# PHYLOGENETIC PROFILING OF BACTERIAL POPULATIONS IN *PERNA PERNA* L. BEFORE AND DURING RED TIDE IN ALGOA BAY, EASTERN CAPE PROVINCE, SOUTH AFRICA

**ELIZABETH BOSEDE FAMEWO**

A Dissertation

Submitted in fulfilment of the requirements for the degree of

**MASTER OF SCIENCE (MSc) MICROBIOLOGY**

**DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY**

**FACULTY OF SCIENCE AND AGRICULTURE**

**UNIVERSITY OF FORT HARE**

**ALICE 5700, SOUTH AFRICA**

**SUPERVISOR: DR AM CLARKE**

**CO-SUPERVISOR: PROF AJ AFOLAYAN**

**APRIL 2015**

## DECLARATION

I, Elizabeth Bosede Famewo, hereby declare that this dissertation has been written by me and is a record of my own research work. It has not been presented in any previous application for a higher degree of this or any other University. All citations and sources of information are clearly acknowledged by means of reference.

10/04/2015

..... .....

Elizabeth Bosede Famewo Date

## DECLARATION ON PLAGIARISM

I, Elizabeth Bosede Famewo student number 201313469 hereby declare that I am fully aware of the University of Fort Hare's policy on plagiarism and I have taken every precaution to comply with the regulations.

10/04/2015

..... .....

Elizabeth Bosede Famewo Date

## CERTIFICATION

This dissertation entitled "Phylogenetic profiling of bacterial populations in *Perna perna* L. before and during red tide in Algoa Bay, Eastern Cape Province, South Africa" by Elizabeth Bosede Famewo meets the regulation governing the award of degree Master of Science of the University of Fort Hare and is approved for its contribution to scientific knowledge and literary presentation.

..... 10/04/2015 .....

Dr AM Clarke Date

..... 10/04/2015 .....

Prof AJ Afolayan Date

## ABSTRACT

Shellfish has been known to harbour bacterial agents of marine origin which cause diseases to individuals who consume raw and lightly cooked molluscs. Some populations of *Perna perna* L. were collected from Algoa Bay in Port Elizabeth, Eastern Cape of South Africa before and during red tide over a period of one month. Samples were dissected, homogenized and their DNA was extracted. A fragment of the bacterial 16S rRNA amplified by PCR with universal primers was visualized on agarose gel electrophoresis. The expected amplicon size (586 bp) from each DNA was further purified using Zymoclean$^{TM}$ Gel DNA recovery kit. The ultra-pure DNA was sequenced in ABI 3500 XL genetic analysers using ABI V3.1 Big Dye kit. The chromatoGrams were converted into text format using Chromas software and blasted on NCBI. The FASTA sequences of each of the bacteria were aligned using Clustal Omega. The evolutionary history was inferred using Neighbour-Joining method and the phylogenetic tree was constructed using MEGA 6 program. Twenty eight bacterial strains were identified in 90% *Perna perna* collected before red tide, of which 25 strains belong to the Phylum Actinobacteria. They were Gram-positive bacteria and members of the genera *Citricoccus*, *Micrococcus*, *Zhihengliuella*, *Arthrobacter*, *Kocuria*, *Nesterenkonia*, *Streptomyces*, *Saccharothrix*, *Lechevalieria*, *Saccharothrix* and *Rhodococcus*, while the remaining three strains belong to Phylum Spirochaetes, genus *Spirochaeta* and were all Gram-negative. Only few bacteria strains were identified in 10% *Perna perna* collected during red tide; of which four strains belong to Phylum Proteobacteria, genera *Paracoccus* and *Cohaesibact*er, and were all Gram-negative. While Phylum Actinobacteria belonging to the genus *Kocuria* was detected in the molluscs, Phylum Spirochaetes was equally detected, with the former being Gram positive and the latter Gram-negative bacteria. This study reveals that *Perna perna* species are potential reservoirs of various bacterial pathogens with associated health threats posed on their consumers. Also, some of these bacterial species are capable of carrying out bioremediation.

## DEDICATION

This dissertation is dedicated to the Almighty God who has been my source of inspiration and strength. He has sustained me right before this research began and throughout the course of this study, from Him every good gift comes. May His name be praise forever and ever.

## ACKNOWLEDGMENTS

My sincere gratitude to my supervisor, Dr AM Clarke for the acceptance and assistance despite all your tight schedules, you are always willing to attend to me. May God continue to bless and keep you. To my co-supervisor, Prof AJ Afolayan, thank you for the availability and assistance, may God continue to bless and keep you. I am grateful to the University of Fort Hare (UFH) for giving me the opportunity to tread on the path where great people have trodden. To my Head of Department, Prof AI Okoh, may you be blessed beyond human imagination. I am equally indebted to South African Institute of Aquatic Biodiversity for providing the samples needed for this research. Many thanks also to National Research Foundations (NRF) for funding this project. God will continue to uphold you all.

My deep appreciation goes to my mother Mrs LB Fameewo, for all your supports, encouragements and efforts to make my dream come to reality. May God continue to protect you, give you good health and sound mind. Special appreciation to my father, a mentor and guidance, Prof AJ Afolayan and my mother Mrs Afolayan for the moral and financial support at all times. Words cannot be sufficient to express my heartfelt gratitude to you Sir and Maam. May God continue to endow you with wisdom and perfect all your ways in Jesus name.

Thanks to my beloved sisters and Nike for their continuous encouragements and constant supports in many ways. May the Lord continue to enrich you in all aspects of your lives. To my good brothers, friends and mentor Mr Cromwell Kibiti and Mr Olayinka Titilawo for all your efforts to make this work a reality, words cannot be sufficient to express my heartfelt gratitude but the good Lord will perfect all that concerns you. He will bless you and grant you all your good heart desires. Thank you all and God bless.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ARDRA: | Amplified Ribosomal DNA Restriction Analysis |
| ARISA: | Automated Ribosomal Intergenic Spacer Analysis |
| ASP: | Amnesic Shellfish Poisoning |
| BLAST: | Basic Local Alignment Search Tool |
| ddNTPs: | Dideoxyribonucleotide Triphosphates |
| DGGE: | Denaturing Gradient Gel Electrophoresis |
| DNA: | Deoxyribonucleic acid |
| dNTPs: | Deoxyribonucleotide Triphosphates |
| DSP: | Diarrhetic Shellfish Poisoning |
| GC: | Guanine-Cytosine |
| HABs: | Harmful Algal Blooms |
| IGS: | Intergenic Spacer Region |
| NCBI: | National Center for Biotechnology Information |
| NCHD: | Nassau County Health Department |
| NSP: | Neurotoxic Shellfish Poisoning |
| PCB: | Polychlorinated Biphenyl |
| PCR: | Polymerase Chain Reaction |
| pNP: | p-nitrophenol |
| PSP: | Paralytic Shellfish Poisoning |
| SSCP: | Single-Strand Conformation Polymorphism |
| SSCP: | Single-Strand Conformation Polymorphism |
| STX: | Saxitoxins |
| TGGE: | Temperature Gradient Gel Electrophoresis |
| T-RFLP: | Terminal Restriction Fragment Length Polymorphism |
| TRFs: | Terminal Restriction Fragments |

# CHAPTER ONE

# GENERAL INTRODUCTION

# TABLE OF CONTENTS

# GENERAL INTRODUCTION

## 1.0 Introduction

Marine environment is the largest habitat on earth, accounting for about 90% of the biosphere by volume and harbouring different kinds of macro and microorganisms (Lauro *et al*., 2009). This ecosystem is contaminated through multifarious human activities especially by the discharge of enormous quantities of sewage and industrial effluents into the ocean. This has resulted into drastic reduction of shallow waters fish populations, ecological imbalance resulting into large scale disappearance of numerous flora and fauna, increased pollution of marine ecosystem with chemicals and toxins which affect the microbial populations. Various researches have been carried out on the microbiota of bivalve molluscs due to their filter-feeding habit; they have been found to concentrate large amount of indigenous and non-indigenous bacteria in their tissues and as such cause diseases to human (Romanenko *et al*., 2008; Gayathri *et al*., 2011). Besides accumulating naturally occurring microorganisms and other pathogenic microorganisms, these shellfish are also prone to contamination by toxin-producing dinoflagellates and faecal pathogens (Rippey, 1994; Wilson and Moore, 1996; Potasman *et al*., 2002).

Bivalve molluscs are economically important marine shellfish species which are abundantly found in estuarine and marine waters that remain below 30 °C. They are usually found on rocky shores in temperate regions (Calci *et al*., 2013). The brown mussel, *Perna perna* L. (Fig. 1) a member of the class Bivalvia, order Mytiloida and family Mytilidae is indigenous to the waters of Africa, Europe, South America and North America up to the Caribbean (Monirith *et al*., 2003; Global Invasive Species Database, 2007). It is prolific on the rocky reefs along the eastern coast of South Africa. These mollusc are usually about 90 mm long but can reach a length of 120 mm. Aside South Africa, it is found in few other African

countries like Namibia, Angola, Cape Verde, Mauritania, Mozambique and Madagascar (Monirith *et al*., 2003).

**Figure 1:** *Perna perna* L.

*Perna perna* is easily recognized by its brown colour and identified by its characteristic divided posterior retractor mussel scar. The presence of pitted resillal ridge differentiates the organism from other species of bivalves (Smith *et al*., 2001). Other species similar to *P. perna* include the mediterranean mussel called *Mytilus galloprovincialis* and the black mussel called *Choromytilus meridionalis*. In about 6–7 months, *P. perna* can grow up to the commercial size of about 60–80 mm, therefore, it is considered suitable for cultivation. It is known to be a significant source of food in different parts of the world since the Roman times (Potasman *et al*., 2002). The species is nutrient-rich part of a healthful diet. The consumption of mussel is associated with health benefits which include neurologic development during gestation and infancy (Hibbeln, 2007), reduced risk of heart diseases and lowering the amount of cholesterol in the blood (Mozaffarian and Rimm, 2006). It is also used in the treatment of rheumatoid arthritis, improved circulation and reduction of the risk of thrombosis. Other benefits include keeping the eyes bright and healthy, protection of the

lungs and relieving the symptoms of asthma in children. The consumption of *P. perna* is reported to protect against serious inflammatory bowel diseases including Crohn's disease and ulcerative colitis, prevents seasonal affective disorder and post-natal depression, provides the body with many essential nutrients including iodine, selenium, zinc and potassium, also it contains omega–3 fatty acids (eicosapentaenioc acid and docosahexaenoic acid), which prevents common chronic diseases (Nesheim *et al*., 2007).

Along with the nutritional benefits associated with the consumption of mussels, however, comes the potential risk of bacterial contamination. *P. perna* are filter-feeders. They feed on plankton, zooplankton, detritus and some microflora. They achieve this by filtering nutrient from surrounding water through their gills by ciliary action. These make them concentrate microorganisms in their tissues and are therefore regarded as reservoirs for diverse microbial pathogens (Iwamato *et al*., 2010). De Donno *et al*. (2008) observed that shellfish may contain high densities of faecal indicator bacteria, even when the surrounding waters are apparently of good quality. This is because, all seafoods are susceptible to surface and tissue contamination originating from the marine environment. Other sources of shellfish contamination include sewage discharge into harvested areas, sewage runoff from points inland after heavy rainfall or flooding. Shellfish may become contaminated during handling, processing or preparation. Contributing factors may include storage and transportation at inappropriate temperatures, or cross-contamination through contact with contaminated seafood or seawater (Iwamato *et al*., 2010). Chemicals, metals, marine toxins and infectious agents such as bacteria, viruses and parasites associated with food-borne illnesses have also been found in shellfish. These agents are able to cause illnesses that range from mild gastroenteritis to life-threatening syndromes (Iwamoto *et al*., 2010).

## 1.1 Occurrence of red tide

Red tide commonly known as Harmful Algal Blooms (HABs) is the discoloration of seawater caused by dense concentrations of marine algae called phytoplanktons (Jorge, 2014). These marine algae contain photosynthetic pigments that vary in colour, from green to brown and to red. Harmful algal blooms occur when colonies of algae that live in the sea and freshwater grow out of control. They produce toxins that can be harmful to marine mammals, birds and human either directly (affecting the nervous and digestive systems) or through the consumption of contaminated shellfish, crustaceans and fish (NOAA, 2014). The factors influencing the occurrence of red tide include over-warm surface water temperature, low salinity, high nutrient content, calm seas, favourable wind and water currents and rain followed by sunny days during the summer months (NOAA, 2014).

Along the South African coast, most red tides are caused by a group of phytoplanktons known as dinoflagellates. These dinoflagellates are single-celled organisms capable of swimming short distances by means of two whip-like appendages called flagella. Their ability to swim from one point to another contributes to their success as red tide organisms. Generally, there are three types of algae associated with red tides; they include, *Alexandrium fundyense* found along the Atlantic coast from the Canadian Maritimes to southern New England, *Alexandrium catenella* found along the Pacific coast from California to Alaska and *Karenia brevis* found in the Gulf of Mexico along the west coast of Florida. The most awful incidence of marine mortality in South Africa was experienced and recorded at the west coast of St. Helena Bay in 1994. The event was caused by the entrapment and subsequent decay of non-toxic *Prorocentrum micans* and *Ceratium furca*, with the toxic species *Alexandrium catanella* and *Dinophysis acuminate* (CDC, 2012).

**Figure 2:** Seawater Red tide.

**(Source: Photo by M. Godfrey, 2013).**

When toxic phytoplanktons are filtered by shellfish, some of the shellfish die while others are usually unaffected by the toxic algae but accumulate algal toxins in their tissues to the level that are potentially harmful to consumers. The consumption of shellfish during red tide algal blooms has been attributed to several mammalian diseases and deaths (Dolah *et al*., 2001). Also, different types of shellfish poisonings have been identified; these include amnesic, neurotoxic, paralytic, diarrhetic shellfish poisoning and aerosol (air-borne) toxins. Each of these syndromes is caused by a different species of toxic algae and has different mechanisms of toxicity and unique symptoms (Dolah *et al*., 2001).

## 1.2 The use of molecular techniques for microbial identification

Over the past century, the identification and differentiation of microorganisms in an environmental sample relied on microbial morphology and growth variables. These traditional methods of microbial identification are based on the phenotypic characteristics such as protein, bacteriophage, and chromatographic profiles as well as biotyping and

susceptibility testing of the organism. Bacterial fermentation, fungal conidiogenesis, parasitic morphology, and viral cytopathic effects are a few phenotypic characteristics commonly used. Though, most phenotypic variables commonly observed are not sensitive enough for strain differentiation (Tang *et al*., 1997), scientists have revealed that, the conventional techniques of culturing and identifying microorganisms have limited usefulness and insufficiency, since diverse microorganisms can share similar morphologies and cultivation limits detection to organisms under different growth conditions (Ward *et al.,* 1990; Ferris *et al*., 1996; Gugliandolo *et al*., 2011; Yendi *et al*., 2012). In order to overcome the limitations of culture based technique, attention has now been focused on molecular techniques with their ability to access information about microbial communities in their natural habitats.

The ability to detect and identify bacteria in environmental samples such as water and soil has been enhanced by the development of molecular techniques for the analysis of 16S rRNA sequences (Olsen *et al.,* 1986; Pace *et al.,* 1986). This is based on the principle that the 16S rRNA genes are directly linked to the phylogeny of microorganisms; they are therefore widely used for bacterial diversity assessments in environmental samples (Olsen *et al*., 1986; Pace *et al*., 1986; Liesack *et al*., 1997). With the development, different techniques have been used to evaluate the microbial genome diversity, analysis of structure and species composition in different ecosystems. In particular, sequence variation in rRNA has been exploited for inferring phylogenetic relationships among microorganisms (Zhu *et al.,* 2002; Scupham *et al.,* 2006) and for designing specific nucleotide probes for the detection of individual microbial taxa in natural habitats (Fuhrman *et al.,* 1993).

In addition, this method allows rapid detection of pathogens especially when a risk assessment is required for human health as soon as the first sign of illness occur in less than 24 h after ingestion of contaminated seafood (Gugliandolo *et al*., 2011). When molecular technique is coupled with polymerase chain reaction and primers that target conserved,

taxonomically significant genes, it then allows the comparison and analysis of total microbial populations (O Callaghan *et al*., 2003). The goal of this project was to determine the evolutionary relationships across different groups of microorganisms found in *Perna perna*, compile and compare the total bacterial populations found and examine the impact of red tides on the microbial population using analysis of PCR-amplified genes coding for 16S rRNA and sequencing.

### 1.3 Problem Statement

The occurrences of red tides along the South African coast are reported periodically in the media, with warnings to the public about the dangers of collecting and consuming shellfish in the affected areas. Some shellfish are susceptible to the toxic red tide while others show a high level of resistance with the accumulation of toxins in their tissues. This toxin can be lethal to humans when consumed and one of such harmful effects is paralysis (Independent Online, 2005). The global consumption of shellfish has increased considerably since the early 1970s with corresponding records of outbreaks of shellfish-associated infection such as diarrhoea, typhoid fever and meningitis (Potasman *et al*., 2002). These outbreaks have been attributed to bacterial and viral agents that are associated either with human wastes (sewage) or bacterial pathogens that are indigenous to coastal marine environments (Rippey, 1994).

Despite the wider distribution and consumption of *Perna perna* in South Africa most especially in the Eastern Cape Province, there is little information available on the profiling of total bacterial populations inhabiting *P. perna*. Hence, there is a need to elucidate the phylogenetics of the bacteria found in the samples. The findings hope to create awareness for the consumers on the harms associated with the consumption of infected brown mussels.

## 1.4 Research Hypothesis

Red tide affects the bacterial flora of *Perna perna* in Algoa Bay.

## 1.5 Aim and objectives

The overall aim of this study is to confirm and profile different bacterial populations in *Perna perna* before and during red tide, using PCR-amplified genes coding for 16S rRNA and sequencing without the need for isolation and identification of pure cultures. The specific objectives are:

(1) To characterise bacteria from *P. perna* before red tide

(2) To characterise bacteria from *P. perna* during red tide

(3) To construct a phylogenetic tree of the bacteria flora of *P. perna*

(4) To determine evolutionary relationships and differences between the bacteria flora found in *P. perna* before and during red tides.

**References**


Calci, K. R., Depaola, A., & Burkhardt III, W. (2013). Molluscan Shellfish: Oysters, Mussels, and Clams Updated September 2013. Accessed 2014/09/08.

Centers for Disease Control and Prevention (2012). http://www.cdc.gov/nceh/hsb/hab/ Accessed 2014/07/10.

De Donno, A., Liaci, D., Bagordo, F., Lugoli, F., & Gabutti, G. (2008). *Mytilus galloprovincialis* as a bio-indicator of microbiological pollution of coastal waters: A study conducted in the Salento Peninsula (Italy). *Journal of Coastal Research 24*(sp1), 216-221.

Dolah, F. M. V., Roelke, D., & Greene, R. M. (2001). Health and ecological impacts of harmful algal blooms: risk assessment needs. *Human and Ecological Risk Assessment: An International Journal*, *7*(5), 1329-1345.

Ferris, M. J., Muyzer, G., & Ward, D. M. (1996). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Applied and Environmental Microbiology*, *62*(2), 340-346.

Fuhrman, J. A., McCallum, K., & Davis, A. A. (1993). Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Applied and EnvironmentalMicrobiology*, 59, 1294-1302.

Gayathri, V., Ganga Baheerathi, C., Revathi, K., & Sethi, S. N. (2011). Identification of Gut Microflora from Green Mussel (Perna viridis).

Global Invasive Species Database (2007). *Perna perna* (mollusc).

Gugliandolo, C., Lentini, V., Spanò, A., & Maugeri, T. L. (2011). Conventional and molecular methods to detect bacterial pathogens in mussels. *Letters in Applied Microbiology*, *52*(1), 15-21

Hibbeln, J. R., Davis, J. M., Steer, C., Emmett, P., Rogers, I., Williams, C., & Golding, J. (2007). Maternal seafood consumption in pregnancy and neuro developmental outcomes in childhood (ALSPAC study): an observational cohort study. *Lancet*, 369, 578–585.

Independent Online (2005). http://www.whoi.edu/science/B/redtide/notedevents/foreign/SouthAfrica/SouthAfrica_3-14-05.html).

Iwamoto, M., Ayers, T., Mahon, B. E., & Swerdlow, D. L. (2010). Epidemiology of seafood-associated infections in the United States. *Clinical Microbiology Reviews*, *23*(2), 399-411.

Jorge R. R. (2008). The Institute of Food and Agricultural Sciences, UF/Extension, Entomology and Nematology Department, Florida Medical Entomology Laboratory, Vero Beach, FL, 32962. (http://edis.ifas.ufl.edu ).

Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., & Cavicchioli, R. (2009). The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences*, *106*(37), 15527-15533.

Liesack, W., Janssen, P. H., Rainey, F. A., Ward-Rainey, N. L, & Stackebrandt, E. (1997). Microbial diversity in soil: the need for a combined approach using molecular and cultivation techniques. *Modern Soil Microbiology*, 375-439.

Monirith, I., Ueno, D., Takahashi, S., Nakata, H., Sudaryanto, A., Subramanian, A., Karuppiah, S., Ismail, A., Muchtar, M., Zheng, J. S., Richardson, B. J., Prudente, M., Hue, N. D., Tana, T. S., Tkalin, A. V., & Tanabe, S. (2003). Asia-PaciWc mussel watch: monitoring contamination of persistent organochlorine compounds in coastal waters of Asian countries. *Marine Pollution Bulletin*, *46*, 281-300.

Mozaffarian, D. & Rimm, E. B. (2006). Fish intake, contaminants, and human health. *JAMA* 296, 1885–1899.

National Oceanic and Atmospheric Administration (NOAA) (2014). http://oceanservice.noaa.gov/facts/redtide.html. Accessed 2014/03/11

Nesheim, M. C. & Yaktine, A. L. (2007). Seafood choices: balancing benefits and risks. *National Academies Press*, (Eds).

O Callaghan, M., Gerard, E. M., Heilig, G. H. J., Zhang, H., Jackson, T. A., & Glare, T. R. (2003). Denaturing gradient gel electrophoresis-a tool for plant protection research. *New Zealand Plant Protection*, 143-150.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., & Pace, N. R. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiol*ogy *40*, 337-365.

Pace, N. R., Stahl, D. A., Lane, D. J., and Olsen, G. J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. *Advance in Microbial Ecology, 9*, 1-55.

Potasman, I., Paz, A., & Odeh, M. (2002). Infectious outbreaks associated with bivalve shellfish consumption: a worldwide perspective. *Clinical Infectious Diseases*, *35*(8), 921-928.

Rippey, S. R. (1994). Infectious diseases associated with molluscan shellfish consumption. *Clinical Microbiology Reviews*, *7*(4), 419-425.

Romanenko, L. A., Uchino, M., Kalinovskaya, N. I., & Mikhailov, V. V. (2008). Isolation, phylogenetic analysis and screening of marine mollusc-associated bacteria for antimicrobial, hemolytic and surface activities. *Microbiological Research, 163*(6), 633-644.

Scupham, A. J., Presley, L. L., & Wei, B. (2006). Abundant and diverse fungal microbiota in the murine intestine. *Appied and Environmental Microbiology 72*, 793-801.

Smith, R. B., Penry, W., Nelson, C., Haddad, K., Kottkamp, J., Ward, W. W., & Perkins, L. D. (2001). Gulf States Marine Fisheries Commission. Retrieved 2013.

Tang, Y. W., Procop, G. W., & Persing, D. H. (1997). Molecular diagnostics of infectious diseases. *Clinical Chemistry*, *43*(11), 2021-2038.

Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* (London) *345*, 63-65.

Wilson, I. G. & Moore, J. E. (1996). Presence of *Salmonella* spp. and *Campylobacter* spp. in shellfish. *Epidemiology and Infection*, *116*(02), 147-153.

Yendi, N., César, H., Juan, C. Z., Beatriz, B., Mario, E. C., Janet J., & Juan, C. C. (2012). 16S rRNA Gene-Based Identification of Bacteria in Postoperative Endophthalmitis by PCR denaturing gradient gel electrophoresis (PCR-DGGE) fingerprinting. *Brazilian Journal of Microbiology*, 283-287.

Zhu, X. Y., Zhong, T. Y., Pandya, Y. G., & Joerger, R. D. (2002). 16SrRNA-based analysis of microbiota from the caecum of broiler chickens. *Applied and Environmental Microbiology*, *68*, 124-137.

# CHAPTER TWO

# LITERATURE REVIEW

# TABLE OF CONTENTS

# LITERATURE REVIEW

## 2.0 Bivalve shellfish and red tide

Bivalve shellfish feed exclusively on phytoplanktons filtered from seawater. This makes shellfish to bio-accumulate toxins in their tissues to the level that can become lethal to humans during red tide outbreak. The news of red tide creates fear among people who consume shellfish because medical records reviews that red tide can cause shellfish to be poisonous (Independent Online, 2005). From literature, it has been evident that red tide has effect on humans, economy and on the environment. For instance, a cluster of respiratory illnesses was reported to the Nassau County Health Department (NCHD) in north-eastern Florida after an outbreak of red tide. People working at a beach restoration worksite reported symptoms of eye or respiratory irritation (such as coughing, sneezing, sniffling and throat irritation) (CDC, 2008). Centers for Disease Control and Prevention reported cases of outbreak in which more than 550 cases of HAB-related illnesses between 1993 and 2002 were recorded. Although this number is considered low estimate but it is believed that seafood-related illnesses are under-reported, misdiagnosed and possibly increasing (Olsen, 2000).

In 2009-2010, Centers for Disease and Control reported eleven freshwater HAB–associated outbreaks in New York, Ohio and Washington where about 61 illnesses, two hospitalizations and no known deaths were recorded. The health effects which varied among outbreak include dermatologic symptoms such as rash, irritation, sores; gastrointestinal symptoms; respiratory symptoms; fever; headache; neurologic symptoms and eye irritation (Hilborn, 2014). The two known forms of red tide toxins-associated with clinical entities in human after the injection of contaminated shellfish in Florida are an acute gastroenteritis with neurologic symptoms and an apparently reversible upper respiratory syndrome after the inhalation of aerosols of the

dinoflagellates and their toxins. In June 2000, toxic red tides were responsible for the death of 30 million mussels in South Africa which were washed ashore at Bloubergstrand in the Western Cape. Also, several people were reported ill with paralytic shellfish poisoning (PSP) symptoms after apparently the consumption of mussels (Jorge, 2008). Different types of shellfish poisonings identified include amnesic, neurotoxic, paralytic, diarrhetic shellfish poisoning and aerosol (air-borne) toxins. Each of these syndromes is caused by a different species of toxic algae and has different mechanisms of toxicity and unique symptoms (Dolah *et al*., 2001).

### 2.1.0 Types of Shellfish Poisonings

### 2.1.1 Paralytic shellfish poisoning (PSP)

Paralytic shellfish poisoning is caused by a group of about 12 structurally similar neurotoxin called saxitoxins (STX). Saxitoxin is a neurotoxin that inhibits the permeability of cell membrane to sodium ions (sodium channel blocker). It acts on the voltage-gated sodium channels of the nerve cells thereby preventing the normal functioning of the cell, hence resulting into the inability of control muscle functions (paralysis) (Yasumoto and Murata, 1993). STX is produced by several species of dinoflagellates of the genus *Alexandrium*, including *A. tamarense, A. cantenella, Gymnodium catenatum and Pyrodinium bahamense* (Yasumoto *et al*., 1993; Lefebvre *et al*., 2008). Of all the types of seafood poisoning, PSP poses the greatest threat to public health and fatal cases have been reported around the world (Faber, 2012). Along the South African coast, dinoflagellate *Gonyaulax* cellular *catenella* is responsible for PSP on the West Coast and it has caused several human deaths. The consumption of shellfish that have been contaminated with saxitoxin also causes paralysis to the consumers (Faber, 2012). The first symptoms that are observed within 5-30 mins of eating poisonous mussels are tingling, prickling, stinging or burning sensation of the lips,

tongue and fingertips. Numbness of the arms, legs and neck, other symptoms that develop later includes dizziness, general muscle in-coordination, headache, vomiting and impaired respiration. Very severe cases resulted into complete paralysis or death by respiratory failure which may occur within 2-24 h. In non-lethal cases, victims begin to recover after 12 h with no effects lasting longer than a few days (Andreson, 1989). The distribution and frequency of STX continues to increase globally with a recorded 2,000 cases reported per year and a 15% human mortality rate (Kellmann *et al*., 2008).

### 2.1.2 Diarrhetic shellfish poisoning (DSP)

Diarrhetic shellfish poisoning is caused by the consumption of shellfish contaminated with okadaic toxins produced by dinoflagellate *Dinophysis acuminata*. The symptoms usually occur between 30 mins to 2 h of consumption and they include diarrhoea, nausea, vomiting, abdominal cramps and chills (Jorge, 2008). Studies show that there has been no case of death recorded and full recovery from most severe cases is usually experienced within few days. Possibly, DSP has gone unreported on many occasions because of the relatively mild nature of the symptoms. In addition, the symptoms may be confused with those of gastroenteritis associated with the consumption of shellfish from polluted waters. DSP has only recently been recorded for the first time on the South African coast but this poisoning was first reported in Netherlands in the 1960's and later then gained prevalence in Europe, Japan and South America (NIEHS, 2000).

### 2.1.3 Amnesic shellfish poisoning (ASP)

Amnesic shellfish poisoning is caused by neurotoxin domoic acid that is produced by *Nitzschia pungens*, a species of phytoplankton belonging to a group of diatoms. ASP was first recorded off the coast of Canada in 1987 where three deaths and over 100 confirmed cases of acute intoxications followed the consumption of cultured mussels were recorded. Poisoning is

potentially fatal within 24 h of consumption and the symptoms include abdominal cramps, vomiting, nausea and more severe neurological symptoms such as dizziness, headache, seizures, disorientation, short-term memory loss, respiratory difficulty and coma occured. Although, along the South African coast ASP has not been recorded (Jorge, 2008).

### 2.1.4 Neurotoxic shellfish poisoning (NSP)

Neurotoxic shellfish poisoning is caused by brevetoxins produced by *Karenia brevis*; a dinoflagellate that occurs principally in the Gulf of Mexico, east and west coasts of Florida and Carolinas. Blooms of *K. brevis* can kill millions of fish and birds within few days and blooms can last for weeks. In addition to transmission through contaminated fish and shellfish, the poison can be disseminated in aerosol form by breaking waves and can cause respiratory distress and asthma-like symptoms (Independent online, 2005). Along South African coast, dinoflagellate *Gymnodinium nagasakiense* is known to cause NSP. Most outbreaks have been reported from False Bay, where they are responsible for the olive-green discolouration of the seawater during autumn. Symptoms of NSP are sensory abnormalities which include dizziness, tingling sensations, dilated pupils, hot-cold reversal, diarrhoea, vomiting abdominal pain, hypersensitivity to cold, inversion of cold and hot sensations, numbness, ataxia, tremors, muscular stiffness, hypersensitivity of the nipples, depression, itching and general muscle pain (Rey, 2007). These symptoms usually disappear within three days and no human deaths have yet been documented. It may also cause irritations of the human eyes, nose and throat through contact with sea spray (Jorge, 2008).

### 2.1.5 Aerosol (air-borne) Toxins

Aerosol can be referred to as any tiny particles or droplets suspended in air. Red tide toxin (brevetoxin) has been recognized to cause both respiratory irritation after inhalation of toxic sea spray and neurotoxic poisoning after consumption of toxic shellfish. Brevetoxins could be

transferred from water to air in the wind-powered whitecapped waves during red tide outbreak (NOAA, 2008). The symptoms include discomfort of coughing, burning of the nasal passages, difficulty in breathing, stinging eyes and irritation to the skin. It has been reported that the discomforts experienced are considerable and symptoms are usually relieved upon leaving the area with no long-term effects (NOAA, 2008).

### 2.2.0 Impacts of red tide (HABs)

According to Jorge (2008), Harmful Algal Blooms have effects on human healths, environments, local and regional economies and can impact natural ecological communities directly or indirectly through the production of excessive biomass. It could cause massive mortalities of fish, birds and marine mammals including endangered and protected species. It could also damage coral reefs, shade submerged vegetation and decrease water quality. HABs have been known to cause a variety of other less obvious effects including decreased reproduction, increased susceptibility to predation and others. Also, it could have economic impacts by affecting seafood industries and businesses, tourism and aquaculture operations. It might impact the quality of life, sometimes for extended periods of time of people living in and around affected areas and could result in disruption of social relationships and cultural practices tied to coastal resources (Jewett *et al*., 2007).

### 2.3.0 Assessing the Impact of red tide on Public Health

In addition to the killing of marine mammals, anyone who consumes contaminated shellfish that has fed on K*arenia brevis* may suffer from neurotoxic shellfish poisoning, a food poisoning that can cause severe gastrointestinal and neurologic symptoms. The human health effects associated with eating brevetoxin-tainted shellfish are well documented. However, scientists know little about how other types of environmental exposures to brevetoxin-such as breathing the air near red tides or swimming in red tides-may affect humans. Anecdotal

evidence suggests that people who swim among brevetoxins or inhale brevetoxins dispersed in the air may experience irritation of the eyes, nose, and throat, as well as coughing, wheezing, and shortness of breath. Additional evidence suggests that people with existing respiratory illness, such as asthma, may experience these symptoms more severely (CDC, 2012).

### 2.4.0 Molecular Detection of Microorganisms

The modern molecular techniques have provided an opportunity to overcome the conventional culture-based method of culturing and identifying microorganisms. They has also increased the understanding of microbial diversity and functionality in the environment (Malik *et al*., 2008). These methods depend on the characterization of cellular constituents such as nucleic acids, proteins, fatty acids and other taxa-specific compounds (Rossello-Mora and Amann, 2001). The aforementioned molecules can be extracted directly from environmental samples without the need for culturing (Tsai and Olson, 1991) and analysis of the molecular composition can be used to elucidate the composition of the microbial community (Amann *et al*., 1995; Greene and Voordouw, 2003). Another advantage of this culture-independent molecular characterization is the capacity to preserve in situ metabolic function and the microbial community composition by immediately preserving samples (Moller *et al*., 1998; Wilson *et al*., 1999a).

These techniques are based on polymerase chain reaction (PCR)-amplification of specific target genes (rRNA genes) in order to obtain a large number of copies of DNA fragments (Gonzalez and Saiz-Jimenez, 2005). The commonly used molecular markers for the identification of genes that encode small subunit ribosomal RNA (rRNA) are 16S rRNA for prokaryotes and 18S rRNA for eukaryotes (Gonzalez and Saiz- Jimenez, 2005; Dakal and Arora, 2012). These genes are considered to be the gold standard and have been used

extensively in molecular detection and identification of microorganisms in environmental samples; because they are ubiquitous in all prokaryotic and eukaryotic organisms, structurally and functionally conserved and both possess highly conserved and variable regions (Rastogi and Sani, 2011). The 16S rRNA gene contains nine hyper-variable regions flanked by relatively conserved regions in the primary and secondary structures (Shah *et al*., 2011; Neelakanta and Sultana, 2013).

### 2.4.1 Polymerase chain reaction (PCR)

The polymerase chain reaction has become most widely used technology in the field of molecular study. This technique is based on the ability of a thermostable Taq DNA polymerase to copy a strand of DNA by elongation of complementary strands initiated from a pair of synthesized oligonucleotide primers over 20–45 cycles, hence yielding billions of copies of a portion of a desired gene, entire gene or gene clusters with high fidelity per reaction (Mullis and Faloona, 1987; Mullis, 1990; Farber, 1996; Wang *et al*., 2000). The basic components involved in PCR include DNA template, two primer sets that are complementary to the 3'end of each of the sense and anti-sense strand of the target DNA, deoxyribonucleotide triphosphate (dNTPs), $MgCl_2$, DNA polymerase which synthesize a new DNA strand, buffer solution which provides a suitable chemical environment for optimum activity and stabilizes the DNA polymerase and finally nuclease free water. PCR is the most widely used technique for the amplification of 16S rRNA, or its gene, prior to fingerprinting studies. PCR-based methods have been used in the detection and quantification of microorganisms found in soil and water (Wilson *et al*., 1999b), analysis of catabolic genes involved in the biodegradation of organic contaminants (Wilson *et al*., 1999a) and the detection of organisms or genes from contaminated environments.

PCR process may involve the use of two variants and these include one set of primer (single PCR) or multiple primers (multiplex PCR) which detect bacterial isolates. Simple PCR uses a pair of primers in a single amplification reaction, while multiplex PCR uses multiple primer pairs simultaneously to amplify several genes in a single reaction (Markoulatos *et al*., 2002). Other polymerase chain reaction assays include real-time PCR, nested PCR, reverse-transcription PCR and many more (Adzitey *et al*., 2013). These techniques have been used to determine the absolute or relative number of bacteria in various samples (Heid *et al*., 1996; Shi *et al*., 2010), to quantify the proportion of microorganisms and assessment of microbial community in Antarctic soil (Powell *et al*., 2006) and detect only viable cells of pathogens (Sharma, 2006; Shi *et al*., 2010).

### 2.4.2 Principle of polymerase chain reaction (PCR)

There are three major steps involved in PCR (Fig. 3), these steps are repeated for 20 to 40 cycles on an automated thermocycler which can heat and cool the tubes with the reaction mixture in a very short time. These steps are;

a) Denaturation at 94 °C: During the denaturation process, the high temperature melts the double-stranded DNA into a single-stranded DNA and this causes all the enzymatic reactions to stop.
b) Annealing at 54 °C: At this stage, the two oligonucleotides bind to the appropriate complementary strands. The temperature for this step varies depending on the size of the primer and its homology to the target DNA.
c) Extension at 72 °C: During this process, the DNA polymerase uses its polymerase activity to extend the primer. The most commonly used enzyme is *Taq* polymerase, the DNA polymerase from the thermophilic bacteria called *Thermus aquaticus*. These steps are repeated from 20- 40 cycles and a large quantity of DNA can be produced in few hours.

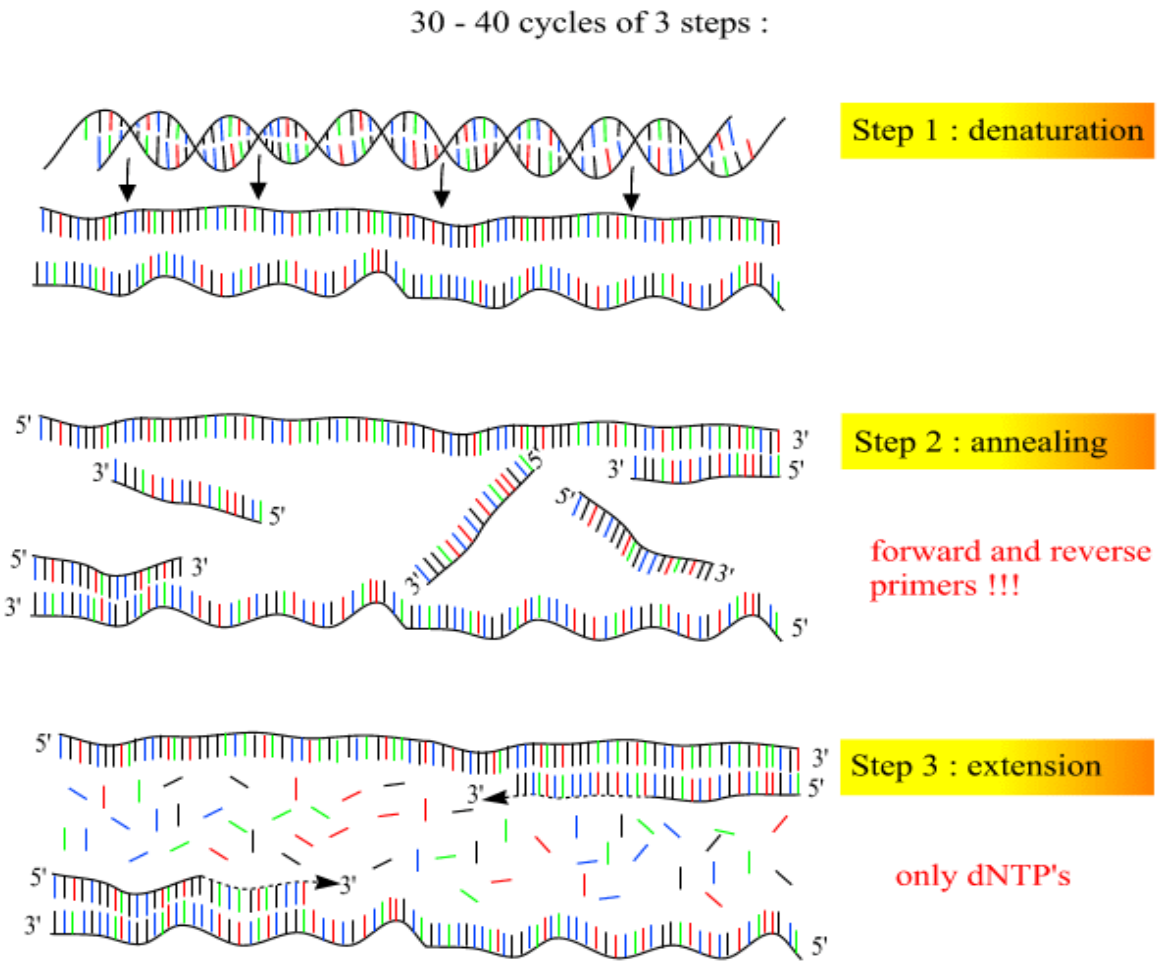


**Figure 3:** Principle of PCR.

**(Source: Pictured by Andy Vierstraete, 1999).**

### 2.5.0 Fingerprinting Techniques and their Application in Microbial Community Profiling

Genetic fingerprinting includes technique based on the electrophoretic analysis of PCR products amplified from metagenomic DNA. Well-characterized molecular techniques used to obtain microbial fingerprints from environmental samples include denaturing and temperature gradient gel electrophoresis (DGGE/TGGE), single-strand conformation polymorphism (SSCP), amplified ribosomal DNA restriction analysis (ARDRA), terminal restriction fragment length polymorphism (T-RFLP), and (automated) ribosomal intergenic

spacer analysis (ARISA)( Justé *et al*., 2008; Otlewska *et al*., 2014 ). These techniques have been extensively used for genetic analysis of microbial communities since they provide a relatively comprehensive description of a given community. Most especially, these techniques are suitable in the comparison of microbial community compositions between different treatments, environments or situations (Justé *et al*., 2008). In addition to the detection of microbial cells, fingerprinting approaches also help in the determination of profiles of quantitative changes in the composition of an ecosystem, but they do not provide direct taxonomic identifications (Otlewska *et al*., 2014).

### 2.5.1 Denaturing and temperature gradient gel electrophoresis (DGGE/TGGE)

Denaturing gradient gel electrophoresis (Fischer and Lerman, 1979, 1983) or temperature gradient gel electrophoresis (Rosenbaum and Riesner, 1987) is used in the separation of amplified rDNA fragments of the same length but with different base pair compositions. The separation of bands in both DGGE and TGGE depends on the decreased electrophoretic mobility of partially melted double stranded DNA molecules in polyacrylamide gels containing a linear gradient of DNA denaturants or a linear temperature gradient (Muyzer and Smalla, 1998). The PCR-amplified DNA fragments are generally limited in size to +/- 500 bp and are separated on the basis of sequence differences, not variation in length. The number of bands produced during DGGE or TGGE is proportional to the number of dominant species in the sample.

PCR-DGGE/TGGE fingerprinting facilitates the identification of individual populations by hybridisation analysis of patterns with specific probes or by sequence analysis of individual bands. It also monitors variations in microbial genetic diversity, providing a minimum estimate of richness of predominant community members. Furthermore, it is used to investigate the diversity of microbial communities (Kawai *et al*., 2002; Theunissen *et al*.,

2005), to determine the spatial and temporal variability of bacterial populations and to monitor community behaviour after natural or induced environmental perturbations (Schäfer, 2001). DGGE is also used to study communities of various habitats, direct amplification and analysis of 16S rRNA genes examined the predominant sequences in mixed PCR products amplified from environmental samples such as activated sludge (Boon *et al*., 2002), rivers (Sekiguchi *et al*., 2002), Asian green-lipped mussel (*Perna viridis*) (Chai *et al*., 2009), sediments, water column, hydrothermal vents, microbial mats, mesocosms or sewage treatment plants and soil (Nakatsu *et al*., 2000). For analyzing total community structures, DGGE has been used to separate 16S ribosomal DNA (rDNA) fragments, which are directly amplified from community DNA by PCR using universal primers (Muyzer and Smalla, 1998; Muyzer, 1999).

The limitations of DGGE/ TGGE in microbial community studies include limitation of 16S rRNA sequences information derived from microbial populations to 500 bp fragments which may lack the specificity required for the phylogenetic identification of some organisms (Gilbride *et al*., 2006); due to the existence of multiple copies of rRNA in an organism, multiple bands for a single species may occur (Nubel *et al*., 1997); and different 16S rRNA sequences may have identical mobilities. Band intensity may not truly reflect the abundance of microbial population (strong band may just mean more copies) and perceived community diversity may be underestimated. Also, DGGE/TGGE analysis of microbial communities produces a complex profile which can be quite sensitive to spatial and temporal sampling variation (Murray *et al*., 1998).

### 2.5.2 Single-strand conformation polymorphism (SSCP)

SSCP is based on separation of PCR product having similar-length DNA fragments according to their differences in mobility in polyacrylamide gel. This separation is caused by

differences in the secondary structure of folded DNA. PCR products are denatured and followed by electrophoretic separation in a non-denaturing polyacrylamide gel (Cetecioglu *et al.*, 2012). This technique is simple and straightforward compared to other fingerprinting methods because it does not require the use of Guanine- Cytosine (GC) clamped primers and gradient gel (Rastogi and Sani, 2011).

This technique have been used for the analysis of microbial communities in different matrices ranging from water ( Lin *et al*., 2007) and soils (Sliwinski and Goodman, 2004; Witzig *et al*., 2006) to anaerobic digestors (Zumstein *et al*., 2000; Leclerc *et al*., 2001) and the detection of mutations most especially in human genetic (Orita *et al*., 1989; Hayashi, 1991; Lee *et al*., 1996; Hamzeiy *et al*., 2002) as well as for the differentiation and typing of isolates (Wagner *et al*., 2000; Moore *et al*., 2000). Also, using SSCP in combination with an automated DNA sequencer, 16S rRNA genes obtained from PCR with several bacterial species of clinical importance could be differentiated (Widjojoatmodjo *et al*., 1995). SSCP have also been used to distinguish between 16S-23S rRNA interspacer regions of selected bacterial strains (Scheinert *et al*., 1996).

One of the disadvantages of SSCP is the formation of more than one stable conformation out of a single-strand DNA sequence; this fragment could be represented by multiple bands on gel (Rastogi and Sani, 2011; Cetecioglu *et al.*, 2012). Furthermore, there is a high rate of re-annealing of DNA strands after an initial denaturation during electrophoresis (Selvakumar *et al*., 1997) especially in the analysis of high-diversity communities since these samples typically require high DNA concentrations for loading onto the gels (Schwieger and Tebbe, 1998). However, this problem could be avoided by using a 5'-phosphorylated primer (Schwieger and Tebbe, 1998); this allows selective removal of the corresponding phosphorylated strand through digestion with lambda exonuclease.

Another disadvantage of SSCP is the appearance of more than one band from a double stranded PCR product after electrophoresis. Typically, three bands are detectable, two DNA single strands and one double stranded DNA molecule, but several conformations of one product may coexist in one gel. Similar conformations of both DNA single strands may result in the detection of fewer than three bands from one organism. Finally, as already described for DGGE and TGGE, the formation of heteroduplex DNA from PCR products with similar sequences occurs frequently (Lee and Kim, 1996).

### 2.5.3 Amplified ribosomal DNA restriction analysis (ARDRA)

Amplified ribosomal DNA restriction analysis (ARDRA), also known as restriction fragment length polymorphism (RFLP) is based on a principle that the recognition site of restriction enzymes is different for different microbial species (Cetecioglu *et al.*, 2012). In this method, the ribosomal RNA genes 16S rRNA or 18S rRNA are amplified by PCR from environmental DNA as a template. The PCR products are then digested into specific DNA fragments with tetracutter restriction endonucleases. Restriction fragments are then separated on agarose or polyacrylamide gels (Gich *et al*., 2000; Rastogi and Sani, 2011; Sklarz *et al.*, 2011; Cetecioglu *et al.*, 2012). The pattern of the bands in ARDRA could be used to screen clones or measure bacterial community structure (Kirk *et al*., 2004), also used in microbial identification (Vaneechoutte *et al*., 1992, 1995; Kita-Tsukamoto *et al*., 2006; Krizova *et al*., 2006) and microbial community studies (Weidner *et al*., 1996; Bai *et al*., 2006; Gich *et al*., 2010).

ARDRA technique is useful for rapid monitoring of microbial communities over time or comparing biodiversity in response to changing environmental conditions. Although sometimes, it becomes impossible to separate the restriction profiles obtained from microbial communities by agarose or polyacrylamide electrophoresis (Rastogi and Sani, 2011). This

method was used to evaluate the biodiversity of cyanobacteria on stone monuments in the Boboli Gardens in Florence (Tomaselli *et al.*, 2000), detection of microorganisms forming a biofilm on Bayon temple sandstone of Angkor Thom, Cambodia (Lan *et al.*, 2010) and those causing discolouration of the walls of Palazzo De Francesco in Campania, Italy (De Felice *et al.*, 2010) and in the screening of 16S and 18S rRNA clones library with appropriate restriction enzymes (e.i. *Rsa*I, *Msp*I) (Otlewska *et al*., 2014).

In contrast to T-RFLP, all digested fragments are detected, increasing the level of resolution, increases the complexity of the profiles, thus obstructing comparison and interpretation of complex patterns. ARDRA utilises the variations in homologous DNA sequences to characterize bacteria. Although a single restriction enzyme generally does not provide sufficient resolution, hence, multiple restriction enzymes have to be used either separately or in combination to obtain the desired resolution. Another setback of this method is the limited staining sensitivity in gels resulting in the suppression of bands from less abundant community members, in a loss of phylogenetic information or provides little or no information about the type of microorganisms present in the sample (Tiedje *et al*., 1999; Kirk *et al*., 2004; Gich *et al.*, 2010). As a result, this technique is only preferred when the community is dominated by a few members (Fries *et al*., 1997). Also, it is of limited use for demonstrating the presence of specific phylogenetic groups or for estimating complexity of communities. Acinas *et al*. (1997) reported the use of ARDRA fingerprinting to study spatial and temporal variation in bacterial marine plankton diversity. Nevertheless, despite its potential, to our knowledge, this method has not yet been used to study microorganisms in food.

### 2.5.4 Terminal restriction fragment length polymorphism (T-RFLP)

Terminal-restriction fragment length polymorphism is an improvement of ARDRA technique. The primers used in T-RFLP analysis are 5'-fluorescently labelled which amplifies the marker genes and the resultant PCR products are visualised and quantified (Liu *et al*., 1997). The use of 5'-fluorescently labelled tagged primers allows only labelled terminal restriction fragments (TRFs) to be detected (Marsh, 1999), which greatly simplifies the restriction profiles; hence enabling the analysis of complex communities as well as providing information on diversity as each visible band represents a single operational taxonomic unit or ribotype (Tiedje *et al*., 1999; Muyzer, 1999; Rastogi and Sani, 2011; Cetecioglu *et al.*, 2012). The number, size and peaks height of resulting terminal restriction fragments is used to analyse the biodiversity of microbial communities (Rastogi and Sani, 2011) and specially configured databases are used to determine the qualitative structure of the microbial population (Więckowicz, 2009; Capodicasa *et al.*, 2010).

T-RFLP allows very sensitive detection and because of its high throughput capacity, it performs well in surveys with large sample numbers, for instance to ascertain spatial or temporal changes in a community structure. Nevertheless, this method also has some limitations to accurately predict microbial community structures. Incomplete or non-specific restriction can lead to overestimation of the diversity since the number of fragments increases. However, restriction efficiency can be checked by inclusion of an internal standard in the restriction step (Marsh *et al*., 2000). Overestimation of diversity can also be generated by pseudoterminal restriction fragments (Egert and Friedrich, 2003). These fragments can be produced upon intramolecular folding of single-stranded products, creating transient structures that are accessible for digestion. However, this problem can be addressed by treatments with single-strand-specific nucleases (Egert and Friedrich, 2003).

Apart from the limitations related to the determination of the diversity within a community, one of the major restrictions for identification is variation between the in silico predicted and the experimentally obtained TRF lengths (Kaplan and Kitts, 2003; Pandey *et al*., 2007). Kaplan and Kitts (2003), measured mobility differences within phylogenetic groups, presumably due to sequence variability and also detected striking mobility variations caused by fluctuations in ambient temperature between runs. As a consequence, tolerance ranges for length assignment from 1 to 2 bases are generally used to allow matching with database entries, which, in turn, increase the number of species associated with a TRF. Another sizing variable is introduced by the choice of the fluorescent dye (Pandey *et al*., 2007). Therefore, the primers and labeling dyes should be carefully evaluated, optimized and standardized for an individual community under investigation.

Despite these shortcomings, T-RFLP has become a valuable molecular tool for microbial community analysis, especially when high throughput and high sensitivity are required without the need for direct sequence information. The usefulness of T-RFLP in microbial ecology has been extensively demonstrated in diverse research domains (Jernberg *et al*., 2005; Rasche *et al*., 2006). Increasingly, also in the food industry, T-RFLP is successfully used to describe microbial populations, e.g. in fishery (Wilson *et al*., 2008), yogurt and cheese production (Rademaker *et al*., 2005, 2006; Sanchez *et al*., 2006) and in sugar industry (Juste´ *et al*., 2008b).

Also, it has been used to study the microbial community composition of lake sediments (Konstantinidis *et al*., 2003), characterize microbial communities recovered from surrogate minerals (Reardon *et al*., 2004) de-chlorinating bacteria from a basalt aquifer (Macbeth *et al*., 2004) and also it was observed that the long-term presence of benzene in groundwater reduced bacterial diversity and community structure compared with that of clean groundwater sources (Fahy *et al*., 2005). In addition, Denaro *et al*. (2005) observed the steadiness of T-

RFLP for monitoring microbial populations characterized by low diversity and high relative abundances of a few dominant groups in a hydrocarbon-polluted marine environment. In contaminated soils, T-RFLP has also been used successfully in describing bacterial communities of polychlorinated biphenyl (PCB) contaminated soils (Fedi *et al*., 2005) and microbial communities that reductively dechlorinate Trichloroethylene to ethene (Richardson *et al*., 2002).

### 2.5.5 (Automated) ribosomal intergenic spacer analysis RISA

Automated ribosomal intergenic spacer analysis (ARISA) is a high resolution, culture-independent method suitable for an analysis of biodiversity and richness of microbial communities. This technique is based on the size and nucleotide sequence variability of the intergenic spacer region (IGS) present between the small (16S) and large (23S) ribosomal subunits (Cardinale *et al.*, 2004; Popa *et al.*, 2009). The IGS fragments are analysed by an automated capillary electrophoresis system containing a laser detector. In general, the intergenic spacer region, depending on the species, has vary between approximately 400 and 1200 base pairs (Fisher and Triplett, 1999) variability and this unique feature facilitates taxonomic identification of organisms (Spiegelman *et al*., 2005).

The main limitation of ARISA is that a large number of peaks are obtained in the cause of using universal primers. In addition, it is very difficult to interpret results for fingerprints obtained for uncultured microorganisms (Popa *et al.*, 2009). Another problems associated with ARISA are the preferential amplification of shorter templates (Fisher and Triplett, 1999) and because IGS length variation within a single genome (Mateos and Markow, 2005), a single organism can contribute to more than one signal. In order to increase reproducibility and standardization, Cardinale *et al*. (2004) evaluated different primer sets with respect to

universality, sensitivity and reproducibility and selected the most suitable primers for ARISA of environmental bacterial communities.

The ARISA technique was applied by Cuzman *et al* (2010) to study the species structure of biofilms formed in historic fountains in Italy and Spain. Also, it has been regularly applied in microbial ecology (Ranjard *et al*., 2001; Schloss *et al*., 2003; Brown *et al*., 2005). However, only one study (Cardinale *et al*., 2004) reports on a community analysis of food samples, i.e. goat milk and ensiled corn. Nevertheless, for the molecular characterization of food isolates, ARISA has been frequently exploited (Kabadjova *et al*., 2002; De Angelis *et al*., 2007).

### 2.6.0 Deoxyribonucleic acid (DNA) sequencing techniques

Deoxyribonucleic acid (DNA) sequencing techniques involve the process by which the precise order of nucleotides within a DNA molecule is determined (adenine, cytosine, guanine and thymine) (Pettersson *et al*., 2009). This technique has been used by different researchers in the identification, typing, characterization and taxonomic classification of unknown or novel pathogens isolates. The 16S rRNA gene is the most targeted gene during PCR amplification, that is, the amplicon sequence is first determined, followed by a DNA sequence-based phylogenetic analysis which is used to specifically identify the species (Dumler *et al*., 1995). The sequence can be compared with databases for identification of the species of a microorganism. The used 16S rRNA gene sequence analysis has been used to identify bacteria isolates from duck houses at the genus level (Martin *et al*., 2010). Martin, (2010) also reported that based on the 16S rRNA gene analyses some isolates were closely related to organisms that may cause pulmonary health effects. Bacterial isolates were also identified using 16S rRNA gene analysis from duck hatcheries and these isolates (more than 50 %) were phylogenetically closely related (Martin and Jackel, 2011).

### 2.6.1 Principles of Sequencing

Sequencing involves the steps in PCR reactions but in this case, only one primer is required so that only one strand copied will be copied. It also requires the use of deoxyribonucleotide triphosphates (dNTPs) and deoxyribonucleotide triphosphates (ddNTP's) in the reaction. During sequencing reaction, the annealing temperature for DNA polymerase will be reduced because of the incorporation of ddNTP's which are chemically modified with a fluorescent label so as to have time to incorporate the 'strange' molecules to the template. The bases complementary to the template are coupled to the primer on the 3' side adding dNTP's or ddNTP's from 5' to 3', reading from the template from 3' to 5' side. When a ddNTP is incorporated, the extension reaction stops because a ddNTP contains an H-atom on the 3rd carbon atom (dNTP's contain an OH-atom on that position). Since the ddNTP's are fluorescently labeled, it is possible to detect the color of the last base of this fragment on an automated sequencer (Pettersson *et al*., 2009).

After the sequencing reactions, the mixture of strands, all of different length and all ending on a fluorescently labelled ddNTP will be separated on an acrylamide gel. The fluorescently labelled fragments that migrate through the gel are passing a laser beam at the bottom of the gel. The laser exited the fluorescent molecule which sends out light of a distinct color. The light will be collected and focused by lenses into a spectrograph. Based on the wavelength, the spectrograph separates the light across a charge-coupled devices camera and each base has its own color, so that the sequencer can detect the order of the bases in the sequenced gene.

**References**


Acinas, S. G., Rodríguez-Valera, F., & Pedrós-Alió, C. (1997). Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. *FEMS Microbiology Ecology*, *24*(1), 27-40.

Adzitey, F., Huda, N., & Ali, G. R. R. (2013). Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks. *3 Biotechnology*, *3*(2), 97-107.

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), 143-169.

Anderson, D. M. (1989). Toxic algal blooms and red tides: a global perspective. *Red tides: Biology, Environmental Science and Toxicology*, 11-16.

Bai, Y., Yang, D., Wang, J., Xu, S., Wang, X., & An, L. (2006). Phylogenetic diversity of culturable bacteria from alpine permafrost in the Tianshan Mountains, northwestern China. *Research in Microbiology*, *157*(8), 741-751.

Boon, N., Windt, W., Verstraete, W., & Top, E. M. (2002). Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different waste-water treatment plants. FEMS *Microbiology Ecology*, *39*, 101-112.

Brown, M. V., Schwalbach, M. S., Hewson, I., & Fuhrman, J. A. (2005). Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *EnvironmentalMicrobiology*, *7*, 1466–1479.

Capodicasa, S., Fedi, S., Porcelli, A. M., & Zannoni, D. (2010). The microbial community dwelling on a biodeteriorated 16th century painting. *International Biodeterioration and Biodegradation*, *64*(8), 727-733.

Cardinale, M., Brusetti, L., Quatrini, P., Borin, S., Puglia, A. M., Rizzi, A., & Daffonchio, D. (2004). Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Applied and Environmental Microbiology*, *70*(10), 6147-6156.

Centers for Disease Control and Prevention (2012). http://www.cdc.gov/nceh/hsb/hab/

Centers for Disease Control and Prevention (CDC. (2008). Illness associated with red tide--Nassau County, Florida, 2007. *MMWR. Morbidity and mortality weekly report*, *57*(26), 717.

Cetecioglu, Z., Ince, O., & Ince, B. (2012) Gel electrophoresis based genetic fingerprinting techniques on environmental ecology. In *Gel electrophoresis– Advanced Techniques*. Magdeldin S, eds, pp 52–66. InTech, Croatia.

Chai, L. C., Fatimah, C. A., Norhisyam, M. S., Rozila, A., Nadzirah, A. S., & Natasha, L. H. Y. (2009). Rapid isolation of genomic DNA from Asian green-lipped mussel (*Perna viridis*) for random amplified microsatellite polymorphism. *International Food Research Journal* 16, 113-118.

Cuzman, O. A., Ventura, S., Sili, C., Mascalchi, C., Turchetti, T., D'Acqui, L. P., & Tiano, P. (2010). Biodiversity of phototrophic biofilms dwelling on monumental fountains. *Microbial Ecology*, *60*(1), 81-95.

Dakal, T. C. & Arora, P. K. (2012). Evaluation of potential of molecular and physical techniques in studying biodeterioration. *Reviews in Environmental Science and Biotechnology*, *11*(1), 71-104.

De Angelis, M., Di Cagno, R., Gallo, G., Curci, M., Siragusa, S., Crecchio, C., Parente, E., & Gobbetti, M. (2007). Molecular and functional characterization of Lactobacillus sanfranciscensis strains isolated from sourdoughs. *International Journal of Food Microbiology*, *114*, 69-82.

De Felice, B., Pasquale, V., Tancredi, N., Scherillo, S., & Guida, M. (2010). Genetic fingerprint of microorganisms associated with the deterioration of an historical tuff monument in Italy. *Journal of Genetics*, *89*(2), 253.

Denaro, R., D'auria, G., Di Marco, G., Genovese, M., Troussellier, M., Yakimov, M. M., & Giuliano, L. (2005). Assessing terminal restriction fragment length polymorphism suitability for the description of bacterial community structure and dynamics in hydrocarbon-polluted marine environments. *Environmental Microbiology*, *7*(1), 78-87.

Dolah, F. M. V., Roelke, D., & Greene, R. M. (2001). Health and ecological impacts of harmful algal blooms: risk assessment needs. *Human and Ecological Risk Assessment: An International Journal*, *7*(5), 1329-1345.

Dumler, J. S., Asanovich, K. M., Bakken, J. S., Richter, P., Kimsey, R., & Madigan, J. E. (1995). Serologic cross-reactions among Ehrlichia equi, Ehrlichia phagocytophila, and human granulocytic Ehrlichia. *Journal of Clinical Microbiology*, *33*(5), 1098-1103.

Egert, M. & Friedrich, M. W. (2003). Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Applied and Environmental Microbiology*, *69*(5), 2555-2562.

Faber, S. (2012). Saxitoxin and the induction of paralytic shellfish poisoning. *Journal of Young Investigators*, *23*(1), 1-7.

Fahy, A., Lethbridge, G., Earle, R., Ball, A. S., Timmis, K. N., & McGenity, T. J. (2005). Effects of long-term benzene pollution on bacterial diversity and community structure in groundwater. *Environmental Microbiology*, *7*(8), 1192-1199.

Farber, J. M. (1996). An introduction to the hows and whys of molecular typing. *Journal of Food Protection*, *59*(10), 1091-1101.

Fedi, S., Tremaroli, V., Scala, D., Perez-Jimenez, J. R., Fava, F., Young, L., & Zannoni, D. (2005). T-RFLP analysis of bacterial communities in cyclodextrin-amended bioreactors developed for biodegradation of polychlorinated biphenyls. *Research in Microbiology*, *156*(2), 201-210.

Fischer, S. G. & Lerman, L. S. (1979). Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell*, *16*(1), 191-200.

Fischer, S. G. & Lerman, L. S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proceedings of the National Academy of Sciences*, *80*(6), 1579-1583.

Fisher, M. M. & Triplett, E. W. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology*, *65*(10), 4630-4636.

Fries, M. R., Hopkins, G. D., McCarty, P. L., Forney, L. J., & Tiedje, J. M. (1997). Microbial Succession during a Field Evaluation of Phenol and Toluene as the Primary Substrates for Trichloroethene Cometabolism. *Applied and Environmental Microbiology*, *63*(4), 1515-1522.

Gich, F. B., Amer, E., Figueras, J. B., Charles A Abella, C. A. A., Balaguer, M. D., & Poch, M. (2010). Assessment of microbial community structure changes by amplified ribosomal DNA restriction analysis (ARDRA). *International Microbiology*, *3*(2), 103-106.

Gilbride, K. A., Lee, D. Y., & Beaudette, L. A. (2006). Molecular techniques in wastewater: understanding microbial communities, detecting pathogens, and real-time process control. *Journal of Microbiological Methods, 66*(1), 1-20.

González Grau, J. M. & Sáiz-Jiménez, C. (2005). Unknown microbial communities on Rock Art paintings. Consequences for conservation and future perspectives. *Coalition 10*, 4–7.

Greene, E. A. & Voordouw, G. (2003). Analysis of environmental microbial communities by reverse sample genome probing. *Journal of Microbiological Methods, 53*(2), 211-219.

Hamzeiy, H., Vahdati-Mashhadian, N., Edwards, H. J., & Goldfarb, P. S. (2002). Mutation analysis of the human CYP3A4 gene 5′ regulatory region: population screening using non-radioactive SSCP. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 500*(1), 103-110.

Hayashi, K. (1991). PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *Genome Research, 1*(1), 34-38.

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research, 6*(10), 986-994.

Hibbeln, J. R., Davis, J. M., Steer, C., Emmett, P., Rogers, I., Williams, C., & Golding, J. (2007). Maternal seafood consumption in pregnancy and neuro developmental outcomes in childhood (ALSPAC study): an observational cohort study. *Lancet 369*, 578-585.

Independent Online (2005). http://www.whoi.edu/science/B/redtide/notedevents/foreign/SouthAfrica/SouthAfrica_3-14-05.html).

Jernberg, C., Sullivan, Å., Edlund, C., & Jansson, J. K. (2005). Monitoring of antibiotic-induced alterations in the human intestinal microflora and detection of probiotic strains by use of terminal restriction fragment length polymorphism. *Applied and Environmental Microbiology*, *71*(1), 501-506.

Jewett, E. B., Lopez, C. B., Dortch, Q., & Etheridge, S. M. (2007). National Assessment of Efforts to predict and Respond to Harmful Algal Blooms in U.S. Waters. InterimReport. Interagency Working Group on Harmful AlgalBlooms, Hypoxia, and Human Health of the Joint Subcommitteeon Ocean Science and Technology. Washington, DC.

Jorge R. (2008). The Institute of Food and Agricultural Sciences, UF/Extension, Entomology and Nematology Department, Florida Medical Entomology Laboratory, Vero Beach, FL, 32962. (http://edis.ifas.ufl.edu ).

Justé, A., Lievens, B., Klingeberg, M., Michiels, C. W., Marsh, T. L., & Willems, K. A. (2008). Predominance of Tetragenococcus halophilus as the cause of sugar thick juice degradation. *Food Microbiology, 25*(2). 413-421.

Justé, A., Thomma, B. P. H. J., & Lievens, B. (2008b). Recent advances in molecular techniques to study microbial communities in food-associated matrices and processes. *Food Microbiology, 25*(6), 745-761.

Kabadjova, P., Dousset, X., Le Cam, V., & Prevost, H. (2002). Differentiation of closely related Carnobacterium food isolates based on 16S–23S ribosomal DNA intergenic spacer region polymorphism. *Applied and Environmental Microbiology, 68,* 5358-5366.

Kaplan, C. W. & Kitts, C. L. (2003). Variation between observed and true terminal restriction fragment length is dependent on true TRF length and purine content. *Journal of Microbiological Methods, 54*(1), 121-125.

Kawai, M., Matsutera, E., Kanda, H., Yamaguchi, N., Tani, K., & Nasu M. (2002). 16S ribosomal DNA-based analysis of bacterial diversity in purified water used in pharmaceutical manufacturing processes by PCR and denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*, *68*, 699-704.

Kellmann, R., Mihali, T. K., Jeon, Y. J., Pickford, R., Pomati, F., & Neilan, B. A. (2008). Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Applied and Environmental Microbiology*, *74*(13), 4044-4053.

Kirk, J. L., Beaudette, L. A., Hart, M., Moutoglis, P., Klironomos, J. N., Lee, H., & Trevors, J. T. (2004). Methods of studying soil microbial diversity. *Journal of Microbiological Methods*, *58*(2), 169-188.

Kita-Tsukamoto, K., Yao, K., Kamiya, A., Yoshizawa, S., Uchiyama, N., Kogure, K., & Wada, M. (2006). Rapid identification of marine bioluminescent bacteria by amplified 16S ribosomal RNA gene restriction analysis. *FEMS Microbiology Letters*, *256*(2), 298-303.

Konstantinidis, K. T., Isaacs, N., Fett, J., Simpson, S., Long, D. T., & Marsh, T. L. (2003). Microbial diversity and resistance to copper in metal-contaminated lake sediment. *Microbial Ecology*, *45*(2), 191-202.

Křížová, J., Španová, A., & Rittich, B. (2006). Evaluation of amplified ribosomal DNA restriction analysis (ARDRA) and species-specific PCR for identification of Bifidobacterium species. *Systematic and Applied Microbiology*, *29*(1), 36-44.

Lan, W., Li, H., Wang, W. D., Katayama, Y., & Gu, J. D. (2010). Microbial community analysis of fresh and old microbial biofilms on Bayon temple sandstone of Angkor Thom, Cambodia. *Microbial Ecology*, *60*(1), 105-115.

Leclerc, M., Delbes, C., Moletta, R., & Godon, J. J. (2001). Single strand conformation polymorphism monitoring of 16S rDNA Archaea during start-up of an anaerobic digester. *FEMS Microbiology Ecology*, *34*(3), 213-220.

Lee, D. H., Zo, Y. G., & Kim, S. J. (1996). Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. *Applied and Environmental Microbiology*, *62*(9), 3112-3120.

Lefebvre, K. A., Bill, B. D., Erickson, A., Baugh, K. A., O'Rourke, L., Costa, P. R., Nance, S., & Trainer, V. L. (2008). Characterization of intracellular and extracellular saxitoxin levels in both field and cultured Alexandrium spp. samples from Sequim Bay, Washington. *Marine Drugs*. *6*, 103-116.

Lin, C. W., Lai, C. Y., Chen, L. H., & Chiang, W. F. (2007). Microbial community structure during oxygen-stimulated bioremediation in phenol-contaminated groundwater. *Journal of Hazardous Materials*, *140*(1), 221-229.

Liu, W. T., Marsh, T. L., Cheng, H., & Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology*, *63*(11), 4516-4522.

Macbeth, T. W., Cummings, D. E., Spring, S., Petzke, L. M., & Sorenson, K. S. (2004). Molecular characterization of a dechlorinating community resulting from in situ biostimulation in a trichloroethene-contaminated deep, fractured basalt aquifer and comparison to a derivative laboratory culture. *Applied and Environmental Microbiology*, *70*(12), 7329-7341.

Malik, S., Beer, M., Megharaj, M., & Naidu, R. (2008). The use of molecular techniques to characterize the microbial communities in contaminated soil and water. *Environment International*, *34*(2), 265-276.

Markoulatos, P., Siafakas, N., & Moncany, M. (2002). Multiplex polymerase chain reaction: a practical approach. *Journal of Clinical Laboratory Analysis*, *16*(1), 47-51.

Marsh, T. L. (1999). Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Current Opinion in Microbiology*, *2*(3), 323-327.

Marsh, T. L., Saxman, P., Cole, J., & Tiedje, J. (2000). Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Applied and Environmental Microbiology*, *66*(8), 3616-3620.

Martin, E., & Jäckel, U. (2011). Characterization of bacterial contaminants in the air of a duck hatchery by cultivation based and molecular methods. *Journal of Environmental Monitoring*, *13*(2), 464-470.

Martin, E., Kämpfer, P., & Jäckel, U. (2010). Quantification and identification of culturable airborne bacteria from duck houses. *Annals of Occupational Hygiene*, *54(*2), 217-227.

Mateos, M., & Markow, T. A. (2005). Ribosomal intergenic spacer (IGS) length variation across the Drosophilinae (Diptera: Drosophilidae). *BMC Evoutionary Biology* 5, 46-59.

Møller, S., Sternberg, C., Andersen, J. B., Christensen, B. B., Ramos, J. L., Givskov, M., & Molin, S. (1998). In situ gene expression in mixed-culture biofilms: evidence of metabolic interactions between community members. *Applied and Environmental Microbiology*, *64*(2), 721-732.

Moore, J. E., Stanley, T., Smithson, R., O'Malley, H., & Murphy, P. G. (2000). Outbreak of campylobacter food-poisoning in Northern Ireland. *Clinical Microbiology and Infection*, *6*(7), 397-398.

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, *262*(4), 56-61.

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, *155*, 335.

Murray, A. E., Preston, C. M., Massana, R., Taylor, L. T., Blakis, A., Wu, K., & DeLong, E. F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Applied and Environmental Microbiology*, *64*(7), 2585-2595.

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology*, *2*(3), 317-322.

Muyzer, G. & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek*, *73*(1), 127-141.

Nakatsu, C. H., Torsvik, V., & Ovreas, L. (2000). Soil community analysis using DGGE of 16S rDNA polymerase chain reaction products. *Soil Science Society of America Journal*, *64*, 1382-1388.

National Institute of Environmental Health Sciences (NIEHS) Marine and Freshwater Biomedical Sciences Center Marine Toxins and Human Health Internet. Mar 3 2000 <http://www.rsmas.miami.edu/groups/niehs/> Accessed April 2014.

National Oceanic and Atmospheric Administration (NOAA), 2008 http://www.noaanews.noaa.gov/stories2008/20080709_redtide.html. Accessed 2014/10/20.

Neelakanta, G. & Sultana, H. (2013). The use of metagenomic approaches to analyze changes in microbial communities. *Microbiology Insights*, *6*, 37.

Nübel, U., Garcia-Pichel, F., & Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, *63*(8), 3327-3332.

Olsen, S. J., MacKinnon, L. C., Goulding, J. S., Bean, N. H., & Slutsker, L. (2000). Surveillance for foodborne-disease outbreaks—United States, 1993–1997. *MMWR CDC Surveillance Summaries, 49*(1): 1-62.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., & Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences*, *86*(8), 2766-2770.

Otlewska, A., Adamiak, J., & Gutarowska, B. (2014). Application of molecular techniques for the assessment of microorganism diversity on cultural heritage objects. *Acta Biochimica Polonica*, *61*(2), 217-225.

Pandey, J., Ganesan, K., & Jain, R. K. (2007). Variations in T-RFLP profiles with differing chemistries of fluorescent dyes used for labeling the PCR primers. *Journal of Microbiological Methods*, *68*(3), 633-638.

Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics*, *93*(2), 105-111.

Popa, R., Mashall, M. J., Nguyen, H., Tebo, B. M., & Brauer, S. (2009). Limitations and benefits of ARISA intra-genomic diversity fingerprinting. *Journal of Microbiological Methods*, *78*(2), 111-118.

Powell, S. M., Ferguson, S. H., Bowman, J. P., & Snape, I. (2006). Using real-time PCR to assess changes in the hydrocarbon-degrading microbial community in Antarctic soil during bioremediation. *Microbial Ecology*, *52*(3), 523-532.

Rademaker, J. L. W., Hoolwerf, J. D., Wagendorp, A. A., & Te Giffel, M. C. (2006). Assessment of microbial population dynamics during yoghurt and hard cheese fermentation and ripening by DNA population fingerprinting. *International Dairy Journal*, *16*(5), 457-466.

Rademaker, J. L. W., Peinhopf, M., Rijnen, L., Bockelmann, W., & Noordman, W. H. (2005). The surface microflora dynamics of bacterial smear-ripened Tilsit cheese determined by T-RFLP DNA population fingerprint analysis. *International Dairy Journal*, *15*(6), 785-794.

Ranjard, L., Poly, F., Lata, J. C., Mougel, C., Thioulouse, J., & Nazaret, S. (2001). Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability. *Applied and Environmental Microbioiology*, 67, 4479-4487.

Rasche, F., Hödl, V., Poll, C., Kandeler, E., Gerzabek, M. H., Van Elsas, J. D., & Sessitsch, A. (2006). Rhizosphere bacteria affected by transgenic potatoes with antibacterial activities compared with the effects of soil, wild-type potatoes, vegetation stage and pathogen exposure. *FEMS Microbiology Ecology*, *56*(2), 219-235.

Rastogi, G. & Sani, R. K. (2011). Molecular techniques to assess microbial community structure, function, and dynamics in the environment. In *Microbes and Microbial Technology* (pp. 29-57). Springer New York.

Reardon, C. L., Cummings, D. E., Petzke, L. M., Kinsall, B. L., Watson, D. B., Peyton, B. M., & Geesey, G. G. (2004). Composition and diversity of microbial communities recovered from surrogate minerals incubated in an acidic uranium-contaminated aquifer. *Applied and Environmental Microbioiology*, *70*(10), 6037-6046.

Rey, J. R. (2007). Ciguatera. University of Florida, IFAS, EDIS IN742; http://edis.ifas.ufl.edu/in742.

Richardson, R. E., Bhupathiraju, V. K., Song, D. L., Goulet, T. A., & Alvarez-Cohen, L. (2002). Phylogenetic characterization of microbial communities that reductively dechlorinate TCE based upon a combination of molecular techniques. *Environmental Science and Technology*, *36*(12), 2652-2662.

Rosenbaum, V. & Riesner, D. (1987). Temperature-gradient gel electrophoresis: thermodynamic analysis of nucleic acids and proteins in purified form and in cellular extracts. *Biophysical Chemistry, 26*(2), 235-246.

Rosselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiology Reviews, 25*(1), 39-67.

Sánchez, J. I., Rossetti, L., Martínez, B., Rodríguez, A., & Giraffa, G. (2006). Application of reverse transcriptase PCR-based T-RFLP to perform semi-quantitative analysis of metabolically active bacteria in dairy fermentations. *Journal of Microbiological Methods, 65*(2), 268-277.

Schäfer, H. (2001). *Dynamics of the genetic diversity of marine bacterial assemblages* (Doctoral dissertation, PhD thesis, Department of Biologie/Chemie. Universitat Bremen, Bremen).

Scheinert, P., Krausse, R., Ullmann, U., Söller, R., & Krupp, G. (1996). Molecular differentiation of bacteria by PCR amplification of the 16S–23S rRNA spacer. *Journal of Microbiological Methods, 26*(1), 103-117.

Schloss, P. D., Hay, A. G., Wilson, D. B., & Walker, L. P. (2003). Tracking temporal changes of bacterial community fingerprints during the initial stages of composting. *FEMS Microbiology Ecology, 46*(1), 1-9.

Schwieger, F. & Tebbe, C.C. (1998). A new approach to utilize PCR-single-strandconformation polymorphism for 16S rRNA gene-based microbial community analysis. *Applied and Environmental Microbioiology, 64*, 4870–4876.

Sekiguchi, H., Watanabe, M., Nakahara, T., Xu, B., & Uchiyama, H. (2002). Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis. *Applied and Environmental Microbioiology*, 68, 5142-5150.

Selvakumar, N., Ding, B. C., & Wilson, S. M. (1997). Separation of DNA strands facilitates detection of point mutations by PCR-SSCP. *Biotechniques 22*(4), 604–606.

Shah, N., Tang, H., Doak, T. G., & Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. In *Pacific Symposium on Biocomputing*, *16*, 165-176.

Sharma, V. K. (2006). Real-time reverse transcription-multiplex PCR for simultaneous and specific detection of rfbE and eae genes of *Escherichia coli* O157: H7. *Molecular and Cellular Probes*, *20*(5), 298-306.

Shi, X. M., Long, F., & Suo, B. (2010). Molecular methods for the detection and characterization of foodborne pathogens. *Pure and Applied Chemistry*, *82*(1), 69-79.

Sklarz, M. Y., Angel, R., & Gillor, O. (2011). Amplified rDNA Restriction Analysis (ARDRA) for Identification and Phylogenetic Placement of 16S-rDNA. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, *1*, 59.

Sliwinski, M. K. & Goodman, R. M. (2004). Spatial heterogeneity of crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-singlestranded conformation polymorphism profiling. *Applied and Environmental Microbiology*, *70*, 1811–1820.

Spiegelman, D., Whissell, G., & Greer, C. W. (2005). A survey of the methods for the characterization of microbial consortia and communities. *Canadian Journal of Microbiology*, *51*(5), 355-386.

Theunissen, J., Britz, T. J., Torriani, S., & Witthuhn, R. C. (2005). Identification of probiotic microorganism in South African products using PCR-based DGGE analysis. *International Journal of Food Microbiology*, *98*, 11-21.

Tiedje, J. M., Asuming-Brempong, S., Nüsslein, K., Marsh, T. L., & Flynn, S. J. (1999). Opening the black box of soil microbial diversity. *Applied* and *Soil Ecology*, *13*(2), 109-122.

Tomaselli, L., Lamenti, G., Bosco, M., & Tiano, P. (2000). Biodiversity of photosynthetic micro-organisms dwelling on stone monuments. *International Biodeterioration and Biodegradation*, *46*(3), 251-258.

Tsai, Y. L. & Olson, B. H. (1991). Rapid method for direct extraction of DNA from soil and sediments. *Applied and Environmental Microbiology*, *57*(4), 1070-1074.

Vaneechoutte, M., Dijkshoorn, L., Tjernberg, I., Elaichouni, A., de Vos, P. A. U. L., Claeys, G., & Verschraegen, G. (1995). Identification of Acinetobacter genomic species by amplified ribosomal DNA restriction analysis. *Journal of Clinical Microbiology*, *33*(1), 11-15.

Vaneechoutte, M., Rossau, R., De Vos, P., Gillis, M., Janssens, D., Paepe, N., .. & Kersters, K. (1992). Rapid identification of bacteria of the Comamonadaceae with amplified ribosomal DNA-restriction analysis (ARDRA). *FEMS Microbiology Letters*, *93*(3), 227-233.

Wagner, M., Lehner, A., Klein, D., & Bubert, A. (2000). Single-strand conformation polymorphisms in the hly gene and polymerase chain reaction analysis of a repeat region in the iap gene to identify and type Listeria monocytogenes. *Journal of Food Protection®*, *63*(3), 332-336.

Wang, H., Ng, L. K., & Farber, J. M. (2000) Detection of Campylobacter jejuni and thermophilic Campylobacter spp. from foods by polymerase chain reaction. In: Spencer JFT, Ragout de Spencer AL (eds) *Methods in Piotechnology, Food Microbiology Protocols*, *14*, 95–106

Weidner, S., Arnold, W., & Puhler, A. (1996). Diversity of uncultured microorganisms associated with the seagrass Halophila stipulacea estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Applied and Environmental Microbiology*, *62*(3), 766-771.

Widjojoatmodjo, M. N., Fluit, A. C., & Verhoef, J. (1995). Molecular identification of bacteria by fluorescence-based PCR-single-strand conformation polymorphism analysis of the 16S rRNA gene. *Journal of Clinical Microbiology*, *33*(10), 2601-2606.

Więckowicz M. (2009). Molekularne metody identyfikacji mikroorganizmóww złożonych ekosystemach. *Post Mikrobiology*, *48*, 67-73 (in Polish).

Wilson, B., Danilowicz, B. S., & Meijer, W. G. (2008). The diversity of bacterial communities associated with Atlantic cod Gadus morhua. *Microbial Ecology*, *55*(3), 425-434.

Wilson, M. S., Bakermans, C., & Madsen, E. L. (1999a). In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. *Applied and Environmental Microbiology*, *65*(1), 80-87.

Wilson, V. L., Tatford, B. C., Yin, X., Rajki, S. C., Walsh, M. M., & LaRock, P. (1999b). Species-specific detection of hydrocarbon-utilizing bacteria. *Journal of Microbiological Methods*, *39*(1), 59-78.

Witzig, R., Junca, H., Hecht, H. J., & Pieper, D. H. (2006). Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. *Applied and Environmental Microbiology*, *72*(5), 3504-3514.

Yasumoto, T. & Murata, M. (1993). Marine toxins. *Chemical Reviews*, *93*(5), 1897-1909.

Zumstein, E., Moletta, R., & Godon, J. J. (2000). Examination of two years of community dynamics in an anaerobic bioreactor using fluorescence polymerase chain reaction (PCR) single-strand conformation polymorphism analysis. *Environmental Microbiology*, *2*(1), 69-78.

# CHAPTER THREE

# MATERIALS AND METHODS

# TABLE OF CONTENTS

## MATERIALS AND METHODS

### 3.1.0 Study Area

This research project was carried out along the coast of Algoa Bay (33°49′36″S, 25°47′42″E) in Port Elizabeth, Eastern Cape Province of South Africa. Algoa Bay is a wide inlet along the South African east coast, about 683 km east of the Cape of Good Hope. It is bounded in the west by Cape Recife, in the east by Cape Padrone and in the South by the India Ocean. The bay is up to 436 m deep (Fig. 4).

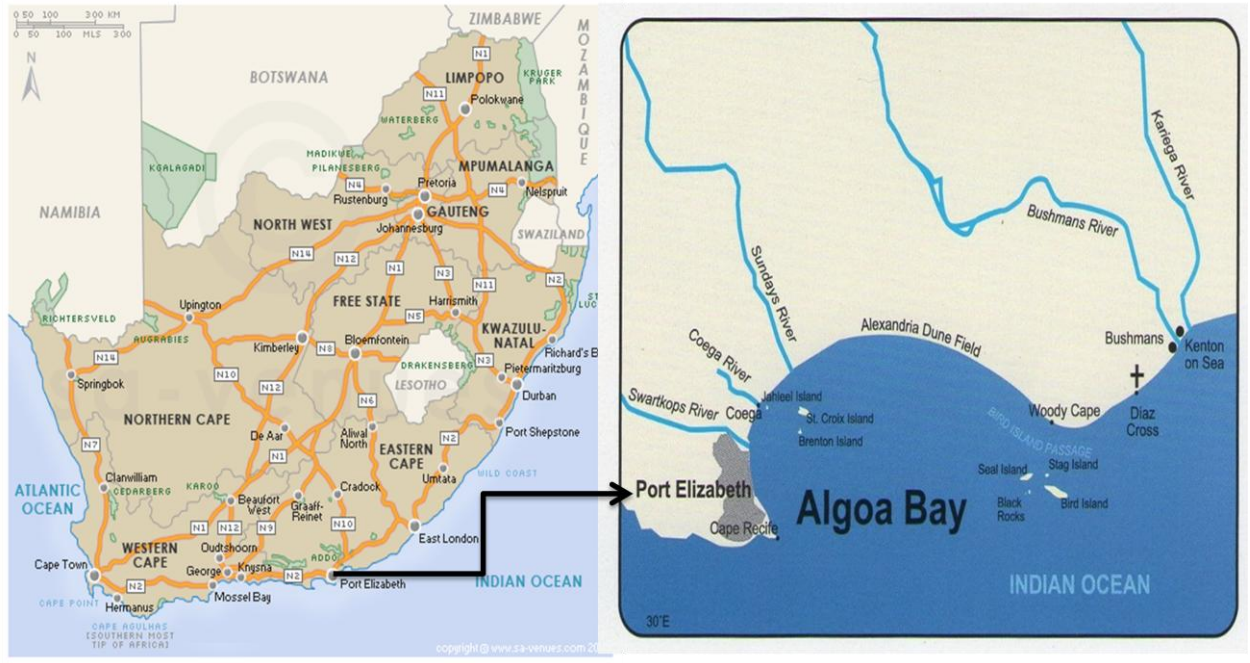


**Figure 4:** Algoa Bay (sampling location) in Port Elizabeth.

### 3.2.0 Sample Collection

Equal number of *Perna perna* (50 each) were collected before and during red tide in February 2014, from the rocky shores of the sampling site and were transported in cooler boxes, with ice packs, to the Microbiology Laboratory of the University of Fort Hare, Alice, for dissection and examination. Samples were surface sterilized with 70% ethanol and rinsed three times with sterile distilled water. The shells were then opened using a sterile scalpel

blade. The shell liquors of the mussels was discarded, tissues were removed and homogenized.

### 3.3.0 DNA Extraction

Total bacteria DNA was extracted from each sample of homogenized *Perna perna* flesh using Thermo scientific GeneJET Genomic DNA Purification kit (EU, Lithuania) thereby yielding 100 DNA templates (50 DNA templates from *Perna perna* collected before red tide and 50 DNA templates from those collected before red tide). The procedures were carried out as follows:

Approximately 0.5 $cm^2$ of each the of tissues was placed in a 1.5 ml micro centrifuge tube and mixed with 180 µl of digestion solution and 20 µl of 20 mg/mL proteinase K solution (Thermo Scientific, USA), vortex and incubated at 56 °C for 1 h. This was followed by addition of 20 µl of RNase A solution, vortex and incubated at room temperature for 10 mins after which 200 µl lysis was added. The mixture was vortex for 15 s and mixed thoroughly with 400 µl of 50% ethanol. The lysate was transferred into GeneJet purification column fitted with a collection tubes. Samples were first centrifuged at 6000 X g for 1 min. The columns were placed into new collection tubes, 500 µl of wash buffer I was added and centrifuged at 8000 x g for 1 min. Followed by the addition of 500 µl of washed buffer II to the column and centrifuged at 12000 x g for 3 mins. The columns were transferred into a sterile 1.5 ml micro centrifuge tubes, 100 µl of elution buffer was added into the centre of each column membrane to elute the DNA, incubated at room temperature for 2 mins and were centrifuged at 8000 x g for 1 min. The eluted DNA templates were quantified using NanoDrop 2000 (Thermo Scientific, USA) and store at 4 °C for further analysis.

### 3.4.0 Confirmatory Identification of the bacteria using PCR-16S rRNA

### 3.4.1 PCR Amplification

Single-target gene amplification was carried out in 25 µL-reactions containing reaction buffer 2×, 4 mM $MgCl_2$, 0.4 mM of each dNTP (dATP, dCTP, dGTP and dTTP), 1.0 µL of each primer, 0.05U/ µL *Taq* DNA polymerase, 5 µL of DNA template and 5.5 µL of nuclease free water to bring the final volume to 25 µL. The amplification was performed on a thermocycler (Bio-Rad Mycycler USA) using the following cycling condition: initial denaturation at 95 °C for 1 min, followed by 40 cycles of denaturation at 94 °C for 30 sec, annealing at 64 °C for 45 sec, extension at 72 °C for 1 min, and was finally extended at 72 °C for 10 mins.

The table below shows the primers used for PCR amplification and the amplicon sizes.

**Table 1: Primer pairs used for PCR amplification of bacteria extracted from *Perna perna* (Clarke *et al*., 2007)**

| S/N | Primer code | Primer sequence (5'– 3') | Amplicon size (bp) |
|---|---|---|---|
| 1 | GM5F | cct acg gga ggc agc ag | 586 |
| 2 | 40 bp GC clamp connected to the 5'end of 907R | cgc ccg ccg cgc ccc gcg ccc gtc ccg ccg ccc ccg ccc g ccgtcaattcctttgagttt | 586 |

### 3.4.2 Gel Electrophoresis

The PCR products (2 µl aliquots) were loaded into 1.0% (w / v) agarose gel stained with ethidium bromide. The amplicons together with 100 bp molecular marker (DNA Ladder Plus) (Promega, USA) were electrophoresed at 100 V for 1 h. The developed bands were visualised using UV trans-illuminator (Alliance 4.7, France) (Muyzer *et al*., 1993).

### 3.5.0 DNA Purification

Amplified DNA fragments were recovered from the agarose gels using a sterile scalpel blade. The excised bands were transfered into a sterile 1.5 ml micro centrifuge tube and the DNA fragments were purified using Zymoclean$^{TM}$ Gel DNA Recovery kit (ZymoResearch Corporation, Irvine, US) according to the manufacturer instructions.

### 3.6.0 Sequence Analysis of PCR Products

Ten (10) microliters of the eluate was used as the DNA template in the PCR using primer GM5F. The Thermocycling reaction sample consisted of DNA template, 1 µl of 10 µM GM5F (5′-CCTACGGGAGGCAGCAG-3′) primer stock solution, 4 µl BigDye, 2µl 5 x dilution buffer and distilled water (to volume of 20 µl). The reaction samples were directly sequenced with an ABI 3500 XL genetic analysers by using an ABI V3.1 Big Dye kit (ZymoResearch Corporation, Irvine, US) according to the manufacturer directions.

### 3.7.0 Analysis of Sequence Results

ChromatoGrams were generated by ABI 3500 XL Genetic Analyser Data Collection system by Applied Biosystems. These chromatoGrams were converted into text format using Chromas software and submitted into the National Center for Biotechnology Information (NCBI) using a Basic Local Alignment Search Tool (BLAST) database (Madden, 2013). The percentage similarity to an identified species was recorded and also the length of the match.

### 3.8.0 Multiple Sequence Alignment

The 16S rRNA FASTA sequence for each of the bacteria species were aligned using Clustal Omega version 1.2.1 on http://www.ebi.ac.uk/Tools/msa/clustalo/. (Chenna et *al*., 2003; Sievers *et al*., 2011), and classified into groupings based on their sequence similarities the ribosomal database project.

### 3.9.0 Phylogenetic Analysis.

The 16S rRNA sequences obtained were inferred using the Neighbor-Joining method (Saitou *et al*., 1987) and the evolutionary analyses was constructed on MEGA 6 (Tamura *et al*., 2013; Reddy, 2013). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al*., 2004). They are in the units of the number of base substitutions per site. The analysis involved 36 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1309 positions in the final dataset.

**References**


Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, *31*(13), 3497-3500.

Clarke, A. M., Kirby, R., & Rose, P. D. (2007). Molecular microbial ecology of lignocellulose mobilisation as a carbon source in mine drainage wastewater treatment. *Water SA*, *30*(5), 558-661.

Madden, T. (2013). The BLAST sequence analysis tool. http://www.ncbi.nlm.nih.gov/books/NBK153387/

Muyzer, G., De Waal, E. C., & Uitter linden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appied and Environmental Microbioiology*. *59*, 695-700.

Reddy, M. B. P. (2013). Basics for the Construction of Phylogenetic Trees. *Webmed Central Biology*, (12): WMC002563.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406-425.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1).

Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(30), 11030-11035.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725-2729.

# CHAPTER FOUR

# RESULTS AND DISCUSSION

# TABLE OF CONTENTS

## RESULTS AND DISCUSSION

### 4.0 PCR analysis

After running the DNA templates on 1% agarose gel, the PCR amplification results indicated that the expected amplicon sizes of 586 bp was found in 90% of DNA templates extracted from *Perna perna* that were collected before red tide, while only 10% of the samples collected during red tide had these amplicon sizes (Figure 5a & 5b).

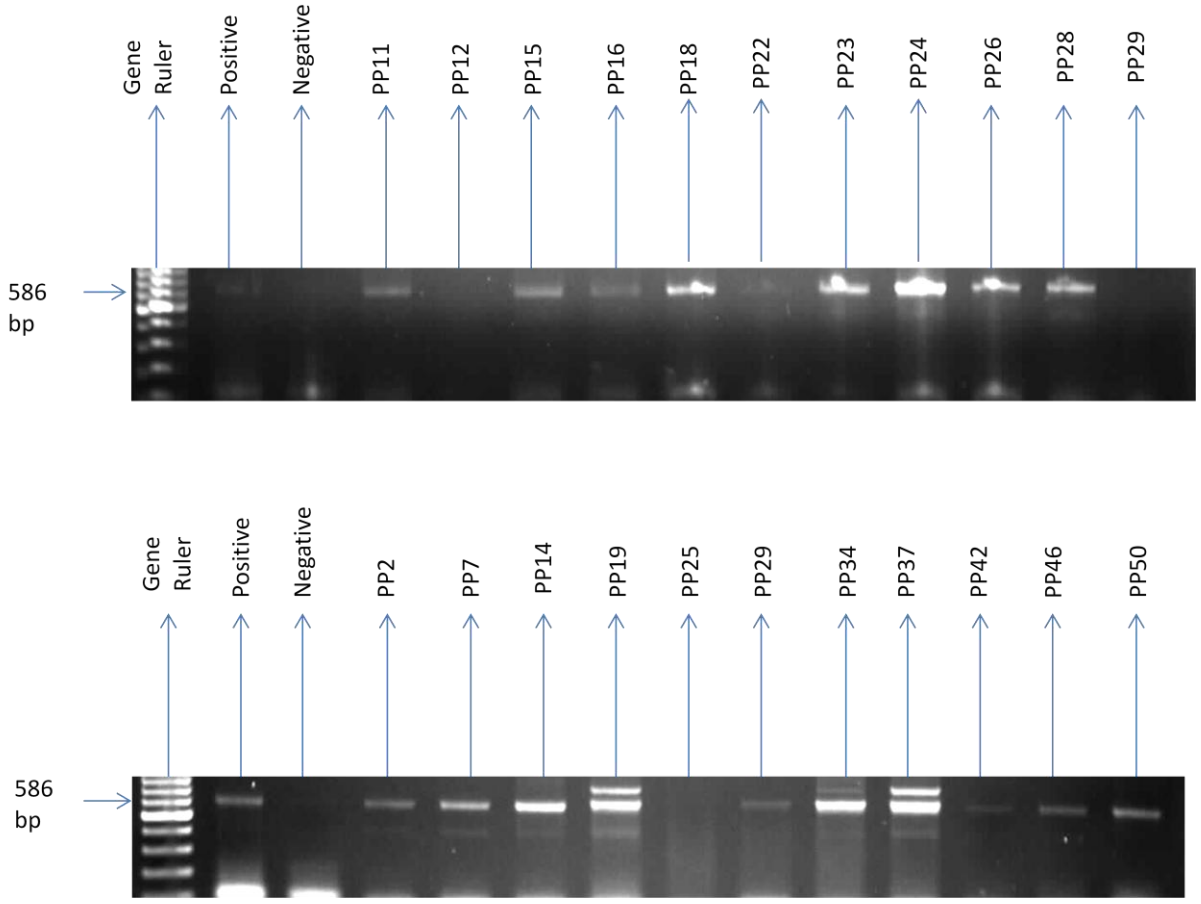


**Figure 5a:** 586 bp amplicon sizes of PCR products from samples collected before red tide with 100 bp gene ruler. Note the presence of clear band (arrow) of 586 bp region.

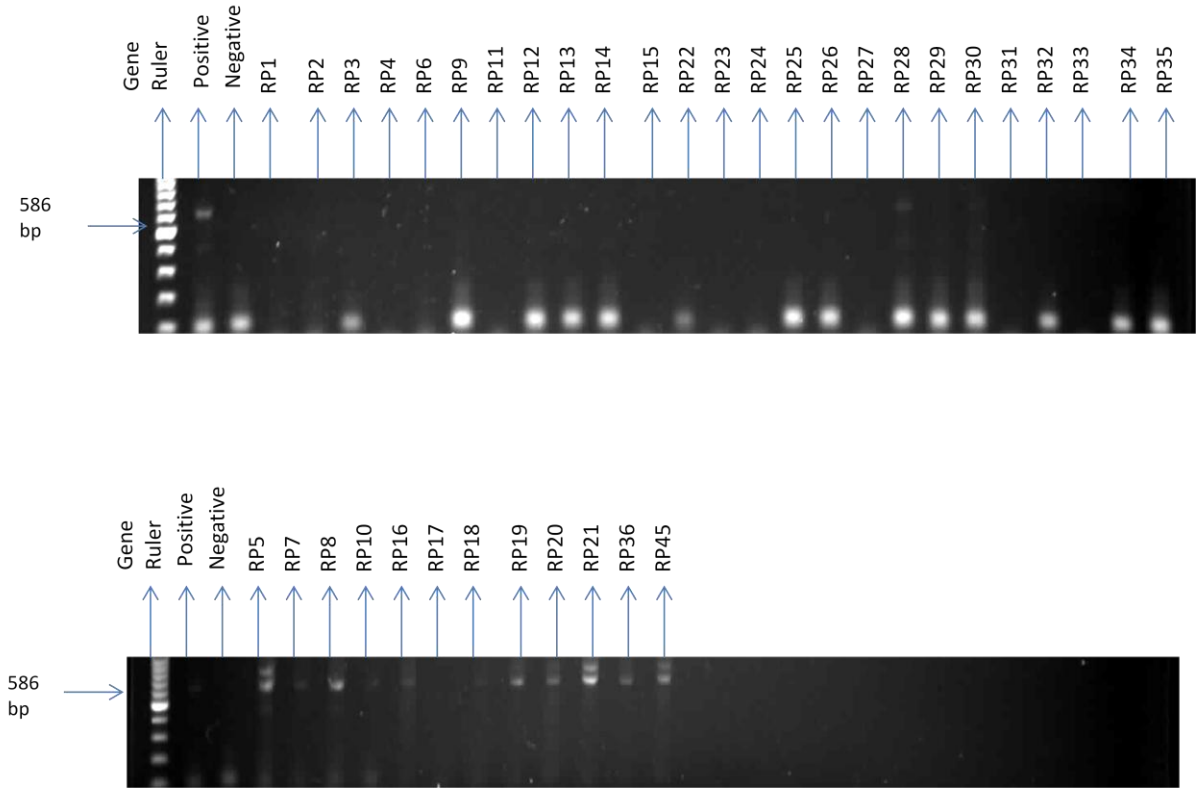


**Figure 5b:** 586 bp amplicon sizes of PCR products from samples collected during red tide with 100 bp gene ruler. Note the absence of 586 bp region in most of the amplicons.

Universal primers 16S rRNA were used in this study because of its presence in almost all bacteria; they also represent the important target of study in bacterial ecology (Větrovský, 2013). Therefore, in this study, 586 bp amplicon sizes were expected to be found in almost all the samples. The limited number of bacteria 586 bp amplicon sizes observed in *Perna perna* collected during red tide may therefore be due to the effect of algae toxins on *Perna perna.* When algal rupture, they release toxins into the gills of the shellfish (or any other marine animal) and within a short period of time, the animal asphyxiate and die (Anderson, 1989).

**Table 2a:** Identification of bacterial populations (Phylum Actinobacteria) from *Perna perna* before red tide along Algoa Bay, South Africa

(The entire organisms belong to the class Actinobacteria, subclass Actinobacteridae and order Actinomycetales)

| Base Pair (Length) | Taxon ID | Strain | Accession Number | Family | Genus | Species | Gram +ve/-ve |
|---|---|---|---|---|---|---|---|
| 1513 | 863573 | PNP1 | NR_117546.1 | Micrococcaceae | *Citricoccus* | *Citricoccus nitrophenolicus* | +ve |
| 1514 | 574650 | V3M1 | NR_116649.1 | Micrococcaceae | *Micrococcus* | *Micrococcus terreus* | +ve |
| 1410 | 556325 | DY66 | NR_116519.1 | Micrococcaceae | *Zhihengliuella* | *Zhihengliuella aestuarii* | +ve |
| 1471 | 199136 | CA15-8 | NR_25665.1 | Micrococcaceae | *Arthrobacter* | *Arthrobacter koreensis* | +ve |
| 1499 | 98672 | CF-25 | NR_25362.1 | Micrococcaceae | *Arthrobacter* | *Arthrobacter luteolus* | +ve |
| 1481 | 1275 | DSM 20447 | NR_44871.1 | Micrococcaceae | *Kocuria* | *Kocuria rosea* | +ve |
| 1412 | 469799 | CAAS 251 | NR_44353.1 | Micrococcaceae | *Nesterenkonia* | *Nesterenkonia flava* | +ve |
| 1467 | 71999 | TAGA27 | NR_26451.1 | Micrococcaceae | *Kocuria* | *Kocuria palustris* | +ve |
| 1439 | 664639 | 104 | NR_117299.1 | Micrococcaceae | *Kocuria* | *Kocuria salsicia* | +ve |
| 1333 | 1272 | ATCC 15306 | NR_114674.1 | Micrococcaceae | *Kocuria* | *Kocuria varians* | +ve |
| 1440 | 223184 | KMM 3905 | NR_25723.1 | Micrococcaceae | *Kocuria* | *Kocuria marina* | +ve |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1480 | 333384 | RCQ1071 | NR_43541.1 | Streptomycetaceae | *Steptomyces* | *Streptomyces lunalinharesii* | +ve |
| 1504 | 103731 | LL-C19004-NS29 | NR_102474.1 | Pseudonocardiaceae | *Saccharothrix* | *Saccharothrix espanaensis* | +ve |
| 1473 | 1128665 | NJ2035 | NR_109528.1 | Pseudonocardiaceae | *Lechevalieria* | *Lechevalieria nigeriaca* | +ve |
| 1510 | 659496 | Hhs.015 | NR_117283.1 | Pseudonocardiaceae | *Saccharothrix* | *Saccharothrix yanglingensis* | +ve |
| 1429 | 173560 | SA 233 | NR_25658.1 | Pseudonocardiaceae | *Saccharothrix* | *Saccharothrix algeriensis* | +ve |
| 1474 | 38311 | DSM 43752 | NR_26183.1 | Nocardiaceae | *Rhodococcus* | *Rhodococcus marinonascens* | +ve |
| 1510 | 103731 | NRRL 15764 | NR_24965.1 | Pseudonocardiaceae | *Saccharothrix* | *Saccharothrix espanaensis* | +ve |
| 1512 | 33919 | NRRL B-16115 | NR_41736.1 | Pseudonocardiaceae | *Saccharothrix* | *Saccharothrix coeruleofusca* | +ve |
| 1413 | 2072 | NRRL 11239 | NR_41735.1 | Pseudonocardiaceae | *Saccharothrix* | *Saccharothrix australiensis* | +ve |
| 1315 | 38311 | ATCC 35653 | NR_119127.1 | Nocardiaceae | *Rhodococcus* | *Rhodococcus marinonascens* | +ve |
| 1478 | 459658 | M1463 | NR_116165.1 | Streptomycetaceae | *Streptomyces* | *Streptomyces samsunensis* | +ve |
| 1469 | 229486 | NBRC 101006 | NR_112601.1 | Streptomycetaceae | *Streptomyces* | *Streptomyces hebeiensis* | +ve |
| 1445 | 92644 | NBRC 16446 | NR_41410.1 | Streptomycetaceae | *Streptomyces* | *Streptomyces malaysiensis* | +ve |
| 1520 | 229486 | YIM 001 | NR_29091.1 | Streptomycetaceae | *Streptomyces* | *Streptomyces hebeiensis* | +ve |

**Table 2b:** Identification of bacterial populations (Phylum Spirochaetes) from *Perna perna* before red tide along Algoa Bay, South Africa

(The entire bacteria belong to the class Spirochaetes, subclass Spirochaetes and order Spirochaetales)

| **Base Pair (Length)** | **Taxon ID** | **Strain** | **Accession Number** | **Family** | **Genus** | **Species** | **Gram +ve/-ve** |
|---|---|---|---|---|---|---|---|
| 1533 | 151 | R 1 | NR_104732.1 | Spirochaetaceae | *Spirochaeta* | *Spirochaeta litoralis* | -ve |
| 1435 | 504483 | SIP1 | NR_44505.1 | Spirochaetaceae | *Spirochaeta* | *Spirochaeta cellobiosiphila* | -ve |
| 1467 | 159291 | ASpG1 | NR_28820.1 | Spirochaetaceae | *Spirochaeta* | *Spirochaeta Americana* | -ve |

**Table 3a:** Identification of bacterial populations (Phylum Proteobacteria) from *Perna perna* during red tide along Algoa Bay, South Africa (The bacteria belong to class Alphaproteobacteria and subclass Alphaproteobacteridae**)**

| **Base pair (length)** | **Taxon ID** | **Strain** | **Accession Number** | **Order** | **Family** | **Genus** | **Species** | **Gram +ve/-ve** |
|---|---|---|---|---|---|---|---|---|
| 1451 | 34003 | DM-15 | NR_121706.1 | Rhodobacterales | Rhodobacteraceae | *Paracoccus* | *Paracoccus aminophilus* | -ve |
| 1388 | 34003 | NBRC 16710 | NR_113863.1 | Rhodobacterales | Rhodobacteraceae | *Paracoccus* | *Paracoccus aminophilus* | -ve |
| 1374 | 372072 | CL-GR15 | NR_43777.1 | Rhizobiales | Cohaesibacteraceae | *Cohaesibacter* | *Cohaesibacter gelatinilyticus* | -ve |
| 1385 | 34003 | NBRC 16710 | NR_115554.1 | Rhodobacterales | Rhodobacteraceae | *Paracoccus* | *Paracoccus aminophilus* | -ve |

**Table 3b:** Identification of bacterial populations (Phylum Actinobacteria) from *Perna perna* during red tide along Algoa Bay, South Africa (The entire bacteria belong to class Actinobacteria, subclass Actinobacteridae and order Actinomycetales)

| **Base pair (length)** | **Taxon ID** | **Strain** | **Accession Number** | **Family** | **Genus** | **Species** | **Gram +ve/-ve** |
|---|---|---|---|---|---|---|---|
| 1533 | 378753 | DC2201(=NBRC103217) | NR_74786.1 | Micrococcaceae | *Kocuria* | *Kocuria rhizophila DC2201* | +ve |
| 1439 | 664639 | 104 | NR_117299.1 | Micrococcaceae | *Kocuria* | *Kocuria salsicia* | +ve |
| 1480 | 136273 | CMS 76or | NR_28924.1 | Micrococcaceae | *Kocuria* | *Kocuria polaris* | +ve |
| 1481 | 1275 | DSM 20447 | NR_44871.1 | Micrococcaceae | *Kocuria* | *Kocuria rosea* | +ve |

**Table 3c:** Identification of bacterial populations (Phylum Spirochaetes) from *Perna perna* during red tide along Algoa Bay, South Africa

(The bacteria belong to class Spirochaetes, subclass Spirochaetes and order Spirochaetales)

| **Base pair (length)** | **Taxon ID** | **Strain** | **Accession Number** | **Family** | **Genus** | **Species** | **Gram +ve/-ve** |
|---|---|---|---|---|---|---|---|
| 1564 | 149 | RS1 | NR_44756.2 | Spirochaetaceae | *Spirochaeta* | *Spirochaeta halophila* | -ve |
| 1467 | 159291 | ASpG1 | NR_28820.1 | Spirochaetaceae | *Spirochaeta* | *Spirochaeta Americana* | -ve |
| 1533 | 151 | R 1 | NR_104732.1 | Spirochaetaceae | *Spirochaeta* | *Spirochaeta litoralis* | -ve |

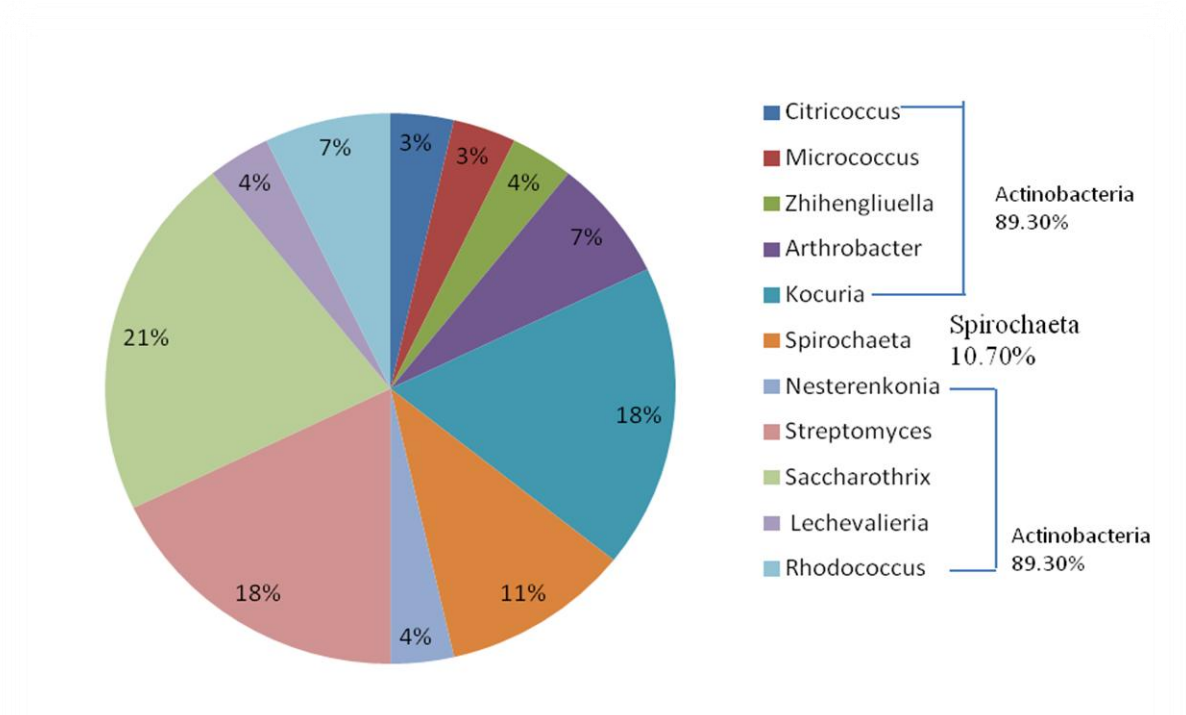


**Figure 6a:** Generic compositions of bacteria found in *Perna perna* before red tide in February, 2014

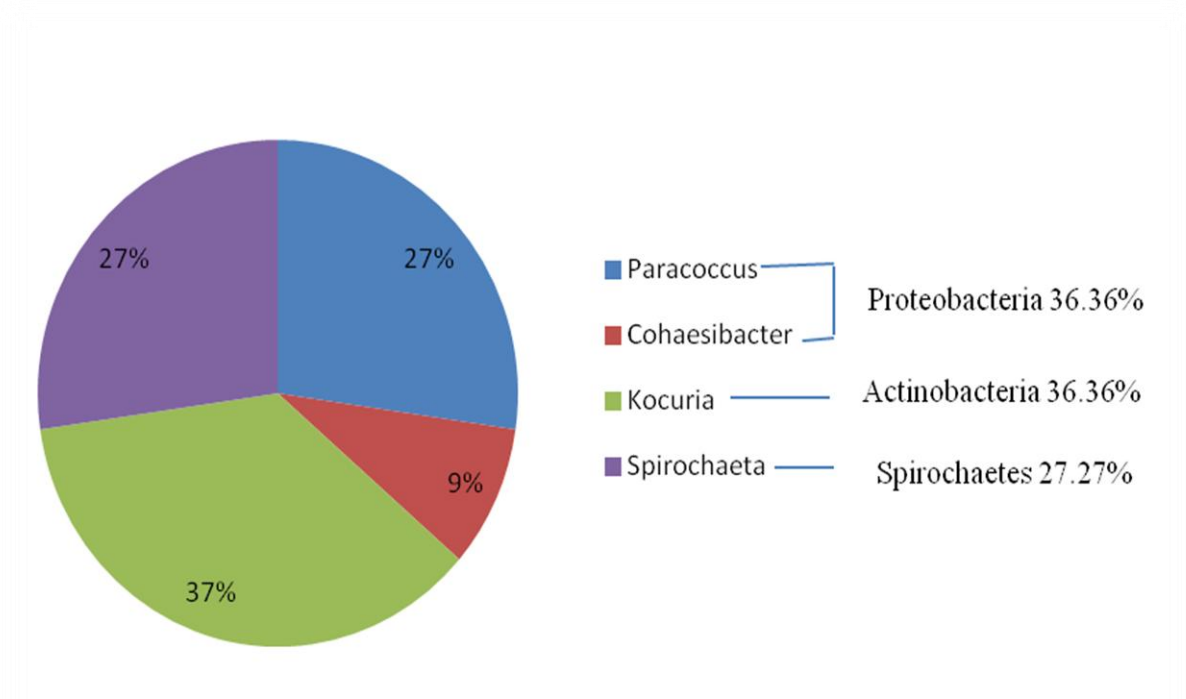


**Figure 6b:** Generic compositions of bacteria found in *Perna perna* during red tide in February, 2014

**Analysis of bacteria diversity in *Perna perna* before and during red tide**

Generally, there were more bacteria species inhabiting *Perna perna* before red tide than during red tide (Table 2a-b, 3a-c). The fewer number of bacteria living on this organism during red tide might be associated with the population explosion of dinoflagellates during the period. Dinoflagellates have been reported to produce toxins whenever they are in high concentrations in the ocean (Jorge, 2014). These toxins have been found to cause shellfish poisonings such as paralytic, diarrhetic, amnesic, neurotoxic and aerosol toxicity (Dolah *et al*., 2001). It is possible that some of these toxic chemicals could affect the organisms growing *Perna perna* including bacteria.

In these findings, different bacteria species were identified in 90% of *Perna perna* collected before red tide (Table 2a-b) and comparative 16S rRNA gene sequence analysis affiliated the bacteria to diverse phylogenetic groups (Fig. 7). 28 bacteria were identified in *Perna perna* collected before red tide. Most of the bacteria were members of genera *Citricoccus*, *Micrococcus*, *Zhihengliuella*, *Arthrobacter*, *Kocuria*, *Nesterenkonia*, *Steptomyces*, *Saccharothrix*, *Lechevalieria*, *Rhodococcus* and *Spirochaeta* (Fig. 6a). The most predominantly bacteria genera observed was *Saccharothrix* with six species (21%) followed by *Kocuria* and *Steptomyces* having five different species each (18%). Also, three species (11%) of *Spirochaeta* was identified while *Rhodococcus* and *Arthrobacter* were having two different species (7%) each. Only one species of *Lechevalieria, Nesterenkonia, Zhihengliuella, Micrococcus* and *Citricoccus* was identified with 4%, 4%, 4%, 3%, and 3% respectively.

All the bacteria belong to Phyla Actinobacteria and are Gram-positive bacteria except a few members of *Spirochaeta* that are Gram-negative and members of Phylum Spirochaetes. Actinobacteria constitute one of the largest phyla of bacteria. Their members have adopted different lifestyle; that is; they can be pathogens, soil inhabitants, plant commensals and

gastrointestinal commensals (Ventura *et al*., 2007). Also, many species of Actinobacteria have been identified from the animal feces (Jiang *et al*., 2013) indicating that these bacteria can adapt and survive in any given condition.

Only 11 bacteria were identified in 10% of *Perna perna* collected during red tide (Table 3a-c), four of which were also identified in the samples collected before red tide; they include *Kocuria rosea* strain DSM 20447, *Kocuria salsicia* strain 104, *Spirochaeta litoralis* R 1 and *Spirochaeta americana* ASpG1. Generally, the bacteria identified in *Perna perna* collected during red tide belong to genera *Paracoccus*, *Cohaesibacter*, *Kocuria* and *Spirochaeta* (Fig. 6b). The most predominantly bacteria genera observed was *Kocuria* with four species (37%) followed by *Spirochaeta* and *Paracoccus* having three different species each (27%). The least bacteria specie identified in this sample was *Cohaesibacter* with one species (9%). They are Gram-positive belonging to Phyla Proteobacteria and Actinobacteria except *Spirochaeta* that are Gram-negative belonging to Phylum Spirochaetes.

**Bacteria identified in *Perna perna* before red tide in February 2014**

*Citricoccus nitrophenolicus* strain PNP1 is Gram-positive bacteria belonging to family Micrococcaceae of the class Actinobacteria. It has been reportedly isolated from wastewater treatment plant at a chemical factory producing methyl-parathion and other pesticides (Nielsen *et al*., 2011). This bacterium may have likely been released into the ocean from a nearby wastewater treatment plants and thus, found their way into *Perna perna*. *Citricoccus nitrophenolicus* is biologically important and capable of degradating p-nitrophenol (pNP) at a high pH. This aromatic compound is listed in the US Environmental Protection Agency priority pollutants in environmental waters, which could be toxic to the soil microflora and invertebrates (PAN, 2008; US EPA, 2012). In the nearest future, *Citricoccus nitrophenolicus* could be an

alternative bacteria used in bioremediation by either in situ or ex situ. The type strain of *Citricoccus nitrophenolicus* PNP1 is DSM 23311=CCUG 59571.

*Micrococcus terreus* strain V3M is Gram-positive, strictly aerobic bacteria belonging to family Micrococcaceae and genus *Micrococcus*. It has been isolated from skin, soil and some marine animals (Kocur *et al*., 2006). The prevalence of these bacteria in water and in shellfish may be due to contamination from the land (Kocur *et al*., 2006). *Micrococci* are commensals but can as well serve as opportunistic pathogens in immunocompromised individuals (Smith *et al*., 1999). They can be involved in the detoxification or biodegeneration of many environmental pollutants (Zhuang *et al*., 2003). The type strain is NBRC 104258=JCM 17489=CGMCC 1.7054.

*Zhihengliuella aestuarii* strain DY66 is Gram-positive bacteria belonging to Phylum Actinobacteria and family Microccaceae. So far, only four species of this genera have been identified namely; *Z. Halotolerans* (Khang *et al*., 2007), *Z. alba* (Tang *et al*., 2009), *Z. salsuginis* (Chen *et al*., 2010) and *Z. aaestuarii* strain DY66 (Baik *et al*., 2011). The first three were isolated from saline soil and subterranean brine samples while the strain Y66 was isolated from marine environment (Baik *et al*., 2011). *Zhihengliuella aestuarii* have DNA GC content of 59.1 mol% and can grow optimally at 30 °C and pH 8.9 in 2% w/v NaCl (Baik *et al*., 2011). The type strain is KCTC 19557=JCM 16364.

The genus *Arthrobacter* is Gram-positive, aerobic, coryneform bacteria belonging to Phylum Actinobacteria. Members of this genus were first isolated from soils in various environments (Holt *et al*., 1994). These bacteria could be mesophilic, pschrophilic, alkaliphilic and alkalitolerant (Lee *et al*., 2003). They are useful in bioremediation of ground water contaminated with pesticides and herbicides; because their genome can adapt easily to any environmental conditions (Mongodin *et al*., 2006). Members of genus *Arthrobacter* cannot be separated phylogenetically from members of genus *Micrococcus* due to similarities between them (Jones

and Keddie, 2007). *Arthrobacter* species identified in *Perna perna* include *Arthrobacter koreensis* and *Arthrobacter luteolus.*

*Arthrobacter koreensis* strain CA15-8 with type strain NBRC 16787=KCTC 9922=JCM 12361=IFO 16787 cells is alkalitolerant, grow best at optimum pH 7.0 to 8.0 and possesses high DNA GC content of about 63±2 mol% (Jones and Keddie, 2007). *Arthrobacter luteolus* CF-25 with type strain CF25=NBRC 107841=JCM 11676=DSM 13067=CIP 106789=CCUG 43811=ATCC BAA-272 has been isolated from clinical specimen (Wauters *et al*., 2000) and from rare earth environment (Emmanuel *et al*., 2012).

*Kocuria* is a member of family Micrococcaceae, order Actinomycetales, suborder Micrococcineae and class Actinobacteria (Stackebrandt *et al*., 1995; Takarada *et al*., 2008; Lee *et al*., 2009; Savini *et al*., 2010). It is Gram-positive, strictly aerobic (except *Kocuria kristinae*, which is facultatively anaerobic, *Kocuria marina*, which may grow in 5% $CO_2$ and *Kocuria rhizophila* strain DC2201, which can proliferate anaerobically). It is ubiquitous in nature and frequently found as normal skin flora in humans and other mammals. *Kocuria* is human pathogens mostly in immunocompromised hosts (Savini *et al*., 2010). This bacteria was first classified into the genus *Micrococcus* but was later re-classified as genus *Kocuria* based on the phylogenetic and chemotaxonomic analyzes of Stackebrant *et al.* (2002). Presently, *Kocuria* is comprises of 18 species (Takarada *et al*., 2008; Stackebrant *et al.,* 2002); out of which seven were identified in this study and they include *K. rosea* (Stackebrant *et al.,* 1995), *K. palustris* (Kovács *et al*., 1999), *K. salsicia* (Yun *et al*., 2010), *K. varians* (Stackebrant *et al.,* 1995), *K. marina* (Kim *et al*., 2004), *K. rhizophila (*Kovács *et al*., 1999) and *K. polaris* (Reddy *et al*., 2003).

*Kocuria rosea* strain DSM 20447 is a non-pathogenic commensal found to colonize the oropharynx, skin and mucosa. It could cause an opportunistic infection in immunocompromised

patients (Ma ES *et al*., 2005). Four cases of *K. rosea* have been reported where it serves as a pathogen in infective endocarditis in immunocompetent patients (Srinivasa *et al*., 2013), in catheter-related bacteremia (Altuntas *et al*., 2004), peritonitis (Kaya *et al*., 2009) and descending necrotizing mediastinitis in an immunocompromised patients (Lee *et al*., 2013). *K. rosea* was identified in *Perna perna* before and during red tide and have the type strain VKM B-1823=NRRL B-2977=NCTC 7523=NBRC 3768=LMG 14224=JCM 11614=IFO 3768=IEGM 394=DSM 20447=CIP 71.15=CCUG 4312=CCUG 4312=ATCC 186.

*Kocuria palustris* strain TAGA27 was first isolated from the rhizoplane of the narrow-leaved cattail (Kovács *et al*., 1999) but now found in *Perna perna*, a musssels. It grows between pH 5.7 to 7.5, and have the type strain NBRC 16318=JCM 11652=IFO 16318=DSM 11925=CIP 105971.

*Kocuria salsicia* strain 104 was first isolated from traditional salt-fermented seafood in Korea. It has the optimal growth temperature between 30 to 36 °C, pH 7 to 8 and DNA GC content of 60.6 mol%. This strain was identified in the samples before and after red tide; and it have the type strain Kocuria sp. 104=strain 104=KACC 21128=JCM 16361.

*Kocuria varians* strain ATCC 15306 grew best at the temperature between 25 to 37 °C. It was first isolated from mammalian skin but now been found in soil and water (Ben-Ami *et al*., 2003). This bacterium have been associated with infections such as endocarditis, arthritis, central nervous system infection, pneumonia, peritonitis, hepatic abscess and nosocomial blood stream infections (Ben-Ami *et al*., 2003; Tsai *et al*., 2010). It has the type strain VKM B-1827=NCTC 7564=NBRC 15358=LMG 14231=JCM 7238=IFO 15358=IEGM 400=HAMBI 40=HAMBI 1951=DSM 20033=CIP 81.73=CCUG 35392=CCM 884=ATCC 15306.

*Kocuria marina* strain KMM 3905 was isolated from a marine sediment sample taken from Troitsa Bay (Kim *et al*., 2004). It contains enzymatic activities, thus have the ability to carry out

bioremediation (Kim *et al*., 2004). This strain is closely related to *K. varians* ATCC15306 on the phylogenetic tree (Fig. 7) and have the type strain *Kocuria* sp. KMM 3905=KCTC 9943=JCM 13363=CCUG 51442.

Genus *Spirochaeta* is Gram-negative, free-living, saccharolytic non-pathogenic, obligate and facultative anaerobic helical shaped bacteria. It is indigenous to aquatic environments including freshwaters, muds of ponds, rivers, marshes, marine waters and halophilic species such as alkaliphiles and thermophiles (Canale- Parola, 1984, 1992; Aksenova *et al*., 1992; Hoover *et al*., 2003; Leshine *et al*., 2006). About 17 species of *Spirochaeta* have been reported of which four were identified in *Perna perna* during the cause of this study; they include *Spirochaeta litoralis* strain R 1, *Spirochaeta cellobiosiphila* strain SIP1, *Spirochaeta americana* strain ASpG1 and *Spirochaeta halophila* strain RS1. These bacteria may harbour enzymes with potential biotechnology applications (Leschine *et al.*, 2006). Various researches have shown that some species of *Spirochaeta* could be used in direct bioconversion of cellulose-containing wastes to fuel such as ethanol or hydrogen gas (Leschine *et al.*, 2006). Despite the economic importance, some of these bacteria are pathogens for humans, causing diseases as syphlis, yaws, leptospirosis, Lyme disease, and relapsing fever.

*Spirochaeta litoralis* strain R 1 was first discovered by Canale Parola and it grows best at the temperature of 30°C (Canale-Parola *et al*., 1992). The re-occurrence of this bacterium was seen in *Perna perna* before and during red tide (Table 2b & 3c). It has the type strain R1=DSM 2029=ATCC 27000.

*Spirochaeta cellobiosiphila* strain SIP1 was first isolated from interstitial water from a cyanobacteria-containing microbial mat. It can optimally grow at 36 °C, pH 7.5 and has the DNA GC content of 41.4 mol% which is the lowest among known *spirochaetas* (Breznak and Warnecke, 2008). It has the type strain SIP1=DSM 17781=ATCC BAA-1285.

*Spirochaeta americana* strain ASpG1 is an extremophile, mesophilic, haloalkaliphilic and grow in the absence of oxygen. This bacterium was first discovered in a highly alkaline anaerobic environment of Mono Lake. It grow best between pH 8.0 to 10 and have the GC content of 55·96 mol% (Hoover *et al*., 2003). This species was identified in both samples (Table 2b & 3c) and it has the type strain ASpG1=DSM 14872=DSM 14872.

*Nesterenkonia flava* strain CAAS 251 is Gram-positive, non-motile, rod-shaped, non-spore-forming bacterium belonging to family *Micrococcaceae* (Stackebrandt *et al*., 1995). It was first isolated from paper-mill effluent. It grew optimally between 40 to 42 °C, pH 9.0 to 10.0 and has the GC content of 65.5 mol% (Luo *et al*., 2008). The type strain is JCM 14814=CCTCC AB 207010.

*Streptomyces* are usually found in soil and are important in soil ecology. They are members of bacterial order Actinomycetales. They are widely studied and well known Actinomycete because of their consequently invaluable in the medical field (Kämpfer, 2006). *Streptomyces* are Gram-positive with a high DNA GC content (Madigan, 2005). The Actinomycetes are unique in forming spores**,** production of numerous antibiotics including streptomycin**,** neomycin**,** chloramphenicol and tetracyclines (Kieser *et al*., 2000). Over 500 species of *Streptomyces* have been identified (Euzéby, 2008). They produce extracellular hydrolytic enzymes, thus, they are considered as agents for bioremediation but few species are pathogenic for animals, and few might cause plant diseases.

*Streptomyces lunalinharesii* strain RCQ1071 is a highly chitinolytic actinomycete (de Souza, *et al*., 2008); isolated from Brazilian cerrado soil. It can be use in biological control against phytopathogenic fungi (Gomes *et al*., 2001). Phylogenetic analyses of 16S rRNA gene sequences showed that strain RCQ1071 formed a long branch in a group related to *Arthrobacter*

*koreensis and Rhodococcus marinonascens* strain DSM 43752 sharing sequence similarity (Fig 7). The type strain is RCQ1071=JCM 16374=DSM 41876=CIP 108852=ATCC BAA-1231.

*Streptomyces samsunensis* strain M1463 was first isolated from the rhizosphere of *Robinia pseudoacacia.* It has a high GC content of 71.8 mol% (Sazak *et al*., 2011). This species formed a short branch in a group related to *Kocuria salsicia* strain104 and *Zhihengliuella aestuarii* strain DY66 on the evolutionary tree (Fig. 7). The type strain is DSM 42010=NRRL B-24803.

*Streptomyces hebeiensis* strain NBRC 101006 and YIM 001 have the same type strain YIM 001=NBRC 101006=NBRC 100914=JCM 12696=DSM 41837=CIP 107974=CCTCC AA 203005, but belong to separate groups on the phylogenetic tree (Fig. 7). This might be as a result of slight differences in their nucleotide sequence arrangement (Appendix 1, Accession number NR_112601.1 and NR_29091.1) bringing about changes in their amino acid and evolutionary trend.

*Streptomyces malaysiensis* strain NBRC 16446 was first isolated from the soil but now found in *Perna perna*. It has the GC content of 72.2 mol% (Al-Tai *et al*., 1999). It belongs to the same group with *Lechevalieria nigeriaca* strain NJ2035 on the evolutionary tree (Fig. 7). NBRC 16446 type strain is ATB-11=NBRC 16446=JCM 10672=IFO 16446=DSM 41697=ATCC BAA-13. While, *Lechevalieria nigeriaca* NJ2035 was isolated from arid soil and have the GC content of 68.4 mol% (Camas *et al*., 2013). It has the type strain NJ2035=NRRL B-24881=KCTC 29057=DSM 45680.

*Saccharothrix* is Gram-positive belonging to the Phylum Actinomycetes and members of the family Pseudonocardiaceae. Different strains of *Saccharothrix* were identified in *Perna perna*; they include *Saccharothrix espanaensis* strain LL-C19004-NS29, *Saccharothrix yanglingensis* strain Hhs.015, *Saccharothrix algeriensis* strain SA 233, *Saccharothrix espanaensis* strain

NRRL 15764, *Saccharothrix coeruleofusca* strain NRRL B-16115 and *Saccharothrix australiensis* strain NRRL 11239.

*Saccharothrix espanaensis* have two different strains, NRRL 15764 and LL-C19004-NS29. They have different nucleotide sequence and appear on a separate group on the phylogenetic tree (Fig. 7). The strain *Saccharothrix coeruleofusca* NRRL B-16115 and *Saccharothrix espanaensis* LL-C19004-NS29 belong to the same group on the evolutionary tree and are closely related to *Streptomyces hebeiensis* YIM 001. Likewise strain Hhs.015, NRRL 11239 and NRRL 15764 belong to same group on the phylogenetic tree. Strain SA 233 is closely related to *Nesterenkonia flava* strain CAAs on the evolutionary tree. The slight changes observed in these bacteria may be due to mutation in the nucleotide sequence resulting into changes in the diversity of the *Saccharothrix* species on the phylogenetic tree (Fig. 7).

**Bacteria identified in *Perna perna* during red tide in February 2014**

*Rhodococcus* is Gram-positive bacteria belonging to Phylum Actinobacteria and members of the family Nocardiaceae. *Rhodococcus marinonascens* is halophilic, psychrotrophic, aerobic and slow-growing *Nocardioform actinomycetes* isolated from marine bottom sediments. It grew well at temperature around 20 °C in a media with seawater content of 75 to 100% (Helmke and Weyland, 1984). *Rhodococcus marinonascens* DSM 43752 and *Rhodococcus marinonascens* ATCC 35653 have the same type strain 3438W=VKM Ac-1182=NRRL B-16940=NBRC 14363=JCM 6241=IFO 14363=DSM 43752=CIP 104177=ATCC 35653 but have different nucleotide sequence. Strain DSM 43752 is closely related to *Arthrobacter korrensis* strain CA 15-8 on the phylogenetic tree while strain ATCC 35653 is related to *Kocuria palustris* strain TAGA 27.

The genus *Paracoccus* is Gram-negative with high GC content of 62 to 68 mol%. It belongs to Phylum Proteobacteria and members of the family Rhodobacteraceae (Urakami *et al*., 1990).

*Paracoccus aminophilus* strain DM-15 and JCM 7686 appear to have the same type strain but different nucleotide sequence. Strain DM-15 stands separate on the phylogenetic tree and it has the type strain VKM B-2141=NBRC 16710=JCM 7686=IFO 16710=DSM 8538=CIP 106077=ATCC 49673. *Paracoccus aminophilus* strain NBRC 16710 and JCM 7686 belong to the sample group on the phylogenetic tree and they are closely related to *Kocuria varians* ATCC 15306 on the evolutionary tree.

*Cohaesibacter gelatinilyticus* strain CL-GR15 is Gram-negative facultative anaerobic (Cohaesibacter) and aerobic (Breoghania) rods bacteria. It has the GC content that ranges between 52 to 64 mol%. Strains of *Cohaesibacter* are phylogenetically group into the subclass Alphaproteobacteria and are members of the family Cohaesibacteraceae (Rua and Thompson, 2014). They were first isolated from coastal waters of the east coast of Korea and China. The type strain is CL-GR15=KCCM 42319=DSM 18289.

As mentioned early, *Kocuria* is a member of the Micrococcaceae family, suborder Micrococcineae, orders Actinomycetales, class Actinobacteria and Gram-positive bacteria (Takarada *et al*., 2008; Lee *et al*., 2009; Stackebrandt *et al*., 1995; Savini *et al*., 2010). *Kocuria rhizophila* DC2201 strain DC2201 having the type strain NBRC 103217 and was derived from IFO 12708 (Fujita *et al*., 2006). It has the ability to exhibit tolerance to wide variety of organic solvents. Thus, it can be use for the development of bacterial bioconversion system which could be used under harsh conditions (Fujita *et al*., 2006).

*Kocuria polaris* strain CMS 76or was first isolated from a cyanobacterial mat sample in a pond located at McMurdo Dry Valley (Reddy *et al*., 2003). It has the type strain DSM 14382=NBRC 103063=MTCC 3702=JCM 12076. It is closely related to *Spirochaeta litoralis* R1 on the evolutionary tree.

*Spirochaeta halophila* RS1 is facultative anaerobe and involved in the bioconversion of cellulose containing wastes to fuel. The type strain is DSM 10522=ATCC 29478.

In general, more information about the epidemiology, application and importance of some these bacteria have not been said in the literature, so this work only provide little details that are relevant to this study.

**Conserved regions of DNA sequences in the identified bacteria before and during red tide**

The conserved regions of the nucleotide sequences in bacteria are highly important sites of evolutionary divergence (Prabhakar *et al*., 2006). These regions are highlighted in red; Appendix 1. Evolution does not tolerate many changes in these regions because the regions are unique to all bacteria. Changes in any conserved regions as a result of mutations could alter the regulation of conserved genes, producing species-specific patterns of gene expression and leads to evolutionary changes (Prabhakar *et al*., 2006). The active sites of enzymes and the binding sites of protein receptors are among the most highly conserved sequences and this sequence often harbour cis- regulatory elements which constrain evolution (Gross, 2007).

In this study, different bacteria have different nucleotides sequence which led to differences under the phylogenetic evolution (Appendix 1; Fig. 7). The differences observed in the nucleotide sequence among bacteria might lead to different coding of the amino acid which could reflect different protein structures. In turn, the protein structures and functions might lead to differences in the phylogenetic evolution (Fig. 7). According to Petsko and Ringe (2004) proteins are the embodiment of the transition from one-dimensional world of sequences to the three-dimensional world of molecules capable of diverse activities. Moreso, evolutionary differences occur in bacteria due to errors during translation to transcription at the molecular level. This explains the current observation in different families and strains identified in *Perna perna* before and during red tide (Table 2a-b & 3a-c). The higher levels of sequence variation

allow differentiation of closely related strains (Ochmann and Wilson, 1987) while the ability to translate DNA to protein sequences permits phylogenetic analysis of distantly related strains and more accurate sequence alignment (Gupta, 1998).

**Phylogenetic relationships of the identified bacteria in *Perna perna* before and during red tide in Algoa Bay, South Africa, February, 2014.**

The evolutionary relationships of the bacterial populations present in *Perna perna* before and during red tide are shown in the phylogenetic (Fig. 7). Changes in the evolutionary trend of the bacteria and the optimal tree has the sum of branch length = 39.21676153. The numbers on the tree indicates the time at which mutation occurs which brings about taxa separation in the bacteria. The bacteria identified from *Perna perna* during red tide are shown in bold letters, while those common to both samples are in blue colour. They include *Kocuria rosea* strain DSM 20447, *Spirochaeta litoralis* strain R1, *Spirochaeta americana* strain A SpG1 and *Kocurria salsicia* strain 104. *Kocuria rosea* strain DSM 20447 shares the same group with *Arthrobacter luteolus* strain CF-25 after taxa separation, *Spirochaeta litoralis* strain R1 also belongs to the same group with *Kocuria polaris* strain CMS 76or while *Spirochaeta americana* strain A SpG1 was separated from *Citricoccus nitrophenolicus* strain PNP1 and *Micrococcus terreus* strain V3M1 on the phylogenetic tree. *Paracoccus aminophilus* strain NBRC-16710 and *Paracoccus aminophilus* strain JCM 686 belong to the same group but got separated from the *Kocuria* spp. Mutation also got *Kocuria rhizophila* DC2201 strain (NBRC 103217) and *Spirochaeta halophila* strain RS1 separated.

**Phylogenetic relationships of the identified bacteria in *Perna perna* before and during red tide in Algoa Bay, South Africa, February, 2014.**

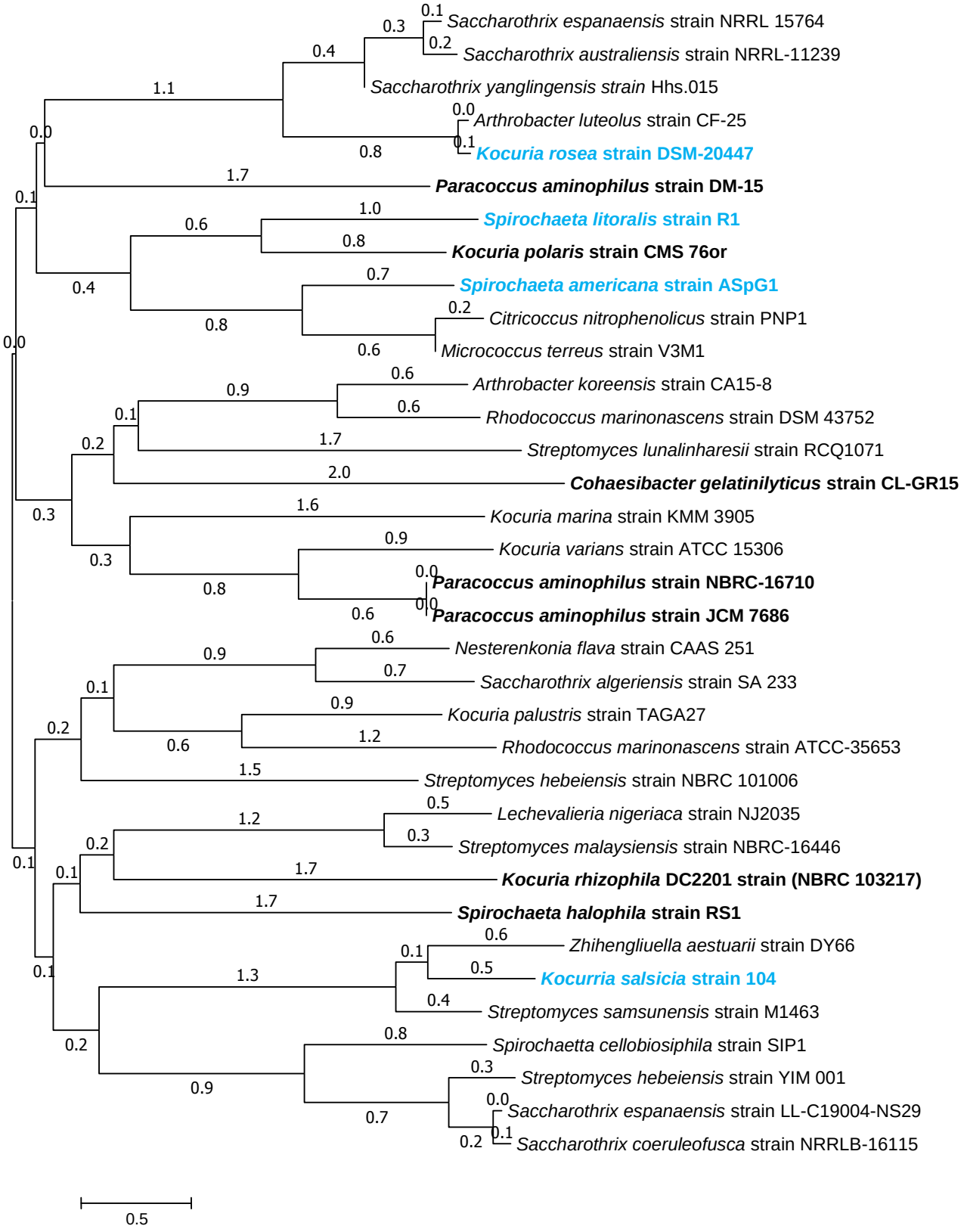


Figure 7: Neighbour-joining tree showing the phylogenetic analysis using MEGA 6 based on 16S rRNA gene sequences of bacteria present in *Perna perna* before and during red tide. The branch length at which each bacteria move before evolution takes place is also shown.

Diversity in the phylogenetic tree might be due to changes in the bacteria nucleotides sequence caused by mutation, which could be substitution, deletion, insertion or frame shift mutation. Mutation usually leads to changes in the amino acid sequences in different bacteria. For instance in Table 2a, *Streptomyces hebeiensis* occurs with two different strains NBRC 101006 and YIM 001. This could be as a result of deletion mutation which occurs in strain NBRC 101006 where about 43 nucleotide sequences are deleted (Appendix 1). The deleted nucleotides codes for arginine, valine, leucine, phenylalanine, glycine, alanine, serine, proline and trptophan on the codon table (Appendix 2) which code for different amino acids; resulting in two strains. Also *Rhodococcus marinonascens* having strain DSM 43752 and ATCC 35653 differs in about 35 nucleotide sequence which code for proline, glycine, serine, arginine, asparagine valine and alanine, which also resulted in bacteria having different strain and appearing in different groups on the phylogenetic tree. This divergence was also noticed in *Saccharothrix espanaensis* strain LL-C19004-NS29 and NRRL 15764, likewise in *Paracoccus aminophilus* strain DM-15, NBRC 16710 and NBRC 16710 (Figure 7) which results in different group on the phylogenetic tree.

From the evolutionary tree, it could be deduced that, bacteria found in *Perna perna* come from the same origin but due to evolutionary changes, there had been led to taxa separation over a period of time.

## CONCLUSION

Globally, shellfish are nutritious and have economical values. Despite these benefits, they harbour bacteria and thus are well known as efficient transmitters of seafood-borne diseases. Using universal primers 16S rRNA gene, the findings of this study reveal that *Perna perna* before and during the red tide collected from Algoa Bay in Port Elizabeth harbour bacteria from different genera. Therefore, it is advisable that individuals, who eat *Perna perna* and seafoods generally, should avoid their consumption either raw or slightly cooked.

Also, in support of the hypothesis, there were more bacteria populations in *Perna perna* that were not affected with red tide compare to the ones that were collected during the red tide; indicating that red tide affects bacterial populations in *Perna perna*. It may even be more dangerous to consume *Perna perna* at the moment of red tide. This is because red algae are believed to release toxins into their environment (Jorge, 2014) and these toxins may lead to paralysis, vomiting, respiratory irritation and amnesia to the consumers.

Lastly, the microbial diversity of marine bivalves should be thoroughly investigated together with the potential microorganisms producing bioactive metabolites. Moreso that most of the bacteria identified in this study may be useful in biodegradation of p-nitrophenol (pNP), detoxification or biodegeneration of many environmental pollutants, bioremediation of ground water contaminated with pesticides and herbicides, biodegradation of Polyethylene and bioconversion of cellulose-containing wastes to fuel such as ethanol or hydrogen gas.

**References**


Aksenova, H. Y., Rainey, F. A., Janssen, P. H., Zavarzin, G. A., & Morgan, H. W. (1992). Spirochaeta thermophila sp. nov., an obligately anaerobic, polysaccharolytic, extremely thermophilic bacterium. *International Journal of Systematic Bacteriology*, *42*(1), 175-177.

Al-Tai, A., Kim, B., Kim, S. B., Manfio, G. P., & Goodfellow, M. (1999). *Streptomyces malaysiensis* sp. nov., a new *Streptomycete* species with rugose, ornamented spores. *International Journal of Systematic and Evolutionary Microbiology*, *49*(4), 1395-1402.

Altuntas, F., Yildiz, O., Eser, B., Gündogan, K., Sumerkan, B., & Çetin, M. (2004). Catheter-related bacteremia due to *Kocuria rosea* in a patient undergoing peripheral blood stem cell transplantation. *BMC Infectious Diseases*, *4*(1), 62.

Anderson, D. M. (1989). Toxic algal blooms and red tides: a global perspective. *Red tides: Biology, Environmental Science and Toxicology*, 11-16.

Baik, K. S., Lim, C. H., Park, S. C., Choe, H. N., Kim, H. J., Kim, D., & Seong, C. N. (2011). *Zhihengliuella aestuarii* sp. nov., isolated from tidal flat sediment. *International Journal of Systematic and Evolutionary Microbiology*, *61*(7), 1671-1676.

Ben-Ami, R., Navon-Venezia, S., Schwartz, D., & Carmeli, Y. (2003). Infection of a ventriculoatrial shunt with phenotypically variable *Staphylococcus epidermidis* masquerading as polymicrobial bacteremia due to various coagulase-negative *Staphylococci* and *Kocuria varians*. *Journal of Clinical Microbiology*, *41*(6): 2444-2447.

Breznak, J. A. & Warnecke, F. (2008). *Spirochaeta cellobiosiphila* sp. nov., a facultatively anaerobic, marine spirochaete. *International Journal of Systematic and Evolutionary Microbiology*, *58*(12), 2762-2768.

Camas, M., Veyisoglu, A., Tatar, D., Saygin, H., Cetin, D., Sazak, A., & Sahin, N. (2013). Lechevalieria nigeriaca sp. nov., isolated from arid soil. *International Journal of Systematic and Evolutionary Microbiology*, *63*(10), 3750-3754.

Canale-Parola, E. (1984). Genus I. Spirochaeta Ehrenberg 1835, 313AL. In Bergey's Manual of Systematic Bacteriology, vol. 1, pp. 39– 46. Edited by N. R. Krieg and J. G. Holt. Baltimore: Williams & Wilkins.

Canale-Parola, E. (1992). Free-living *Saccharolytic spirochetes*: the genus *Spirochaeta*. In The Prokaryotes, 2nd edn, pp. 3524–3536. Edited by A. Balows, H. G. Tru¨per, M. Dworkin, W. Harder and K. H. Schleifer. New York: Springer.

Chen, Y. G., Tang, S. K., Zhang, Y. Q., Liu, Z. X., Chen, Q. H., He, J. W., & Li, W. J. (2010). *Zhihengliuella salsuginis* sp. nov., a moderately halophilic actinobacterium from a subterranean brine. *Extremophiles*, *14*(4), 397-402.

de Souza, R. F., Coelho, R. R. R., Macrae, A., Soares, R. M. A., Nery, D. D. C. M., Semêdo, L. T. D. A. S., and Gomes, R. C. (2008). Streptomyces lunalinharesii sp. nov., a chitinolytic streptomycete isolated from cerrado soil in Brazil. *International Journal of Systematic and Evolutionary Microbiology*, *58*(12), 2774-2778.

Emmanuel, E. C., Ananthi, T., Anandkumar, B., & Maruthamuthu, S. (2012). Accumulation of rare earth elements by siderophore-forming *Arthrobacter luteolus* isolated from rare earth environment of Chavara, India. *Journal of Biosciences*, *37*(1), 25-31.

Euzéby, J. P. (2008). "Genus *Streptomyces*". *List of Prokaryotic names with Standing in Nomenclature*. Retrieved 2008-09-28.

Fujita, K., Hagishita, T., Kurita, S., Kawakura, Y., Kobayashi, Y., Matsuyama, A., & Iwahashi, H. (2006). The cell structural properties of *Kocuria rhizophila* for aliphatic alcohol exposure. *Enzyme and Microbial Technology*, *39*(3), 511-518.

Gomes, R. C., Semedo, L. T. A. S., Soares, R. M. A., Linhares, L. F., Ulhoa, C. J., Alviano, C. S., & Coelho, R. R. R. (2001). Purification of a thermostable endochitinase from *Streptomyces* RC1071 isolated from a cerrado soil and its antagonism against phytopathogenic fungi. *Journal of Applied Microbiology*, *90*(4), 653-661.

Gross, L. (2007). Are "Ultraconserved" Genetic Elements Really Indispensable?. *PLoS Biology*, *5*(9), e253.

Gupta, R. S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiology and Molecular Biology Reviews*, *62*(4), 1435-1491.

Helmke, E. & Weyland, H. (1984). *Rhodococcus marinonascens* sp. nov., an Actinomycete from the Sea. *International Journal of Systematic Bacteriology*, *34*(2), 127-138.

Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley J. T., & Williams, S. T. (1994). Bergey's Manual of Determinative Bacteriology. Nineth Edition. 787 pp, Baltimore, Maryland: Williams & Wilkins.

Hoover, R. B., Pikuta, E. V., Bej, A. K., Marsic, D., Whitman, W. B., Tang, J., & Krader, P. (2003). *Spirochaeta americana* sp. nov., a new haloalkaliphilic, obligately anaerobic spirochaete isolated from soda Mono Lake in California. *International Journal of Systematic and Evolutionary Microbiology*, *53*(3), 815-821.

Jiang, Y., Han, L., Chen, X., Yin, M., Zheng, D., Wang, Y., & Huang, X. (2013). Diversity and Bioactivity of Cultivable Animal Fecal *Actinobacteria*. *Advances in Microbiology*, *3*, 1.

Jones, D. & Keddie, R. M. (2007). The Prokaryotes. (M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt, Eds.) (pp. 945–960). Springer New York. doi:10.1007/0-387-30743-5.

Kämpfer, P. (2006). The family Streptomycetaceae, part I: taxonomy. In *The Prokaryotes* (pp. 538-604). Springer New York.

Kaya, K. E., Kurtoğlu, Y., Cesur, S., Bulut, C., Kinikli, S., Irmak, H., & Karakoc, E. (2009). [Peritonitis due to Kocuria rosea in a continuous ambulatory peritoneal dialysis case]. *Mikrobiyoloji Bulteni*, *43*(2), 335-337.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). *Practical Streptomyces Genetics* (2nd ed.). Norwich, England: John Innes Foundation

Kim, S. B., Nedashkovskaya, O. I., Mikhailov, V. V., Han, S. K., Kim, K. O., Rhee, M. S., & and Bae, K. S. (2004). *Kocuria marina* sp. nov., a novel actinobacterium isolated from marine sediment. *International Journal of Systematic and Evolutionary Microbiology*, *54*(5), 1617-1620.

Kocur, M., Kloos, W. E., & Schleifer, K. H. (2006). The genus *Micrococcus*. In *The Prokaryotes* (pp. 961-971). Springer New York.

Kovács, G., Burghardt, J., Pradella, S., Schumann, P., Stackebrandt, E., & Màrialigeti, K. (1999). *Kocuria palustris* sp. nov. and *Kocuria rhizophila* sp. nov., isolated from the rhizoplane of the narrow-leaved cattail (Typha angustifolia). *International Journal of Systematic Bacteriology*, *49*(1), 167-173.

Lee, J. S., Lee, K. C., Pyun, Y. R., & Bae, K. S. (2003). *Arthrobacter koreensis* sp. nov., a novel alkalitolerant bacterium from soil. *International Journal of Systematic and Evolutionary Microbiology*, *53*(5), 1277-1280.

Lee, J. Y., Kim, S. H., Jeong, H. S., Oh, S. H., Kim, H. R., Kim, Y. H., & Shin, J. H. (2009). Two cases of peritonitis caused by *Kocuria marina* in patients undergoing continuous ambulatory peritoneal dialysis. *Journal of Clinical Microbiology*, *47*(10), 3376-3378

Lee, M. K., Choi, S. H., & Ryu, D. W. (2013). Descending necrotizing Mediastinitis caused by Kocuria rosea: a case report. *BMC Infectious Diseases*, *13*(1), 475.

Leschine, S., Paster, B. J., & Canale-Parola, E. (2006). Free-living saccharolytic spirochetes: the genus Spirochaeta. In *The prokaryotes* (pp. 195-210). Springer New York.

Luo, H. Y., Miao, L. H., Fang, C., Yang, P. L., Wang, Y. R., Shi, P. J., & Fan, Y. L. (2008). *Nesterenkonia flava* sp. nov., isolated from paper-mill effluent. *International Journal of Systematic and Evolutionary Microbiology*, *58*(8), 1927-1930.

Ma, E. S., Wong, C. L., Lai, K. T., Chan, E. C., Yam, W. C., & Chan, A. C. (2005*). Kocuria kristinae* infection associated with acute cholecystitis. *BMC Infectious Diseases*, *5*(1), 60.

Madigan, M. T. (2005). Brock Biology of Microorganisms, 11th edn.

Mongodin, E. F., Shapir, N., Daugherty, S. C., DeBoy, R. T., Emerson, J. B., Shvartzbeyn, A., & Sadowsky, M. J. (2006). Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. *PLoS Genetics*, *2*(12): e214.

Nielsen, M. B., Kjeldsen, K. U., & Ingvorsen, K. (2011). Description of *Citricoccus nitrophenolicus* sp. nov., a para-nitrophenol degrading actinobacterium isolated from a wastewater treatment plant and emended description of the genus *Citricoccus* Altenburger *et al*., 2002. *Antonie van Leeuwenhoek*, *99*(3), 489-499.

Ochman, H. & Wilson, A. C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution*, *26*(1-2), 74-86.

Pesticide Action Network (2008). The pesticide action network (PAN) pesticide database: www.pesticideinfo.org. Accessed 2014/12/01.

Petsko, G. A. & Ringe, D. (2004). *Protein Structure and Function*. New Science Press.

Prabhakar, S., Noonan, J. P., Pääbo, S., & Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science*, *314*(5800), 786-786.

Reddy, G. S., Prakash, J. S., Prabahar, V., Matsumoto, G. I., Stackebrandt, E., & Shivaji, S. (2003). Kocuria polaris sp. nov., an orange-pigmented psychrophilic bacterium

isolated from an Antarctic cyanobacterial mat sample. *International Journal of Systematic and Evolutionary Microbiology*, *53*(1), 183-187.

Rua, C. P. & Thompson, F. (2014). The Family Cohaesibacteraceae: The Genera *Cohaesibacter* and *Breoghania*. *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, 207-212.

Savini, V., Catavitello, C., Masciarelli, G., Astolfi, D., Balbinot, A., Bianco, A., & D'Antonio, D. (2010). Drug sensitivity and clinical impact of members of the genus *Kocuria*. *Journal of Medical Microbiology*, *59*(12), 1395-1402.

Sazak, A., Şahin, N., Güven, K., Işık, K., & Goodfellow, M. (2011). *Streptomyces samsunensis* sp. nov., a member of the *Streptomyces violaceusniger* clade isolated from the rhizosphere of Robinia pseudoacacia. *International Journal of Systematic and Evolutionary Microbiology*, *61*(6), 1309-1314.

Smith, K. J., Neafie, R., Yeager, J., & Skelton, H. G. (1999). Micrococcus folliculitis in HIV-1 disease. *British Journal of Dermatology*, *141:* 558-561.

Srinivasa, K. H., Agrawal, N., Agarwal, A., & Manjunath, C. N. (2013). Dancing vegetations: *Kocuria rosea* endocarditis. *BMJ Case Reports*, 2013.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kämpfer, P., Maiden, M. C., & Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *52*(3), 1043-1047.

Stackebrandt, E., Koch, C., Gvozdiak, O., & Schumann, P. (1995). Taxonomic Dissection of the Genus *Micrococcus*: *Kocuria* gen. nov., *Nesterenkonia* gen. nov., *Kytococcus* gen. nov., *Dermacoccus* gen. nov., and *Micrococcus* Cohn 1872 gen. emend. *International Journal of Systematic Bacteriology*, *45*(4), 682-692.

Takarada, H., Sekine, M., Kosugi, H., Matsuo, Y., Fujisawa, T., Omata, S., & Harayama, S. (2008). Complete genome sequence of the soil actinomycete Kocuria rhizophila. *Journal of Bacteriology, 190*(12), 4139-4146.

Tang, S. K., Wang, Y., Chen, Y., Lou, K., Cao, L. L., Xu, L. H., & Li, W. J. (2009). *Zhihengliuella alba* sp. nov., and emended description of the genus *Zhihengliuella*. *International Journal of Systematic and Evolutionary Microbiology*, *59*(8), 2025-2032.

Tsai, C. Y., Su, S. H., Cheng, Y. H., Chou, Y. L., Tsai, T. H., & Lieu, A. S. (2010). *Kocuria varians* infection associated with brain abscess: A case report. *BMC Infectious Diseases*, *10*(1), 102.

Urakami, T., Araki, H., Oyanagi, H., Suzuki, K. I., & Komagata, K. (1990). *Paracoccus aminophilus* sp. nov. and *Paracoccus aminovorans* sp. nov., which utilize N, N-dimethylformamide. *International Journal of Systematic Bacteriology*, *40*(3), 287-291.

U.S. Environmental Protection Agency (EPA) (2012). (http:// www.epa.gov/waterscience/methods/pollutants.htm)

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007). Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews*, *71*(3), 495-548.

Větrovský, T. & Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One*, *8*(2): e57923.

Wauters, G., Charlier, J., Janssens, M., & Delmée, M. (2000). Identification of *Arthrobacter oxydans, Arthrobacter luteolus* sp. nov., and *Arthrobacter albus* sp. nov., isolated from human clinical specimens. *Journal of Clinical Microbiology*, *38*(6), 2412-2415.

Yun, J. H., Roh, S. W., Jung, M. J., Kim, M. S., Park, E. J., Shin, K. S., & Bae, J. W. (2011). *Kocuria salsicia* sp. nov., isolated from salt-fermented seafood. *International Journal of Systematic and Evolutionary Microbiology*, *61*(2), 286-289.

Zhang, Y. Q., Schumann, P., Yu, L. Y., Liu, H. Y., Zhang, Y. Q., Xu, L. H., & Li, W. J. (2007). *Zhihengliuella halotolerans* gen. nov., sp. nov., a novel member of the family Micrococcaceae. *International Journal of Systematic and Evolutionary Microbiology*, *57*(5), 1018-1023.

Zhuang, W. Q., Tay, J. H., Maszenan, A. M., Krumholz, L. R., & Tay, S. L. (2003). Importance of Gram-positive naphthalene-degrading bacteria in oil-contaminated tropical marine sediments. *Letters in Applied Microbiology*, *36*(4), 251-257.

**Appendix 1: Multiple Sequence Alignment of the Bacteria Identified**

Multiple sequence alignment was used to align the biological sequences having similar length. The inferrs homology and studied the evolutionary relationship between sequences. The table below shows the conserved regions of the bacteria present in *Perna perna* before and during red tide in October 2014 and are marked in colour red. The asterisk "*" represent the conserved regions in the bacteria

CLUSTAL 2.1 multiple sequence alignment

```
NR_025665.1 --------------------------------------------------GATGAACGCTG 11
NR_025362.1 ------------------------------------CCTGGCTCAGGATGAACGCTG 21
NR_117546.1 --------------------------------TTTGATCCTGGCTCAGGATGAACGCTG 27
NR_116649.1 ----------------------------AGAGTTTGATCCTGGCTCAGGATGAACGCTG 31
NR_116519.1 ------------------------------------------------------------ 0
NR_117299.1 ------------------------------------------------------------ 0
NR_114674.1 ------------------------------------------------------------ 0
NR_74786.1  ----------------------TTGACGGAGAGTTTGATCCTGGCTCAGGACGAACGCTG 38
NR_25723.1  --------------------------------------------------GACGAACGCTG 11
NR_044871.1 ----------------------------------ATCCTGGCTCAGGACGAACGCTG 23
NR_44871.1  ----------------------------------ATCCTGGCTCAGGACGAACGCTG 23
NR_28924.1  ----------------------------TGAGTTTGATCCTGGCTCAGGACGAACGCTG 31
NR_26451.1  ------------------------------------------------------------G 1
NR_44353.1  ------------------------------------------------------------ 0
NR_116165.1 ------------------------------------------------------------ 0
NR_41410.1  --------------------------------------------GACGAACGCTG 11
NR_112601.1 --------------------------------------------GACGAACGCTG 11
NR_29091.1  ----------------------------AGAGTTTGATCC-TGGCTCAGGACGAACGCTG 31
NR_43541.1  ----------------------------TAGAGTTTGATACATGGCTCAGGACGAACGCTG 33
```

```
NR_102474.1 ----------------------------AGAGTTTGATCCTGGCTCAGGACGAACGCTG 31
NR_24965.1  ----------------------------GAGTTTGATCCTGGCTCAGGACGAACGCTG 30
NR_41735.1  ----------------------------GAGTTTGATCCTGGCTCAGGACGAACGCTG 30
NR_41736.1  ----------------------------GAGTTTGATCCTGGCTCAGGACGAACGCTG 30
NR_25658.1  ------------------------------------------------------GAACGCTG 8
NR_117283.1 ----------------------------AGAGTTTGATCCTGGCTCAGGACGAACGCTG 31
NR_109528.1 ------------------------------------------------TCAGGACGAACGCTG 15
NR_26183.1  ---------------------------------------CCTGGCTCAGGACGAACGCTG 21
NR_119127.1 ----------------------------------------------------ACGAACGCTG 10
NR_121706.1 ----------------------------AGAGTTTGATCCTGGCTCAGAACGAACGCTG 31
NR_113863.1 ------------------------------------------------AACGAACGCTG 11
NR_115554.1 ------------------------------------------------AACGAACGCTG 11
NR_43777.1  ------------------------------------------------------CAG 3
NR_104732.1 -------------------------------GTTTGATCCTGGCTCAGAACGAACGCTG 28
NR_44505.1  ---------------------------------------TCCTGGCTCAGAACGAACGCTG 22
NR_28820.1  ------------------------------------------------AACGAACGCTG 11
NR_44756.2  CGTCCCTTCGGGGACGTACAAAATCATGGAGAGTTTGATCCTGGCTCAGAACGAACGCTG 60

NR_025665.1 GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGACTTCT-GTGCTTGCA-CAGA-A 64
NR_025362.1 GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGACTTCT-GTGCTTGCA-CAGA-A 74
NR_117546.1 GCGGCGTGCTTAACACATGCAAGTCGAACGC----TGAAGCCCCA-GCTTGC--TGGG-G 79
NR_116649.1 GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGAAGCCC-A-GCTTGC--TGGG-- 81
NR_116519.1 --------CTTAACACATGCAAGTCGAACGA----TGATGCCT-A-GCTTGC--TAGG-- 42
NR_117299.1 ------------------TGCAAGTCGAACGC----TGAAGCTTGGTGCTTGCACTGG--G 37
NR_114674.1 ----------------------------CGC----TGAAGCTTGGTGCTTGCACTGG--G 26
NR_74786.1  GCGGCGTGCTTAACACATGCAAGTCGAACGC----TGAAGCTTGGTGCTTGCACTGG--G 92
NR_25723.1  GCGGCGTGCTTAACACATGCAAGTCGAACGC----TGAAGCTTGGTGCTTGCACTGG--G 65
NR_044871.1 GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGATGCCCA--GCTTGC--TGG--G 73
NR_44871.1  GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGATGCCCA--GCTTGC--TGG--G 73
NR_28924.1  GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGATGCCCA--GCTTGC--TGG--G 81
NR_26451.1  GCGGCGTGCTTAACACATGCAAGTCGAACGC----TGAAGCACCA-GCTTGC--TGGT-G 53
NR_44353.1  ----CGTGCTTAACACATGCAAGTCGAACGA----TGAAGCCC-GTGCTTGCA-CGG--G 48
NR_116165.1 ------TGCTTAACACATGCAAGTCGAACGA----TGAACCGG----TTT----CGGCCG 42
NR_41410.1  GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGAACCGG----TTT----CGGCCG 59
NR_112601.1 GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGAAGCCC----TTC----GGG--G 57
NR_29091.1  GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGAAGCCC----TTC----GGG--G 77
NR_43541.1  GCGGCGTGCTTAACACATGCAAGTCGAACGA----TGAACCTC----CTT----CGGGAG 81
NR_102474.1 GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 77
NR_24965.1  GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 76
NR_41735.1  GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 76
NR_41736.1  GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 76
NR_25658.1  GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 54
NR_117283.1 GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 77
NR_109528.1 GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 61
NR_26183.1  GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 67
NR_119127.1 GCGGCGTGCTTAACACATGCAAGTCGAGCGG----TAAGGCCC-----TT----CGGG-G 56
NR_121706.1 GCGGCAGGCCTAACACATGCAAGTCGAGCGC--------GCCC-----TT----CGGG-- 72
NR_113863.1 GCGGCAGGCCTAACACATGCAAGTCGAGCGC--------GCCC-----TT----CGGG-- 52
NR_115554.1 GCGGCAGGCCTAACACATGCAAGTCGAGCGC--------GCCC-----TT----CGGG-- 52
NR_43777.1  GCG--AG--CTACC--ATGCA-GTCGAACGG--------GCTC-----TT----CGGA-- 37
NR_104732.1 GCGGCGCGTTTTAAGCATGCAAGTCGAACGGGATCTGAGGTGC-----TT-----GCACC 78
NR_44505.1  GCGGCGCGTTTTAAGCATGCAAGTCGAACGGGAT-TCAAGTGC-----TT-----GCAC- 70
NR_28820.1  GCGGCGCGTTTTAAGCATGCAAGTCGA--GGGGTAGAAGCGCT-----TC-----GGTGC 59
NR_44756.2  GCGGCGCGTTTTAAGCATGCAAGTCGA--GGGGTAAGATGCCT-----TC-----GGGTA 108
                                         *                   *
NR_025665.1 TG-ATTAG--TGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCTTAACTTCGGGATA 121
NR_025362.1 TG-ATTAG--TGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCTTAACTTCGGGATA 131
NR_117546.1 TGGATGAG--TGGCGAACGGGTGAGTAACACGTGAGTAACCTTCCCTTGACTCTGGGATA 137
NR_116649.1 TGGATTAG--TGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCTTGACTCTGGGATA 139
NR_116519.1 TGGATTAG--TGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCTGACTCTGGGATA 100
NR_117299.1 TGGATGAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 95
```

```
NR_114674.1 TGGATGAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTAACTCTGGGATA 84
NR_74786.1  TGGATGAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 150
NR_25723.1  TGGATGAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 123
NR_044871.1 CGGATTAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 131
NR_44871.1  CGGATTAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 131
NR_28924.1  CGGATTAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 139
NR_26451.1  TGGATGAG--TGGCGAACGGGTGAGTAATACGTGAGTAACCTGCCCTTGACTCTGGGATA 111
NR_44353.1  TGGATTAG--TGGCGAACGGGTGAGTATCACGTGAGTAACCTGCCCTTAACTCTGGGATA 106
NR_116165.1 GGGATTAG--TGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACA 100
NR_41410.1  GGGATTAG--TGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACA 117
NR_112601.1 TGGATTAG--TGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACA 115
NR_29091.1  TGGATTAG--TGGCGAACGGGTGAGTAACACGTAGGCAATCTGCCCTGCACTCTGGGACA 135
NR_43541.1  GGGATTAG--TGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACA 139
NR_102474.1 TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCTGTACTCCGGGATA 135
NR_24965.1  TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCTGTACTCCGGGATA 134
NR_41735.1  TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCTGTACTCCGGGATA 134
NR_41736.1  TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGCAACCTGCCCCGTACTCTGGGATA 134
NR_25658.1  TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCTGTACTCCGGGATA 112
NR_117283.1 TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCTGTACTCCGGGATA 135
NR_109528.1 TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCTGTACTCTGGGATA 119
NR_26183.1  TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTCGGGATA 125
NR_119127.1 TACACGAG--CGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTCGGGATA 114
NR_121706.1 ---GTGAG--CGGCGGACGGGTGAGTAACACGTGGG-AACATACCCTTTTCTACGGAATA 126
NR_113863.1 ---GTGAG--CGGCGGACGGGTGAGTAACACGTGGG-AACATACCCTTTTCTACGGAATA 106
NR_115554.1 ---GTGAG--CGGCGGACGGGTGAGTAACACGTGGG-AACATACCCTTTTCTACGGAATA 106
NR_43777.1  ---GCTAG--TGGCAGACGGGTGAGTAACGCGTGGG-AACCTACCTATAAGTACGGAACA 91
NR_104732.1 TTGGTGAGAGTGGCGAACGGGTGAGTAACGCGTAGGTGATCTACCTTTTAGTTGGGGATA 138
NR_44505.1  TTGATGAGAGTGGCGAACGGGTGAGTAACACGTAGGTGATCTACCTAGAAGTTGGGGATA 130
NR_28820.1  TT--TGAGACCGGCGAACGGGTGAGTAACACGTAGGTGATCTGCCGTGAGGTTGGGGATA 117
NR_44756.2  TC--TGAGACCGGCGAACGGGTGAGTAACACGTAGGTGATCTACCTTGAGGTTGGGGATA 166
                  **   ***  ***********   ***  *  *  * **      *  ** * *

NR_025665.1 AGCCTGGGAAACCGGGTCTAATACCGGATACGACCTCCTGGCGCATGCCATGGTGGTGGA 181
NR_025362.1 AGCCTGGGAAACTGGGTCTAATACCGGATACGACCACCTGGCGCATGCCATGGTGGTGGA 191
NR_117546.1 AGCCCGGGAAACTGGGTCTAATACCGGATACGACCTCCTGCCGCATGGTGTGGGGGTGGA 197
NR_116649.1 AGCCTGGGAAACTGGGTCTAATACTGGATATGACTTCTCACCGCATGGTG--GGGGTGGA 197
NR_116519.1 AGCCTGAGAAATTGGGTCTAATACTGGATATGAGCATCTATCGCATGGTG-GGTGTTGGA 159
NR_117299.1 AGCCTGGGAAACTGGGTCTAATACTGGATATGACATGCCACTGCATGGTGGTGTGTGGAA 155
NR_114674.1 AGCCTGGGAAACTGGGTCTAATACTGGATATGACATGTCACTGCATGGTGGCGTGTGGAA 144
NR_74786.1  AGCCTGGGAAACTGGGTCTAATACTGGATACGACGGCGCATCGCATGGTGTGTTGTGGAA 210
NR_25723.1  AGCCTGGGAAACTGGGTCTAATACTGGATATGACCTCCTGTCGCATGGTGGGGGGTGGAA 183
NR_044871.1 AGCCTGGGAAACTGGGTCTAATACTGGATACTACCTCTTACCGCATGGTGGGTGGTGGAA 191
NR_44871.1  AGCCTGGGAAACTGGGTCTAATACTGGATACTACCTCTTACCGCATGGTGGGTGGTGGAA 191
NR_28924.1  AGCCTGGGAAACTGGGTCTAATACTGGATACTACCGTCCACCGCATGGTGGGTGGTGGAA 199
NR_26451.1  AGCCCGGGAAACTGGGTCTAATACTGGATGCTACATGTCACCGCATGGTGGTGTGTGGAA 171
NR_44353.1  AGCCTGGGAAACTGGGTCTAATACCGGATACGACCAGTCACCGCATGGTGTGCTGGTGGA 166
NR_116165.1 AGCCCTGGAAACGGGGTCTAATACCGGATATGACGCGTTCCCGCATGGGATACGTGTGGA 160
NR_41410.1  AGCCCTGGAAACGGGGTCTAATACCGGATANGACGCGTTCCCGCATGGGATACGTGTGGA 177
NR_112601.1 AGCCCTGGAAACGGGGTCTAATACCGGATATCACTTTTCCTCGCATGGGGGAGG-GTTGA 174
NR_29091.1  AGCCCTGGAAACGGGGTCTAATACCGGATATCACTTTTCCTCGCATGGGGGAGG-GTTGA 194
NR_43541.1  AGCCCTGGAAACGGGGTCTAATACCGGATATGACACGGGATCGCATGGTCTCCGTGTGGA 199
NR_102474.1 AGCCTGGGAAACTAGGTCTAATACCGGATACGACCCCATGGGGCATCTTG-TGGGGTGGA 194
NR_24965.1  AGCCTGGGAAACTAGGTCTAATACCGGATACGACCCCATGGGGCATCTTG-TGGGGTGGA 193
NR_41735.1  AGCCTGGGAAACTAGGTCTAATACCGGATACGACCTTCCATCGCATGGTG-GGGGGTGGA 193
NR_41736.1  AGCCTGGGAAACTAGGTCTAATACCGGATATGACTTCGCATCGCATGGTGGTGGGGTGGA 194
NR_25658.1  AGCCTGGGAAACTAGGTCTAATACCGGATACGACCCTCCATCGCATGGTG-GGGGGTGGA 171
NR_117283.1 AGCCTGGGAAACTAGGT-TAATACCGGATACGACCTGTCACCGCATGGTG-GTGGGTGGA 193
NR_109528.1 AGCCCGGGAAACTGGGTCTAATACCGGATACGACCACTTGACGCATGTCT-GGTGGTGGA 178
NR_26183.1  AGCCTGGGAAACTGGGTCTAATACCGGATACGACCTTCGNCTGCATGACT-GGGGGTGGA 184
NR_119127.1 AGCCTGGGAAACTGGGTCTAATACCGGATACGACCTTCGGCTGCATGGCT-GGGGGTGGA 173
NR_121706.1 GCCTCGGGAAACTGAGAGTAATACCGTATACGCCCTTC----------------GGGGGA 170
```

```
NR_113863.1 GCCTCGGGAAACTGAGAGTAATACCGTATACGCCCTTC----------------GGGGGA 150
NR_115554.1 GCCTCGGGAAACTGAGAGTAATACCGTATACGCCCTTC----------------GGGGGA 150
NR_43777.1  ACACAGAGAAATTTGTGCTAATACCGTATGTGCCCTTT----------------GGGGTA 135
NR_104732.1 GCCCATGGAAACATGGGGTAATACCGAATA-ATCTCATTGA--TTTTGTTTAATGA-GGA 194
NR_44505.1  GCTCATGGAAACATGGGGTAATACCGAATG-AGCTTTTAGAACTATGGTTTTAAAA-GGA 188
NR_28820.1  GCCTGTGGAAACACAGGGTAATACCGAATG-AGCTTCTGACACTGTGGTG-ACAGAAGGA 175
NR_44756.2  GCCCATGGAAACATGGGGTAATACCGAATG-AGCTCTATGTACTGTGGTGTATAGA-GGA 224
                   ****       ****** * **                           *

NR_025665.1 AAGC----TTT-A-GCGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTTGGGGTA 235
NR_025362.1 AAGC----TTT-ATGCGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGT-GGGGTA 245
NR_117546.1 AAGA----TTT--ATCGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGT-GAGGTA 250
NR_116649.1 AAGA----TTT--ATTGGTCTTGGATGGACTCGCGGCCTATCAGCTTGTTGGT-GAGGTA 250
NR_116519.1 AAGA----TTT--TTTGGTTAGGGATGGACTCGCGGCCTATCAGCTTGTTGGT-GAGGTA 212
NR_117299.1 AGGG----TTT-TACTGGTTTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 209
NR_114674.1 AGGG----TTT-TACTGGTTTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 198
NR_74786.1  AGGG----TTT-TACTGGTTTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 264
NR_25723.1  AGGG----TTTGTACTGGTTGTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 238
NR_044871.1 AGGG----TTT-TACTGGTTTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 245
NR_44871.1  AGGG----TTT-TACTGGTTTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 245
NR_28924.1  AGGG----TTT-TACTGGTTTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GGGGTA 253
NR_26451.1  AGGG----TTT--ACTGGTCTTGGATGGGCTCACGGCCTATCAGCTTGTTGGT-GAGGTA 224
NR_44353.1  AAGC----TTT--TGCGGTTTTGGATGGACTCGCGGCCTATCAGCTAGACGGT-GGGGTA 219
NR_116165.1 AAGC----TCC--GGCGGTGCAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 213
NR_41410.1  AAGC----TCC--GGCGGTGCAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 230
NR_112601.1 AAGC----TCC--GGCGGTGCAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 227
NR_29091.1  AAGC----TCC--GGCGGTGCAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 247
NR_43541.1  AAGC----TCC--GGCGGTGAAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GAGGTA 252
NR_102474.1 AAGT----TCC--GGCGGTATGGGATGGACCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 247
NR_24965.1  AAGT----TCC--GGCGGTATGGGATGGACCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 246
NR_41735.1  AAGT----TCC--GGCGGTATGGGATGGACCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 246
NR_41736.1  AAGT----TCC--GGCGGTACGGGATGGGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTA 247
NR_25658.1  AAGT----TCC--GGCGGTATGGGATGGACCCGCGGCCTATCAGCTTGTTGGT-GGGGTG 224
NR_117283.1 AAGT----TCC--GGCGGTATGGGATGGACCCGCGGCCTATCAGCTTGTTGGT-GGGGTA 246
NR_109528.1 AAGT----TCC--GGCGGTATGGGATGGACCCGCGGCCTATCAGCTTGTTGGT-GGGGTA 231
NR_26183.1  AAGG----TTT--ACTGGTGCAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTA 237
NR_119127.1 AAGG----TTT--ACTGGTGCAGGATGAGCCCGCGGCCTATCAGCTTGTTGGT-GGGGTA 226
NR_121706.1 AAGA----TTT--ATCGGAGAAGGATTGGCCCGCGTTGGATTAGGTAGTTGGT-GGGGTA 223
NR_113863.1 AAGA----TTT--ATCGGAGAAGGATTGGCCCGCGTTGGATTAGGTAGTTGGT-GGGGTA 203
NR_115554.1 AAGA----TTT--ATCGGAGAAGGATTGGCCCGCGTTGGATTAGGTAGTTGGT-GGGGTA 203
NR_43777.1  AAGA----TTT--ATCGCTTATAGATGGGCCCGCGTTAGATTAGCTAGTTGGT-GAGGTA 188
NR_104732.1 AAGGGGCGTTTGCCTCGCTGAAAGATGAGCCTGCGTATGATTAGCTAGTTGGT-GGGGTA 253
NR_44505.1  AAGGAGCTTAGGCTTCGCTTTTAGATGAGCCTGCGCACTATTAGCTAGTTGGT-GGGGTA 247
NR_28820.1  AAGGGGCGAGTGCCCCGCCTTACGATGAGCCTGCGGCTGATTAGCTAGTTGGT-GGGGTA 234
NR_44756.2  AAAGGGCTTTGGCCCTGCCTTGAGATGAGCCTGCGGCTGATTAGGTAGTTGGT-GGGGTA 283
            *              *      ***   *   **    ** ** * *  *** * ***

NR_025665.1 ATGGCCCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 295
NR_025362.1 ATGGCCCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 305
NR_117546.1 ATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 310
NR_116649.1 ATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 310
NR_116519.1 GTGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 272
NR_117299.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 269
NR_114674.1 ATGGCCCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 258
NR_74786.1  ATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 324
NR_25723.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 298
NR_044871.1 ATGGCTCACCAAGCCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 305
NR_44871.1  ATGGCTCACCAAGCCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 305
NR_28924.1  ATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 313
NR_26451.1  ATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 284
NR_44353.1  ATGGCCTACCGTGGCGATGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 279
NR_116165.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGAC 273
```

```
NR_41410.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGAC 290
NR_112601.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGAC 287
NR_29091.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGAC 307
NR_43541.1  GTGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGAC 312
NR_102474.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 307
NR_24965.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 306
NR_41735.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 306
NR_41736.1  GTGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 307
NR_25658.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 284
NR_117283.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 306
NR_109528.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 291
NR_26183.1  ATGGCCTACCAAGGCGACGACGGGTAGCCGACCTGAGAGGGTGACCGGCCACACTGGGAC 297
NR_119127.1 ATGGCCTACCAAGGCGACGACGGGTAGCCGACCTGAGAGGGTGACCGGCCACACTGGGAC 286
NR_121706.1 ATGGCCTACCAAGCCGACGATCCATAGCTGGTTTGAGAGGATGATCAGCCACACTGGGAC 283
NR_113863.1 ATGGCCTACCAAGCCGACGATCCATAGCTGGTTTGAGAGGATGATCAGCCACACTGGGAC 263
NR_115554.1 ATGGCCTACCAAGCCGACGATCCATAGCTGGTTTGAGAGGATGATCAGCCACACTGGGAC 263
NR_43777.1  ATGGCTCACCAAGGCGACGATCTATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGAC 248
NR_104732.1 ATTGCCTACCAAGGCGACGATCATTAGCCGGCCTGAGAGGGTGACCGGCCACATTGGGAC 313
NR_44505.1  AGAGCCTACCAAGGCAACGATAGTTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 307
NR_28820.1  AAGGCCTACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGATCGGCCACACTGGGAC 294
NR_44756.2  AAGGCCTACCAAGCCGACGATCAGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGAC 343
               **  ***  * * * **    **** *   *******  ** * ****** ******

NR_025665.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 355
NR_025362.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 365
NR_117546.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCA 370
NR_116649.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCA 370
NR_116519.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 332
NR_117299.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 329
NR_114674.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 318
NR_74786.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 384
NR_25723.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCA 358
NR_044871.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 365
NR_44871.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 365
NR_28924.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 373
NR_26451.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 344
NR_44353.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 339
NR_116165.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCA 333
NR_41410.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCA 350
NR_29091.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 367
NR_43541.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 372
NR_102474.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 367
NR_112601.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGA 347
NR_24965.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 366
NR_41735.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 366
NR_41736.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 367
NR_25658.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 344
NR_117283.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 366
NR_109528.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 351
NR_26183.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 357
NR_119127.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAA 346
NR_121706.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTAGACAATGGGGGCA 343
NR_113863.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTAGACAATGGGGGCA 323
NR_115554.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTAGACAATGGGGGCA 323
NR_43777.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGAGGAATATTGGACAATGGGGGCA 308
NR_104732.1 TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGACGCA 373
NR_44505.1  TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGTCGCA 367
NR_28820.1  TGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGACGAA 354
NR_44756.2  TGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGACGAA 403
            ***** ****************************    **** **   ******  * *
```

```
NR_025665.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 414
NR_025362.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 424
NR_117546.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 429
NR_116649.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 429
NR_116519.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 391
NR_117299.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 388
NR_114674.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 377
NR_74786.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 443
NR_25723.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 417
NR_044871.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 424
NR_44871.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 424
NR_28924.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 432
NR_26451.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 403
NR_44353.1  AGCCTGATGCAGCGACGCCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTCA- 398
NR_116165.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 392
NR_41410.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 409
NR_112601.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 406
NR_29091.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 426
NR_43541.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 431
NR_102474.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 426
NR_24965.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 425
NR_41735.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 425
NR_41736.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 426
NR_25658.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 403
NR_117283.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 425
NR_109528.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 410
NR_26183.1  AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 416
NR_119127.1 AGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCA- 405
NR_121706.1 ACCCTGATCTAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTCA- 402
NR_113863.1 ACCCTGATCTAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTCA- 382
NR_115554.1 ACCCTGATCTAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTCA- 382
NR_43777.1  ACCCTGATCCAGCCATGCCGCGTGAGTGATGACGGCCTTAGGGTTGTAAAGCTCTTTCG- 367
NR_104732.1 AGTCTGACGGAGCGACGCCGCGTGTGCGAAGAATGTCGAGAGATTGTAAAGCACTTTTAT 433
NR_44505.1  AGACTGACGGAGCGACGCCGCGTGGACGAAGAATGTCGAGAGATTGTAAAGTCCTTTTAT 427
NR_28820.1  AGTCTGACGGAGCGACGCCGCGTGGGTGAAGAAGGCCGTGAGGTTGTAAAGCCCTTTTCT 414
NR_44756.2  AGTCTGACGGAGCGACGCCGCGTGGGTGAAGAAGGCCGTGAGGTTGTAAAGCCCTTTCTT 463
            *  ****   *** * ********   ** **  * *    * *******   ****

NR_025665.1 --GCAGGGAA------------GAAGCGA-----AAGTGACGGTACCT-GCA-GAAGAAG 453
NR_025362.1 --GCAGGGAA------------GAAGCGA-----AAGTGACGGTACCT-GCA-GAAGAAG 463
NR_117546.1 --GTAGGGAA------------GAAGCGA-----AAGTGACGGTACCT-GCA-GAAGAAG 468
NR_116649.1 --GTAGGGAA------------GAAGCGA-----AAGTGACGGTACCT-GCA-GAAGAAG 468
NR_116519.1 --GTAGGGAA------------GAAGCGA-----AAGTGACGGTACCT-GCA-GAAGAAG 430
NR_117299.1 --GCACGGAA------------GAAGCGA-----GAGTGACGGTACGT-GCA-GAAGAAG 427
NR_114674.1 --GCACGGAA------------GAAGCGA-----GAGTGACGGTACGT-GCA-GAAGAAG 416
NR_74786.1  --GCACGGAA------------GAAGCGA-----GAGTGACGGTACGT-GCA-GAAGAAG 482
NR_25723.1  --GCACGGAA------------GAAGCGA-----AAGTGACGGTACGT-GCA-GAAGAAG 456
NR_044871.1 --GTAGGGAA------------GAAGCGA-----GAGTGACGGTACCT-GCA-GAAGAAG 463
NR_44871.1  --GTAGGGAA------------GAAGCGA-----GAGTGACGGTACCT-GCA-GAAGAAG 463
NR_28924.1  --GTAGGGAA------------GAAGCGA-----GAGTGACGGTACCT-GCA-GAAGAAG 471
NR_26451.1  --GCAGGGAA------------GAAGCCAC----AAGTGACGGTACCT-GCA-GAAGAAG 443
NR_44353.1  --GCAGGGAA------------GAAGCGC-----AAGTGACGGTACCT-GCA-GAAGAAG 437
NR_116165.1 --GCAGGGAA------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 431
NR_41410.1  --GCAGGGAA------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 448
NR_112601.1 --GCAGGGAA------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 445
NR_29091.1  --GCAGGGAA------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 465
NR_43541.1  ---CAGG-AA------------GAAGCGA-----GAGTGAAGGTAC-T-GCA-GAAGAAG 467
NR_102474.1 --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 465
NR_24965.1  --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 464
NR_41735.1  --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 464
NR_41736.1  --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 465
NR_25658.1  --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 442
```

```
NR_117283.1 --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 464
NR_109528.1 --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 449
NR_26183.1  --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 455
NR_119127.1 --GCAGGGAC------------GAAGCGT-----GAGTGACGGTACCT-GCA-GAAGAAG 444
NR_121706.1 --GCTGGGAA------------GAT----------AATGACGGTACC-AGCA-GAAGAAG 436
NR_113863.1 --GCTGGGAA------------GAT----------AATGACGGTACC-AGCA-GAAGAAG 416
NR_115554.1 --GCTGGGAA------------GAT----------AATGACGGTACC-AGCA-GAAGAAG 416
NR_43777.1  --CTAGGGAA------------GAT----------AATGACGGTACCTAGTA--AAGAAG 401
NR_104732.1 ATGTGAGGAATAACAGTATCAGGAAATGGGTATTGGATGACGTTAGC--ATATGAATAAG 491
NR_44505.1  ACATGAGGAATAACCCTTAGAGGAAATGCTAGGGGGATGACGTTAAT--GTATGAATAAG 485
NR_28820.1  GCTTGAGGAATAAGCCTGGGAGGAAATGCCTGGGTGATGACGGTAGG--GCAGGAATAAG 472
NR_44756.2  ACCTGAGGAATAAGTATCGGAGGGAATGCCGGTATGATGACGGTAGG--GTAGGAATAAG 521
                  * *             *              *** * **      *  ** ***

NR_025665.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGA 512
NR_025362.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGA 522
NR_117546.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCGAGCGTTATCCGGA 527
NR_116649.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCGAGCGTTATCCGGA 527
NR_116519.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGA 489
NR_117299.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 486
NR_114674.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 475
NR_74786.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 541
NR_25723.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 515
NR_044871.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 522
NR_44871.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 522
NR_28924.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 530
NR_26451.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 502
NR_44353.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCGAGCGTTATCCGGA 496
NR_116165.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 490
NR_41410.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 507
NR_112601.1 CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 504
NR_29091.1  CGCCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGA 524
NR_43541.1  CCCCGG-TAA-TACGTGCCAGAAAGCCGCGGTAATACGTAGGGCGCGAGCGTTGTCCGGA 525
NR_102474.1 CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 524
NR_24965.1  CACCGGCTAATTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 523
NR_41735.1  CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 523
NR_41736.1  CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 524
NR_117283.1 CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 523
NR_109528.1 CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 508
NR_26183.1  CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCAAGCGTTGTCCGGA 514
NR_119127.1 CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCAAGCGTTGTCCGGA 503
NR_121706.1 CCCCGGCTAACTCCGTGCCAGCA-GCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGA 495
NR_25658.1  CACCGGCTAACTACGTGCCAGCA-GCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA 501
NR_113863.1 CCCCGGCTAACTCCGTGCCAGCA-GCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGA 475
NR_115554.1 CCCCGGCTAACTCCGTGCCAGCA-GCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGA 475
NR_43777.1  CCCCGGCTAACTTCGTGCCAGCA-GCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGA 460
NR_104732.1 CTCCGGCCAATTACGTGCCAGCA-GCCGCGGTAACACGTAAGGAGCGAGCGTTGTTCGGA 550
NR_44505.1  CCCCGGCTAATTACGTGCCAGCA-GCCGCGGTAATACGTAAGGGGCGAGCGTTGTTCGGA 544
NR_28820.1  CCCCGGCTAATTACGTGCCAGCA-GCCGCGGTAACACGTATGGGGCGAGCGTTGTTCGGA 531
NR_44756.2  CCCCGGCTAATTACGTGCCAGCA-GCCGCGGTAACACGTATGGGGCGAGCGTTGTTCGGA 580
              ****  ** * ******** * ********** *** * ** ** ****** * ****

NR_025665.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 571
NR_025362.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 581
NR_117546.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGC 586
NR_116649.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGC 586
NR_116519.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGC 548
NR_117299.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 545
NR_114674.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 534
NR_74786.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 600
NR_25723.1  ATTATTGGGCGTAAAGAGCT-CGTAAGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 574
NR_044871.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 581
NR_44871.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 581
```

```
NR_28924.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 589
NR_26451.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 561
NR_44353.1  ATTATTGGGCGTAAAGAGCT-TGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGGGGC 555
NR_116165.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGCTTGTCGCGTCGGATGTGAAAGCCCGGGGC 549
NR_41410.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGCTTGTCGCGTCGGATGTGAAAGCCCGGGGC 566
NR_112601.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGCTTGTCGCGTCGGTTGTGAAAGCCCGGGGC 563
NR_29091.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGCTTGTCGCGTCGGTTGTGAAAGCCCGGGGC 583
NR_43541.1  ATTATTGGGCGTAAAGAGCTTCGTAGGCGGCTTGTCACGTCGGATGTGAAAGCCCGGGGC 585
NR_102474.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTTGCGTCGGCCGTGAAAACTTCACGC 583
NR_24965.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTTGCGTCGGCCGTGAAAACTTCACGC 582
NR_41735.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTTGCGTCGGCCGTGAAAACTTCACGC 582
NR_41736.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTTGCGTCGGCCGTGAAAACTTCACGC 583
NR_25658.1  ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTTGCGTCGGCCGTGAAAACTTCACGC 560
NR_117283.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTTGCGTCGGCTGTGAAAACTTCACGC 582
NR_109528.1 ATTATTGGGCGTAAAGAGCT-CGTAGGCGGTTTGTCGCGTCGGCCGTGAAAACTTGGGGC 567
NR_26183.1  ATTACTGGGCGTAAAGAGTT-CGTAGGCGGTTTGTCGCGTCGTTTGTGAAAACTCACAGC 573
NR_119127.1 ATTACTGGGCGTAAAGAGTT-CGTAGGCGGTTTGTCGCGTCGTTTGTGAAAACTCACAGC 562
NR_121706.1 ATTACTGGGCGTAAAGCGCA-CGTAGGCGGACTGGAAAGTTGGAGGTGAAATCCCAGGGC 554
NR_113863.1 ATTACTGGGCGTAAAGCGCA-CGTAGGCGGACTGGAAAGTTGGAGGTGAAATCCCAGGGC 534
NR_115554.1 ATTACTGGGCGTAAAGCGCA-CGTAGGCGGACTGGAAAGTTGGAGGTGAAATCCCAGGGC 534
NR_43777.1  ATCACTGGGCGTAAAGCGCG-CGTAGGCGGACTTTTAAGTCAGGGGTGAAATCCCGGGGC 519
NR_104732.1 ATTATTGGGCGTAAAGGGCA-TGTAGGCGGTTTTGTAAGTCCGATGTGAAAGCGTGCAGC 609
NR_44505.1  ATTATTGGGCGTAAAGGGAG-CGCAGGCTGTTTCATAAGTCTGATGTGAAAGACCGCAGC 603
NR_28820.1  ATTATTGGGCGTAAAGGGCG-CGTAGGCGGATGGACAAGTCTGGTGTGAAAGACCGCAGC 590
NR_44756.2  ATTATTGGGCGTAAAGGGCG-TGTAGGCGGATATGCAAGTCTGATGTAAAAGACCGCAGC 639
            ** * *********** *    * * ** *        **     ** ***       **

NR_025665.1 TCAACCCC-GGGTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 630
NR_025362.1 TCAACCCC-GGGTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 640
NR_117546.1 TTAACCCC-GGATCTGCGGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 645
NR_116649.1 TTAACCCC-GGATCTGCGGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 645
NR_116519.1 TTAACTCC-GGATCTGCGGTGGGTACGGGCAGGCTAGAGTGCAGTAGGGGAGACTGGAAT 607
NR_117299.1 TTAACCCC-GGGTGTGCAGTGGGTACGGGCAGACTTGAGTGCAGTAGGGGAGACTGGAAT 604
NR_114674.1 TTAACCCC-GGGTGTGCAGTGGGTACGGGCAGACTTGAGTGCAGTAGGGGAGACTGGAAT 593
NR_74786.1  TTAACCCC-GGGTGTGCAGTGGGTACGGGCAGACTTGAGTGCAGTAGGGGAGACTGGAAT 659
NR_25723.1  TTAACCCC-GGGTGTGCAGTGGGTACGGGCAGACTTGAGTGCAGTAGGGGAGACTGGAAC 633
NR_044871.1 TTAACCCC-GGGTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 640
NR_44871.1  TTAACCCC-GGGTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 640
NR_28924.1  TCAACCCAAGGGTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 649
NR_26451.1  TTAACCCC-GGGTGTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 620
NR_44353.1  TTAACCCC-GGGTGTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAAT 614
NR_116165.1 TTAACTCC-GGGTCTGCATTCGATACGGGCAGGCTAGAGTTCGGTAGGGGAGATCGGAAT 608
NR_41410.1  TTAACTCC-GGGTCTGCATTCGATACGGGCAGGCTAGAGTTCGGTAGGGGAGATCGGAAT 625
NR_112601.1 TTAACCCC-GGGTCTGCAGTCGATACGGGCAGGCTAGAGTTCGGTAGGGGAGATCGGAAT 622
NR_29091.1  TTAACCCC-GGGTCTGCAGTCGATACGGGCAGGCTAGAGTTCGGTAGGGGAGATCGGAAT 642
NR_43541.1  TTAACCCC-GGGTCTGCCATCGATACGGGCAGGCTAGAGTTCGGTAGGGGAGATCGGAAT 644
NR_102474.1 TTAACGTG-GAGCCTGGGTCCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 642
NR_24965.1  TTAACGTG-GAGCCTGCGGTCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 641
NR_41735.1  TTAACGTG-GAGCCTGCGGTCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 641
NR_41736.1  TTAACGTG-GAGCCTGCGGTCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 642
NR_25658.1  TTAACGTG-GAGCCTGCGGTCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 619
NR_117283.1 TTAACGTG-GAGCCTGCAGTCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 641
NR_109528.1 TTAACCCC-GAGCCTGCGGTCGATACGGGCAGACTTGAGTTCGGCAGGGGAGACTGGAAT 626
NR_26183.1  TCAACTGT-GAGCTTGCAGGCGATACGGGCAGACTTGAGTACTGCAGGGGAGACTGGAAT 632
NR_119127.1 TCAACTGT-GAGCTTGCAGGCGATACGGGCAGACTTGAGTACTGCAGGGGAGACTGGAAT 621
NR_121706.1 TCAACCTT-GGAACTGCCTTCAAAACTATCAGTCTGGAGTTCGAGAGAGGTGAGTGGAAT 613
NR_113863.1 TCAACCTT-GGAACTGCCTTCAAAACTATCAGTCTGGAGTTCGAGAGAGGTGAGTGGAAT 593
NR_115554.1 TCAACCTT-GGAACTGCCTTCAAAACTATCAGTCTGGAGTTCGAGAGAGGTGAGTGGAAT 593
NR_43777.1  TCAACCTC-GGAACTGCCTTTGATACTGGAAGTCTTGAGTTCGAGAGAGGTGAGTGGAAT 578
NR_104732.1 TTAACTGC-ATTGTTGCGTTGGAAACTGCGAGACTAGAATCCGAGAGGGGTAAGTGGAAT 668
NR_44505.1  TTAACTGC-GGAAACGCGTTGGAAACTGTGTAACTTGAATCACGGAGGGGGAGATGGAAT 662
NR_28820.1  TTAACTGC-GGGAGCGCGCTGGAAACTGTACATCTTGAATCCTTGAGGGGAAACTGGAAT 649
```

```
NR_44756.2  TCAACTGC-GGGCATGCGTTGGAAACTGTATGTCTTGAATCCTTGAGGGGGAACTGGAAT 698
            * ***          **       **       ** ** *     ** **     ****
NR_025665.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 690
NR_025362.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 700
NR_117546.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 705
NR_116649.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 705
NR_116519.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 667
NR_117299.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 664
NR_114674.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAAGAACACCGATGGCGAAGGCAGGTC 653
NR_74786.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 719
NR_25723.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAAGAACACCGATGGCGAAAGCAGGTC 693
NR_044871.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 700
NR_44871.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 700
NR_28924.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 709
NR_26451.1  TCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 680
NR_44353.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC 674
NR_116165.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATC 668
NR_41410.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATC 685
NR_112601.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATC 682
NR_29091.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATC 702
NR_43541.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATC 704
NR_102474.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 702
NR_24965.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 701
NR_41735.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 701
NR_41736.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 702
NR_25658.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 679
NR_117283.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 701
NR_109528.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 686
NR_26183.1  TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 692
NR_119127.1 TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTC 681
NR_121706.1 TCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTC 673
NR_113863.1 TCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTC 653
NR_115554.1 TCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTC 653
NR_43777.1  ACCGAGTGTAGAGGTGAAATTCGTAGATATTCGGTGGAACACCAGTGGCGAAGGCGGCTC 638
NR_104732.1 TCCTAGTGTAGGGGTGAAATCTGTTGATATTAGGAAGAACACCAGAGGCGAAGGCGACTT 728
NR_44505.1  TCCTAGTGTAGGGGTGAAATCTGTTGATATTAGGAAGAACACCGGAGGCGAAGGCGATCT 722
NR_28820.1  TCCTGGTGTAGGGGTGAAATCTGTAGATATCAGGAAGAACACCGGTGGCGAAGGCGAGTT 709
NR_44756.2  TCCTGGTGTAGGGGTGAAATCTGTAGATATCAGGAAGAACACCGGTGGCGAAGGCGAGTT 758
             **  ****** **** ***  *  *****  **  *******   ****** **

NR_025665.1 TCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 749
NR_025362.1 TCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 759
NR_117546.1 TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 764
NR_116649.1 TCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 764
NR_116519.1 TCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCATGGGT-AGCGAACAGGATTAGATACC 726
NR_117299.1 TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 723
NR_114674.1 TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 712
NR_74786.1  TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 778
NR_25723.1  TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 752
NR_044871.1 TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 759
NR_44871.1  TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 759
NR_28924.1  TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 768
NR_26451.1  TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGG-AGCGAACAGGATTAGATACC 739
NR_44353.1  TCTGGGCTGTTACTGACGCTGAGAAGCGAAAGCATGGGGGAGCGAACAGGATTAGATACC 734
NR_116165.1 TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 727
NR_41410.1  TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 744
NR_112601.1 TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 741
NR_29091.1  TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 761
NR_43541.1  TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 763
NR_102474.1 TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 761
NR_24965.1  TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 760
```

```
NR_41735.1  TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 760
NR_41736.1  TCTGGGCCGATACTGACGCTGAAGAACGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 761
NR_25658.1  TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 738
NR_117283.1 TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 760
NR_109528.1 TCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 745
NR_26183.1  TCTGGGCAGTAACTGACGCTGAGGAACGAAAGCGTGGGT-AGCAAACAGGATTAGATACC 751
NR_119127.1 TCTGGGCAGTAACTGACGCTGAGGAACGAAAGCGTGGGT-AGCAAACAGGATTAGATACC 740
NR_121706.1 ACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGG-AGCAAACAGGATTAGATACC 732
NR_113863.1 ACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGG-AGCAAACAGGATTAGATACC 712
NR_115554.1 ACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGG-AGCAAACAGGATTAGATACC 712
NR_43777.1  ACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGG-AGCAAACAGGATTAGATACC 697
NR_104732.1 GCTGGCTCTGGATTGACGCTGAGGTGCGAAAGCGTAGGT-AGCGAACAGGATTAGATACC 787
NR_44505.1  CCTGGCCGATGATTGACGCTGAGGCTCGAAAGCGTAGGG-AGCGAACAGGATTAGATACC 781
NR_28820.1  TCTGGCAAGTGATTGACGCTGAGGCGCGAAAGCGTGGGG-AGCGAACAGGATTAGATACC 768
NR_44756.2  CCTGGCAAGTGATTGACGCTGAGGCGCGAAAGCGTGGGT-AGCAAACAGGATTAGATACC 817
             ****      * *********    ******* * **  *** ****************

NR_025665.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 809
NR_025362.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 819
NR_117546.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 824
NR_116649.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 824
NR_116519.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 786
NR_117299.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 783
NR_114674.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 772
NR_74786.1  CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGAACATTCCACGTTTTCCGC 838
NR_25723.1  CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 812
NR_044871.1 CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 819
NR_44871.1  CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 819
NR_28924.1  CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 828
NR_26451.1  CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 799
NR_44353.1  CTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGC 794
NR_116165.1 CTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTTGTCCGT 787
NR_41410.1  CTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTTGTCCGT 804
NR_112601.1 CTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTCGTCCGT 801
NR_29091.1  CTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTCGTCCGT 821
NR_43541.1  CTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTCGTCCGT 823
NR_102474.1 CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGGC-TTCCACGTCCTCCGT 820
NR_24965.1  CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGGC-TTCCACGTCCTCCGT 819
NR_41735.1  CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGGC-TTCCACGTCCTCCGT 819
NR_41736.1  CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGGC-TTCCACGTCCTCCGT 820
NR_25658.1  CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGGC-TTCCACGTCCTCCGT 797
NR_117283.1 CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGAC-TTCCACGTCCTCCGT 819
NR_109528.1 CTGGTAGTCCACGCCGTAAACGGTGGGTGCTAGGTGTGGGGGAC-TTCCACGTTCTCCGT 804
NR_26183.1  CTGGTAGTCCACGCCGTAAACGGTGGGCGCTAGGTGTGGGTTCC-TTCCACGGGATCCGT 810
NR_119127.1 CTGGTAGTCCACGCCGTAAACGGTGGGCGCTAGGTGTGGGTTCC-TTCCACGGGATCCGT 799
NR_121706.1 CTGGTAGTCCACGCCGTAAACGATGAATGCCAGTCGTCGGGTAGCATGC---TATTCGGT 789
NR_113863.1 CTGGTAGTCCACGCCGTAAACGATGAATGCCAGTCGTCGGGTAGCATGC---TATTCGGT 769
NR_115554.1 CTGGTAGTCCACGCCGTAAACGATGAATGCCAGTCGTCGGGTAGCATGC---TATTCGGT 769
NR_43777.1  CTGGTAGTCCACGCCGTAAACGATGAATGCTAGTTGTTGGTGGGTTTAC---TCATCAGT 754
NR_104732.1 CTGGTAGTCTACGCCGTAAACGATGTACACTTGGTGTCTGGA---TCATGAGATCTAGGT 844
NR_44505.1  CTGGTAGTCTACGCCGTAAACGATGTGCACTTGGTGTTGGGGAG-TTAT----TCTCAGT 836
NR_28820.1  CTGGTAGTCCACGCTGTAAACGATGTGCACTAGGTGTTGGGG---TGTTGAAACCTCAGT 825
NR_44756.2  CTGGTAGTCCACGCTGTAAACGATGTGCACTAGGCGTTGGGG---TGTTGAAACCTCAGT 874
            ********* * ** ******* **    *  *  **  *               *  *

NR_025665.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 868
NR_025362.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 878
NR_117546.1 GCCGCAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 883
NR_116649.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 883
NR_116519.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 845
NR_117299.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 842
NR_114674.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 831
NR_74786.1  GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 897
```

```
NR_25723.1  GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 871
NR_044871.1 GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 878
NR_44871.1  GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 878
NR_28924.1  GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 887
NR_26451.1  GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 858
NR_44353.1  GCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGGCCGCAAGGCTAAAACTCA 854
NR_116165.1 GCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 846
NR_41410.1  GCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 863
NR_112601.1 GCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 860
NR_29091.1  GCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 880
NR_43541.1  GCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 882
NR_102474.1 GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 879
NR_24965.1  GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 878
NR_41735.1  GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 878
NR_41736.1  GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 879
NR_25658.1  GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 856
NR_117283.1 GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 878
NR_109528.1 GCCGCAGCTAACGCATTAAGCACCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 863
NR_26183.1  GCCGTAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 869
NR_119127.1 GCCGTAGCTAACGCATTAAGCGCCCNGCCTGGGGAGTACGG-CCGCAAGGCTAAAACTCA 858
NR_121706.1 GACACACCTAACGGATTAAGCATTCCGCCTGGGGAGTACGG-TCGCAAGATTAAAACTCA 848
NR_113863.1 GACACACCTAACGGATTAAGCATTCCGCCTGGGGAGTACGG-TCGCAAGATTAAAACTCA 828
NR_115554.1 GACACACCTAACGGATTAAGCATTCCGCCTGGGGAGTACGG-TCGCAAGATTAAAACTCA 828
NR_43777.1  GACGCAGCTAACGCATTAAGCATTCCGCCTGGGGAGTACGG-TCGCAAGATTAAAACTCA 813
NR_104732.1 GCCGTAGCTAACGCGTTAAGTGTACCGCCTGGGGAGTATGC-CGGCAACGGTGAAACTCA 903
NR_44505.1  GCCGAAGTAAACGCGATAAGTGCACCGCCTGGGGAGTATGC-CGGCAACGGTGAAACTCA 895
NR_28820.1  ACCGGAGCTAACGCGGTAAGTGCACCGCCTGGGGAGTATGC-CGGCAACGGTGAAACTCA 884
NR_44756.2  GCCGGAGCTAACGTGGTAAGTGCACCGCCTGGGGAGTATGC-CGGCAACGGTGAAACTCA 933
              *  *   ****   ****    * ************ *    ****   * *******

NR_025665.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 928
NR_025362.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 938
NR_117546.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 943
NR_116649.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 943
NR_116519.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 905
NR_117299.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 902
NR_114674.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 891
NR_74786.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 957
NR_25723.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 931
NR_044871.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 938
NR_44871.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 938
NR_28924.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 947
NR_26451.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 918
NR_44353.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCG 914
NR_116165.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGCTTAATTCGACGCAACGCG 906
NR_41410.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGCTTAATTCGACGCAACGCG 923
NR_112601.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGCTTAATTCGACGCAACGCG 920
NR_29091.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGCTTAATTCGACGCAACGCG 940
NR_43541.1  AAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGCTTAATTCGACGCAACGCG 942
NR_102474.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 939
NR_24965.1  AAGGAATTGACGGGGGCCCGCACAAGCCGCGGAGCATGTGGATTAATTCGATGCAACGCG 938
NR_41735.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 938
NR_41736.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 939
NR_25658.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 916
NR_117283.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 938
NR_109528.1 AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 923
NR_26183.1  AAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 929
NR_119127.1 AAGGAATTGACGGGGGCCNGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCG 918
NR_121706.1 AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCG 908
NR_113863.1 AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCG 888
NR_115554.1 AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCG 888
```

```
NR_43777.1  AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCG 873
NR_104732.1 AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGATACGCG 963
NR_44505.1  AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGATACGCG 955
NR_28820.1  AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGATACGCG 944
NR_44756.2  AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGATACGCG 993
            ****************** ******** * ******** ** ********* *  *****

NR_025665.1 AAGAACCTTACCAAGGCTTGACATG-AACCGGTAAGACCTGGAAACAG-GTCCCCC-ACT 985
NR_025362.1 AAGAACCTTACCAAGGCTTGACATG-AACCGGAAAGGCCTGGAAACAG-GTCCCCC-ACT 995
NR_117546.1 AAGAACCTTACCAAGGCTTGACATG-TTCCCGACCGGGCCAGAGATGG-TCTTTCC-CCT 1000
NR_116649.1 AAGAACCTTACCAAGGCTTGACATG-TTCCTGACCGCCGTAGAGATAC-GGTTTCC-CCT 1000
NR_116519.1 AAGAACCTTACCAAGGCTTGACATG-GACTAGATCGCCGCAGAGATGT-GGTTTCC-C-T 961
NR_117299.1 AAGAACCTTACCAAGGCTTGACATA-CACCGGACCGGTTCAGAGATGT-TCCTTCCCCCT 960
NR_114674.1 AAGAACCTTACCAAGGCTTGACATA-CACCGGACCGGTCCAGAGATGG-ATCTTCCCCCT 949
NR_74786.1  AAGAACCTTACCAAGGCTTGACATA-CACCGGACCGGGCCAGAGATGG-TCTTTCCCCCT 1015
NR_25723.1  AAGAACCTTACCAAGGCTTGACATA-CACCGGATCGGCTCAGAGATGG-GTTTTCCTCCT 989
NR_044871.1 AAGAACCTTACCAAGGCTTGACATT-CACCGGACCGCCCCAGAGATGG-GGTTTC--CCT 994
NR_44871.1  AAGAACCTTACCAAGGCTTGACATT-CACCGGACCGCCCCAGAGATGG-GGTTTC--CCT 994
NR_28924.1  AAGAACCTTACCAAGGCTTGACATT-CACCGGACCGCCCCAGAGATGG-GGTTTC--CCT 1003
NR_26451.1  AAGAACCTTACCAAGGCTTGACATA-TACCGGATCGTTCCAGAGATGG-TTCTTCC-CCT 975
NR_44353.1  AAGAACCTTACCAAGGCTTGACATG-GACCGGACCGCTGCAGAGATGT-AGCTTC--CCT 970
NR_116165.1 AAGAACCTTACCAAGGCTTGACATA-CATCGGAAAACTCTGGAGACAG-GGTCCC--CCT 962
NR_41410.1  AAGAACCTTACCAAGGCTTGACATA-CACCGGAAACATCCAGAGATGG-GTGCCC--CCT 979
NR_112601.1 AAGAACCTTACCAAGGCTTGACATA-CACCGGAAAACCCTGGAGACAG-GGTCCC--CCT 976
NR_29091.1  AAGAACCTTACCAAGGCTTGACATA-CACCGGAAAACCCTGGAGACAG-GGTCCC--CCT 996
NR_43541.1  AAGAACCTTACCAAGGCTTGACATA-CGCCGGAAAACCGTGGAGACAC-GGTCCC--CCT 998
NR_102474.1 AAGAACCTTACCTGGGCTTGACATG-CACCGGAAACCTGCAGAGATGTAGGCCTC--T-T 995
NR_24965.1  AAGAACCTTACCTGGGCTTGACATG-CACCGGAAACCTGCAGAGATGTAGGCCTC--T-T 994
NR_41735.1  AAGAACCTTACCTGGGCTTGACATG-CACCGGAAACCTCCAGAGATGGGGGCCTC--T-T 994
NR_41736.1  AAGAACCTTACCTGGGCTTGACATG-CACCGGAAACCTGCAGAGATGTAGGCCTC--T-T 995
NR_25658.1  AAGAACCTTACCTGGGCTTGACATG-CACCGGAAACCCACAGAGATGTGGGCCTC--T-T 972
NR_117283.1 AAGAACCTTACCTGGGCTTGACATG-CACTGGAAACCAGTAGAGATATTGGCCCC--CCT 995
NR_109528.1 AAGAACCTTACCTGGGCTTGACATG-CACTGGAAACCGGTAGAGATATCGGCCCC--C-T 979
NR_26183.1  AAGAACCTTACCTGGGTTTGACATA-TACCGGAAAGCTGCAGAGATGT-GGCCCC--CCT 985
NR_119127.1 AAGAACCTTACCTGGGTTTGACATA-TACCGGAAAGCTGCAGAGATGT-GGCCCC--CCT 974
NR_121706.1 CAGAACCTTACCAACCCTTGACATCGTAG--GACAACTCCAGAGATGGAGCTTTC--TCG 964
NR_113863.1 CAGAACCTTACCAACCCTTGACATCGTAG--GACAACTCCAGAGATGGAGCTTTC--TCG 944
NR_115554.1 CAGAACCTTACCAACCCTTGACATCGTAG--GACAACTCCAGAGATGGAGCTTTC--TCG 944
NR_43777.1  CAGAACCTTACCAGCCCTTGACATACCGGTCGCGGATATTGGAGACGATATCCTT--CAG 931
NR_104732.1 AGAAACCTTACCAGGGCTTGACATA-TACAGGACGGCAYTAGAGATAGTGTTTTCT-CT- 1020
NR_44505.1  AGGAACCTTACCAGGGCTTGACATA-TACAGGACGGCAGTAGAGATATTGCTTCC--CT- 1011
NR_28820.1  AGGAACCTTACCAGGGTTTGACATA-TACCGGGAGCTATGAGAGATCATGGTGCC--CT- 1000
NR_44756.2  AGGAACCTTACCAGGGTTTGACATA-TACTGGAAGTCGTGGGAGACCACGGTGCC--CT- 1049
               *********     *******       *          ** *

NR_025665.1 T--GT-GGCCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1041
NR_025362.1 T--GT-GGCCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1051
NR_117546.1 T-TGG-GGCGGGTTCACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1057
NR_116649.1 T-TGG-GGCAGGTTCACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1057
NR_116519.1 T-CGG-GGCTGGTTCACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1018
NR_117299.1 TGTGG-GGCTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1018
NR_114674.1 TGTGG-GGCTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTAAGATGTTG 1007
NR_74786.1  TGTGG-GGCTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1073
NR_25723.1  TGTGG-GGCTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1047
NR_044871.1 T-CGG-GGCTGGTGGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1051
NR_44871.1  T-CGG-GGCTGGTGGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1051
NR_28924.1  T-CGG-GGCTGGTGGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1060
NR_26451.1  T-TGG-GGTCGGTATACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1032
NR_44353.1  T-CGG-GGCTGGTTCACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG 1027
NR_116165.1 T--GT-GGTCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1018
NR_41410.1  T--GT-GGTCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1035
```

```
NR_112601.1 T--GT-GGTCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1032
NR_29091.1  T--GT-GGTCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1052
NR_43541.1  T--GT-GGTCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1054
NR_102474.1 C--G--GACTGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1050
NR_24965.1  C--GC-GACTGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1050
NR_41735.1  C--G--GACTGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1049
NR_41736.1  C--G--GACTGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1050
NR_25658.1  C--G--GACTGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1027
NR_117283.1 T--GT-GGCCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1051
NR_109528.1 T--GT-GGCCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 035
NR_26183.1  T--GT-GGTCGGTATACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1041
NR_119127.1 T--GT-GGTCGGTATACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 1030
NR_121706.1 TAAGA-GACCTATAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTC 1022
NR_113863.1 TAAGA-GACCTATAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTC 1002
NR_115554.1 TAAGA-GACCTATAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTC 1002
NR_43777.1  TTTGG-GGACCGGATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG 991
NR_104732.1 TCGGA-GACTTGTATACAGGTGCTGCATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTTG 1078
NR_44505.1  TCGGG-G-CTTGTATACAGGTGCTGCATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTTG 1068
NR_28820.1  TCGGG-G-CCGGTATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGCCGTGAGGTGTTG 1057
NR_44756.2  TCGGG-G-CCGGTATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGCCGTGAGGTGTTG 1106
               *   *       ******* ******* *************** *** ** ****

NR_025665.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTGTATGTTGCCAGCGGGTT-------ATG 1094
NR_025362.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCGCGTT-------ATG 1104
NR_117546.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCGGGTT-------ATG 1110
NR_116649.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACGTA-------GTG 1110
NR_116519.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTT-------ATG 1071
NR_117299.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTG-------ATG 1071
NR_114674.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTNTATGTTGCCAGCACGTG-------ATG 1060
NR_74786.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTG-------ATG 1126
NR_25723.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTG-------ATG 1100
NR_044871.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTG-------ATG 1104
NR_44871.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTG-------ATG 1104
NR_28924.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCACGTG-------ATG 1113
NR_26451.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACGTG-------ATG 1085
NR_44353.1  GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTATGTTGCCAGCACGTA-------AAG 1080
NR_116165.1 GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTGTGTTGCCAGCGGGTT-------ATG 1071
NR_41410.1  GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTGTGTTGCCAGCGGGTT-------ATG 1088
NR_112601.1 GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCCGTGTTGCCAGCAGGCCCTT-GTGGTG 1091
NR_29091.1  GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCCGTGTTGCCAGCAGGCCCTT-GTGGTG 1111
NR_43541.1  GGTTAAGTCCCGCAACGAGCGCAACCCTTGTTCTGTGTTGCCAGCATGCCTTTCGGGGTG 1114
NR_102474.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCGCGTT-------ATG 1103
NR_24965.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCGCGTT-------ATG 1103
NR_41735.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCGCGTG-------ATG 1102
NR_41736.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCGCGTT-------ATG 1103
NR_25658.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCAAGTA-------AAG 1080
NR_117283.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCGCGTT-------ATG 1104
NR_109528.1 GGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACGTT-------ATG 1088
NR_26183.1  GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCTTATGTTGCCAGCACGTA-------ATG 1094
NR_119127.1 GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCTTATGTTGCCAGCACGTA-------ATG 1083
NR_121706.1 GGTTAAGTCCGGCAACGAGCGCAACCCACGTCCCTAGTTGCCAGCATTCA---------G 1073
NR_113863.1 GGTTAAGTCCGGCAACGAGCGCAACCCACGTCCCTAGTTGCCAGCATTCA---------G 1053
NR_115554.1 GGTTAAGTCCGGCAACGAGCGCAACCCACGTCCCTAGTTGCCAGCATTCA---------G 1053
NR_43777.1  GGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCATCATTTA---------G 1042
NR_104732.1 GGTTAAGTCCCGCAACGAGCGCAACCCCTGTTGTATGTTACCATCA--TT-------AAG 1129
NR_44505.1  GGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTCTGTTGCCAGCA--TT-------AAG 1119
NR_28820.1  GGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTCTGTTGCTAACAGGTA-------AAG 1110
NR_44756.2  GGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTCTGTTGCTAACAG-TT-------CGG 1158
            ********** ****************         *** * * *             *

NR_025665.1 CCGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGAAGACGTCAAAT 1154
```

```
NR_025362.1 GCGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAAT 1164
NR_117546.1 CCGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAAT 1170
NR_116649.1 GTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAAT 1170
NR_116519.1 GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAAT 1131
NR_117299.1 GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1131
NR_114674.1 GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1120
NR_74786.1  GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1186
NR_25723.1  GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1160
NR_044871.1 GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1164
NR_44871.1  GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1164
NR_28924.1  GTGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1173
NR_26451.1  GTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1145
NR_44353.1  GTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAAT 1140
NR_116165.1 CCGGGGACTCACAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1131
NR_41410.1  CCGGGGACTCACAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1148
NR_112601.1 CTGGGGACTCACGGGAGACCGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1151
NR_29091.1  CTGGGGACTCACGGGAGACCGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1171
NR_43541.1  ATGGGGACTCACAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1174
NR_102474.1 GCGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1163
NR_24965.1  GCGGGGACTCATGGGAGACTCCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1163
NR_41735.1  GCGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1162
NR_41736.1  GCGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1163
NR_25658.1  TTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1140
NR_117283.1 GCGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1164
NR_109528.1 GTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGT 1148
NR_26183.1  GTGGGGACTCGTAAGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1154
NR_119127.1 GTGGGGACTCGTAAGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGT 1143
NR_121706.1 TTGGGCACTCTATGGAAACTGCCGATGATAAGTCGGAGGAAGGTGTGGATGACGTCAAGT 1133
NR_113863.1 TTGGGCACTCTATGGAAACTGCCGATGATAAGTCGGAGGAAGGTGTGGATGACGTCAAGT 1113
NR_115554.1 TTGGGCACTCTATGGAAACTGCCGATGATAAGTCGGAGGAAGGTGTGGATGACGTCAAGT 1113
NR_43777.1  TTGGGCACTCTAGGGGGACTGCCGGTGATAAGCCGGAGGAAGGTGGGGATGACGTCAAGT 1102
NR_104732.1 TTGGGGACTCATGCAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAAT 1189
NR_44505.1  TTGGGGACTCAGGCAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAAT 1179
NR_28820.1  CTGAGGACTCAGGCGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGACGACGTCAAGT 1170
NR_44756.2  CTGAGGACTCAGATGAGACTGCCGGTGATAAACCGGAGGAAGGTGGGGACGACGTCAAGT 1218
              * * ****       **  ***  *  **  ************ *** ******** *

NR_025665.1 CATCATGCCCCTTATGTTTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1214
NR_025362.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1224
NR_117546.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAATGGGTTGC 1230
NR_116649.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAATGGGTTGC 1230
NR_116519.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1191
NR_117299.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAATGGGTTGC 1191
NR_114674.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAATGGGTTGC 1180
NR_74786.1  CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCAGTACAATGGGTTGC 1246
NR_25723.1  CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCAGTACAATGGGTTGC 1220
NR_044871.1 CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1224
NR_44871.1  CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1224
NR_28924.1  CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1233
NR_26451.1  CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGC 1205
NR_44353.1  CATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAGTGGGTTGC 1200
NR_116165.1 CATCATGCCCCTTATGTCTTGGGCTGCACACGTGCTACAATGGCCGGTACAATGAGCTGC 1191
NR_41410.1  CATCATGCCCCTTATGTCTTGGGCTGCACACGTGCTACAATGGCCGGTACAATGAGCTGC 1208
NR_112601.1 CATCATGCCCCTTATGTCTTGGGCTGCACACGTGCTACAATGGCCGGTACAATGAGCTGC 1211
NR_29091.1  CATCATGCCCCTTATGTCTTGGGCTGCACACGTGCTACAATGGCCGGTACAATGAGCTGC 1231
NR_43541.1  CATCATGCCCCTTATGTCTTGGGCTGCACACGTGCTACAATGGCCGGTACAATGAGCTGC 1234
NR_102474.1 CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1223
NR_24965.1  CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1223
NR_41735.1  CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1222
NR_41736.1  CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1223
NR_25658.1  CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1200
NR_117283.1 CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1224
```

```
NR_109528.1 CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGC 1208
NR_26183.1  CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1214
NR_119127.1 CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAGAGGGCTGC 1203
NR_121706.1 CCTCATGGCCCTTACGGGTTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGGT--- 1190
NR_113863.1 CCTCATGGCCCTTACGGGTTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGGT--- 1170
NR_115554.1 CCTCATGGCCCTTACGGGTTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGGT--- 1170
NR_43777.1  CCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTAGTGACAGAGGGCAGC 1162
NR_104732.1 CATCATGGCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCCGGTACAGAGTGATGC 1249
NR_44505.1  CATCATGGCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCCAGTACAGAGTGATGC 1239
NR_28820.1  CATCATGGCCCTTATATCCTGGGCTACACACGTGCTACAATGGCCGGTACAAAGTGAGGC 1230
NR_44756.2  CATCATGGCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCCGGTACAAAGCGAAGC 1278
            * ***** ******      ***** *** * ***********     ***  * *

NR_025665.1 GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGAGGTCTGCAA 1274
NR_025362.1 GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGAGGTCTGCAA 1284
NR_117546.1 GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1290
NR_116649.1 GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1290
NR_116519.1 GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1251
NR_117299.1 GATGCCGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATCGTGGTCTGCAA 1251
NR_114674.1 GATACCGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATCGTGGTCTGCAA 1240
NR_74786.1  GATGCCGTGAGGTGGAGCTAATCCCAAAAAGCTGGTCTCAGTTCGGATCGTGGTCTGCAA 1306
NR_25723.1  GATACCGTGAGGTGGAGCTAATCCCAAAAAGCTGGTCTCAGTTCGGATCGTGGTCTGCAA 1280
NR_044871.1 GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGAGGTCTGCAA 1284
NR_44871.1  GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGAGGTCTGCAA 1284
NR_28924.1  GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGAGGTCTGCAA 1293
NR_26451.1  GATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGGTCTCAGTTCGGATTGAGGTCTGCAA 1265
NR_44353.1  GATACTGTGAGGTGGAGCTAATCCCTAAAAGCCGGTCTCAGTTCGGATCGAAGTCTGCAA 1260
NR_116165.1 GAAGCCGTGAGGTGGAGCGAATCTCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1251
NR_41410.1  GAAGCCGTGAGGTGGAGCGAATCTCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1268
NR_112601.1 GATGCCGTGAGGTGGAGCGAATCTCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1271
NR_29091.1  GATGCCGTGAGGTGGAGCGAATCTCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1291
NR_43541.1  GATACCGCGAGGTGGAGCGAATCTCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAA 1294
NR_102474.1 TAAGCCGTGAGGTGGAGCGAATCCCACAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1283
NR_24965.1  TAAGCCGTGAGGTGGAGCGAATCCCACAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1283
NR_41735.1  TAAGCCGTGAGGTGGAGCGAATCCCAGAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1282
NR_41736.1  TAAGCCGTGAGGTGGAGCGAATCCCAGAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1283
NR_25658.1  TAAGCCGTGAGGTGGAGCGAATCCCACAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1260
NR_117283.1 GAAGCCGTGAGGTGGAGCGAATCCCACAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1284
NR_109528.1 TAAACCGTGAGGTGGAGCGAATCCCATAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1268
NR_26183.1  GATACCGTGAGGTGGAGCGAATCCCTTAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1274
NR_119127.1 GATACCGTGAGGTGGAGCGAATCCCTTAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA 1263
NR_121706.1 ------------------TAATCCCCAAAAGCCA-TCTCAGTTCGGATTGGGGTCTGCAA 1231
NR_113863.1 ------------------TAATCCCCAAAAGCCA-TCTCAGTTCGGATTGGGGTCTGCAA 1211
NR_115554.1 ------------------TAATCCCCAAAAGCCA-TCTCAGTTCGGATTGGGGTCTGCAA 1211
NR_43777.1  GACTCCGCGAGGGGAAGCTAATCTCCAAAAACTA-TCTCAGTTCGGATTGTTCTCTGCAA 1221
NR_104732.1 GAAGCCGCGAGGTGAAGCAAAACGCATAAAGCCGGTCTCAGTTCGGATTGGAGTCTGAAA 1309
NR_44505.1  GAAGCCGCGAGGTGGAGCCAAACGCAGAAAGCTGGTCTCAGTTCGGATTGGAGTCTGAAA 1299
NR_28820.1  GAAGCCGCGAGGTGGAGCAAATCGCAAAAAACCGGTCTCAGTTCGGATTGGAGTCTGAAA 1290
NR_44756.2  AAGACCGCGAGGTGGAGCGAATCGCAAAAAACCGGTCTCAGTTCGGATTGGAGTCTGAAA 1338
                              ** * *  *** *   ************ *   **** **

NR_025665.1 CTCGACCTCATGAAGTTGGAGTCGATAGTAATCGCAGATCAGCAATCGATACGGTGAATA 1334
NR_025362.1 CTCGACCTCATGAAGTTGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1343
NR_117546.1 CTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1349
NR_116649.1 CTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1349
NR_116519.1 CTCGACCCCATGAAGTTGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1310
NR_117299.1 CTCGACCACGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1310
NR_114674.1 CTCGACCACGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1299
NR_74786.1  CTCGACCACGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1365
NR_25723.1  CTCGACCACGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1339
NR_044871.1 CTCGACCTCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1343
NR_44871.1  CTCGACCTCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1343
```

```
NR_28924.1  CTCGACCTCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1352
NR_26451.1  CTCGACCTCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1324
NR_44353.1  CTCGACTTCGTGAAGTTGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1319
NR_116165.1 CTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAT-TGCTGCGGTGAATA 1310
NR_41410.1  CTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAT-TGCTGCGGTGAATA 1327
NR_112601.1 CTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAT-TGCTGCGGTGAATA 1330
NR_29091.1  CTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCGGCAT-TGCTGCGGTGAATA 1350
NR_43541.1  CTCGACCCCATGAAGTCGGAGTTGCTAGTAATCGCAGATCAGCAT-TGCTGCGGTGAATA 1353
NR_102474.1 CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1342
NR_24965.1  CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1342
NR_41735.1  CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1341
NR_41736.1  CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1342
NR_25658.1  CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1319
NR_117283.1 CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1343
NR_109528.1 CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1327
NR_26183.1  CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCGGTGAATA 1333
NR_119127.1 CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAA-CGCTGCG------- 1315
NR_121706.1 CTCGACCCCATGAAGTTGGAATCGCTAGTAATCGCGGAACAGCAT--GCCGCGGTGAATA 1289
NR_113863.1 CTCGACCCCATGAAGTTGGAATCGCTAGTAATCGCGGAACAGCAT--GCCGCGGTGAATA 1269
NR_115554.1 CTCGACCCCATGAAGTTGGAATCGCTAGTAATCGCGGAACAGCAT--GCCGCGGTGAATA 1269
NR_43777.1  CTCGAGAGCATGAAGTTGGAATCGCTAGTAATCGCAGATCAGCAT--GCTGCGGTGAATA 1279
NR_104732.1 CTCGACTCCATGAAGGTGGAATCGCTAGTAATCGCGCATCAGCAT--GGCGCGGTGAATA 1367
NR_44505.1  CTCGACTCCATGAAGGTGGAATCGCTAGTAATCGCGCATCAGCAT--GGCGCGGTGAATA 1357
NR_28820.1  CTCGACTCCATGAAGGTGGAATCGCTAGTAATCGCATATCAGCAT--GATGCGGTGAATA 1348
NR_44756.2  CCCGACTCCATGAAGGTGGAATCGCTAGTAATCGCATATCAGCAT--GATGCGGTGAATA 1396
            * ***   * *****  *** * * **********  * * ***   *   **
```

## Appendix 2: Codon Table

| 1 | 2 | | | | | | | | 3 |
|---|---|---|---|---|---|---|---|---|---|
| | T | | C | | A | | G | | |
| T | TTT | Phe | TCT | Ser | TAT | Tyr | TGT | Cys | T |
| | TTC | Phe | TCC | Ser | TAC | Tyr | TGC | Cys | C |
| | TTA | Leu | TCA | Ser | TAA | stop | TGA | stop | A |
| | TTG | Leu | TCG | Ser | TAG | stop | TGG | Trp | G |
| C | CTT | Leu | CCT | Pro | CAT | His | CGT | Arg | T |
| | CTC | Leu | CCC | Pro | CAC | His | CGC | Arg | C |
| | CTA | Leu | CCA | Pro | CAA | Gln | CGA | Arg | A |
| | CTG | Leu | CCG | Pro | CAG | Gln | CGG | Arg | G |
| A | ATT | Ile | ACT | Thr | AAT | Asn | AGT | Ser | T |
| | ATC | Ile | ACC | Thr | AAC | Asn | AGC | Ser | C |
| | ATA | Ile | ACA | Thr | AAA | Asn | AGA | Arg | A |
| | ATG | Met | ACG | Thr | AAG | Lys | AGG | Arg | G |
| G | GTT | Val | GCT | Ala | GAT | Lys | GGT | Gly | T |
| | GTC | Val | GCC | Ala | GAC | Asp | GGC | Gly | C |
| | GTA | Val | GCA | Ala | GAA | Glu | GGA | Gly | A |
| | GTG | Val | GCG | Ala | GAG | Glu | GGG | Gly | G |