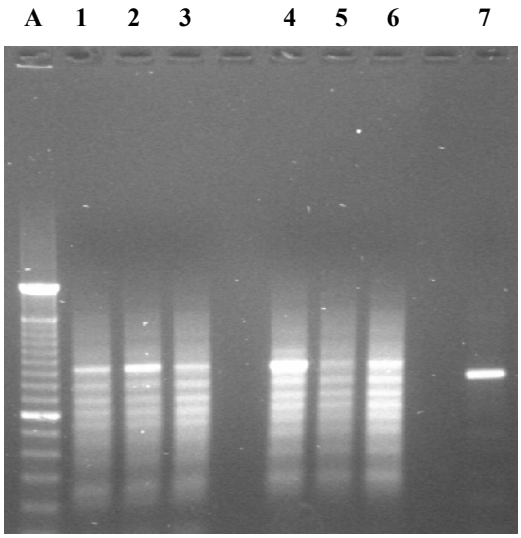


Figure 3.3-A: Small TBE gel displaying PCR products obtained using human mucosal surface sample DM20 and 1AF/927R or 927R NEW primer sets. A, 100bp ladder; 1-3, 1AF/927R; 4-6, 1AF/927R NEW; 7, Positive control.

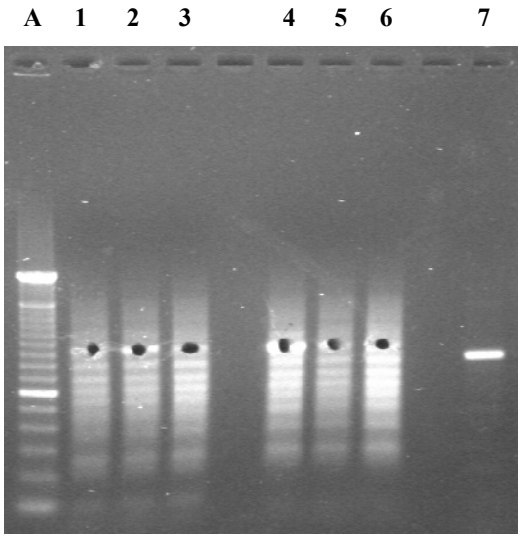


Figure 3.3-B: The gel from Figure 3.3-A after the excision of the PCR product of interest at ~900bp.

3.3.1.1.B Verification of 1AF/927R and 1AF/927R NEW Bandstabs and their PEG/$MgCl_2$ Precipitation

To verify whether the excision of the PCR products and their purity were successful a further round of PCR with the same primer sets from Figure 3.3-A was carried out using the supernatant of the re-suspended gel excision as a template. The results for each of the primer sets can be seen below in Figure 3.3-C.

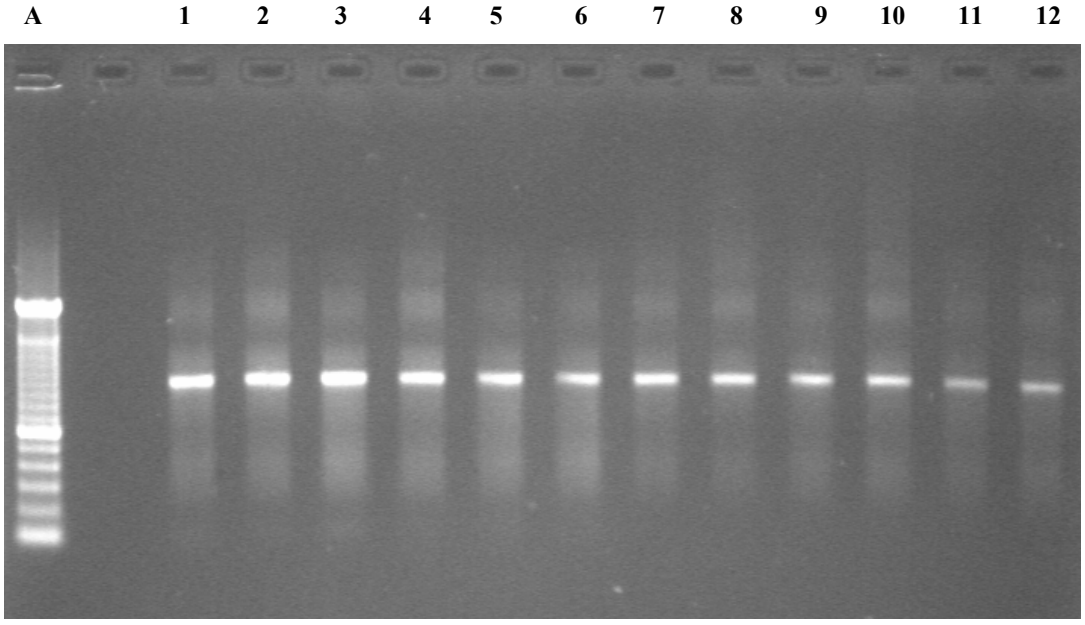


Figure 3.3-C: Large TAE gel depicting: 1-6, PCR amplicons produced for the six 'bandstabs' carried out in Figure 3.3-B; 7-12 PEG/MgCl2 PCR product precipitations of 1-6 respectively.

3.3.1.1.C 1AF/927R and 927R NEW Direct Colony PCR Screening

The 'screening method' outline in Chapter 4 was used here, an example is seen below in Figure 3.3-D. 1AF/927R transformants underwent PCR with the 109F/927R primer set (1AF/927R NEW transformants were amplified with the 109F/927R NEW primer set), to amplify sequences that contained either the known sequence or those of a similar nature. These methods were repeated for all applicable primer sets (1AF/927R or 927R NEW; 109F/1391R), building a small library of sequences to analyse.

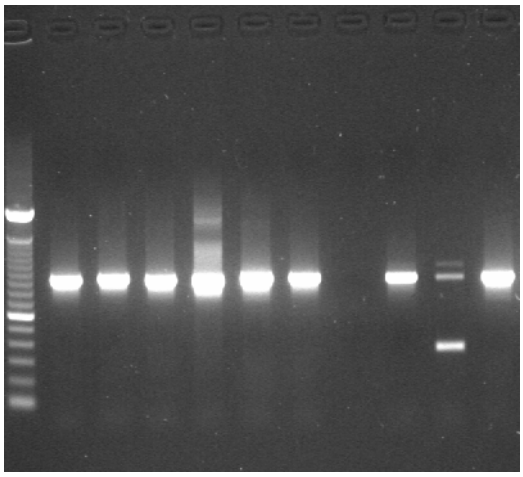


Figure 3.3-D: Small TAE gel depicting a direct colony PCR result of 1AF/927R transformants with the primer set 109F/927R. A, 100kb ladder; 1-10, 1AF/927R transformants amplified with the 109F/927R primer set.

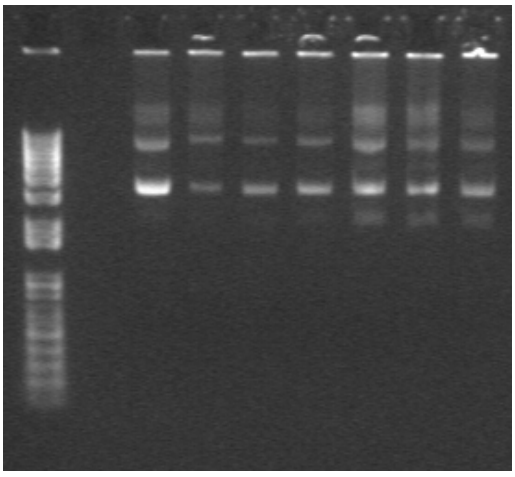


Figure 3.3-E: TAE gel displaying the plasmid isolation of the selected transformants from Figure 3.3-D. NB: The selected transformant 6 was lost during the precipitation procedure and is absent.

3.3.1.2 3' End Extension

3.3.1.2.A 109F and 109F NEW /1391R Bandstabs and Verification

Figure 3.3-F depicts the results of the initial PCR carried out using the human mucosal surface sample DM20 and the primer sets 109F/1391R and 109F NEW/1391R NEW. The labelled bands (A-F) in Figure 3.3-E were subsequently excised by 'bandstab'. The amplification of these 'bandstabs' were inconclusive for both of the aforementioned primer sets (Figure 3.3-G). It was this result that led to the precipitation of the entire original PCR result from Figure 3.3-F, with the objective being to clone a number of amplicons, ideally containing the 'known sequence'.

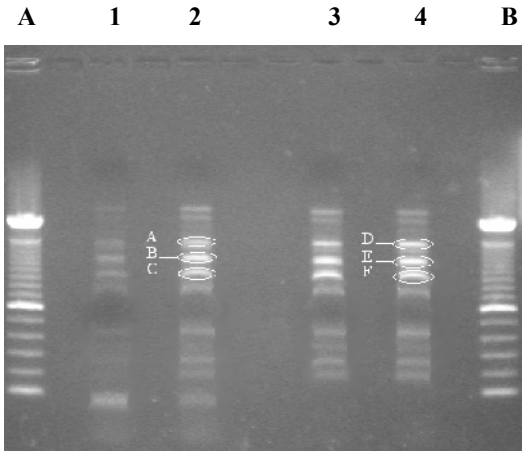


Figure 3.3-F: Small TBE gel displaying PCR amplicons obtained using the human mucosal surface sample DM20 and the 109F or 109F NEW/1391R primer sets. A, 100bp ladder; 1-2, 109F/1391R; 3-4, 109F/1391R NEW; B, 100kb ladder; A-F, bands excised by 'bandstab' for PCR.

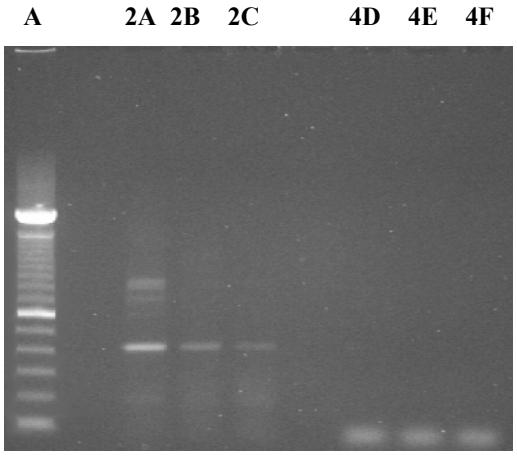


Figure 3.3-G Small TBE gel displaying PCR results of 'bandstabs' from Figure 3.3-E. A, 100bp ladder; 2A-2C, PCR results of excisions A-C; 4D-4F, PCR results of excisions D-F.

3.3.1.2.B 109F/1391R PCR Product Precipitation

Amplification of the excised bands (A-F) in Figure 3.3-F, proved to be unreliable (Figure 3.3-G, the entire raw PCR result from the 109F/1391R primer set was PEG/$MgCl_2$ precipitated for cloning. The results of the PEG/$MgCl_2$ precipitation is below in Figure 3.3-H.

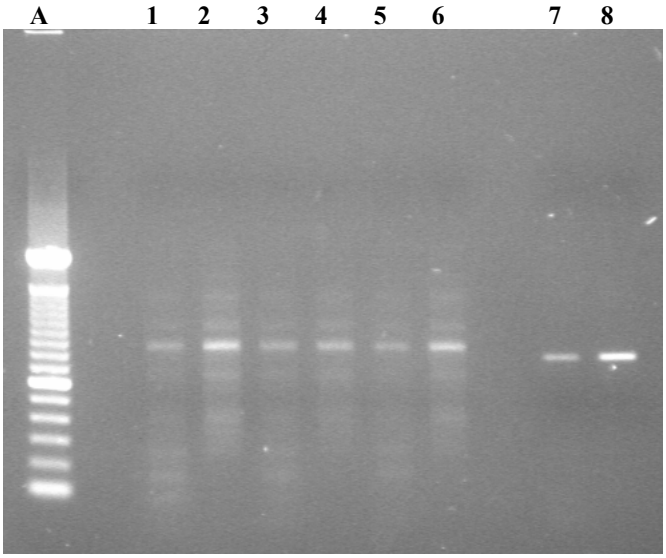


Figure 3.3-H: Small TAE gel depicting PCR and PEG/$MgCl_2$ PCR product precipitations : A, 100kb ladder; 1,3,5, 109F/1391R; 2,4,6 PEG/$MgCl_2$ PCR product precipitations of 1,3,5 respectively; 7, Positive control: DM20(1:10 dilution) with the primer set 109F/927R; 8, PEG/$MgCl_2$ PCR product precipitations of 7. All PCR products were obtained using human mucosal surface sample DM20.

3.3.1.2.C 109F/1391R Direct Colony PCR

109F/1391R transformants were subjected to the same 'screening' process as the 1AF/927R or 927R NEW transformants in 3.3.1.1. Direct colony PCR was carried out with the primers 109F/927R so that those that contained potentially relevant sequences (brighter bands) could be selected for plasmid precipitation and sequencing. The selected colonies were: 1, 4, 5, 6, 13 and 20.

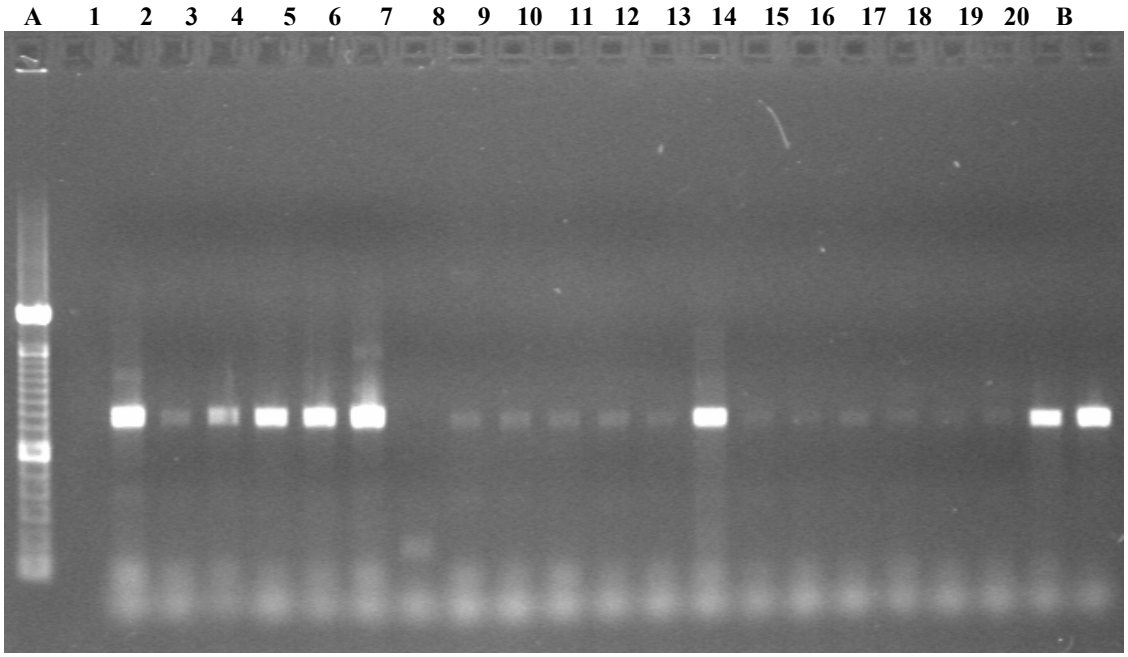


Figure 3.3-I: Large TAE gel displaying the direct colony PCR of 109F/1391R transformants with the 109F/927R primer set: A, 100kb ladder; 1-20, 109F/1391R transformants amplified with109F/927R; B, Positive control: DM20(1:10 dilution) with the primer set 109F/927R.

3.3.1.2.D 109F/1391R Plasmid Isolation Verification

To verify the quality and give an insight into the quantity of the plasmid present following precipitation of the chosen transformant colonies in duplicate (A and B) from Figure 3.3-J small amounts of each plasmid was gel electrophoresed.

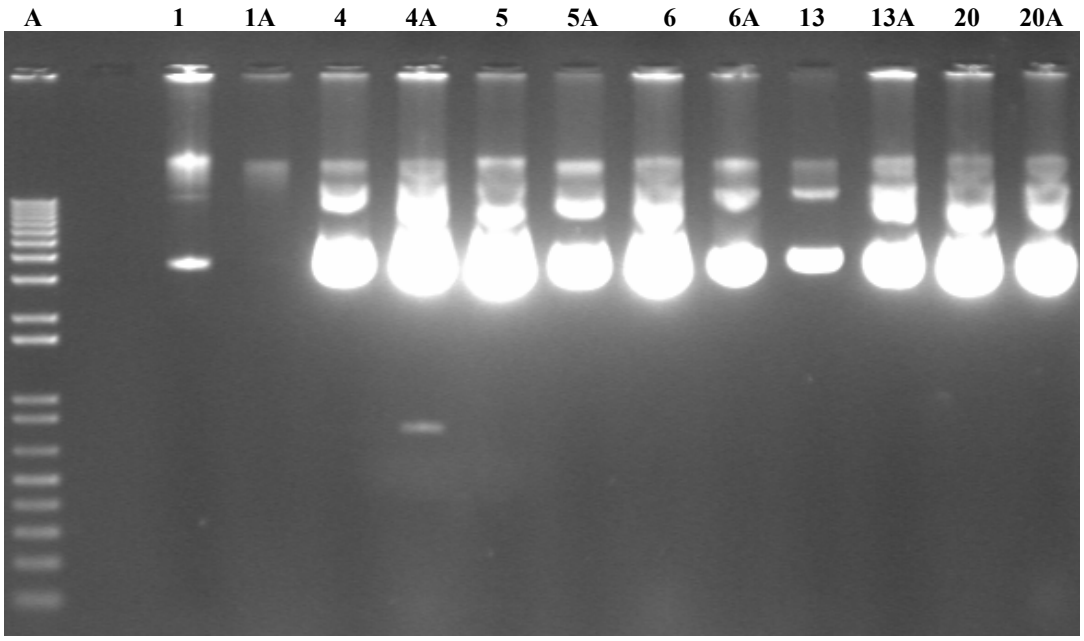


Figure 3.3-J: Large TAE gel depicting the plasmid DNA precipitated from the 109F/1391R transformants selected from the 'screening method' in Figure 3.3-I. Each selected (1-20A) transformant was isolated and precipitated in duplicate.

3.4 Discussion

3.4.1 5' End Extension of 'Known' Sequence

To extend the 5' end of the ~830bp 'known' sequence from the aforementioned preliminary studies, PCR was carried out using the primer sets: 1AF/927R or 927R NEW, with the human mucosal surface sample DM20 as a template. The theoretical size of the manufactured amplicon was ~927bp, therefore it was our aim to isolate this PCR product, or any within this range, for cloning and transformation, and from there preparation for sequencing and analysis could begin.

3.4.1.1 1AF/927 R or 927R NEW PCR Product Isolation for Cloning and Transformation

PCR was carried out in triplicate using both primer sets: 1AF/927R or 927R NEW, with the human mucosal surface sample DM20 as a template. Results are shown in Figure 3.3-A, both primer sets amplified similarly, with a bright band in all amplifications just above the 900bp mark illustrated by the 100kb ladder in track A. These PCR products were believed to be candidates for the 5' extended version of the 'known' sequence, and were therefore flagged for isolation, cloning and transformation.

Isolated PCR products were attained using the 'bandstab' method described in 3.2.1.3. The outcome of the excision performed on the applicable bands is presented in Figure 3.3-B. These excised products were then checked for purity by repeating the PCR with the respective primer sets, the PCR products were also PEG/$MgCl_2$ precipitated, in order to enhance purification and increase concentration prior to cloning; the results are exhibited in Figure 3.3-C. In Figure 3.3-C (1-6), the excised band from Figure 3.3-B 're-amplified' brighter and with a marked decrease in non-specific 'noise'. The precipitation of the 're-amplified' excised bands (7-12) showed little if any decrease in intensity, indicating a successful precipitation. The purity of

the isolated PCR products were also slightly increased as the precipitations displayed some decrease in non-specific 'noise'.

These purified and concentrated PCR products were inserted into a pCR® 2.1-TOPO plasmid vector (Figure 3.2-A) following the TOPO® Cloning Reaction outlined in the TOPO TA Cloning® Instruction Manual (Version R), Invitrogen ™ (see 3.2.2). The vector was then transformed into One Shot® TOP10 Competent Cells adhering to the method described in TOPO TA Cloning® Instruction Manual (Version R), Invitrogen ™ (see 3.2.2.2). Transformants were then spread out on LB/Kan plates containing X-gal (see 3.2.2.3) with white cell colonies being selected to undergo the subsequent 'screening' process, in an effort to find transformants that harbored relevant PCR product sequences.

3.4.1.2 Screening Transformant Plasmids for Sequencing

The 'screening method' outlined in 3.2.3 and validated in Chapter 4, was employed to filter out transformant colonies that contained PCR products that were not of interest. This was accomplished using the 109F/927R or 109F/927R NEW primer sets in direct colony PCR, in the expectation that, only colonies containing relevant sequences would be targeted and produce a PCR product ~830bp. This would narrow down the sum of transformant colonies that would ultimately have their plasmids precipitated and their PCR product insert sequenced.

An example of the 'screening method' for the 5' end extension sequence is shown in Figure 3.3-D and 3.3-E. Figure 3.3-D displays the outcome of a direct colony PCR (see 3.2.3.1), with 1AF/927R transformant colonies and the primer set: 109F/927R. All bright bands located ~830bp (1, 2, 3, 4, 5, 6, 8 and 10) were of interest and their corresponding transformant colonies went on to have their plasmids precipitated for sequencing. Bands (7 and 9) from Figure 3.3-D that displayed non-specific binding or no amplification whatsoever were excluded.

Selected transformant colonies had their plasmids precipitated, checked and measured following the protocols detailed in 3.2.4. Figure 3.3-E is an example of the plasmid

DNA check (3.2.4.1.A) carried out on plasmids precipitated from transformants. In this case Figure 3.3-E displays the plasmid precipitations of the selected transformants from Figure 3.3-D. Due to the sensitive nature of the plasmid preparation procedure it was not uncommon to lose the precipitated plasmid DNA pellet during its final ethanol wash (see 3.2.4.1), in this example 6 was lost and is absent from Figure 3.3-E.

After the plasmids were precipitated they were checked and measured by both gel and Nanodrop® ND-1000 Spectrophotometer following the methods in 3.2.4.1.A and 3.2.4.1.B. Using this information plasmid DNA was diluted to a concentration of (~100ng/μL) (see 3.2.4.1.C) and sent to the Waikato DNA Sequencing Facility for sequencing (Refer to Chapter 5 for sequencing results and analysis).

3.4.2 3' End Extension of 'Known' Sequence

To extend the 3' end of the 'known' sequence from aforementioned preliminary studies, PCR was carried out using the primer sets: 109F or 109F NEW/1391R, with the human mucosal surface sample DM20 as a template. The theoretical size of the manufactured amplicon was ~1282bp, therefore it was our aim to isolate this PCR product, or any within this range, for cloning and transformation, and from there, preparation for sequencing and analysis could begin.

3.4.2.1 109F/1391R PCR Product Isolation for Cloning and Transformation

PCR was carried out in duplicate using both primer sets: 109F/1391R and 109F NEW/1391R, with the human mucosal surface sample DM20 as a template. Results are shown in Figure 3.3-F, both primer sets amplified similarly, with bright bands at ~1000, 1200 and 1400bp as illustrated by the 100kb ladder in track A. These 6 PCR products (labeled A-F respectively) were believed to be potential candidates for the 5' extended version of the 'known' sequence, and were therefore flagged for isolation, cloning and transformation.

The same method of 'bandstabbing' candidate bands as in 3.3.1.1.A proved to be unsuccessful. The outcome of the excision and subsequent 're-amplification' with the primer sets 109F/1391R or 109F NEW/1391R performed on the applicable bands (A-F) is presented in Figure 3.3-G. It can be seen that there is some amplification for bands A-C from Figure 3.3-F (2A-2C in Figure 3.3-G), with a major product at ~400bp and some accompanying non-specific products. Reamplified bands D-F in Figure 3.3-F displayed little or no amplification, as seen in Figure 3.3-F (4D-4F), with either a small product under 100bp or a primer dimer having been produced. In light of this failure it was decided that the entire raw PCR product for only 109F/1391R would be PEG/$MgCl_2$ precipitated, and not 109F NEW/1391R as it was considered superfluous.

Figure 3.3-H portrays the entire raw PCR product for the 109F/1391R primer set with the human mucosal surface sample DM20 as a template (1, 3 and 5) and the PEG/$MgCl_2$ precipitation of each (2, 4, and 6), this was prepared in triplicate. The precipitations were successful and allowed for cloning, it was thought that all of the non-specific products would by cloned into the vector, including some of the relevant sequences that may have belonged to the potential candidate bands circled in Figure 3.3-F.

The precipitated and concentrated raw PCR product from Figure 3.3-H (2) was inserted into a pCR® 2.1-TOPO plasmid vector (Figure 3.2-A) following the TOPO® Cloning Reaction outlined in the TOPO TA Cloning® Instruction Manual (Version R), Invitrogen ™ (see 3.2.2). The vector was then transformed into One Shot® TOP10 Competent Cells adhering to the method described in TOPO TA Cloning® Instruction Manual (Version R), Invitrogen ™ (see 3.2.2.2). Transformants were then spread out on LB/Kan plates containing X-gal (see 3.2.2.3) with white cell colonies being selected to undergo the subsequent 'screening' process, in an effort to find transformants that harbored relevant PCR product sequences.

3.4.2.2 Screening Transformant Plasmids for Sequencing

The 'screening method' outlined in 3.2.3 and validated in Chapter 4, was applied to filter out transformant colonies that contained PCR products that were not of interest. This was accomplished using the 109F/927R primer set in direct colony PCR, in the expectation that, only colonies containing relevant sequences would be targeted and produce a PCR product ~830bp. This would narrow down the sum of colonies that would have their plasmids precipitated and their PCR product insert sequenced.

Unfortunately, the 'screening method' used in the 5' extension (3.3.1.1) did not function in the same manner as when applied to the 3' extension (3.3.1.2). The application of the 'screening method' to the 3' extension was tested in Chapter 4 and illustrated the method, in respect to the 3' extension, to be flawed. To maintain consistency and avoid constant sequencing or totally irrelevant and unwanted sequences the 'screening method' was used throughout the project.

Use of the 'screening method' for the 3' end extension sequence is shown in Figure 3.3-I and 3.3-J. Figure 3.3-I displays the outcome of a direct colony PCR (see 3.2.3.1), with 109F/1391R transformant colonies and the primer set: 109F/927R. All bright bands located ~830bp (1, 4, 5, 6, 13 and 20) were of interest and these transformant colonies went on to have their plasmids precipitated for sequencing. Bands (2, 3, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18 and 19) showed weaker or no amplification whatsoever and were consequently excluded.

Selected transformant colonies had their plasmids precipitated, checked and measured following the protocols detailed in 3.2.4. Figure 3.3-J is an example of the plasmid DNA check (3.2.4.1.A) carried out on plasmids precipitated from transformants. In this case Figure 3.3-J displays the plasmid precipitations of the selected transformants from Figure 3.3-I.

After the plasmids were precipitated they were checked and measured by both gel and Nanodrop® ND-1000 Spectrophotometer following the methods in 3.2.4.1.A and 3.2.4.1.B. Using this information plasmid DNA was diluted to a concentration of

(~100ng/μL) (see 3.2.4.1.C) and sent to the Waikato DNA Sequencing Facility for sequencing (refer to Chapter 5 for sequencing results and analysis).

3.5 Conclusion

Isolation of the amplicons produced by the 1AF/927R or 927R NEW primers sets with the DM20 human gastrointestinal mucosal DNA template was relatively successful. The isolation of the 109F/1391R amplicons was not successful; therefore the use of precipitated, raw PCR products in cloning was to be carried out in this case. The hope was that at least some of the relevant amplicons would be inserted into plasmids and eventually be sequenced.

Chapter 4: Validation of Screening Methodology

4.1 Introduction

The 'screening method' was devised to filter out colonies that contained PCR products that were not of interest. This was accomplished using the 109F/927R or 109F/927R NEW primer sets in the expectation that, only colonies containing relevant sequences would be targeted. This would narrow down the sum of colonies that would have their plasmids precipitated and their PCR product insert sequenced.

4.2 Materials and Methods

4.2.1 Direct Colony PCR

Direct colony PCRs were carried out following the procedure outlined in (2.6.3.1), with the 109F/927R or 109F/927R NEW primer sets.

4.2.2 Gel electrophoresis

Gel electrophoresis of PCR products was carried out following the procedure outlined in

4.2.3 Preparation of Plasmid DNA for Sequencing

4.2.3.1 Gel Electrophoresis Check

A qualitative and quantitative check of the re-suspended plasmid DNA were carried out using gel electrophoresis as explained in 3.2.4.1.A.

4.2.3.2 Determination of DNA concentration using a Nanodrop spectrophotometer

The quantity of re-suspended plasmid DNA was determined using a Nanodrop® ND-1000 Spectrophotometer following the method detailed in 3.2.4.1.B.

4.2.3.3 Dilution of Plasmid DNA

The plasmid DNA concentration requirements for sequencing were reached adhering to the procedure in 3.2.4.1.B.

4.2.4 Sequencing

Plasmid DNA was sequenced by the Waikato DNA Sequencing Facility, Hamilton, New Zealand. Adhering to the requirements stipulated on their website (http://bio.waikato.ac.nz/sequence/), sequences were determined using the M13 Forward Primer provided with the TOPO® Cloning Reaction Kit (binding site shown in Fig.1 above) and Applied Biosystems Big Dye v3.1 chemistry.

4.2.5 Analysis of Sequencing Results

4.2.5.1 Chromas

Electropherogram sequence results were edited in Chromas 2.33, Technelysium Pty Ltd. Flanking vector and primer sequences were manually removed from the sequence of interest and the integrity of the sequence was also monitored and sequence of low integrity was also removed prior to further analysis being carried out.

4.2.5.2 ClustalW2

Edited sequences were submitted for analysis, along with the DM1 consensus sequence obtained from preliminary work, to the DNA/Protein multiple sequence

alignment program ClustalW2, located on the European Bioinformatics Institute (EBI) website (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

4.2.5.3 BLAST

Edited sequences were submitted to BLAST located on the National Biocenter for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) for analysis. The parameters for BLAST were: Nucleotide BLAST; Non-redundant Database (Nucleotide Collection (nr/nt); Megablast (optimised for highly similar sequences).

4.2.5.4 Classifier

Edited sequences were submitted for analysis to the Classifier program located on the Ribosomal Database Project (RDP) website (http://rdp.cme.msu.edu/classifier/classifier.jsp). The classification algorithm that aligned 16S rRNA sequences was designed by (Q. Wang et al., 2007)

4.3 Results

4.3.1 Direct Colony PCR of 1AF/927R Transformants with 109F/927R Primers and Analysis of Sequences

Transformant colonies presumably containing 1AF/927R PCR product inserts underwent direct colony PCR (Figure 4.3-A), the plasmids were then precipitated and the inserts sequenced and analyzed (Table 4.3-A).

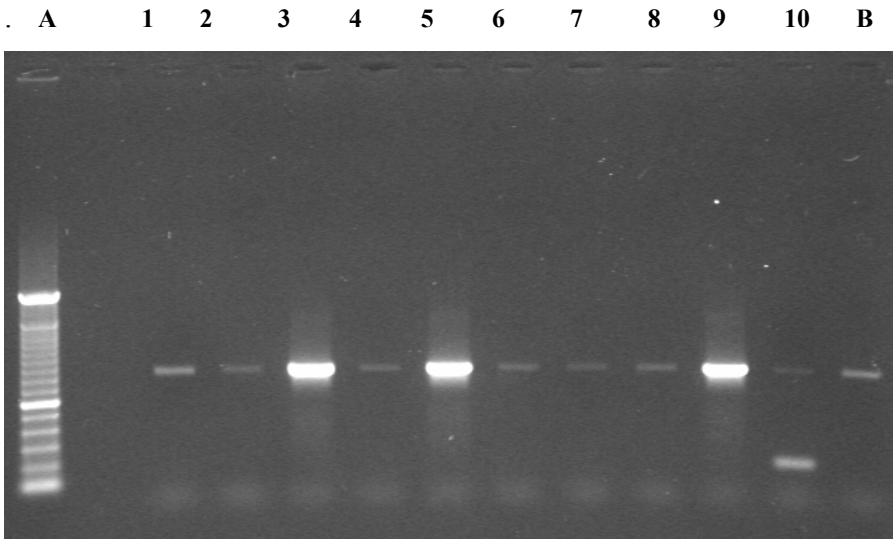


Figure 4.3-A: A, 100kb ladder; 1-10, 1AF/927R transformant colonies from reference plate amplified with the primer set: 109F/927R; 11, Positive control: DM20(1:10 dilution) with the primer set 109F/927R. 1-9 were selected for sequencing.

Table 4.3-A: Sequence results and analysis for 1-9 from figure 4.3-B above, the highlighted rows are 'hits' found in BLASTand Classifier that come from *Firmicutes* and sequences obtained from the human gastrointestinal tract.

Primers 109F/927R	BLAST No.	Best Blast Hit	Reference	Classifier Result	Division
(Date_Primer Set_No.)					
Product: 1AF/927R					
300908_1AF_927R_ 1	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Gordon et al. 2006	Lachnospiraceae Incertae Sedis	Firmicute
300908_1AF_927R _ 2	AC010093.8	Homo sapiens BAC clone RP11-323O5	Levy et al. 2001	Unclassified Root	N/A
300908_1AF_927R _ 3	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Gordon et al. 2006	Unclassified Lachnospiraceae	Firmicute
300908_1AF_927R _ 4A	AC106827.2	Homo sapiens chromosome 3, RP11-320P11	Haugen et al. 2002	Unclassified Root	N/A
300908_1AF_927R _ 5	AY977855.1	Uncultured bacterium clone LF53	Relman et al. 2005	Roseburia	Firmicute
300908_1AF_927R _ 6	AC011597.27	Homo sapiens 3 BAC RP11-85F14	Worley et al. 1999	Unclassified Root	N/A
300908_1AF_927R _ 7	AC011597.27	Homo sapiens 3 BAC RP11-85F14	Worley et al. 1999	Unclassified Root	N/A
300908_1AF_927R _ 8	AC093014.11	Homo sapiens 12 BAC RP11-734E19	Worley et al. 2001	Unclassified Bacteria	N/A
300908_1AF_927R _ 9	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Gordon et al. 2006	Lachnospiraceae Incertae Sedis	Firmicute

4.3.2 Direct Colony PCR of 109F/1391R Transformants with 109F/927R Primers and Analysis of Sequences

After completing the validation above in 4.3.1, it was thought that the same 'screening method' would be applied to transformant colonies presumably containing 109F/1391R PCR product inserts. In this case only the 'brighter' bands in a direct colony PCR (Figure 4.3-B), had their corresponding transformant colony's plasmids precipitated and the inserted PCR products sequenced and analyzed (Table 4.3-B).

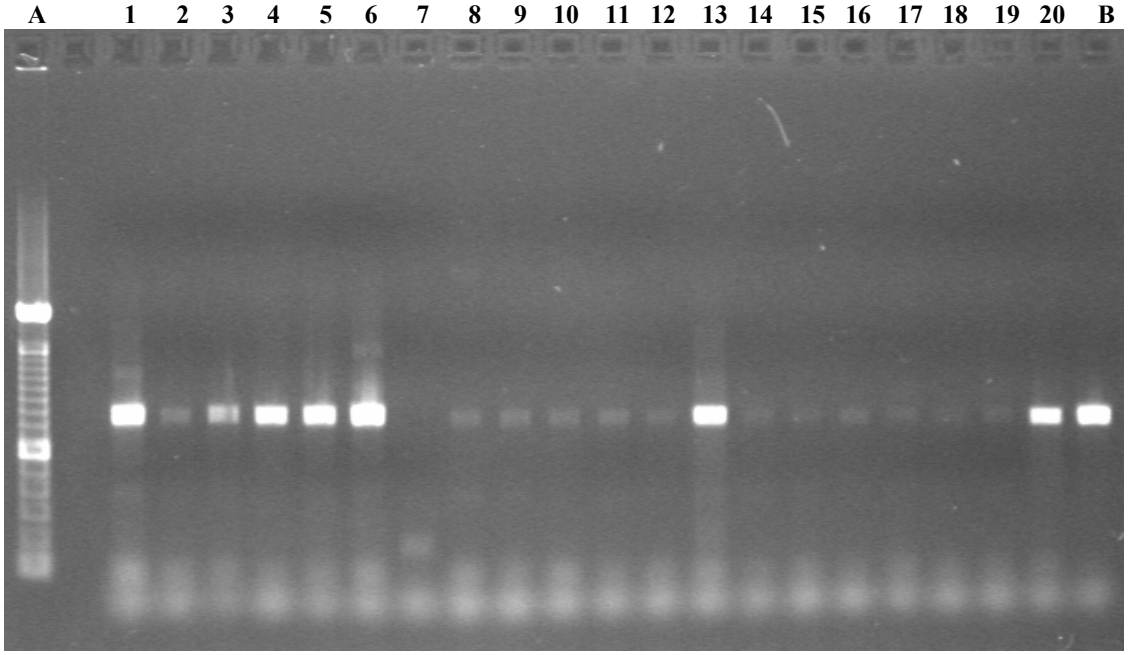


Figure 4.3-B: Large TAE gel displaying the direct colony PCR of 109F/1391R transformants with the 109F/927R primer set: A, 100kb ladder; 1-20, 109F/1391R transformants amplified with109F/927R; B, Positive control: DM20(1:10 dilution) with the primer set 109F/927R. 4, 5, 6, 13 and 20 were selected for sequencing and analysis.

Table 4.3-A: Sequence results and analysis for 1-9 from figure 4.3-B above, the highlighted rows are 'hits' found in BLASTand Classifier that come from *Firmicutes* and sequences obtained from the human gastrointestinal tract.

Sequence	BLAST No.	Best Blast Hit	Reference	Classifier Result	Division
109F/1391R:					
270108_109F_1391R _4	AC209419.4	Homo sapiens FOSMID clone ABC9-43925600O21	Wilson,R.K. 2007	Unclassified Root	N/A
270108_109F_1391R _5	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	Unclassified Root	N/A
270108_109F_1391R _6	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	Unclassified Root	N/A
270108_109F_1391R _20	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	Unclassified Root	N/A

4.4 Discussion

4.4.1 1AF/927R and 1AF/927R NEW Screening Method

1AF/927R transformant colonies were selected for direct colony PCR (Figure 4.3-A), plasmids were then precipitated from the selected transformants and sequenced. Sequences were edited in Chromas and submitted to BLAST and Classifier to better identify them. These results can be viewed in Table 4.3-A.

BLAST and Classifier results designated 'brighter' bands from tracks: 1, 3, 5, and 9 as relevant *Firmicute* sequences that were similar to those described in the literature that were also sourced from the human gastrointestinal tract. The 'less bright' bands from tracks: 2, 4, 6, 7, and 8 gave unclassified roots and divisions in Classifier and BLAST 'hits' with Homo sapiens DNA and are thus expected to be the result of non-specific amplification.

As our known sequence belonged to the *Mollicute* Class and in turn the bacterial Division of the *Firmicutes*, it was reasoned that all the 'brighter' bands that gave *Firmicute* sequences, would be selected and sequenced; 'less bright' bands would be excluded.

4.4.2 109F/1391R Screening Method

109F/1391R transformant colonies were selected for direct colony PCR (Figure 4.3-B), selected transformant colonies then had their plasmids precipitated and sequenced. These sequences were edited in Chromas and submitted to BLAST and Classifier to better identify them. These results can be viewed in Table 4.3-B.

BLAST and Classifier results designated the selected 'brighter' bands from tracks: 4, 5, 6, 13 and 20 (13 was lost in the precipitation process and is absent from Table 4.3-B). None of the analysed sequences were Classified as *Firmicutes*, in fact all the

sequences were Classified as having ‘Unclassified Roots’ and having BLAST ‘hits’ with Homo sapiens DNA; these results are expected to be the result of non-specific amplification by the 109F/1391R primer set.

4.5 Conclusion

The ‘screening method’ outlined in this chapter displayed itself as a viable selection method for 1AF/927R transformant plasmids that contained PCR product inserts that belonged to potential *Firmicutes* inhabiting the human gastrointestinal tract.

Due to similarity between the reverse primers 927R and 927R NEW, the ‘screening method’ was employed throughout the project, against the PCR product inserts attained from the template DM20 (sample isolated from the human gastrointestinal tract) and the primer sets: 1AF/927R or 927R NEW. This method was seen as means of saving time, labour and laboratory resources.

Although the ‘screening method’ was inefficient when applied to transformants containing 109F/1391R amplicons, all transformant colonies underwent the same process, to keep consistency throughout the project.

Chapter 5: Sequence Analysis

5.1 Introduction

The PCR products were isolated, cloned and transformed, these transformants then underwent the 'screening process' outlined in Chapter 4. 'Selected' transformants had their plasmids precipitated and prepared for sequencing. Chapter 5 gives the sequencing results and analysis of those inserted PCR product.

5.2 Material and Methods

5.2.1 Sequencing

Plasmid DNA was sequenced by the Waikato DNA Sequencing Facility, Hamilton, New Zealand. Adhering to the requirements stipulated on their website (http://bio.waikato.ac.nz/sequence/), sequences were determined using the M13 Forward Primer provided with the TOPO® Cloning Reaction Kit (binding site shown in Fig.1 above) and Applied Biosystems Big Dye v3.1 chemistry.

5.2.2 Analysis of Sequencing Results

5.2.2.1 Chromas

Electropherogram sequence results were edited in Chromas 2.33, Technelysium Pty Ltd. Flanking vector and primer sequences were manually removed from the sequence of interest and the integrity of the sequence was also monitored and sequence of low integrity was also removed prior to further analysis being carried out.

5.2.2.2 ClustalW2

Edited sequences were submitted for analysis, along with the DM1 consensus sequence obtained from preliminary work, to the DNA/Protein multiple sequence alignment program ClustalW2, located on the European Bioinformatics Institute (EBI) website (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

5.2.2.3 BLAST

Edited sequences were submitted to BLAST located on the National Biocenter for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) for analysis. The parameters for BLAST were: Nucleotide BLAST; Non-redundant Database (Nucleotide Collection (nr/nt); Megablast (optimised for highly similar sequences).

5.2.2.4 Classifier

Edited sequences were submitted for analysis to the Classifier program located on the Ribosomal Database Project (RDP) website (http://rdp.cme.msu.edu/classifier/classifier.jsp). The classification algorithm that aligned 16S rRNA sequences was designed by (Q. Wang et al., 2007).

5.3 Results

Table 5.3-A: Relevant results from the sequence analysis of 1AF/927R amplicons. The DM1 Consensus ('known') sequence analysis is presented.

Sequence	BLAST No.	Best Blast Hit	Blast Reference	Query Cover/ Identity (%)	Classifier Result	Division
109F/927R						
DM1 Consensus	DQ797049.1	Uncultured bacterium clone RL386_aao86c09	Ley et al. 2006	91/99	Anaeroplasma	Firmicute
1AF/927R:						
(Date_Primer Set_No.)						
040608_1AF_927R_3	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Ley et al. 2006	99/99	Lachnospiraceae Incertae Sedis	Firmicute
210808_1AF_927R_2	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Ley et al. 2006	88/98	Unclassified Lachnospiraceae	Firmicute
210808_1AF_927R 6A	DQ802576.1	Uncultured bacterium clone RL188_aan93f02	Ley et al. 2006	100/98	Faecalibacterium	Firmicute
300908_1AF_927R_ 1	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Ley et al. 2006	99/98	Lachnospiraceae Incertae Sedis	Firmicute
300908_1AF_927R_ 3	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Ley et al. 2006	99/97	Unclassified Lachnospiraceae	Firmicute
210808_1AF_927R _8A	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Ley et al. 2006	100/98	Lachnospiraceae Incertae Sedis	Firmicute
300908_1AF_927R_ 9	DQ807184.1	Uncultured bacterium clone RL244_aaj47c08	Ley et al. 2006	99/98	Lachnospiraceae Incertae Sedis	Firmicute
040608_1AF_927R_5	AY977855.1	Uncultured bacterium clone LF53	Eckburg et al. 2005	99/97	Roseburia	Firmicute
210808_1AF_927R _10	EF403800.1	Uncultured bacterium clone SJTU_A2_03_21	Zhao et al. 2007	99/98	Faecalibacterium	Firmicute
210808_1AF_927R _5	EF404855.1	Uncultured bacterium clone SJTU_G_09_34	Zhao et al. 2007	99/95	Ruminococcaceae Incertae Sedis	Firmicute
210808_1AF_927R _7A	EF404855.1	Uncultured bacterium clone SJTU_G_09_34	Zhao et al. 2007	100/95	Ruminococcaceae	Firmicute
210808_1AF_927R _3A	EU139255.1	Ruminococcus gnavus strain A2	Nicoli et al. 2007	100/97	Unclassified Lachnospiraceae	Firmicute
300908_1AF_927R_ 5	AY977855.1	Uncultured bacterium clone LF53	Relman et al. 2005	99/99	Roseburia	Firmicute
210808_1AF_927R _9	EU762168.1	Uncultured bacterium clone Ax2_123	Relman et al. 2008	99/99	Faecalibacterium	Firmicute
040608_1AF_927R_7	CP000383.1	Cytophaga hutchinsonii ATCC 33406	McBride et al. 2006	99/97	Cytophaga	Bacteriodetes
040608_1AF_927R_2	DQ807609.1	Uncultured bacterium clone RL307_aam08h03	Ley et al. 2006	99/97	Bacteriodes	Bacteriodetes
040608_1AF_927R_4	DQ447856.1	Uncultured bacterium clone MSFC_4M9I	Venkateswaran et al. 2006	99/99	Cloacibacterium	Bacteriodetes

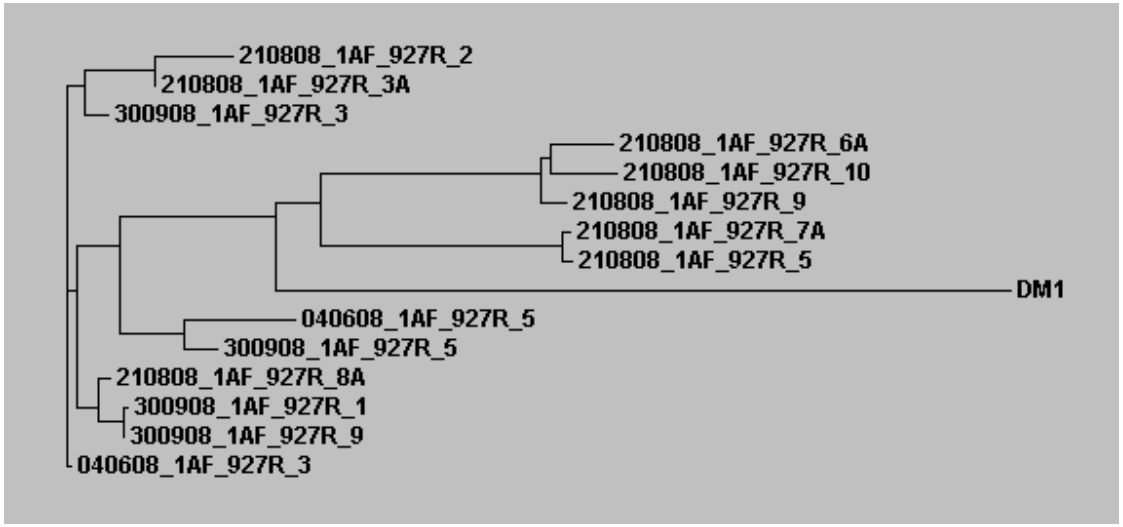


Figure 5.3-A: Phylogram output from ClustalW2 of the highlighted 1AF/927R sequences from Table 5.3-A and the DM1 'known' Consensus Sequence

Table 5.3-B: Relevant results from the sequence analysis of 1AF/927R NEW amplicons. The DM1 Consensus ('known') sequence analysis is presented.

Sequence	BLAST No.	Best Blast Hit	Blast Reference	Query Cover/ Identity (%)	Classifier Result	Division
109F/927R						
DM1 Consensus	DQ797049.1	Uncultured bacterium clone RL386_aao86c09	Ley et al. 2006	91/99	Anaeroplasma	Firmicute
1AF/927RNEW:						
(Date_Primer Set_No.)						
040608_1AF_927Rn_19	DQ798083.1	Uncultured bacterium clone RL302_aal95c11	Ley et al. 2006	99/97	Faecalibacterium	Firmicute
200808_1AF.927Rn_8	DQ807308.1	Uncultured bacterium clone RL244_aaj48h11	Ley et al. 2006	94/99	Lachnospiraceae Incertae Sedis	Firmicute
200808_1AF.927Rn_5	DQ807308.1	Uncultured bacterium clone RL244_aaj48h11	Ley et al. 2006	100/99	Lachnospiraceae	Firmicute
200808_1AF.927Rn_6A	EU766834.1	Uncultured bacterium clone C1_757	Relman et al.2008	100/99	Faecalibacterium	Firmicute
200808_1AF.927Rn_4A	EU766834.1	Uncultured bacterium clone C1_757	Relman et al.2008	100/100	Faecalibacterium	Firmicute
040608_1AF_927Rn_14	EU765493.1	Uncultured bacterium clone B3_632	Relman et al.2008	100/96	Roseburia	Firmicute
200808_1AF.927Rn_2	EU139255.1	Ruminococcus gnavus strain A2	Nicoli et al. 2007	99/99	Unclassified Lachnospiraceae	Firmicute
040608_1AF_927Rn_16	EU130959.1	Hydrogenophaga sp. BAC120	Van der Kooij et al. 2007	100/93	Unclassified Root	N/A
200808_1AF.927Rn_9A	AF035052.1	Beta proteobacterium B6	Szewzyk et al.	99/98	Aquabacterium	Proteobacteria

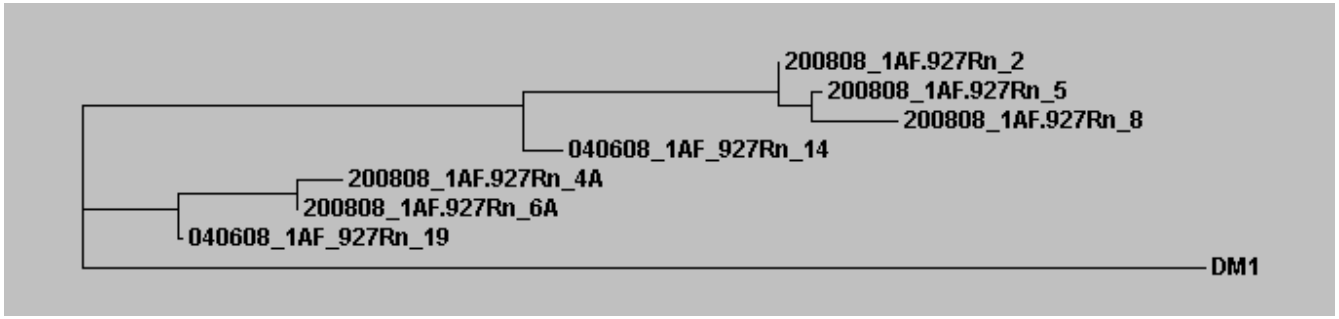


Figure 5.3-B.: Phylogram output from ClustalW2 of the highlighted 1AF/927R NEW sequences from Table 5.3-B and the DM1 'known' Consensus Sequence

Table 5.3-C: Sequence analysis of the 109F/1391R amplicons.

Sequence	BLAST No.	Best Blast Hit	Blast Reference	Query Cover/ Identity (%)	Classifier Result	Division
109F/1391R:						
131008_109F_1391R_R3A	EU881319.1	Uncultured bacterium clone KMS200711-017	Chen,X.B. 2008	99/99	Clostridium	Firmicute
270109_109F_1391R _48	EU778429.1	Uncultured bacterium clone RP_2aaa01f05	Gordon,J.I. 2008	100/98	Clostridium	Firmicute
131008_109F_1391R_R2A	AC091022.4	Homo sapiens chromo 8, RP11-65A5	Birren,B. 2001	100/98	Unclassified Root	N/A
131008_109F_1391R_R4	AC087521.10	Homo sapiens chromo 11, RP11-613D13	Birren,B. 2001	100/99	Unclassified Bacteria	N/A
131008_109F_1391R_R7	AC087521.10	Homo sapiens chromo 11, RP11-613D13	Birren,B. 2001	100/98	Unclassified Root	N/A
270109_109F_1391R _62	CU634019.4	Human DNA CH507-236L23 on chromo 21	Collins,J. 2009	99/99	Unclassified Root	N/A
270109_109F_1391R _62A	CU634019.4	Human DNA CH507-236L23 on chromo 21	Collins,J. 2009	98/99	Unclassified Root	N/A
270109_109F_1391R _65	AC112205.2	Homo sapiens chromo 5 RP11-80G7	DOE Inst.. 2002	98/98	Unclassified Root	N/A
131008_109F_1391R_R8A	AC009119.10	Homo sapiens chromoe 16 RP11-483P21	DOE Inst. 2002	100/86	Unclassified Root	N/A
131008_109F_1391R_R5	AC114485.2	Homo sapiens chromosome 1 RP11-202K23	Kaul,R.K. 2002	98/98	Unclassified Bacteria	N/A
270109_109F_1391R _45A	NM_000599.3	Homo sapiens insulin-like grwth fac protein5	Park,S.Y. 2008	100/99	Unclassified Root	N/A
131008_109F_1391R_B5A	AL359457.12	Human DNA sequence from RP11-76K19	Pelan,S. 2009	89/99	Unclassified Root	N/A
270109_109F_1391R _59	AC079882.6	Homo sapiens BAC clone RP11-577O18	Wilson,R. 1998	100/99	Unclassified Bacteria	N/A
270109_109F_1391R _59A	AC079882.6	Homo sapiens BAC clone RP11-577O18	Wilson,R. 1998	100/99	Unclassified Bacteria	N/A
270109_109F_1391R _61A	AC079882.6	Homo sapiens BAC clone RP11-577O18	Wilson,R. 1998	84/97	Unclassified Bacteria	N/A
270109_109F_1391R _66	AC079882.6	Homo sapiens BAC clone RP11-577O18	Wilson,R. 1998	100/98	Unclassified Bacteria	N/A
270109_109F_1391R _4	AC209419.4	Homo sapiens FOSMID ABC9-43925600O21	Wilson,R.K. 2007	100/98	Unclassified Root	N/A
270109_109F_1391R _30A	AC209419.4	Homo sapiens FOSMID ABC9-43925600O21	Wilson,R.K. 2007	99/84	Unclassified Bacteria	N/A
270109_109F_1391R _64	AC024887.22	Homo sapiens 3 BAC RP11-307B9	Worley,K.C. 2000	100/98	Unclassified Bacteria	N/A
131008_109F_1391R_B4A	AC024887.22	Homo sapiens 3 BAC RP11-307B9	Worley,K.C. 2000	91/98	Unclassified Bacteria	N/A
131008_109F_1391R_B7	AC024887.22	Homo sapiens 3 BAC RP11-307B9	Worley,K.C. 2000	99/97	Unclassified Bacteria	N/A
131008_109F_1391R_B9	AC024887.22	Homo sapiens 3 BAC RP11-307B9	Worley,K.C. 2000	99/98	Unclassified Bacteria	N/A
131008_109F_1391R_R9	AC078889.20	Homo sapiens 12 BAC RP11-335I12	Worley,K.C. 2000	100/98	Unclassified Root	N/A
131008_109F_1391R_B1A	AC092185.3	Homo sapiens 3q BAC RP11-148O7	Worley,K.C. 2001	99/86	Unclassified Bacteria	N/A
270109_109F_1391R _5	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	99/97	Unclassified Root	N/A
270109_109F_1391R _5A	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	99/98	Unclassified Root	N/A

Sequence	BLAST No.	Best Blast Hit	Blast Reference	Query Cover/ Identity (%)	Classifier Result	Division
109F/1391R:						
270109_109F_1391R _6	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	91/98	Unclassified Root	N/A
270109_109F_1391R _20A	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	99/99	Unclassified Root	N/A
270109_109F_1391R _44	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	97/98	Unclassified Bacteria	N/A
270109_109F_1391R _52A	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	99/98	Unclassified Root	N/A
270109_109F_1391R _53A	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	99/99	Unclassified Root	N/A
270109_109F_1391R _58	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	100/99	Unclassified Root	N/A
270109_109F_1391R _58A	AC023593.14	Homo sapiens 3 BAC RP11-427I6	Worley,K.C. 2002	98/99	Unclassified Root	N/A

Figure 5.3-C: Phylogram output from ClustalW2 of the highlighted 109F/1391R sequences from Table 5.3-C and the DM1 'known' Consensus Sequence

5.4 Discussion

The sequences obtained from the WDSF were checked and edited in Chromas, where vector, primer and erroneous sequences were removed or if required, excluded. After editing the remaining sequences were submitted to both BLAST and Classifier (5.2.2). Sequences that showed promise i.e. were sequences that belonged to the bacterial division: *Firmicutes*, were submitted for alignment in the ClustalW2 with the 'known' sequence representative: DM1 consensus sequence.

5.4.1 5' End Extension Sequences

The primer sets 1AF/927R and 1AF/927R NEW were designed with the purpose of amplifying the 5' end of the known sequence, extending it from ~830bp to ~930bp. If this was achieved the 'known' sequence could then become better defined, giving us a potential insight into its phylogeny and possibly its functionality and role within the human gastrointestinal tract.

5.4.1.1 Sequences from the 1AF/927R Primer Set

Sequences belonging to the inserted PCR products from the 'selected' 1AF/927R transformants were analysed, with the results placed in Table 5.3-A. Sequences that belonged to *Firmicutes* were highlighted, these sequences were then picked to undergo alignment in ClustalW2 with the 'known' sequence representative: DM1 consensus. The result of this alignment is given in the form of a phylogram in Figure 5.3-A, which depicts the 'relatedness' of the sequences.

Of the 17 sequences found in Table 5.3-A, 14 (82%) belonged to the *Firmicutes* Division, 11 (65%) of these belonged to the Family *Lachnospiraceae* and 7 (41%) had BLAST 'hits' with 'Uncultured bacterium' 16S rRNA sequences obtained from

the human gastrointestinal tract by (R. E. Ley et al., 2006), interestingly this was the same work that the DM1 consensus sequence 'hit' when submitted to BLAST (Table 5.3-A).

Despite 'hitting' the same work as the DM1 consensus (R. E. Ley et al., 2006), none of the submitted sequences from Table 5.3-A actually 'hit' the same BLAST sequence as the DM1 consensus (DQ797049.1). Although the DM1 consensus was determined by Classifier to belong to the bacterial Division: *Firmicutes* like 82% of the 1AF/927R amplified sequences, it was placed within the Genus of *Anaeroplasma*. None of the sequences in Table 5.3-A were classified as belonging to *Anaeroplasma* Genus, in fact the majority of the *Firmicute* sequences (79%) belonged to the Family: *Lachnospiraceae*. *Lachnospiraceae* are part of the *Clostridia* Class, whereas *Anaeroplasma* is a member of the *Mollicutes* Class.

The ClustalW2 alignment of all the highlighted *Firmicute* sequences in Table 5.3-A are represented in the phylogram in Figure 5.3-A. The phylogram supports the BLAST and Classifier results from Table 5.3-A, the DM1 sequence is similar to the others, yet, it is still grouped alone. The other sequences that show similar results in Table 5.3-A group together, the DM1 consensus shows some 'relatedness' to the other sequences but is grouped alone, this is also concordant with the information in Table 5.3-A.

Designed to target and 5' extended version of the 'known' sequence it would appear that the 1AF/927R primer set did not attain its overall objective. But in sight of this it managed to amplify 16S rRNA sequences that belonged to *Firmicutes* and primarily 'hit' sequences that had been acquired from the human gastrointestinal tract, where of course all of the samples from this project were obtained.

5.4.1.2 Sequences from the 1AF/927R NEW Primer Set

Sequences belonging to the inserted PCR products from the 'selected' 1AF/927R NEW transformants were analysed, with the results placed in Table 5.3-B. Sequences that belonged to *Firmicutes* were highlighted, these sequences were then picked to undergo alignment in ClustalW2 with the 'known' sequence representative: DM1 consensus. The result of this alignment is given in the form of a phylogram in Figure 5.3-B, which depicts the 'relatedness' of the sequences.

Of the 9 sequences found in Table 5.3-B, 7 (78%) belonged to the *Firmicutes* Division, of which 4 (44%) belonged to the Family *Lachnospiraceae* and 3 (33%) had BLAST 'hits' with 'Uncultured bacterium' 16S rRNA sequences obtained from the human gastrointestinal tract by (R. E. Ley et al., 2006), the same work that the DM1 consensus sequence 'hit' when submitted to BLAST (Table 5.3-B).

Despite 'hitting' the same work as the DM1 consensus (R. E. Ley et al., 2006), none of the submitted sequences from Table 5.3-B actually 'hit' the same BLAST sequence as the DM1 consensus (DQ797049.1). Although the DM1 consensus was determined by Classifier to belong to the bacterial Division: *Firmicutes* like 78% of the 1AF/927R amplified sequences, it was placed within the Genus of *Anaeroplasma*. None of the sequences in Table 5.3-B were classified as belonging to *Anaeroplasma* Genus, in fact the majority of the *Firmicute* sequences (57%) belonged to the Family: *Lachnospiraceae. Lachnospiraceae* are part of the *Clostridia* Class, whereas *Anaeroplasma* is a member of the *Mollicutes* Class.

The ClustalW2 alignment of all the highlighted *Firmicute* sequences in Table 5.3-B are represented in the phylogram in Figure 5.3-B. The phylogram supports the BLAST and Classifier results from Table 5.3-B, the DM1 sequence is similar to the others, yet, it is still grouped alone. The other sequences that show similar results in Table 5.3-A are grouped together; DM1 consensus shows some 'relatedness' to the

other sequences but is grouped alone, this is also concordant with the information in Table 5.3-B.

Equivalent to the 1AF/927R primer set, the 1AF/927R NEW primer set was designed to target and 5' extended version of the 'known' sequence, and as with the 1AF/927R primer set it would appear that it did not attain its overall objective. Despite this, it managed to amplify 16S rRNA sequences that belonged to *Firmicutes* and primarily 'hit' sequences that had been acquired from the human gastrointestinal tract, where of course all of the samples from this project were obtained.

5.4.2 3' Extension Sequences

The primer sets 109F/1391R and 109F NEW/1391R primer sets were designed with the purpose of amplifying the 3' end of the known sequence, extending it from ~830bp to ~1280bp (only the 109F/1391R primer set was used to give sequence data). If this was achieved the 'known' sequence could then become better defined, giving us a potential insight into its phylogeny and possibly its functionality and role within the human gastrointestinal tract.

5.4.2.1 Sequences from the 109F/1391R Primer Set

Sequences belonging to the inserted PCR products from the 'selected' 109F/1391R transformants were analysed, with the results placed in Table 5.3-C. Sequences that belonged to *Firmicutes* were highlighted, these sequences were then picked to undergo alignment in ClustalW2 with the 'known' sequence representative: DM1 consensus. The result of this alignment is given in the form of a phylogram in Figure 5.3-C, which depicts the 'relatedness' of the sequences.

Of the 33 sequences found in Table 5.3-C, just 2 (6%) belonged to the *Firmicutes* Division with both belonged to the *Clostridium* Genus. The ClustalW2 alignment of

all the highlighted *Firmicute* sequences in Table 5.3-C are represented in the phylogram in Figure 5.3-C. The phylogram supports the BLAST and Classifier results from Table 5.3-C, the DM1 sequence is grouped alone; the other sequences are both *Clostridia* and group together.

Designed to target and 3' extended version of the 'known' sequence it would appear that the 109F/1391R primer set failed to reach its goal. Unfortunately the primer set amplified relevant *Firmicute* sequences poorly.

5.5 Conclusion

Overall the attempt to amplify and sequence an extended version of the ~830bp novel *Mollicute*-like 16S rRNA gene sequence uncovered in the preliminary work carried out by (Kubs & Musgrave, 2007) was unsuccessful. Without this extended version our aim to further characterise the novel *Mollicute*-like sequence was severely hindered.

Primer specificity is a prime candidate for the reasoning behind our failure, with non-specific amplification prevalent with both primer sets. Despite the high specificity of at least one primer in each set (109F/ 927R or 927R NEW) and our attempts to select, purify and 'screen' PCR products prior to sequencing, only a small amount of relevant sequences were found, the majority of which were amplified by the primer sets 1AF/927R or 927R NEW, which were designed to amplify an extended 5' end of the 'known' sequence.

The primer set 109F/1391R which was designed to target the extended 3' end of the 'known' sequence found only a small percentage of relevant sequences. This is most likely due to the size of the desired amplicon, the large amount of non-specific amplification, the lack of a clear cut amplicon at the desired size and the fact that the entire raw PCR product was cloned into transformants, with many inserts being irrelevant.

It is also important to keep in mind that the segments of the sequences found in this study, that were submitted may have contained part of the 'known' sequence. But, when aligned with other sequences in the BLAST or Classifier databases, the other regions of the sequence aligned better with results less related to the results 'hit' by the 'known' sequence representative, DM1 consensus. This is only a possibility and is less than likely to be true.

The inability to amplify any full-length versions of the *Mollicute*-like sequences was surprising as the 109F/927R primer set gave consistently strong results: an amplicon

at ~830bp (Kubs & Musgrave, 2007). Even though one of these primers (or modifications of: 927R NEW) was used in conjunction with a newly designed primer (1AF or 1391R), we were still unable to characterise a full-length 16S sequence.

The 'consensus' sequence was classified as belonging to the *Mollicute* Genus, *Anaeroplasma*, with 94% confidence, even so, this still suggests this organism could be somewhat more distantly related to the *Anaeroplasmas* than previously thought. It could be that the outer regions of sequence (5' and 3' ends) could have little homology with *Anaeroplasma* and this organism is in fact even more novel than earlier contemplated. With this in mind, the design of our primers may have been misguided as we used *Anaeroplasma* and other sequences that were within their Class and also sequences that were best BLAST 'hits' with the 'consensus' sequence.

(Eckburg et al., 2005) amplified and sequenced a large percentage of *Mollicute* sequences in the human gut microflora using universal primers. (R. E. Ley et al., 2005; R. E. Ley et al., 2006) also found similar sequences in the human and mouse gut. In preliminary work by (Kubs & Musgrave, 2007) these *Mollicute*-like sequences of interest were found in nearly all of the samples from different regions of the human gastrointestinal tract. All of these results suggest that *Mollicutes* are a major component of the human microflora and are there and available to amplify.

The plethora of different environments within the gastrointestinal tract and the diversity of the species that inhabit them, are just some of the factors that contribute to the difficulty associated with studying the gut microflora. This study was metagenomic in nature, and regrettably our task may have been on a too grander scale. Possibly, with more time and labour we would have been able to sequence more inserts, as only a small amount was done, and possibly this would have led to the discovery of the extended version of the 'known' sequence amongst these, allowing us to further characterise the targeted novel *Mollicute*-like sequence and ultimately help define its functionality.

The human gut microflora is poorly explored and ill-defined to date, with more studies carried out with the same intent as this one, the microflora and its functionality will surely become better defined, opening up numerous options in respect to human health.

5.6 Future Work

The primers were the most likely reason that the attempt to amplify an extended version of the novel *Mollicute*-like sequence at both the 5' and 3' ends was ultimately unsuccessful. It would be best for future work to focus on designing and testing better primers at each of the 5' and 3' ends.

Future primer design could involve: finding a larger quantity of relevant sequences and sequences found in the human gut; it might even be interesting to look into some generic primers and modifying them; reducing the size of the extended ends and looking for other primer target regions closer to the 'known' sequence could also be advantageous.

Perhaps, if more sequencing was carried out on a larger scale it could be possible that the 5' end could be deduced, as this study showed that the 'screening' of 1AF/927R or 927R NEW transformants was valid and gave relevant, related results.

The 'screening' method did not work as well for the 109F/1391R transformants, it could be possible that total abandonment of this method could assist in determining the 3' end of the 'known' sequence. Thorough sequencing all of the PCR products inserted into transformants could reveal the targeted 3' extended sequence.

Pyrosequencing of the microflora and searching the sequences for the 'known' sequence could also be possible. Use of probes and hybridization could also assist in finding and defining 'full-length' versions of these *Mollicute*-like sequences.

With the increased interest in defining the human gut microflora, and an ever increasing range of metagenomic techniques, it is only a matter of time before this novel *Mollicute*-like sequence can be determined.

Appendix

1AF/927R Edited Sequences Submitted to BLAST and Classifier

>210808_1AF_927R_2

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGAAGCACTTTTGCGGATTTC
TTCGGATTGAAGCAATTGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACAGTACCG
CATGGTACCGTGTGAAAAACTCCGGTGGTATGAGACGGACCCGCGTCTGATTAGCTAGTT
GGTGGGGTAACGGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCTTGATGCAGCGACGCCGCGTGAGCGATGAAGTATTTCGGTATGTAAAG
CTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGC
CAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCTGGATTTACTAGGTGTAAAGGGA
GCGTAGACGGCATGGCAAGCCAGACGTGAAAGCCCGGGGCTCAACCCCGGGACTGCATTT
GGAACTGTCAGGCTAGAGTGTCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATG
CGTAGATATTAGAGGAACACCAGTGGCGAAAGCGGCTTTCTGGACGATGACTGACGTTGA
GGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCAACGCTGTAAACGA
TGGATACTAGTGTGGGGGGGTCTGACCCCCCTCCGTGCCCGAAGTTAACCACAAATAAAG
ATATCCCTACCTTGGGGAAGTACCCGATCGCCAAGCTGGAAACCTCATGGATGACCG

>210808_1AF_927R_6A

GACGAACGCTGGTGGCGAAGCCTAACACATGCAAGTCGAACGAGAGATAAGGAGCTTGCT
CCTCTGATTGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGA
CAACAGTTGGAAACGACTGCTAATACCGCATAAGCCCACGGCTCGGCATCGAGCAGAGGG
AAAAGGAGCAATCCGCTTTGAGATGGCCTCGCGTCCGATTAGCTAGTTGGTGAGGTAACG
GCCCACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGA
GACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACC
CTGATGCAGCGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGA
GGAAGATAATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGCCAGCAGCCGCGG
TAAAACGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGA
GATCAAGTTGGAAGTGAAATCCATGGGCTCAACCCCATGAACTGCTTTCAAAACTGTTCT
TCTTGAATAGTGC

>210808_1AF_927R_7A

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGAGAGCTAATGATTGAGGA
TTCGTCCAAATGATTTAGTTTCTTAGTGGCGGACGGGTGAGTAACGCGTGAGGAACCTGC
CTCGGAGTGGGGAATAACACAGTGAAAATTGTGCTAATACCGCATAATGCAGTTGAGTCG
CATGGCTCTGACTGCCAAAGATTTATCGCTCTGAGATGGCCTCGCGTCTGATTAGATAGT
TGGCGGGGTAACGGCCCACCAAGTCGACGATCGGTAGCCGGACTGAGAGGTTGACCGGCC
ACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGGC
AATGGGCGCAAGCCTGACCCAGCAACGCCGCGTGAAGGAAGAAGGCTTTCGGGTTGTAAA
CTTCTTTTGTCAGGGACGAATCAATGACGGTACTTGACGAACAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGATTTACTGGGTGTAAAG
GGCGTGTAGGCGGGACTGCAAGTCAGGCGTGAAAACCAGAGGCTCAACCTCTGGCCTGCG
TTTGAAACTGTAGTTCTTGAGTGCTGGAGAGGCAATCGGAATTCCGTGTGTAGCGGTGAA
ATGCGTAGATATACGGAGGAACACCAGTGGCGAAGGCGGATTGCTGGACAGTAACTGACG
CTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAA
ACGATGGATACTAGGTGTGGGGGGGTCTGACCCCCTCCGTGCCGCAGTTAACGCAATAAG
TATCCCCACCTGGGGAGTACGATCGCAAGGTTGAAACTTAAGGAATTGAACGG

>210808_1AF_927R_8A

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGGAGCACTTTTGCGGATTTC
TTCGGATTGAAGCAATTGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACAGTACCG
CATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTT
GGTGGGGTAACGGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCTTGATGCAGCGACGCCGCGTGAGCGATGAAGTATTTCGGTATGTAAAG
CTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGC
CAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGA
GCGTAGACGGCATGGCAAGCCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCATTT
GGAACTGTCAGGCTAGAGTGTCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATG
CGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGATGACTGACGTTG
AGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGAATACTAGGTGTCGGGTGGCAAAGCCATTCGGTGCCGCAGCAAACGCAATAAGTATT
CCACCTGGGGAGTACGTTCGCAAGAATGAACTTTAAGGAATTTGAACGGG

>210808_1AF_927R_9

GACGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGAGAGAGAGGGAGCTTGCTT
CCTTGATCGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGAC
AACAGTTGGAAACGACTGCTAATACCGCATAAGCCCACGACCCGGCATCGGGAAGAGGGA
AAAGGAGCAATCCGCTTTGAGATGGCCTCGCGTCCGATTAGCTAGTTGGTGAGGTAACGG
CCCACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGAG
ACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCC
TGATGCAGCGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGAG
GAAGATAATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGCCAGCAGCCGCGGT
AAAACGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGCG
ATCAAGTTGGAAGTGAAATCCATGGGCTCAACCCATGAACTGCTTTCAAAACTGGTCGTC
TTGAGTAGTGCAGAGGTAGGCGGAATTCCCGGTGTAGCGGTGGAATGCGTAGATATCGGG
AGGAACACCAGTGGCGAAGGCGGCCTACTGGGCACCAACTGACGCTGAGGCTCGAAAGTG
TGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCACACCGTAAACGATGATTACTAGTG
TTGGGAGATTGACCCTCTCAGTGCCGCAGTTAACACAATAAGGTAATCCACCTGGGGAGT
ACGACCGCAAGGTTGAAACTTAAAGGAATTGACCGGA

>210808_1AF_927R_10

GACGAACGCTGGTGGCGCGCCTAACACATGCAAGTCGAACGAGCGAGAGAGAGTTTACTT
TTTCGAGCGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGAC
AACAGTTGGAAACGGCTGCTAATACCGCATAAGCCCACGACCCGGCATCGGGTAGAGGGA
AAAGGAGCAATCCGCTTTGAGATGGCCTCGCGTCCGATTAGCTAGTTGGTGAGGTAACGG
CCCACCAAGGCGACGATCGGTAGCCGGACTGAGAGGCTGAACGGCCACATTGGGACTGAG
ACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCC
TGATGCAGCGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGAG
GAAGATAATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGCCAGCAGCCGCGGT
AAAACGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGAA
GACAAGTTGGAAGTGAAATCCATGGGCTCGACCCATGAACTGCTTTCAAAACTGTTTTTC
TTGAGTAGTGCAGAGGTAGGCGGAATTCCCCGGTGTAGCGGTGGAATGCGTAGATATCGG
GAGGAACACCAGTGGCGAAGCGGCCTACTGGGCACCAACTGACGCTGAGGCTCGAAAGTG
TGGGTAGCAAAACAGGATTAGATACCCTGGTAAGTCCACACTGTAAACGAATGAATTACT
AGGTGTTGGGAGGAATTGACCCCCTT

>210808_1AF_927R_5

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGAGAGCTAATGATTGAGGA
TTCGTCCAAATGATTTAGTTTCTTAGTGGCGGACGGGTGAGTAACGCGTGAGGAACCTGC
CTCGGAGTGGGGAATAACACAGTGAAAATTGTGCTAATACCGCATAATGCAGTTGAGTCG
CATGGCTCTGACTGCCAAAGATTTATCGCTCTGAGATGGCCTCGCGTCTGATTAGATAGT
TGGCGGGGTAACGGCCCACCAAGTCGACGATCGGTAGCCGGACTGAGAGGTTGACCGGCC
ACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGGC
AATGGGCGCAAGCCTGACCCAGCAACGCCGCGTGAAGGAAGAAGGCTTTCGGGTTGTAAA
CTTCTTTTGTCAGGGACGAATCAATGACGGTACTTGACGAACAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGATTTACTGGGTGTAAAG
GGCGTGTAGGCGGGACTGCAAGTCAGGCGTGAAAACCAGAGGCTCAACCTCTGGCCTGCG
TTTGAAACTGTAGTTCTTGAGTGCTGGAGAGGCAATCGGAATTCCGTGTGTAGCGGTGAA
ATGCGTAGATATACGGAGGAACACCAGTGGCGAAGGCGGATTGCTGGACAGTAACTGACG
CTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGCAGTCCACGCTGTAA
ACGATGGATACTATGTGTGGGGGGGTCTGACCCCCTCCGTGCCGCAGTTTAACACAATAA
GTATCCCACCCTGGGGGAGTACGATCGCAAGGTTGAAAACTTAAAGGAATTTGACCGGGA

>210808_1AF_927R_3A

GATGAACGCTGGAGGCGTGCTTAACACATGCAAGTCGAGCGAAGCACTTTTGCGGATTTC
TTCGTATTGACACAATTGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACAGTACCG
CATGGTAACGTGTGAAAAACTCGGGTGGAATGAGAGGGACCCGCGTCTAATTAGCTAGTT
GGTGGGGTAACGGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCTTGATGCAGCGACTCCGCGTGAGCGATGAAGTATTTCGGTATGTAAAG
CTCTATCAACAGGGAAGAAAATGACGGTACCTGA

>040608_1AF_927R_2

GATGAGCGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATGGTCTTAGCTTG
CTAAGGCTGATGGCGACCGGCGCACGGGTGAGTAACACGTATCCAACCTGCCGTCTACTC
TTGGCCAGCCTTCTGAAAGGAAGATTAATCCAGGATGGGATCACGAGTTCACATGTCCGC
ATGATTAAAGGTATTTTCCGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGG
TAACGGCCCACCTAGTCAACGATGGATAGGGGTTCTGAGAGGAAGGTCCCCCACATTGGA
ACTGAGACACGGTCCAAACTCCTACGGGAGGCAGCAGTGAGGAATATTGGTCAATGGGCG
ATGGCCTGAACCAGCCAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAACTTCTTT
TATAAGGGAATAAAGTCGGGTATGCATACCCGTTTGCATGTACTTTATGAATAAGGATCG
GCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGATCCGAGCGTTATCCGGATTTATT
GGGTTTAAAGGGAGCGTAGATGGATGTTTAAGTCAGTTGTGAAAGTTTGCGGCTCAACCG
TAAAATTGCAGTTGATACTGGATGTCTTGAGTGCAGTTGAGGCAGGCGGAATTCGTGGTG
TAGCGGTGAAATGCTTAGATATCACGAAGAACTCCGATTGCGAAGGCAGCCTGCTAAGCT
GCAACTGACATTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTC
CACGCTGTAAACGATGAATACTAGGTGTTGGAGAGCATTGCTCTTCGGTGCCGCAGCAAA
CGCAATAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTTAAAGGAATTGACC
GG

>040608_1AF_927R_4

GATGAACGCTAGCGGGAGGCCTAACACATGCAAGCCGAGCGGTATTGTTTCTTCGGAAAT
GAGAGAGCGGCGTACGGGTGCGTAACACGTGTGCAACCTGCCTTTATCTGGGGGATAGCC
TTTCGAAAGGAAGATTAATACTCCATAATATATTGAACGGCATCGTTTAATATTGAAAGC
TCCGGCGGATAGAGATGGGCACGCGCAAGATTAGCTAGTTGGTGAGGTAACGGCTCACCA
AGGCGATGATCTTTAGGGGGCCTGAGAGGGTGATCCCCCACACTGGTACTGAGACACGGA
CCAGACTCCTACGGGAGGCAGCAGTGAGGAATATTGGTCAATGGGTGCAAGCCTGAACCA
GCCATCCCGCGTGAAGGACGACTGCCCTATGGGTTGTAAACTTCTTTTGTATAGGGATAA
ACCTACCCTCGTGAGGGTAGCTGAAGGTACTATACGAATAAGCACCGGCTAACTCCGTGC
CAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGT
CCGTAGGCGGACTTATAAGTCAGTGGTGAAAGCCTGTCGCTTAACGATAGAACTGCCATT
GATACTGTAAGTCTTGAGTATATTTGAGGTAGCTGGAATAAGTAGTGTAGCGGTGAAATG
CATAGATATTACTTAGAACACCAATTGCGAAGGCAGGTTACCAAGATATAACTGACGCTG
AGGGACGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGCTAACTCGTTTTTGGGCTTTAGGGTTCAGAGACCAAGCGAAAGTGATAAGTTAGCCA
CCTGGGGAGTACGCTCGCAAGAGTGAAACTTAAAGGAATTGACCGG

>040608_1AF_927R_3

GAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAG
CGAAGCACTTTTGCGGATTTCTTCGGATTGAAGCAATTGTGACTGAGCGGCGGACGGGTG
AGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATA
CCGCATAAGCGCACAGTACCGCATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATGGA
CCCGCGTCTGATTAGCTAGTTGGTGGGGTAACGGCCTACCAAGGCGACGATCAGTAGCCG
ACCTGAGAGGGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGC
AGCAGTGGGGAATATTGCACAATGGGGGAAACCTTGATGCAGCGACGCCGCGTGAGCGAT
GAAGTATTTCGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAG
AAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCG
GATTTACTGGGTGTAAAGGGAGCGTAGACGGCATGGCAAGCCAGATGTGAAAGCCCGGGG
CTCAACCCCGGGACTGCATTTGGAACTGTCAGGCTAGAGTGTCGGAGAGGAAAGCGGAAT
TCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTT
TCTGGACGATGACTGACGTTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCC
TGGTAGTCCACGCCCGTAAACGATGAACACTAGGTGTCGGGTGGCAAAGCCATTTCGGTG
CCGCAGCAAACGCAATAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACCTTAA
AGGAATTGACTGGA

>040608_1AF_927R_5

GAGAGTTTGATCATGGCTCAGGATGAACGTTGGTGGCGTGCTTAACACATGCAAGTCGAA
CGAAGCACTCTATTTGATTTTCTTCGGAGATGAAGATTTTGTGATTGAGTGGCGGACGGG
TGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAACAGTTGGAAACGACTGCTAA
TACCGCATAAGCGCACGGTACCGCATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATG
GACCCGCGTCTGATTAGCTAGTTGGTGGGGTAACAGCCTACCAAGGCGACGATCAGTAGC
CGACCTGAGAGGGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAG
GCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGCG
AAGAAGTATTTCGGTATGTAAAGCTCTATCAGCAGGGAAGAAAAAATGACGGTACCTGAC
TAAGAAGCACCGGCTAAATACGTGCCAGCAGCCGCGGTAATACGTATGGTGCAAGCGTTA
TCCGGATTTACTGGGTGTAAAGGGAGCGCAGGCGGTGCGGCAAGTCTGATGTGAAAGCCC
GGGGCTCAACCCCGGTACTGCATTGGAAACTGCCTACTAGAGTGTCGGAAGGGTAAGTGG
AATTCCTAGTGTACCGGTGAAATGCATATATATTTACGGAGGACACCCGTGG

>040608_1AF_927R_7

GATGAACGCTAGCGGCAGGCCTAATACATGCAAGTCGAGGGGCAGCGGGGTAGCAATACT
GCCGGCGACCGGCGCACGGGTGCGTAACACGTATGCAACCTACCTTAAACTGGGAGATAG
CCCGGGGAAACTCGGATTAATACCCCATAACACAGTGGATCCTCATGGTTTTATTGTTAA
AGATTTATCGGCTTAAGATGGGCATGCGTTCGATTAGCTAGTTGGTAAGGTAACGGCTTA
CCAAGGCTACGATCGATAGGGGGTCTGAGAGGATGATCCCCCACACTGGTACTGAGATAC
GGACCAGACTCCTACGGGAGGCAGCAGTAGGGAATATTGGTCAATGGAGGCAACTCTGAA
CCAGCCATGCCGCGTGTAGGAAGAAGGCGTTCTGCGTTGTAAACTACTTTTGATTGGGAA
CGAATGACTCTTGCGAGAGTAGCTGAGTGTACCAATAGAATAAGCCACGGCTAACTACGT
GCCAGCAGCCGCGGTAATACGTAGGTGGCAAACGTTGTCCGGATTTATTGGGTTTAAAGG
GTGCGTAGGCGGCTTGATAAGTCAGTGGTGAAAGCCGATCGCTTAACGATCGAACTGCCA
TTGATACTGTCGGGCTTGAGTATAGATGAGGTAGGCGGAATTGACAGTGTAGCGGTGAAA
TGCATAGATATTGTCAAGAACACCAATTGCGAAGGCAGCTTACTAAGTTTATAACTGACG
CTGAGGCACGAAAGTGCGGGGATCAAAACAGGATTAGATACCCTGGTAGTCCGCACCGTA
AACGATGATCACTCGCTGTGTGCGATAAACAGGTACGCGGCCAAGCGAAAGCGATAAGTG
ATCCACCTGGGGAGTACGCCCGCAAGGGTGAAACTTAAAGGAATTGACCG

>300908_1AF_927R_1

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGAAGCACTTTTGCGGATTTC
TTCGGATTGAAGCAATCGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACGGTACCG
CATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTT
GGTGGGGTAACAGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGTATTTCGGTATGTAAAG
CTCTATCAGCAGGGAAGAAAAAAATGACGGTACCTGACTAAGAAGCACCGGCTAAATACG
TGCCAGCAGCCGCGGTAATACGTATGGTGCAAGCGTTATCCGGATTTACTGGGTGTAAAG
GGAGCGTAGACGGCATGGCAAGCCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCA
TTTGGAACTGTCAGGCTAGAGTGTCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAA
ATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGATGACTGACG
TTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAA
ACGATGAATACTAGGTGTCGGGTGGCAAAGCCATTCGGTGCCGCAGCAAACGCAATAAGT
ATTCCACCTGGGGGAGTACGTTCGCAAGAATGAACTTTAAGGGAATTGGAACGG

>300908_1AF_927R_ 3

GATGAGCGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGGAGCACTTTTGCGGATTTC
TTCGGATTGAAGCAATTGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACAGTACCG
CATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTT
GGTGGGGTAACGGGCTGGGCCAAGGTGACGATCAGTAGCCGACCTGAGAGGGTGACCGGC
CACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
CAATGGGGGAAACCTTGATGCAGCGACGCCGCGTGAGCGATGAAGTATTTCGGTATGTAA
AGCACTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGT
GCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGG
GAGCGTAGACGGCATGGCAAGCCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCAT
TTGGAACTGTCAGGCTAGAGTGTCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAA
TGCGTAGATATTAGGAGGAACACCAGTGGCGAAGCGGCTTTCTGGACGATGACTGACGTT
GAGCTCGAAGCGTGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGAT
GAATACTAGTTGTCCGGTGGCAAAGCCATTTCGGTGCCGCAGCAACGCATAGTATTCCAC
CTTGGGGAGTACGTTCGCAGATTGAAACTTAAAGGAATTGACCG

>300908_1AF_927R_ 5

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGAAGCACTCTATTTGATTTT
CTTCGGAGATGAAGATTTTGTGACTGAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCT
GCCTCATACAGGGGGATAACAGTTGGAAACGACTGCTAATACCGCATAAGCGCACGGTAC
CGCATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAG
TTGGTGGGGTAACAGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGC
CACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
CAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGTATTTCGGTATGTAA
AGCTCTATCAGCAGGGAAGAAAAAAATGACGGTACCTGACTAAGAAGCACCGGCTAAATA
CGTGCCAGCAGCCGCGGTAATACGTATGGTGCAAGCGTTATCCGGATTTACTGGGTGTAA
AGGGAGCGCAGGCGGTGCGGCGAATCTGATGTGAAAGCCCGGGGCTCAACCCCGGTACTG
CATTGGAAACTGTCGTACTAGAGTGTCGGAGGGGTAAGTGGAATTCCTAGTGTAGCGGTG
AAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACGATAACTGA
CGCTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGT
AAACGATGAATACTAGGTGTTGGAGAGCATTGCTCTTCGGTGCCGCAGCAAACGCAATAA
GTATTCCACCTGGGGGAGTACGTTCGCAAGAATGAAACTTAAGGGAAATTGGGCCCGGG

>300908_1AF_927R_ 9

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGAAGCACTTTTGCGGATTTC
TTCGGATTGAAGCAATCGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACGGTACCG
CATGGTACCGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTT
GGTGGGGTAACAGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGTATTTCGGTATGTAAAG
CTCTATCAGCAGGGAAGAAAAAAATGACGGTACCTGACTAAGAAGCACCGGCTAAATACG
TGCCAGCAGCCGCGGTAATACGTATGGTGCAAGCGTTATCCGGATTTACTGGGTGTAAAG
GGAGCGTAGACGGCATGGCAAGCCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCA
TTTGGAACTGTCAGGCTAGAGTGTCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAA
ATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGATGACTGACG
TTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAA
ACGATGAATACTAGGTGTCGGGTGGCAAAGCCATTCGGTGCCGCAGCAAACGCAATAAGT
ATTCCACCTGGGGGAGTACGTTCGCAAGAATGAAACTTTAAAGGGAATTGGGACGGGA

1AF/927R NEW Edited Sequences Submitted to BLAST and Classifier

>200808_1AF.927Rn_2

GATGAACGCTGGCGGCGTGCCTAACACATGCGAGTCGAGCGAAGCGATTTAGCGGAAGTT
TTCGGATGGAAGTTAAATTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTAGAAATGACTGCTAATACCGCATAAGCGCACAGCATCG
CATGATGCAGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTT
GGTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGATCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCCTGATGCAGCAACGCCGCGTGAGTGAAGAAGTATTTCGGTATGTAAAG
CTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGC
CAGCAGCCGCGGTAATACGTAAGGGGCAAGCGTTATCCGGATTTACTGGGTGA

>200808_1AF.927Rn_4A

GACGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGAGCGAGAGAGAGCTTGCTT
TCTCAAGCGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGAC
AACAGTTGGAAACGACTGCTAATACCGCATAAGCCCACGGGTCGGCATCGACCAGAGGGA
AAAGGAGCAATCCGCTTTGAGATGGCCTCGCGTCCGATTAGCTAGTTGGTGAGGTAATGG
CCCACCAAGGCAACGATCGGTAGCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGAG
ACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCC
TGATGCAGCGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGAG
GAAGATAATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGCCAGCAGCCGCGGT
AAAACGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGAA
GACAAGTTGGAAGTGAAATCTATGGGCTCAACC

>200808_1AF.927Rn_5

GATGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAAGCGATTTAGCGGAAGTT
TTCGGATGGAAGTTAAATTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCATACAGGGGGATAACAGTTAGAAATGACTGCTAATACCGCATAAGCGCACAGCATCG
CATGATGCAGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGACTAGCTAGTT
GGTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGATCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCCTGATGCAGCAACGCCGCGTGAGTGAAGAAGTATTTCGGTATGTAAAG
CTCTATCAGAGGGAAGAAAATGACGGTACCTG

>200808_1AF.927Rn_6A

GACGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGAGCGAGAGAGAGCTTGCTT
TCTCAAGCGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGAC
AACAGTTGGAAACGACTGCTAATACCGCATAAGCCCACGGGTCGGCATCGACCAGAGGGA
AAAGGAGCAATCCGCTTTGAGATGGCCTCGCGTCCGATTAGCTAGTTGGTGAGGTAATGG
CCCACCAAGGCAACGATCGGTAGCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGAG
ACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCC
TGATGCAGCGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGAG
GAAGATAATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGCCAGCAGCCGCGGT
AAAACGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGAA
GACAAGTTGGAAGTGAAATCTATGGGCTCAACCCATAAACTGCTTTCAAAACTGTTTTTC
TTGAGTAGTGCAGAGGTAGGCGGAATTCCCGGTGTAGCGGTGGAATGCGTAGATATCGGG
AGGAACACCAGTGGCGAAGGCGGCCTACTGGGCACCAACTGACGCTGAGGCTCGAAAGTG
TGGGTAGCAAACAGGATTAGATACCCTGGTAGTCTACACCGTAAACGATGATTACTAGGT
GTTGGAGGATTGACCCCTTCAGTGCCGCAGTTAACACAATAAGTAATCCACCTGGGGAGT
ACGACCGCAAG

>200808_1AF.927Rn_8

GATGAACGCTGGTTAGAAAGAAAAACACATGCAAGTCGAGCGAAGCGATTTAGCGGAAGT
TTTCGGATGGAAGTAAAATTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTG
CCTCATACAGGGGGATAACAGTTAGAAATGACTGCTAATACCGCATAAGCGCACAGCATC
GCATGATGCAGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGT
TGGTGAGGTAACGGCTCACCAAGGCGACGATCACTAGCCGACCTGAGAGGGTGATCGGCC
ACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCAC
AATGGGGGAAACCCTGATGCAGCAACGCCGCGTGAGTGAAGAAGTATTTCGGTATGTAAA
GCTCTATCAGCAGGGAAGAAAATGACGGTACCTGAC

>040608_1AF_927Rn_14

GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGAAGCACTTTTACAGATTTC
TTCGGAATGAAGTTTTGGTGACTGAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGC
CTCACACAGGGGGATAACAGTTGGAAACGGCTGCTAATACCGCATAAGCGCACAGTACCG
CATGGTACGGTGTGAAAAACTCCGGTGGTGTGAGATGGACCCGCGTCTGATTAGCTAGTT
GGCAGGGTAACGGCCTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGTATTTCGGTATGTAAAG
CTCTATCAGCAGGGAAGAAGAAATGACGGTACCTGACTAAGAAGCACCGGCTAAATACGT
GCCAGCAGCCGCGGTAATACGTATGGTGCAAGCGTTATCCGGATTTACTGGGTGTAAAGG
GAGCGCAGGCGGAAGGCTAAGTCTGATGTGAAAGCCCGGGGCTCAACCCCGGTACTGCAT
TGGAAACTGGTCATCTAGAGTGTCGGAGGGGTAAGTGGAATTCCTAGTGTAGCGGTGAAA
TGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACGATAACTGACGC
TGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAA
CGATGTCAACTGGTTGTTGGACGGCTTGCTGTTCAGTAACGAAGCTAACGCGTGAAGTTG
ACCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTAAAAGGAATTGACGG

>040608_1AF_927Rn_19

ACTGAACGCTGGCGGCATGCCTTACACATGCAAGTCGAACGGTAACAGGTCTTCGGATGC
TGACGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGACAACA
GTTGGAAACGACTGCTAATACCGCATAAGCCCACGGGTCGGCATCGACCAGAGGGAAAAG
GAGCAATCCGCTTTGAGATGGCCTCGCGTCCGATTAGCTAGTTGGTGAGGTAATGGCCCA
CCAAGGCAACGATCGGTAGCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGAT
GCAGCGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGAGGAAG
ATAATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGCCAGCAGCCGCGGTAAAA
CGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGAAGACA
AGTTGGAGGTGAAATCTATGGGCTCAACCCATAAACTGCTTTCAAAACTGTTTTTCTTGA
GTAGTGCAGAGGTAGGCGGAATTCCCGGTGTAGCGGTGGAATGCGTAGATATCGGGAGGA
ACACCAGTGGCGAAGGCGGCCTACTGGGCACCAACTGACGCTGAGGCTCGAAAGTGTGGG
TAGCAAACAGGATTAGATACCCTGGTAGTCCACACCGTAAACGATGATTACTAGGTGTTG
GAGGATTGACCCCTTCAGTGCCGCAGTTAACACAATTAAGTAATCCACCTGGGGAGTACG
ACCGCAAGGTTGAAACTAAAGGAATTGACGG

>040608_1AF_927Rn_16

TCTCAAACTAGGACCGAGTCTAACTTGCGACCGCCGTACGGAAATGTGTACGTTCAGCT
TGCCATTGTCCAGAAGCCTACGACTGCTCACCGCTTGCCCACTCATTATGTAGCCTTGC
ACGGGCTAGCACCCCCTATTGCTCCGCTTTCGAAACGATTATGGCGTATGCTAGATGCC
TACTTTCGCCCCCTAGAAGCCTGGAGCGCGCCTGCCTCGCCGGCTACCGTCTAATCCAT
CAACCACCCTATTTTCGAATGGTTCGGCTGCTAGACATCGACCAGACTCTCCTACTAGT
CGGTGTGACCCTGACTCTGTGCCGGGTCTGAGGATGCCCTCCGTCGTCACCCCTTAAAA
CCTGTTACCCGCTTTCGGACTAGGTCGGTACGGCGCACGTCCTACTTCCGGAAGCCCAA
CATTTGACGAAAACATGCCTTGCTTTTCGGAGAAAGATTATTTCTCTCCAGTACTGCCA
TGGCATTCTTATTCGTGGCCGATTGATGCACGGTCGTCGGCGCCATTATGCATCCCACG
TTCGCAATTAGCCTTAATGACCCGCATTTCGCACGCGTCCGCCAAAACATTCTGTCTCC
ACTTTAGGGGCCCGAGTTGGACCCTTGACGGAAACACTGACGTTCCGACCTCACGCCGT
CTCCCCCTACCTTAAGGCGCACATCGTCACTTTACGCATCTATACGCCTCCTTGTGGCT
ACCGCTTCCGTTAGGGGACCCGGACGTGACTGCGAGTACGTGCTTTCGCACCCCTCGTT
TGTCCTAATCTATGGGACCATCAGGTGCGGGATTTGCTACAGTTGACCAACAACCCAGA
AGTGACTGAGTCATTGCTTCGATTGCGCACTTCAACTGGCGGACCCCTCATGCCGGCGT
TCCAA

>200808_1AF.927Rn_9A

ATTGAACGCTGGCGGCATGCCTTACACATGCAAGTCGAACGGTAACAGGTCTTCGGATGC
TGACGAGTGGCGAACGGGTGAGTAATGCATCGGAACGTGCCCAGTAGTGGGGGATAGCCC
GGCGAAAGCCGGATTAATACCGCATACGACCTGAGGGTGAAAGGGGGGGATCGCAAGACC
TCTCGCTATTGGAGCGGCCGATGTCAGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAG
GCAACGATCTGTAGTTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCC
AGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGC
AATGCCGCGTGCAGGAAGAAGGCCTTCGGGTTGTAAACTGCTTTTGTCAGGGAAGAAATC
TTCTGGGTTAATACCCCGGGAGGATGACGGTATCTGAAGAATAAGCACCGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAG
CGTGCGCAGGCGGCTTTGCAAGACAGAGGTGAAAATCCCCGGGCTCAACCTGGGAACTGC
CTTTGTGACTGCAAGGCTAGAGTACGGCAGAGGGGGATGGAATTCCGCGTGTAGCAGTGA
AATGCGTAGATATGCGGGGGAACACCAATGGCGAATGGCAATCCCCCTGGGGCCTGTACT
GACGCTCATGCACGAAAGCCGTGGGGGAAGCAAACAGGATTAAAATA

109F/1391R Edited Sequences Submitted to BLAST and Classifier

>131008_109F_1391R_B1A

GGGAAATCTGTTGCAAAAAAAACGCACAAGAGAAAGCAGGAAAGATCTAAAATCAACACC
ATAACATTACAATTAAAAGAACTAGGGAAGCAAGAGCAAACAAATTCAAAAGCTAGCAGA
AGGCAAGAAATAACTAAGATCAGAGCAGAACTGAAAGAGATAGAGACACAAAAAACCCTT
CAAAAAAAAATCAATGGGGGGGGAGAGCTGGTTTTTTGAAGAGCTCAAGGGGGTTTTTTT
TACCTCTTAACCCCCCCCCAAAAAAAACAAAAGAGAGAAGAATCAAATACACTCTATAAA
AAATGATAAAGGGGATATCCCCACCTATCCCACAGAAATACAAACTGCATAAAATAATAC
TATAAACACCTCTATGCAAATAAACTAGAAAATCTAGAAGAAATGGATAAATTCCTGGAC
TCATATACCCTCCCAAGACTAAACCAGGAAGAAATTGAATCTCTGAATAGACCAATT

>131008_109F_1391R_B4a

GGGAAATTTGTTGCACTAAACACCCACAAGAGAAAGCAGGAAAGATCTAAAATCAACACC
ATAACATCACAATTAAAAGAACTAGGGAAGCAAGAGCAAACAAATTCAAAAGCTAGCAGA
AGGCAAGAAATAACTAAGATCAGAGCAGAACTGAAAGAGATAGAGACACAAAAACCCTTC
AAAAAAAAATCAATGAATCTGAGAGCTGGTTTTTTGAAATGATCAACAAAATTGATAGAC
CGCTAGCAAGATTAATAAAGAAGAAAAGAGAGAAGAATCAAATAGACTCAATAAAAAATG
ATAAAGGGGATATCACCACCAATCCCACAGAAATACAAACTGCATCAGAGAATACTATAA
ACACCTCTATGCAAATAAACTAGAAAATCTAGAAGAAATGGATAAATTCCTGGACACATA
TACCCTCCCAAGACTAAACCAGGAAGAAGTTGAATCTCTGAATAGACCAATAACAGGCTC
TGAAATTGAGGCAATACTTAATAGCCTACCAACCAAAAAAATTCCAGGACCAGACGGATT
CACAGCCAAATTCTATCAGAGGCACAAAGAGGAGCTGGTAACATTCCTTCTGAAACTATT
CCAATCAATAGAAAAAGAGGGAATCCTCCCTAACTTATTTTATGAGGCCAGCATCATCCT
GATACCAAAGCCTGGCAGAGACACAACAAAAAAGGAGAATTTTAGACCAATACCCCAGAT
GAACATTGATGCAAAATCCTCAATAAAAACACACTGGGCAAACCAAATCCAGCAGCACAT
CAAAAAGCTTATCCACCATGATCAAGTGGGGCTGCATCTTTGGGAAACAAGGCTGGTTCA
AACTTACCACTAATTCACTAAAACATAATTCCATCATGTAACCGGATCATGGACAAACCA
CAGTAGGAGATATCTATTCTCAACGAATGGCAGAAAGGCTTTGACAAATTCCAGAGGCT

>131008_109F_1391R_B5A

CTGTAATCCCAACACTTTCAGAGGCTGAGGCAGGAGGATCACTTGAAGCCAGGGGTTCAA
GACCAGCCTGGGCAACAAAGCAAGACCCTGTCTCTCTCTAAAAACAAACAAACAAAAACC
CCAAAAAAGTGGTCTTTGGATGAGAATTTTGGAATAGTGCTAGACAGAAGGAAGCATGAG
ATAATACTAGTGAGAGAGTAGACTGCAGAGTGAGATAGGAAGATAAGCCACCCTGGACAA
ATCTGTAAAGGTTAGGAATGAACCCAGTGACGGCTGAAGGGCAGGTGACAAAGTGAGGTT
GAGAGATGCGTTTTTTTCTTTAATTTGTTTAAAATGGGAAATGATACCAATTGCTCAGTT
GCCTATTCCCCTTGCCCAGCCTCAGGTCTGAGTATAATTCTCAGGCCTGATGGGGTTTGC
AGTGTGGGCTTTATCTCCTTGTAGTCCTGGGGAGGAATGGAGGAGTAGCGTTGGCGCCAA
TATGGTGGCAATACTTAATAGACCTACCAACTGCTGTTTTTCCTAGACCATACGGATTCA
CAGCC

>131008_109F_1391R_B7

GGGAAATTTGTTGCACTAAACACCCACAAGAGAAAGCAGGAAAGATCTAAAATCAACACC
ATAACATCACAATTAAAAGAACTAGGGAAGCAAGAGCAAACAAATTCAAAAGCTAGCAGA
AGGCAAGAAATAACTAAGATCAGAGCAGAACTGAAAGAGATAGAGACACAAAAAACCCTT
CAAAAAAAATCAATGAATCTGAGAGCTGGTTTTTTGAAATGATCAACAAAATTGATAGAC
CGCTAGCAAGATTAATAAAGAAGAAAAGAGAGAAGAATCAAATAGACTCAATAAAAAATG
ATAAAGGGGATATCACCACCAATCCCACAGAAATACAAACTGCATCAGAGAATACTATAA
ACACCTCTATGCAAATAAACTAGAAAATCTAGAAGAAATGGATAAATTCCTGGACACATA
TACCCTCCCAAGACTAAACCAGGAAGAAGTTGAATCTCTGAATAGACCAATAACAGGCTC
TGAAATTGAGGCAATACTTAATAGCCTACCAACCAAAAAAATTCCAGGACCAGACAGATT
CACAGCCAAATTCTATCAGAGGCACAAAGAGGAGCTGGTAACATTCCTTCTGAAACTATT
CCAATCAATAGAAAAAGAGGGAATCCTCCCTAACTTATTTTATGAGGCCAGCATCATCCT
GATACCAAAGCCTGGCAGAGACACAACAAAAAAAGAGAATTTTAGACCAATACCCCAGAT
GAACATTGATGCAAAAATCCTCAATAAAACACACTGGCAAACCAAATCCAGCAGCACATC
AAAAAGCTTATCCACCATGATCAAGTGGGCTTCATCTTTGGGACACAAAGGCTGGTTCAA
CATACACAAATCACTAAACATAATCCATCATGTAAACAGAACCAATGACAAAAACACGTG
ATTATCTCAACAGATGCAGAAAAGCCTTTGGACAATTTCACAGCGGCTTTCATGCTAAAA
CTATCAATGAACTAGTATGGAATGGAATGTTATTTCAAATTATAGAGCTTATTTTATGGA
CACAAGACTCA

>131008_109F_1391R_B9

GGGAAATTTGTTGCACTAAACACCCACAAGAGAAAGCAGGAAAGATCTAAAATCAACACC
ATAACATCACAATTAAAAGAACTAGGGAAGCAAGAGCAAACAAATTCAAAAGCTAGCAGA
AGGCAAGAAATAACTAAGATCAGAGCAGAACTGAAAGAGATAGAGACACAAAAAACCCTT
CAAAAAAAATCAATGAATCTGAGAGCTGGTTTTTTGAAATGATCAACAAAATTGATAGAC
CGCTAGCAAGATTAATAAAGAAGAAAAGAGAGAAGAATCAAATAGACTCAATAAAAAATG
ATAAAGGGGATATCACCACCAATCCCACAGAAATACAAACTGCATCAGAGAATACTATAA
ACACCTCTATGCAAATAAACTAGAAAATCTAGAAGAAATGGATAAATTCCTGGACACATA
TACCCTCCCAAGACTAAACCAGGAAGAAGTTGAATCTCTGAATAGACCAATAACAGGCTC
TGAAATTGAGGCAATACTTAATAGCCTACCAACCAAAAAAATTCCAGGACCAGACAGATT
CACAGCCAAATTCTATCAGAGGCACAAAGAGGAGCTGGTAACATTCCTTCTGAAACTATT
CCAATCAATAGAAAAAGAGGGAATCCCTCCCTAACTTATTTTATGAGGCCAGCATCATCC
TGATACCAAAGCCTGGCAGAGACACAAACAAAAAAAGAGAATTTTAGACCAAATACCCCA
AGATGAACATTTGATGCAAAAAATCCTCAATAAAAACACACTTGACCA

>131008_109F_1391R_R3A

AGGCCCGGGAATGTATTCACCGCGACATGCTGATTCGCGATTACTAGCAACTCCAGCTTC
ATGTAGGCGAGTTTCAGCCTACAATCCGAACTGAGACAAGTTTTTGAGTTTTGCTCCACC
TCGCGGTCTTGCTTCTCTTTGTACTTGCCATTGTAGCACGTGTGTAGCCCTAGACATAAG
GGGCATGATGATTTGACGTCATCCCCACCTTCCTCCCGGTTAACCCGGGCAGTCTCGCTA
GAGTGCTCAACTAAATGGTAGCAACTAACAATAAGGGTTGCGCTCGTTGCGGGACTTAAC
CCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTTCCTGCCACCGAA
GTGGCTTCCCCAGTTACGGGTAATTCAGGAGATGTCAAGTCTAGGTAAGGTTCTTCGCGT
TGCTTCGAATTAAACCACATGCTCCGCTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGT
TTTAATCTTGCGACCGTACTCCCCAGGCGGAATACTTAATGCGTTAGCGGCGGCACAGAG
GTCATGACAACCCCTACACCTAGTATTCATCGTTTACGGCGTGGACTACCAGGGTATCTA
ATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAGTCAAGAAG

>131008_109F_1391R_R2A

CCCCATGAGAGCCAACACATGCAAACCTCATCTTTGGTTCTGGAAGGGGGGATGGAGTTA
CCCAGTGACTCGAAGGGGCACGAGGGGCCCTGGGTTCTGTTCCTGCCCGTCTCTGATCTG
GGTGTTGATTATAGCGGTGGCCACTTGGTGAGAATTTACTGAGCCACATTCCTATGAGTC
GTGCACTCCTCTATGTAAATGTTATGCTTTGACATTTATATTTTTAAAAAGGCAGGCTAA
GGAAATGGGAGGCAACACAGGAGGCTGAAAAGAAGCCTGAGAGGTGGAGGGAAAAGCCAT
GGAATGTGGCGGCCCAAGAGCAAGAGGAGATGGCGTGAA

>131008_109F_1391R_R4

TCATTTAGGGACTCGGACTAATGATGGTGCTGCTATCTTTAACGTGAGAGTTCCAGAATC
TCTCTGGTTATTGCTCTTCTGGCCAGTGTGAGGGGAAAAACACAGAAGTCTAGGGCAGGA
ATTTCTTCTTAATCGAGTGAGGCAGAAGTTGAGCATACTTCTCACATTCCACAGATGAGA
GTTTAGACACATGGCCTCCTAGCTGCAAGGAGAGTTGGGTCATGCCACAGCTGGGTGGAT
AGCTAGCCATCTCTGTCACAGTCTGTCCTCGAGAACAGTGAGGACTTGCCTGCCCTGGCA
GCGGAGGAGGTTCTCTGGGGGGGTGA

>131008_109F_1391R_R7

TCATTTAGGGACTCGGACTAATGATGGTGCTGCTATCTTTAACGTGAGAGTTCCAGAATC
TCTCTGGTTATTGCTCTTCTGGCCAGTGTGAGGGGAAAAACACAGAAGTCTAGGGCAGGA
ATTTCTTCTTAATCGAGTGAGGCAGAAGTTGAGCATACTTCTCACATTCCACAGATGAGA
GTTTAGACACATGGCCTCCTAGCTGCAAGGAGAGTTGGGTCATGCCACAGCTGGGTGGAT
AGCTAGCCATCTCTGTCACAGTCTGTCCTCGAGAACAGTGAGGACTTGCCTGCCCTGGCA
GCGGAGGAGGTTCTCTGGGGGGGTGACCCCTTATTCTGTGTCTCCCTAACCTTGGCTTTG
CCCTTCAACATTGCTTGCCTGTCATACTCCACAACCACCGACAGGGTGTTGAGAGTCTAC
CCTCCTTAGATGCTGACTCGGGCTTGAGCAATCATGGGCTTTTGCAGTCCAGCCTTGCAA
TTCTGTTGGTGATAACAATTCCTTCAGAAACATAGTAGGCTTCATCTTTAGTCAGTTCCT
TGAGCCGTGACCACATCCAAAGTTCCTTTCTTGACAGAGTTATGAAGTCTATCTTCCATA
CCCTTCTGTTTTCCCATTCATTTAAAATGAAAGTGGCTGCGTGCAGTGACTCACGCCCTG
TAAATCCCCAGCACTTTTGGGAGGCTGAAGGCAGGTG

>131008_109F_1391R_R8A

GAGCGCGGACTGCACCTGGGGCGAGTGCCACGGCCTTCCGTGGAGGGGCTCACTTCACCT
CCCAGGACCCCTGCGAGGAAGAGACCCGAGGAGCCATTTTACTGCCTGGGAGATGGGAAT
TCACACGTTATGGCACTGTCCCAAGATCACACGACGGGCAAGCAACAGCACCATGTTCCC
ACCCAGGCCACCTGCACCCCACAGTCAAGAAACTGAGCCTCCTGCCTCCCGGGCCGGATC
CCTAGACCTCCTCCTCATGCCTGGCCCTAGAGGCCCCCAGGACTTGCCACCCCTTCCTGG
AGAAGACGTTCTCTGGCGGGGTTCCAGGCCCTCCTTACCTGACAAAAAATGGGCCGGGTG
CATCAAAGGTGACAG

>131008_109F_1391R_R5

GTCTCATTGAAGAGTAAAAGGGGTGGAAAAGAGGATGCTAGGAAACTCTCCGCCTTACAC
ATTGGCATAAATTAAAGGTTGTAATGTGCTTCTGCATGTAGAGAATTATTGAACACATTT
TCTTAGTAGAACAGTCTGTAACAACCATGTCAAATAGATTATATATTATAGCTGAGTTAC
TCAACTCATGTAGCAAGGAGATGGCAGTGCTTTTAACTGCAACCCGCTTCCAAGTTTTGA
TAACCTGTTTTTGTGCTTTAGGACACACTGATAATCTTTGGTAGACCCGGATTTTTTTTT
CAAGGCCAATGTTATTTACAAGCAGAGTTAGCAAATAGTTAATGAGCTATAATTTGAAGG
TAAATTGAAAAAAATCTTGCCTCACGTGAATTCCATTATTAATGTTGAGGTTAATAGGAA
AATAACAGATGCTTTTTTTTTTGAGACGGAGTCTCTCTGTCGCCCAGGCTGGAGTGCA

>131008_109F_1391R_R9

AAAATGGGGGGCCTAAATATAGTGCCATGAAAGGTACTGAAATATGAAACCGTTTCCTTC
TGCAGTCTTTCTCTCTCTTGACCTGTCATGGTGTTTTTATTTGATATTTCATTTCTTGTT
CCCTCAGACATGGGGATATTTGCAGAGGGAGTGCAGAGACACCCAGTGGGGGCTCTGTCC
CCTGACTCAGTGCATGGCCCACTGGCTTCTGAGTTCTACCTCCCCTGAAGTGCTGCACCC
CAGCTGGAGACAGGGAGATCCACCTCCTCATTCCCACAGGCCACCGCCCCCACCCACAGC
AGAAGGGGAAGAATCAAGGAGTGGCAACCTCCACATCAGGATGCCCTTGGGACCTGGATC
AGTATTGGCAAGAGTCTGTCTGTTTTGAATGAAGTGTTTTGAGTTTTTCTAGCAACTTGG
CTGGTAGCTACTGTTTCTAGCAGCCATCAGCAGACCAGACAGTAGCACATCAGTAGGGGA
TCCACTTGCATGCATCTCCTTGAGCTCACCATTAGGGGTTGCATGCAAGCTTCAATGACA
CAACCCATCATGTGAAAAAGTTAAAAAGTTTTTAGT

>270108_109F_1391R _4

ATTTGTATCAGGTTGTTAGTTGTCTTATTTCTGGGCGACGAGTTCCTCGTCTATTCTAGA
TACAAGTCCCTTCCAGATACGGGCATCGCAGACAGTTTCACCTGCTCAGGTGCTTGCCTA
TTCGTTCTCTTAACCGCACCTTTCGAAGGGCAAACATTTTTAACTGCGATAAAGGCTGAT
TGACCAATATTTTCTTTTATTGTAAATAACTTTCTCATTCTTGATATTGTTATTTTGTGT
TCCCTCCTTTTCTTCTGCTCATTGCAGCTGGAGACTTATCCGTTGTATTGATATTCTCAA
AGAGCCAGATTTCGGCTTCATCCGTTTTCTCTGTTGTTTTTCTGTTTTCTGTGTCACTGC
TTTTTGCTCTGATCTCTGTTATTCACTTTCTTTCTGCTTGCTCTGGATTTAATTTGCTCT
TTTTTCCCCACTTTCCTACAGTGAGAGCTGAGCTCATTGATGGGAGGCACCTTCCTTTGC
TAAACAGGCATTTTCCACCACGTACATCCTGCCCTTCTCGCCCTCCTCTCCCAAGACCCT
GAGTGTGAGGCGGGTCCCGCCAGCCCGAGACCCTGAGTGTGAGACGGGTCCCGCCAGCCC
CGAGACCCCTGAGTGTGAGACGGGTCCCGCCAGCCCCGAGACCCTGAGTGTGTGACGGCT
CCCGCCAGCCCCGAGACCCCACGTGCATGTGGCCACATGTGTGACAGGTCCCGCCCAGCC
CCTCGGGCTCCATCCCTCGTTCGAAGAGGACA

>270108_109F_1391R _5

TGGTTTGTTACACACATAATAAGGGGTCCTCAGCAGCAGCAGCAGCAGCAGCAGCACCCC
CAGGGGCTGTTAGAAATGCACATTCTTGGGCCCCACAGCAGACCAACTGAATCTGAATCC
TGGGGTGGGCCCAGCCATTTGGCTTTCAACAGGGGATGCTGCTGCCACTGGAGTTTCAGA
CCCTGCGCTATCTCATTTCATCTCCACGGCCACAGAGCAACCTTGTTATTATCATGGTCC
TCATTTTACAAACGAGCAACTGAGCCACATGGAGGTTTTTGTTTTGTTCAACAGAGGCAA
ATGACCTCATATCAGAGAGGGCACACAAACTCATATCCGAGACTTACATCAAATAAACTC
AACAGGCTTTCATCAAAGGTAGAGCATTCACTATGCTTAATAGTGAATGAAGAAGGAAGA
GGAGGAGGCAAGGGGACGAAGGGAATGAGTTGGGGAATAGCAGAGACTATGGGGGTCTCA
GGGTGGCACAGTCCACTCTCCCACACTCAACAAGTTGCAGCAAGTTCACTGTCACCCACC
CTCGGAGTTGGGCCCCCAAGGAACTCTTGAAGACTTAAAGGGCAACTCACTCCATTATTT
TAATTCTCATATTTACTAAGGCAAGAAAAACGGTTAACACATTTTTCAAACGTTTCAAAC
AAAAATATATTACTTACAAGAAATAAAATAAACCTATGGCAGTTTATTTTAGAAAACTTC
CAGCAAGCGACCTTGACTCAAATCTGTCTCCCTGGGATGAAAAACACCCATATACAATTA
TTCCCCCCTGAATTTCCTAAAATACTTCTTCTTCTGACTGGGTTAGGACCCCCCCCCGCC
AAAGGAGAACCACGCGTTACTTCAGCCCGTTA

>270108_109F_1391R _5A

TGGTTTGTTACACACACTTTTTCAGGGGTCCTCAGCAGCAGCAGCAGCAGCAGCAGCACC
CCCAGGGGCTGTTAGAAATGCACATTCTTGGGCCCCACAGCAGACCAACTGAATCTGAAT
CCTGGGGTGGGCCCAGCCATTTGGCTTTCAACAGGGGATGCTGCTGCCACTGGAGTTTCA
GACCCTGCGCTATCTCATTTCATCTCCACGGCCACAGAGCAACCTTGTTATTATCATGGT
CCTCATTTTACAAACGAGCAACTGAGCCACATGGAGGTTTTTGTTTTGTTCAACAGAGGC
AAATGACCTCATATCAGAGAGGGCACACAAACTCATATCCGAGACTTACATCAAATAAAC
TCAACAGGCTTTCATCAAAGGTAGAGCATTCACTATGCTTAATAGTGAATGAAGAAGGAA
GAGGAGGAGGCAAGGGGACGAAGGGAATGAGTTGGGGAATAGCAGAGACTATGGGGGTCT
CAGGGTGGCACAGTCCACTCTCCCACACTCAACAAGTTGCAGCAAGTTCACTGTCACCCA
CCCTCGGAGTTGGGCCCCCAAGGAACTCTTGAAGACTTAAAGGGCAACTCACTCCATTAT
TTTAATTCTCATATTTACTAAGGCAAGAAAAACGGTTAACACATTTTTCAAACGTTTCAA
ACAAAAATATATTACTTACAAGAAATAAAATAAACCTATGGCAGTTTATTTTAGAAAACT
TCCAGCAAGCGACCTTGACTCAAATCTGTCTCCCTGGATGAAAAACACCCATATACAATT
ATTCCCCCTGAATTTCCTAAATACTCTTCTTCTGACTGGGTAGGACCCCCCCCGCCCAAG
GAGGAACCACGCGTTACTCAGCCCGT

>270109_109F_1391R _6

TTCTCCTTGGCGGGGGGGTCCTACCCAGTCAGAAGAAGAGTATTTAGGAAATTCAGGGGG
AATAATTGTATATGGGTGTTTTTCATCCAGGGAGACAGATTTGAGTCAAGGTCGCTTGCT
GGAAGTTTTCTAAAATAAACTGCCATAGGTTTATTTTATTTCTTGTAAGTAATATATTTT
TGTTTGAAACGTTTGAAAAATGTGTTAACCGTTTTTCTTGCCTTAGTAAATATGAGAATT
AAAATAATGGAGTGAGTTGCCCTTTAAGTCTTCAAGAGTTCCTTGGGGGCCCAACTCCGA
GGGTGGGTGACAGTGAACTTGCTGCAACTTGTAGAGTGTGGGAGAGTGGACTGTGCCACC
CTGAGACCCCCATAGTCTCTGCTATTCCCCAACTCATTCCCTTCGTCCCCTTGCCTCCTC
CTCTTCCTTCTTCATTCACTATTAAGCATAGTGAATGCTCTACCTTTGATGAAAGCCTGT
TGAGTTTATTTGATGTAAGTCTCGGATATGAGTTTGTGTGCCCTCTCTGATATGAGGTCA
TTTGCCTCTGTTGAACAAAACAAAAACCTCCATGTGGCTCAGTTGCTCGTTCGTAAAATG
AGGACCATGATAATAACAAGGTTGCTCTGTGGCCGTGGAGATGAAATGAGATAGCGCATG
GTCTGAAACTCCAGTGGCAGCAGCATCCCCTGTTGAAAGCCAGATGGCTGGGCCCACCCC
AGGATTCAGATTCAGTTGGTCTGCTGTGGGGGCCCAAGAATGTGCATTTCTAACAGCCCC
TGGGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGAGGACCCCTGAGAGAGTGTGTGTATC
AAAGCATACACACTGCCCGTCAAACATGTCGAACTCCTAGCACACTGGCTGGCGTTACCT
AGTTGGAATGCCGATGCTGTCGGGTTCAC

>270109_109F_1391R _20A

TTCTCCTTGGCGGGGGGGTCCTACCCAGTCAGAAGAAGAGTATTTAGGAAATTCAGGGGG
AATAATTGTATATGGGTGTTTTTCATCCAGGGAGACAGATTTGAGTCAAGGTCGCTTGCT
GGAAGTTTTCTAAAATAAACTGCCATAGGTTTATTTTATTTCTTGTAAGTAATATATTTT
TGTTTGAAACGTTTGAAAAATGTGTTAACCGTTTTTCTTGCCTTAGTAAATATGAGAATT
GAAATAATGGAGTGAGTTGCCCTTTAAGTCTTCAAGAGTTCCTTGGGGGCCCAACTCCGA
GGGTGGGTGACAGTGAACTTGCTGCAACTTGTAGAGTGTGGGAGAGTGGACTGTGCCACC
CTGAGACCCCCATAGTCTCTGCTATTCCCCAACTCATTCCCTTCGTCCCCTTGCCTCCTC
CTCTTCCTTCTTCATTCACTATTAAGCATAGTGAATGCTCTACCTTTGATGAAAGCCTGT
TGAGTTTATTTGATGTAAGTCTCGGATATGAGTTTGTGTGCCCTCTCTGATATGAGGTCA
TTTGCCTCTGTTGAACAAAACAAAAACCTCCATGTGGCTCAGTTGCTCGTTTGTAAAATG
AGGACCATGATAATAACAAGGTTGCTCTGTGGCCGTGGAGATGAAATGAGATAGCGCAGG
GTCTGAAACTCCAGTGGCAGCAGCATCCCCTGTTGAAAGCCAGATGGCTGGGCCCACCCC
AGGATTCAGATTCAGTTGGTCTGCTGTGGGGCCCAAGAATGTGCATTTCTAACAGCCCCT
GGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGAGGACCCCTGAGAGAGTGTGTGTAACAA
ACCATACACACCGCCCGTCAAAC

>270109_109F_1391R _30A

TTTTCCTTGACAGGGAGGTTGTTGTGGGCCTCAAGACTAATGGGAATGGATTCCCCTCCT
TTGATTGTATCTCCGTGTTTAAGGGCTTCATTAACATTAAAAAGTAATGTAAAATTTTCT
TACATTTATCATGGGGAATTGGATGTAATTGGGTAAAGATGCTTTTGGAATATGTTTATC
CGACAGTGAAACCACTGGGCACAGTGCCCATTGCTGAGCCTTACATCTGCCCGGAAAACA
CATAATGTGTGGGTCTCCTCCTTCCACTGATGCAGAGTTCCTTGTGTCCTCTTCTAAC

>270109_109F_1391R _44

TGGTTTGTGACAGGGGGGTTCTCAGGGGTCCTCAGAAGAAGCAGCAGCAGCAGCAGGACC
CCCAGGGGCTGTTAGAAATGCACATTCTTGGGCCCCACAGCAGACCAACTGAATCTGAAT
CCTGGGGTGGGCCCAGCCATCTGGCTTTCAACAGGGGATGCTGCTGCCACTGGAGTTTCA
GACCCTGCGCTATCTCATTTCATCTCCACGGCCACAGAGCAACCTTGTTATTATCATGGT
CCTCATTTTACAAACGAGCAACTGAGCCACATGGAGGTTTTTGTTTTGTTCAACACAGGC
AAATGACCTCATATCAGAGAGGGCACACAAACTCATATCCGAGACTTACATCAAATAAAC
TCAACAGGCTTTCATCAAAGGTACAGCATTCACTATGCTTAATAGTGAATGAAGAAGGAA
GAGGAGGAGGCAAGGGGACGAAGGGAATGAGTTGGGGAATAGCAGAGACTATGGGGGTCT
CATGGTGGCACAGTCCACTCTCCCACACTCTACAAGTTGCAGCAAGTTCACTGTCACCCA
CCCTCGGAGTTGGGCCCCCAAGGAACTCTTGAAGACTTAAAGGGCAACTCACTCCGTTAT
TTTAATTCTCATATTTACTAAGGCAAGACAAACGGTTAACACATGTTTTCAAACGTTTCA
CACAAGAATATATTACTTACAAGAAATAAAATAAACCTATGGCAGTTTATTTTAGAAAAC
TTCCAGCAAGCGACCTTGACTCATATCTGTCTCCCTGGAT

>270109_109F_1391R _45A

TTTTAGATTTTAAGCAAAAATTTTAAAGATAAATCCATTTTTCTCTCCCACCCCCAACGC
CATCTCCACTGCATCCGATCTCATTATTTCGGTGGTTGCTTGGGGGTGAACAATTTTGTG
GCTTTTTTTCCCCTATAATTCTGACCCGCTCAGGCTTGAGGGTTTCTCCGGCCTCCGCTC
ACTGCGTGCACCTGGCGCTGCCCTGCTTCCCCCAACCTGTTGCAAGGCTTTAATTTTTGC
AACTGGGACCTGCTCGCAGGCACCCCAGCCCTCCACCTCTCTCTACATTTTTGCAAGTGT
CTGGGGGAGGGCACCTGCTCTACCTGCCAGAAATTTTAAAACAAAAACAAAAACAAAAAA
ATCTCCGGGGGCCCTCTTGGCCCCTTTATCCCTGCACTCTCGCTCTCCTGCCCCACCCCG
AGGTAAAGGGGGCGACTAAGAGAAGATGGTGTTGCTCACCGCGGTCCTCCTGCTGCTGGC
CGCCTATGCGGGGCCGGCCCAGAGCCTGGGCTCCTTCGTGCACTGCGAGCCCTGCGACGA
GAAAGCCCTCTCCATGTGCCCCCCCAGCCCCCCTGGGCTGCGAAGCTGGTCAAGGAGCCG
GGGCTGCGGCTGCTG

>270109_109F_1391R _48

CAAGGCCCGGGAACGTAAAAACCGCGACATGCTGATTCGCGATTACTAGCAACTCCAGCT
TCATGTAGGCGAGTTTCAGCCTACAATCCGAACTGAGACAAGTTTTTGAGTTTTGCTCCA
CCTCGCGGTCTTGCTTCTCTTTGTACTTGCCATTGTAGCACGTGTGTAGCCCTAGACATA
AGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCCGGTTAACCCGGGCAGTCTCGC
TAGAGTGCTCAACTAAATGGTAGCAACTAACAATAAGGGTTGCGCTCGTTGCGGGACTTA
ACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTTCCTGTCCCCG
AAGGAACTTCTCCGATTAAGGGTAGTGCAGGAGATGTCAAGTCTAGGTAAGGTTCTTCGC
GTTGCTTCGAATTAAACCACATGCTCCGCTGCTTGTGCGGGCCCCCGTCAATTCCTTTGA
GTTTTAATCTTGCGACCGTACTCCCCAGGCGGAATACTTAATGCGTTAGCGGCGGCACAG
AGGTCATGACAACCCCTACACCTAGTATTCATCGTTTACGGCGTGGACTACCAGGGTATC
TAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAGTCCAGAGAATCGC
CTTCGCCACTGGTGTTCTTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCAT
TCTCCTCTCCTGCACTCTAGACTTCCAGTTTGAAAATGCAGCACCCCAAGTTGAGCCCCG
GGTATTTCACATCTCACTT

>270109_109F_1391R _52A

TTCTCCTTGGCGGGGGGGTCCTACCCAGTCAGAAGAAGAGTATTTAGGAAATTCAGGGGG
AATAATTGTATACGGGTGTTTTTCATCCAGGGAGACAGATTTGAGTCAAGGTCGCTTGCT
GGAAGTTTTCTAAAATAAACTGCCATAGGTTTATTTTATTTCTTGTAAGTAATATATTTT
TGTTTGAAACGTTTGAAAAATGTGTTAACCGTTTTTCTTGCCTTAGTAAATATGAGAATT
AAAATAATGGAGTGAGTTGCCCTTTAAGTCTTCAAGAGTTCCTTGGGGGCCCAACTCCGA
GGGTGGGTGACAGTGAACTTGCTGCAACTTGTAGAGTGTGGGAGAGTGGACTGTGCCACC
CTGGA

>270109_109F_1391R _53A

CGCGTGGTTCTCCTTGGCGGGGGGGTCGGAAGCAGTCAGAAGAAGAGTATTTAGGAAATT
CAGGGGGAATAATTGTATATGGGTGTTTTTCATCCAGGGAGACAGATTTGAGTCAAGGTC
GCTTGCTGGAAGTTTTCTAAAATAAACTGCCATAGGTTTATTTTATTTCTTGTAAGTAAT
ATATTTTTGTTTGAAACGTTTGAAAAATGTGTTAACCGTTTTTCTTGCCTTAGTAAATAT
GAGAATTAAAATAATGGAGTGAGTTGCCCTTTAAGTCTTCAAGAGTTCCTTGGGGGCCCA
ACTCCGAGGGTGGGTGACAGTGAACTTGCTGCAACTTGTAGAGTGTGGGAGAGTGGACTG
TGCCACCCTGAGACCCCCATAGTCTCTGCTATTCCCCAACTCATTCCCTTCGTCCCCTTG
CCTCCTCCTCTTCCTTCTTCATTCACTATTAAGCATAGTGAATGCTCTACCTTTGATGAA
AGCCTGTTGAGTTTATTTGATGTAAGTCTCGGATATGAGTTTGTGTGCCCTCTCTGATAT
GAGGTCATTTGCCTCTGTTGAACAAAACAAAAACCTCCATGTGGCTCAGTTGCTCGTTTG
TAAAATGAGGACCATGATAATAACAAGGTTGCTCTGTGGCCGTGGAGATGAAATGAGATA
GCGCAGGGTCTGAAACTCCAGTGGCAGCAGCATCCCCTGTTGAAAGCCAGATGGCTGGGC
CCACCCCAGGATTCAGATTCAGTTGGTCTGCTGTGGGGCCCAAGAATGTGCATTTCTAAC
AGCCCCTGGGGGTGCTGCTGCTGCTGCTGCTGCTGAGGACCCTGAGAGAGTGTGTGT
AACAAACCATACACACCGCCCGTCAAAC

>270109_109F_1391R _58

TTCTCCTTGGCGGGGGGGTCCTACCCAGTCAGAAGAAGAGTATTTAGGAAATTCAGGGGG
AATAATTGTATATGGGTGTTTTTCATCCAGGGAGACAGATTTGAGTCAAGGTCGCTTGCT
GGAAGTTTTCTAAAATAAACTGCCATAGGTTTATTTTATTTCTTGTAAGTAATATATTTT
TGTTTGAAACGTTTGAAAAATGTGTTAACCGTTTTTCTTGCCTTAGTAAATATGAGAATT
AAAATAATGGAGTGAGTTGCCCTTTAAGTCTTCAAGAGTTCCTTGGGGGCCCAACTCCGA
GGGTGGGTGACAGTGAACTTGCTGCAACTTGTAGAGTGTGGGAGAGTGGACTGTGCCACC
CTGAGACCCCCATAGTCTCTGCTATTCCCCAACTCATTCCCTTCGTCCCCTTGCCTCCTC
CTCTTCCTTCTTCATTCACTATTAAGCATAGTGAATGCTCTACCTTTGATGAAAGCCTGT
TGAGTTTATTTGATGTAAGTCTCGGATATGAGTTTGTGTGCCCTCTCTGATATGAGGTCA
TTTGCCTCTGTTGAACAAAACAAAAAACCTCCATGTGGCTCAGTTGCTCGTTTGTAAAAT
GAGGACCATGATAA

>270109_109F_1391R _58A

TTCTCCTTGGCGGGGGGGTCCTACCCAGTCAGAAGAAGAGTATTTAGGAAATTCAGGGGG
AATAATTGTATATGGGTGTTTTTCATCCAGGGAGACAGATTTGAGTCAAGGTCGCTTGCT
GGAAGTTTTCTAAAATAAACTGCCATAGGTTTATTTTATTTCTTGTAAGTAATATATTTT
TGTTTGAAACGTTTGAAAAATGTGTTAACCGTTTTTCTTGCCTTAGTAAATATGAGAATT
AAAATAATGGAGTGAGTTGCCCTTTAAGTCTTCAAGAGTTCCTTGGGGGCCCAACTCCGA
GGGTGGGTGACAGTGAACTTGCTGCAACTTGTAGAGTGTGGGAGAGTGGACTGTGCCACC
CTGAGACCCCCATAGTCTCTGCTATTCCCCAACTCATTCCCTTCGTCCCCTTGCCTCCTC
CTCTTCCTTCTTCATTCACTATTAAGCATAGTGAATGCTCTACCTTTGATGAAAGCCTGT
TGAGTTTATTTGATGTAAGTCTCGGATATGAGTTTGTGTGCCCTCTCTGATATGAGGTCA
TTTGCCTCTGTTGAACAAAACAAAAACCTCCATGTGGCTCAGTTGCTCGTTTGTAAAATG
AGGACCATGATAATAACAAGGTTGCTCTGTGGCCGTGGAGATGAAATGAGATAGCGCAGG
GTCTGAAACTCCAGTGGCAGCAGCATCCCCTGTTGAAAGCCAGATGGCTGGGCCCACCCC
AGGATTCAGATTCAGTTGGTCTGCTGTGGGGCCCAAGAATGTGCATTTCTAACAGCCCCT
GGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGAGGACCCCTGAAGAGAGTGTGTGTAAAC
AAAACCATACCACACCCGCCCCGTCAAACA

>270109_109F_1391R _59

GCAAGAAGTGAATAACACAGAAATCAATGATGAAAGCACTAGAAGACAATTGAGTTTATC
AGAACTGCAAAATATTGCTGAGTGTGGATTGCTCTGAAATCTGAAAACATGACTTGTGAA
TTGCTTCTATTCAAAATGCAGACACAATCAATGCCGGGTGTTGGTTTACTTGTTTCCCAT
TTTGCAACCCTCTTTTCTAGGCAAAAGGTATCCAAACTATGCAGAGCCACAGAATCTAAC
AGATGTCTCTCTATTCCTTCTCCTAGAACTCTCAGGGGATCCAGAACTGCAGCTGGTCCT
CGCTGGGCTGTTCCTGTCCATCTGCCTGCTCATGGTGCTGGGGAACCTGCTCATCATCCT
GGCCATCAGCCCTGGCTCCCACCTCCACACCCCCATGTACTTCTTCCTCTCCAACCTGTC
CTTGCCTGACATCAGTTTCACCTCCACCACAGTCCCCAAGATGATTGTGGACATCCAATC
TCACAGCAGTCATCTCCTATGCAGGCTGCCTGACTCAGATGTCTCTCTTTGTCATTTTTG
GAGGCATGGAAGAGAGGCATGCTCCTGAGTGTGATGGCCTATGACTGGTTTGTAGCCATC
TGTCACCCTCTATATCATTCAGCCATCATGAACCCGTGTTTCTGTGGCTTCCTAGTTTTG
TTGTCTTTTTTTTTCTCAGTCTTTTAGATGCCCAGCTGCACAACTTGATTGCCTTACAAG
TGACCTGCTTCAAGGATGTG

>270109_109F_1391R _59A

GCAAGAAGTGAATAACACAGAAATCAATGATGAAAGCACTAGAAGACAATTGAGTTTATC
AGAACTGCAAAATATTGCTGAGTGTGGATTGCTCTGAAATCTGAAAACATGACTTGTGAA
TTGCTTCTATTCAAAATGCAGACACAATCAATGCCGGGTGTTGGTTTACTTGTTTCCCAT
TTTGCAACCCTCTTTTCTAGGCAAAAGGTATCCAAACTATGCAGAGCCACAGAATCTAAC
AGATGTCTCTCTATTCCTTCTCCTAGAACTCTCAGGGGATCCAGAACTGCAGCTGGTCCT
CGCTGGGCTGTTCCTGTCCATCTGCCTGCTCATGGTGCTGGGGAACCTGCTCATCATCCT
GGCCATCAGCCCTGGCTCCCACCTCCACACCCCCATGTACTTCTTCCTCTCCAACCTGTC
CTTGCCTGACATCAGTTTCACCTCCACCACAGTCCCCAAGATGATTGTGGACATCCAATC
TCACAGCAGTCATCTCCTATGCAGGCTGCCTGACTCAGATGTCTCTCTTTGTCATTTTTG
GAGGCATGGAAGAGAGGCATGCTCCTGAGTGTGATGGCCTATGACTGGTTTGTAGCCATC
TGTCACCCTCTATATCATTCAGCCATCATGAACCCGTGTTTCTGTGGCTTCCTAGTTTTG
TTGTCTTTTTTTTTCTCAGTCTTTTAGATGCCCAGCTGCACAACTTGATTGCCTTACAAG
TGACCTGCTTCAAGGATGTG

>270109_109F_1391R _61A

CCTTCACGATGATAATGAAAGAAAATGGGGGAGAAGATGAATTACTCTGATTTTAGAAGT
GTAAAAAATGATGAAGTTGTTGCTCTGTAATCAAAAAAATGACTTGTGATTTGCTTCTAT
TCAAAATGCGTACACAATCAATGCCGGGTGTTGGTTTACTTGTTTCCCATTTTGCAACCC
TCTTTTCTAGGCAAAAGGTATCCAAACTATGCAGAGCCACAGAATCTAACAGATGTCTCT
CTATTCCTTCTCCTAGAACTCTCAGGGGATCCAGAACTGCAGCTGGTCCTCGCTGGGCTG
TTCCTGTCCATCTGCCTGCTCATGGTGCTGGGGAACCTGCTCATCATCCTGGCCATCAGC
CCTGGCTCCCACCTCCACACCCCCATGTACTTCTTCCTCTCCAACCTGTCCTTGCCTGAC
ATCAGTTTCACCTCCACCACAGTCCCCAAGATGATTGTGGACATCCAATCTCACAGCAGT
CATCTCCTATGCAGGCTGCCTGACTCAAATGTCTCTCTTTGTC

>270109_109F_1391R _62

CGCTGGAGGGAAGGAAAGAGAGGCGCGTCCCTGGCCAGGGATGCTGGGAGTTATGGTCTC
TTAACGGTTTAAAGCGATGGCCCCCGGCCTGCAGACCACAATCCCAGCAGCCACCGGGCT
TCGAGGCGGTGTGTAGCACTGAAGGGAAGGATAGGGAGGTGCGTCCTAGCCAGGCGTGCT
GGGAGTTATGGTCTCTTAACAGTTTCCAGTCAGTTGGTCCCAGGACTACCTGACTACAAT
CCCAGTATGCGTTGGGCTTGGGGGCGGTGCGCAGCCCTAGAGGAAGGATCGGGACGGCGG
GTACCTCGCAAAGCATCCTGGGAGTCATAGTCCTTTCAGTATTTCCAGCCCATTGGTCGC
GAGGCTAACGGACTACAATCTCAGCATGCGCTGGGTTTGGGGGCGGTGTGTAGTATGGAA
GCGAAGGATAGGGAGGCGCGTCCCTAGCTAGGAGTGCTGGGAGTTATGGTGTCTTAACGG
TTTCCAGCCCATTGGTCGCCGACCTGCTAACTACAAAACCAGCATGCGCTGTCTGTCCTA
CCCCGTGGTGCGCAGCCCTGGAGGGAGGGACAGGGCGGTGTGGACTCGTCCTTTCCTAAG
CGATGCCACATGCTGATTCTGTGCCACCCCCTCGCCAAGGGAGTCCGCAGAAGGACTTGA
GGGGCAGGTCTAGGCTGGGCGATGAGGACGGTGTGACCCTGCGAAGTGCACCTCCCTTGC
TCAAATCGGAGGGGTCTGGTCCCTCACTGAACAGCCCGCTGAACATCTCGGTGTCCTCTC
ACATACACACCGC

>270109_109F_1391R _62A

CGCTGGAGGGAAGGATAGAGAGGCGCGTCCCTGGCCAGGGATGCTGGGAGTTATGGTCTC
TTAACGGTTTTAAGCGATGGCCCCCGGCCTGCAGACCACAATCCCAGCAGCCACCGGGCT
TCGAGGCGGTGTGTAGCACTGAAGGGAAGGATAGGGAGGTGCGTCCTAGCCAGGCGTGCT
GGGAGTTATGGTCTCTTAACAGTTTCCAGTCAGTTGGTCCCAGGACTACCTGACTACAAT
CCCAGTATGCGTTGGGCTTGGGGGCGGTGCGCAGCCCTAGAGGAAGGATCGGGACGGCGG
GTACCTCGCAAAGCATCCTGGGAGTCATAGTCCTTTCAGTATTTCCAGCCCATTGGTCGC
GAGGCTAACGGACTACAATCTCAGCATGCGCTGGGTTTGGGGGCGGTGTGTAGTATGGAA
GCGAAGGATAGGGAGGCGCGTCCCTAGCTAGGAGTGCTGGGAGTTATGGTGTCTTAACGG
TTTCCAGCCCATTGGTCGCCGACCTGCTAACTACAAAACCAGCATGCGCTGTCTGTCCTA
CCCCGTGGTGCGCAGCCCTGGAGGGAGGGACAGGGCGGTGTGGACTCGTCCTTTCCTAAG
CGATGCCACATGCTGATTCTGTGCCACCCCCTCGCCAAGGGAGTCCGCAGAAGGACTTGA
GGGGCAGGTCTAGGCTGGGCGATGAGGACGGTGTGACCCTGCGAAGTGCACCTCCCTTGC
TCAAATCGGAGGGGTCTGGTCCTCACTGAACAGCCCGCTGAACATCTCGGTGTCCTCTCA
CATACACACCGCCCGTCAAACA

>**270109_109F_1391R _64**

GGGAAATTTGTTGAAAAAAACACCCACAAGAGAAAGCAGGAAAGATCTAAAATCAACACC
ATAACATCAAAATTAAAAGAACTAGGGAAGCAAGAGCAAACAAATTCAAAAGCTAGCAGA
AGGCAAGGAATAACTAAGATCAGAGCAGAACTGAAAGAGATAGAGACACAAAAAACCCTT
CAAAAAAAAATCAATGAATCTGAGAGCTGGTTTTTTGAAATGATCAACGAAATTGATAGA
CCGCTAGCAAGATTAATAAAGAAGAAAAGAGAGAAGAATCAAATAGACTCAATAAAAAAT
GATAAAGGGGATATCACCACCAATCCCACAGAAATACAAACTGCATCAGAGAATACTATA
AACACCTCTATGCAAATAGACTAGAAAATCTAGAAGAAATGGATAAATTCCTGGACACAT
ATACCCTCCCAAGACTAAACCAGGAAGAAGTTGAATCTCTGAATAGACCAATAACAGGCT
CTGAAATTGAGGCAATACTTAATAGCCTACCAACCAAAAAAATTCCAGGACCGGACAGAT
TCACAGCCAAATTCTATCAGAGGCACAAAGAGGAGCTGGTAACATTCCTTCTGAAACTAT
TCCAATCAATAGAAAAAGAGGGAATCCTCCCTAACTTATTTTATGAGGCCAGCATCATCC
TGATACCAAAGCCTGGCAGAGACACAACAAAAAAAGAGAATTTTAGACCAATACCCCAGA
TGAACATTGATG

>270109_109F_1391R _65

GTACAAAGAGAGAGGAGAGTTTTTACGGTGATTGAACTCTTCAATACTCTGACTGTGGAG
GTGTTTATTATATCTATACTTAAAATTCACCAAACCATACACCAAAAAAGTTAATTTTAC
TTTACTCAAAAAATTTTGTTAAATGTCAAAGAGAATAAAATTAAAAGAAAAATAGCCTCT
TCCCCCGACAATTCATTCCTAAAATGAGCAAGATAGGAATCACAGGGTGAATGGACTGCC
ATGGCCTGGAGTGAGGTATTAGGGTCCAGGCAGGATGAGGAGGACATTCCTGCTCAGCAT
GGGATGCGAAGCCTAAGCAGGGTGAGGAAGACATTCATGTGGGGAGATGACCTACAATAG
AGAACCAGGCCAGGCGCAGTGGCTCACACCTGTAATCCCAGCAATTTGGGAGGCCCAGGC
GGGCGGATCACCTGATGTCCGGAGTTCGAGACCAGCCTGACCAACATGGAGAAACCCCGT
CTCTACCAAAAATTAAAAATTAGCCAGGTGTGGCAGCGCATGCCTGTAATCCCAGCTACT
TGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTAAGCCGAG
ATCGCGCGACTGCACTCCAGTCTGGGCAACAAGAGTGAAACTCCACCTTAAAAAAAAAAC
CCCACAAATAATCAAATAGGAATCAGGTTTCTCACTGGTACTGAAAACGGTTATAAATGT
GGAGAAGGAAAAAAATGAAAATGAAGCCCATGGGATCAGAATTTATATGTGAATTCATGG
TTTTCAATATACATAGATGCAGAAATATAGATGTAAAAGACATGCATTATATATACACAC
CCGCCCGTTCAAAAC

>270109_109F_1391R _66

GCAAGAAGTGAATAACACAGAAATCAATGATGAAAGCACTAGAAGACAATTGAGTTTATC
AGAACTGCAAAATATTGCTGAGTGTGGATTGCTCTGAAATCTGAAAACATGACTTGTGAA
TTGCTTCTATTCAAAATGCAGACACAATCAATGCCGGGTGTTGGTTTACTTGTTTCCCAT
TTTGCAACCCTCTTTTCTAGGCAAAAGGTATCCAAACTATGCAGAGCCACAGAATCTAAC
AGATGTCTCTCTATTCCTTCTCCTAGAACTCTCAGGGGATCCAGAACTGCAGCTGGTCCT
CGCTGGGCTGTTCCTGTCCATCTGCCTGCTCATGGTGCTGGGGAACCTGCTCATCATCCT
GGCCATCAGCCCTGGCTCCCACCTCCACACCCCCATGTACTTCTTCCTCTCCAACCTGTC
CTTGCCTGACATCAGTTTCACCTCCACCACAGTCCCCAAGATGATTGTGGACATCCAATC
TCACAGCAGTCATCTCCTATGCAGGCTGCCTGACTCAGATGTCTTCTCTTTTGTCATTTT
TGGAGGCATG

References

Backhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., et al. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A, 101*(44), 15718-15723.

Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science, 307*(5717), 1915-1920.

Blaut, M., Collins, M. D., Welling, G. W., Dore, J., van Loo, J., & de Vos, W. (2002). Molecular biological methods for studying the gut microbiota: the EU human gut flora project. *British Journal of Nutrition, 87*(Suppl. 2), S203-S211.

Bove, J. M. (1993). Molecular features of mollicutes. *Clin Infect Dis, 17 Suppl 1*, S10-31.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science, 308*(5728), 1635-1638.

Fredricks, D., & Relman, D. A. (1998). Improved Amplification of Microbial DNA from Blood Cultures by Removal of the PCR Inhibitor Sodium Polyanetholesulfonate *Journal of Clinical Microbiology, 36*(10), 2810-2816.

Garrity, G. M., Bell, J. A., & Lilburn, T. G. (2004). *Taxonomic Outline of the Prokaryotes*.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science, 312*(5778), 1355-1359.

Godoy-Vitorino, F., Ley, R. E., Gao, Z., Pei, Z., Ortiz-Zuazaga, H., Pericchi, L., et al. (2008). Bacterial Community in the Crop of the Hoatzin, a Neotropical Folivorous Flying Bird. *Applied and Environmental Microbiology, 74*(19), 5905-5912.

Handelsman, J. (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms *Microbiology and Molecular Biology Reviews, 68*(4), 17.

Hartwell, L. H., Hood, L., Goldberg, M. L., Reynolds, A. E., Silver, L. M., & Veres, R. C. (2004). *Genetics: From Genes to Genomes* (Second ed.): McGraw-Hill.

Invitrogen. (2004). *TOPO TA Cloning Manual* (R ed.): Invitrogen.

Kubs, C., & Musgrave, D. (2007). A novel *Mollicute* species inhabiting the human gastrointestinal tract. University of Waikato, New Zealand.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A, 102*(31), 11070-11075.

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature, 444*, 1023-1024.

Lilburn, T. G., & Garrity, G. M. (2004). Exploring prokaryotic taxonomy. *Int J Syst Evol Microbiol, 54*, 7-13.

Madigan, M. T., Martinko, J. M., & Parker, J. (2003). *Brock Biology of Microorganisms* (Tenth ed.). New Jersey: Pearson Education, Inc.

Maniloff, J. (1996). The minimal cell genome: "on being the right size". *Proc Natl Acad Sci U S A, 93*(19), 10004-10006.

Nanodrop-Technologies. (2004). *Nanodrop ND-1000 Spectrometer V3.1.0 User's Manual*.

Naomi, W. (2006). New directions and interactions in metagenomics research (Vol. 55, pp. 331-338).

Pettersson, B., Tully, J. G., Bolske, G., & Johansson, K. E. (2001). Re-evaluation of the classical Mycoplasma lipophilum cluster (Weisburg et al. 1989) and description of two new clusters in the hominis group based on 16S rDNA sequences. *Int J Syst Evol Microbiol, 51*(Pt 2), 633-643.

Razin, S. (1992). Mycoplasma taxonomy and ecology. In J. Maniloff, R. N. McElhaney, L. R. Finch & J. B. Baseman (Eds.), *Mycoplasmas: molecular biology and pathogenesis* (pp. 3–22). Washington, D.C: American Society for Microbiology.

Razin, S. (1995). Molecular Properties of Mollicutes: A Synopsis. In *Molecular and Diagnostic Procedures in Mycoplasmology* (pp. 1-25). San Diego: Academic Press.

Razin, S. (1997). The minimal cellular genome of mycoplasma. *Indian J Biochem Biophys, 34*(1-2), 124-130.

Razin, S., Yogev, D., & Naot, Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev, 62*(4), 1094-1156.

Rieu-Lesme, F., Delbes, C., & Sollelis, L. (2005). Recovery of Partial 16S rDNA sequences Suggests the Presence of *Crenarchaeota* in the Human Digestive Ecosystem. *Curr. Microbiology*(51), 317-321.

Rottem, S., & Kahane, I. (1993). *Mycoplasma Cell Membranes* (Vol. 20). New York: Plenum Press.

Sambrook, J., & Russell, D. (2001). *Molecular Cloning: A Laboratory Manual* (3rd ed.): Cold Spring Harbor Laboratory Press.

Sirand-Pugnet, P., Lartigue, C., Marenda, M., Jacob, D., Barre, A., Barbe, V., et al. (2007). Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS Genet, 3*(5), e75.

Stakenborg, T., Vicca, J., Verhelst, R., Butaye, P., Maes, D., Naessens, A., et al. (2005). Evaluation of tRNA gene PCR for identification of mollicutes. *J Clin Microbiol, 43*(9), 4558-4566.

Streit, W. R., & Schmitz, R. A. (2004). Metagenomics--the key to the uncultured microbes. *Curr Opin Microbiol, 7*(5), 492-498.

Turnbaugh, P., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R. D., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature, 449*, 804-810.

Turnbaugh, P., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature, 444*, 5.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature, 444*(7122), 1027-1031.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science, 304*(5667), 66-74.

Volokhov, D. V., George, J., Liu, S. X., Ikonomi, P., Anderson, C., & Chizhikov, V. (2006). Sequencing of the intergenic 16S-23S rRNA spacer (ITS) region of Mollicutes species and their identification using microarray-based assay and DNA sequencing. *Appl Microbiol Biotechnol, 71*(5), 680-698.

Wang, H., Kong, F., Jelfs, P., James, G., & Gilbert, G. L. (2004). Simultaneous detection and identification of common cell culture contaminant and pathogenic mollicutes strains by reverse line blot hybridization. *Appl Environ Microbiol, 70*(3), 1483-1486.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology, 73*(16), 5261-5267.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev, 51*(2), 221-271.

Woese, C. R., Maniloff, J., & Zablen, L. B. (1980). Phylogenetic analysis of the mycoplasmas. *Proc Natl Acad Sci U S A, 77*(1), 494-498.

Worliczek, H. L., Kampfer, P., Rosengarten, R., Tindall, B. J., & Busse, H.-J. (2007). Polar lipid and fatty acid profiles - Re-vitalizing old approaches as a modern tool for the classification of mycoplasmas? *Systematic and Applied Microbiology, 30*(5), 355-370.