ISOLATION OF NOVEL SPECIES FROM THE GENUS *DERMACOCCUS* BY PHENOTYPIC AND PHYLOGENETIC ANALYSES

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Biology

By

Erin Clingerman

Advisor: Dr. Sean O'Connell
Associate Professor of Biology Biology
Department

Committee Members: Dr. Anjana Sharma, Department of Biology
Dr. Katherine Mathews, Department of Biology
Dr. Robert Youker, Department of Biology

April 2019

ACKNOWLEDGEMENTS

Thank you to my committee members Seán, Anjana, Kathy, and Bob for everything you have done to help me reach my goals and for working with me through this season of my life. To my fellow graduate students, thank you for continuously building me up and encouraging me for the last two years. Lastly, thank you to my loved ones for pushing me to always be the best version of myself, I could not have done this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CDC- Centers for Disease Control and

Prevention HKY- Hasegawa Kishino Yano

iTOL- Interactive Tree of Life

MEGA- Molecular Evolutionary Genetics Analysis

ML- Maximum likelihood

NCBI- National Center for Biotechnology Information

NJ- Neighbor-joining

NIH- National Institutes of Health

PCA- Principal components analysis

R2A- Reasoner's 2 agar

R2B- Reasoner's 2 broth

RDP- Ribosomal Database Project

WCU-Western Carolina University

ABSTRACT

ISOLATION OF NOVEL SPECIES FROM THE GENUS *DERMACOCCUS* BY PHENOTYPIC AND PHYLOGENETIC ANALYSES

Erin Michelle Clingerman, M.S.
Western Carolina University (April, 2019)
Director: Seán O'Connell

There are estimated to be 10^{30} total bacterial cells on earth and approximately 10^{11} cells/m^2 of bacteria on human skin. Microorganisms are ubiquitous, and there are countless services that microbes provide, such as breaking down harmful pollutants in the environment, providing nitrogen fixation to plant root systems, allowing for fermentation of foods and beverages, and assisting the human body in innumerable ways. The use of 16S rDNA sequencing has allowed bacterial taxonomists to presumptively identify many species by grouping similar species via a 97.5% threshold of sequence homology. This analysis coupled with metabolic and phenotypic tests is instrumental in the identification of novel species, such as species in the *Dermacoccus* genus. *Dermacoccus nishinomiyaensis* is a common resident of the skin and is the type species within the genus found in the family *Dermacoccaceae.* The genus contains only four species described to date; *D. barathri*, *D. abyssi*, *D. profundi*, and *D. nishinomiyaensis*. Species within this genus can often have very similar 16S rDNA sequences, thus they require more extensive study with bacteriological laboratory testing and additional methods including phylogenetics. Seventeen cultures that closely align with the genus *Dermacoccus* were obtained in Cullowhee, North Carolina at Western Carolina University from a variety of human-associated environments and were tested using a combination of DNA sequencing and phenotypic testing to determine if

novel species were isolated. Based on this work, it appears that at a minimum, three new species have been discovered.

CHAPTER ONE: INTRODUCTION

Novel species are defined by the CDC as organisms that have not yet been viewed or described by the scientific community [1]. With 10^{30} total bacterial cells on Earth, and approximately 10^{11} cells/m^2 of skin [2], new bacterial species are being discovered every day [3]. Human skin contains a diverse range of bacterial communities, with specific metacommunities found over zones of the epidermis. For example, the human forearm is thought to contain 182246 unique bacterial species [4], however, the exact number of species on the human skin is unclear. Many organizations have tried to compose a list of novel species and their relation to the human microbiome, such as the NCBI [5], NIH [6], and the CDC [7]. The human microbiome describes the normal microflora of the human body and can classify what bacteria can be found on the parts of the body based on sex, age, and area sampled to determine what microorganisms form a symbiotic relationship with their human host, and which organisms can be opportunistically pathogenic. *Dermacoccus* species are typically isolated from the surface of human skin, and three species have also been discovered in deep marine trenches [8].

Understanding bacterial species found on human skin and the contributions of these cultures to the human microbiome can help explain the services they provide to allow our bodies to function properly. The number of bacteria within the human microbiome was thought to outnumber human cells 10:1, however, according to Sender et al. the number of human cells is likely equal to the number of bacterial cells present.[11]. Understanding the

taxonomy of these microflora are important to the medical community, especially from an immunological standpoint. Bacteria often assist the human body by stopping other harmful microorganisms from causing quandaries for the immune system [12]. However, helpful microorganisms can switch roles and act as opportunistic pathogens if the host is immunologically compromised [12] *Dermacoccus* is not inherently pathogenic, however, due to a recent infection caused by *Dermacoccus barathri*, understanding and classifying the bacteria is necessary to further study the capabilities of these microorganisms [13].

In addition to protection, bacterial species can also be credited with retaining skin moisture [14]. The link between *Dermacoccus* species and humans is usually a commensal or mutualistic relationship. *Dermacoccus* spp. utilize nutrients found within the skin of the human body from sources such as lipids, urea, and amino acids. Lipids are found in the sebum, apocrine glands, and eccrine sweat as free fatty acids as well as in sphingolipids on the stratum corneum. Urea is in eccrine sweat of human skin. Amino acids are found in eccrine sweat secreted as well as on the stratum corneum [15]. The exact ecology of *Dermacoccus* on and in the skin is not well understood, however, these links to possible sources of nutrition are worth exploring further.

The family *Dermacoccaceae* is defined by Gram-positive, coccus-shaped bacteria that are non-motile, nonsporeforming, and nonencapsulated [8]. *Dermacoccus* spp. are chemoorganotrophic, which means that they obtain energy from the oxidation of reduced organic compounds. *Dermacoccus nishinomiyaensis* is mainly found on the outermost layer of the skin, also known as the stratum corneum but have also been found in water, humanassociated environments, and in salt used for dry-cured ham production [8]. According to Szczerba [9], *Dermacoccus* is most often isolated from the skin of the palms, and the oral cavity in both sexes. There are currently four species within the genus: *D. barathri*, *D. abyssi*, *D.*

profundi, and *D. nishinomiyaensis*. *Dermacoccus nishinomiyaensis* was first described as *Micrococcus nishinomiyaensis* but was changed to *Dermacoccus* based on the characterization of strains isolated from a human sample [8]. These strains were tested using "chemotaxonomic" and physiological properties such as metabolic functions and the combination of these properties has been shown to be useful in determining intra-species diversity, which can lead to discovering novel species of bacteria. Possible novel species stemming from intra-species diversity can arise from horizontal gene transfer or by gene mutations.

Novel bacterial species are classified by four criteria; they must differ from known strains by specific traits and genetic sequence(s), including the 16S rDNA, they must be included in a publication in the International Journal of Systematic and Evolutionary Microbiology (IJSEM), the proposed name must be appropriate with regard to nomenclature guidelines, and the organisms must be the first validly published organism bearing the taxon name [2]. The organisms must also be submitted as cultures to at least two culture collections such as the American Type Culture Collection (ATCC) and the Deutsche Sammlung von Mikroorganismen (DSMZ).

One of the primary ways of distinguishing a novel species from a known species is via 16S rDNA sequencing and analyses. 16S rDNA sequencing is used due to the slow evolutionary change to this section of DNA over time, allowing it to serve as a molecular chronometer to infer taxonomic relationships. The interpretation of 16S rDNA can be simplified by use of a variety of tools including those found on the Ribosomal Database Project [10] to determine how similar a sequence is to over three million sequences in this database, which is populated with data from GenBank. Tools within RDP such as Classifier and

SeqMatch can help elucidate whether a culture is likely a known species or whether further testing could be used to discover a novel species. The threshold to determine whether a species is considered a known species is generally if it has at least a 97.5% sequence homology to another organism. SeqMatch can be used to compare unknown organisms to well-described type cultures to aid in this species-level determination. Classifier can be used to apply confidence limits to sequences down to the level of genus and can often detect organisms from novel genera, families, and even orders.

Alongside sequencing, phenotypic and metabolic traits can assist in determining if a species is new. Examining, for example, such things as colony color, and temperature, salt, and pH growth tolerances and optima, can be effective ways to detect a novel species when 16S rDNA sequences are similar [16] [17]. Other phenotypic tests, including gelatin hydrolysis, urea hydrolysis, and glucose fermentation can also be useful in determining new species [2]. Additional tests that could be done include benzidine testing to detect the iron porphyrin systems of aerobic bacteria, assessment of G+C content, sequencing of housekeeping genes, and whole genome sequencing. The suite of testing required for an individual isolate is determined by which taxonomic group it is most closely related to.

This study examined the phenotypic diversity of 17 *Dermacoccus* strains and used metabolic testing such as presence of catalase and oxidase; salt, pH, and oxygen requirements and tolerances; EnteroPluri testing; and commonly used bacteriological tests such as gelatin hydrolysis, starch hydrolysis, nitrate reduction, antibiotic resistance, and lipid hydrolysis in order to better understand diversity within *Dermacoccus* and to determine if any of the 17 isolates were novel bacterial species. Antibiotic resistance is of particular interest as these species were isolated from human environments. Although *Dermacoccus nishinomiyaensis* is nonpathogenic,

further studies are especially important as one *Dermacoccus* species has shown pathogenic capabilities within immunocompromised patients [13]. Another aim of this study was to fill gaps within the knowledge of *Dermacoccus* spp. to better understand these organisms since the literature does not include many parameters such as a pH range or oxidase results, nor reports the same biochemical and metabolic tests across each of the species. Lastly, phylogenetic analyses were performed on the 16S rDNA of the isolates to better gauge how similar the cultures were to each other and to their closest relatives already included in the literature.

CHAPTER TWO: METHODS AND MATERIALS

2.1: Sample Collection and Maintenance

Sterile cotton swabs were used to aseptically collect samples from multiple environments as part of a laboratory experiment in a microbiology course. The swabs were used to inoculate R2A media plates and subsequent colonies grown were streaked until pure cultures had been obtained. All plates were incubated upside down at room temperature for one week. The cultures were labeled as WCU 1-43 for this experiment and individual cultures were characterized by microscopy, metabolic and physiological testing, and had their 16S rDNA sequenced. Seventeen of the cultures were isolated from various human environments and were identified as *Dermacoccus* spp. as described below (Table 1).

2.2: Partial 16S rDNA Sequencing

The unidentified cultures on R2A plates were sent out for off-site Sanger sequencing of their 16S rDNA at Genewiz, Inc. (South Plainfield, NJ). Sequences were completed from both ends of the 16S rRNA gene for all organisms and were aligned and ambiguous bases resolved or removed prior to further analysis. Analysis of the sequences was completed using RDP Classifier and RDP SeqMatch tools [10] to establish the closest related organisms (Table 1). Based on this initial analysis, 17 of the cultures were identified as *Dermacoccus* spp. Isolates 10, 15, 21, and 26 were sent back to Genewiz Inc. for resequencing due to initial short sequence length (<1,000 nucleotides). Resequencing was completed from the center of the 16S rDNA sequence outward based on new methods used by that lab.

2.3: Physiological Growth Characteristics

All 17 of the *Dermacoccus* cultures were grown in various conditions to determine their optimal growth preferences and tolerances. Salt, pH and temperature ranges permitting growth were tested using a 24-well plate with the broth culture medium R2B used as a base recipe. For salt testing, the WCU cultures were subjected to 0, 1, 2, 3, 4, 5, 7.5, 10, and 15% NaCl. The samples were inoculated from R2A plates (n = 3) by suspending cells from colonies in sterile saline and transferring biomass by pipetting or through the use of an inoculating loop. The pH range allowing growth was tested in a similar manner, at pH 3, 5, 7, 9 and 10. Temperature was tested at 4, 15, 20, 25, 37, 40, 42, 45, 50 and 55°C. All salt, pH, and temperature samples were performed in triplicate and incubated for one week before observing results.

Other tests were performed to compare WCU samples against the type species, *D. nishinomiyaensis*, and the other three previously described species. The cultures were negative stained to determine cell shape and sizes using a nigrosin stain and were also Gram stained. The cultures were tested in triplicate to see whether they could grow in anaerobic conditions by the shake culture technique: A 24-hour culture of each sample was prepared and inoculated into molten agar and quickly cooled using ice; the samples were then incubated for 48 hours to determine if the cultures could grow anaerobically based on their growth location after incubation (e.g., anaerobes would be found deep within the media in a tube and obligate aerobes at the top).

2.4: Metabolic Capabilities

The EnteroPluri metabolic test system was run (n=3) to determine metabolic capabilities of the isolates including the fermentation of glucose, with or without gas production; ability to decarboxylate lysine and ornithine; use of sulfur as a terminal electron acceptor and production of hydrogen sulfide; fermentation of adonitol, lactose, sorbitol, and dulcitol; production of acetoin; the ability to hydrolyze urea; and the ability to utilize citrate [18].

Nutrient gelatin medium was used to check for the activity of gelatinase [19]. Starch hydrolysis was tested by inoculating the culture onto starch media plates, storing for one week, and assessing for starch loss using iodine. Lipid hydrolysis was tested using a lipid agar medium and one week of incubation. DNAse media was tested to determine whether the *Dermacoccus* cultures could break down DNA and cause a clearing on the media when hydrochloric acid was introduced. Nitrate reduction was tested using a pre-prepared broth with an inverted Durham tube in the broth [20]. All tests were performed in triplicate fashion.

Antibiotic resistance was assessed (n=3) using antibiotic disks that were placed onto spread-plated culture on R2A plates. Clindamycin, colistin, chloramphenicol, tetracycline, ampicillin, nitrofurantoin, and nalidixic acid were used at a concentration of 30µg/disc. The plates were stored at room temperature for one week to allow for growth before observation [4]. Resistance was characterized by no zone of inhibition surrounding the antibiotic disc. Susceptibility was characterized by a zone of inhibition surrounding the disc.

Oxidase testing was performed in triplicate using cotton swabs containing N,N,N',N'tetramethyl-p-phenylenediamine dihydrochloride which rapidly detects the presence of the cytochrome c oxidase enzyme. A positive result of this procedure resulted in the swabs

changing color to a dark blue-purple. A catalase test was also completed in triplicate to test the organism's ability to catalyze hydrogen peroxide into water and oxygen. If the organisms were able to catalyze H2O2, O2 bubbles were present upon addition of 3% hydrogen peroxide to colony material

2.5: Principal Components Analysis (PCA)

A principal components analysis (PCA) was run to analyze the physiological and metabolic data, grouping isolates based on similarities across these variables. The dataset was analyzed by Systat 10 software [27].

2.6: Phylogeny

2.6.1: MOLE-BLAST Phylogenetic Tree

A preliminary phylogenetic tree was created using the MOLE-BLAST program located on the National Center for Biotechnology Information (NCBI) website [21] [22]. MOLE- BLAST was used to find similar organisms as well as an outgroup for further analysis. Edited sequences from Genewiz were inputted into the MOLE-BLAST program in a fasta format and 16S ribosomal RNA sequences (Bacteria and Archaea) was chosen for the database search set. A distance tree was created using the neighbor-joining algorithm (NJ). The outgroup was identified as *Austwickia chelonae.*

2.6.2: MEGA Phylogenetic Trees

The MOLE-BLAST tree was also used to download a Newark tree alignment file and the alignment file was inputted into MEGA [23]. The sequences were realigned using the MEGA muscle alignment section and the ends were trimmed to remove excess ambiguous and missing bases. The aligned MOLE-BLAST sequences were used to recreate a neighbor-joining tree. The best-fitting model (HKY+G+I) [24] was chosen using the Bayesian information criterion (BIC) value [25]. Hasegawa Kishino Yano (HKY) is a statistical analysis which assumes that transitions and transversions occur at different rates. Gamma distribution (+G) is used due to nonuniformity of evolutionary rates among sites. Invariability (+I) assumes some sites are evolutionarily invariable. The evolutionary distances were computed using the Maximum Composite Likelihood method [38] and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). One thousand bootstrap replications were performed to evaluate the confidence of tree structures.

2.6.3: Annotated Phylogenetic Tree

An annotated phylogenetic tree was created using iTOL software [26]. A dataset was created by manually inputting characteristic data using arbitrary numeric values to classify the physiological and metabolic test results (Table 2). A Newark tree file from MOLE-BLAST was uploaded into the iTOL software to create the tree, then the dataset was applied to the phylogenetic tree, creating a heatmap. Settings were adjusted to root the tree to *Austwickia chelonae* and *Kineosphaera limosa* and the dataset color values were set from 0-5 with an intermediate at 2.48.

2.7: Pairwise Distance

Pairwise distance calculations for the 16S rDNA sequences were completed using MEGA X software [23]. The 17 *Dermacoccus* isolates were contrasted with *Dermacoccus nishinomiyaensis* strain DSM 20448, *D. abyssi* strain MT1.1, *D. barathri* MT2.1, and *D. profundi* MT2.2. Analysis preferences were chosen as follows: no variation method, maximum composite likelihood model, transitions and transversion substitutions, uniform rates among sites, homogeneous patterns among lineages, and a pairwise deletion gaps/missing data treatment were used

2.9: Sequencher Variance Report

A variance report based on nucleotide substitutions in the 16S rDNA sequences was created using Sequencher version 5.4.1[28]. *Dermacoccus nishinomiyaensis* DSM20448 was used as a reference sequence to compare the 17 isolates to determine the positions and amount of 16S rDNA nucleotide changes. WCU 1 was also utilized as a reference sequence based on its 100% 16S rDNA homology compared to *D. barathri* and *D. profundi*.

CHAPTER THREE: RESULTS

3.1: 16S rDNA Sequencing Analyses

Each of the isolates obtained from the microbiology course were given a number from 143 based on the order that they were obtained. All samples were obtained in a broader study, which included other taxa of bacteria. Bacteria that were not *Dermacoccus* spp. were not used in

this study. The 16S rDNA sequences for 17 samples showed that the isolates were within the domain Bacteria, phylum Actinobacteria, class Actinobacteria, order Actinomycetales, family Dermacoccaceae, and genus *Dermacoccus* with 100% confidence at each level, based on the RDP Classifier tool (Table 1) [10]. When RDP BLAST searches were performed, the highest species match found belonged to WCU 1 with a 100% alignment equally with both *Dermacoccus barathri* and *Dermacoccus profundi*. The lowest percent match was found to be WCU 10 with a

96.2% match to *Dermacoccus nishinomiyaensis* (Table 1). All sequences can be found in Appendix A (as used in this study) and Appendix B (as submitted to GenBank).

3.2: Phenotypic Data

The 17 *Dermacoccus* cultures varied from orange to bright yellow in color. The colonies were all shiny, however, the consistency ranged from translucent (3 of 17 cultures) to opaque (14/17). The colony size varied from 0.5mm to 2.25mm, with a mean size of 1.35mm. Edge characteristics were all entire, and the colony shape was circular. Colonies ranged from flat to convex, and the Gram stain results were positive for all (Table 3). Cell shape was coccus, and the size ranged from 1 μm to 1.5μm in diameter.

Temperature, salt, and pH ranges were also established for each isolate to determine optimal growth conditions and tolerances. Each of the seventeen cultures were found to grow from 20-42^{o}C, with samples WCU 4 and WCU 30 reaching 45^{o}C and WCU 8 showing the most heat tolerance at 50^{o}C. Salt tolerances were 0-10% with most of the cultures (10/17) growing up to 5% NaCl tolerance, 4/17 could tolerate salt concentrations up to 4% NaCl, and 3/17 reaching 10% NaCl. pH tolerances ranged from 5-9 (4/17), 5-10 (1/10), 7-9 (7/17), and 7-10 (5/17) (Table 4).

3.3: Metabolic Data

EnteroPluri results were negative for most tests, but urea hydrolysis was positive for all 17 isolates and citrate production was positive for 11/17 (Table 5). Gelatin hydrolysis results showed that most species could hydrolyze gelatin (12/17). Starch hydrolysis was positive for 11 of 17 cultures while nitrate reduction was negative for all isolates. Cytochrome c oxidase was present for one organism, lipid hydrolysis was positive for 14 of 17 cultures, and DNAse activity was positive for 6 of 17. Catalase was positive for all samples, however, WCU 17 showed a lower rate of bubble formation with smaller bubbles (Table 6).

Antibiotic resistance showed variable results for each antibiotic tested. Fourteen of 17 cultures were shown to be resistant against nalidixic acid and colistin. Clindamycin was resisted by 11 of 17 isolates, while ampicillin was resisted by 8 of 17. Chloramphenicol inhibited 11 of 17 samples, whereas tetracycline inhibited 12 of 17 isolates (Table 7).

3.4: Phylogeny

3.4.1: MOLE-BLAST phylogenetic tree

The MOLE-BLAST software constructed a neighbor-joining tree with 61 closely related organisms. Figure 1 displays 31 organisms on the MOLE-BLAST tree used to generate a tree file for use in MEGA X software. This tree showed the grouping of *Dermacoccus* spp. separating the clade containing *Dermacoccus nishinomiyaensis* from the clade containing *Dermacoccus abyssi*,

Dermacoccus barathri, and *Dermacoccus profundi*. *Austwickia chelonae* was also established as the outgroup based on this tree.

3.4.2: MEGA Phylogenetic Tree

The neighbor-joining (NJ) tree created using data from the MOLE-BLAST tree in MEGA software showed two distinct *Dermacoccus* clades, one containing *D. nishinomiyaensis* and the second clade containing *D. abyssi, D. barathri,* and *D. profundi* (Figure 1). The clade containing

D. nishinomiyaensis had two species with less strong relatedness to the others; WCU 23 and WCU 9. WCU 29 was an outlier of the clade containing *D. abyssi, D. barathri,* and *D. profundi.* Clade 1 contained *D. nishinomiyaensis* and WCU isolates 4, 8, 10, 15, 17, 20, 21, 22, 26, 30, 39, and 43. WCU 23 and WCU 9 were sisters to this clade. Clade 2 contained *D. abyssi, D. barathri,* and *D. profundi* and WCU 1 and 29 (Figure 1).

The ML tree showed identical topography to the NJ tree. There was 99% bootstrap support for Clade 2 and ≤92% for Clade 1.

3.4.3: Annotated Phylogenetic Tree

The annotated phylogenetic tree used phenotypic data to supplement the tree file retrieved from MOLE-BLAST (Figure 3). The annotation showed color changes ranging from red as a 0, blue for an intermediate value 2.48 and green at values of approximately 5. The most noticeable effects occurred in pH tolerances and colony color.

3.5: Principal Components Analysis

The PCA plot of factors 1 versus 2 showed three distinct groupings of organisms and three outlying isolates (21, 22, 29) (Figure 4a). Factor 1 accounted for 24.4% of all the variation in the dataset while Factor 2 accounted for 15.2% of the variance. The three major groupings of isolates (10, 15, 20, 39, and 43; 17, 23, and 30; 1, 4, 8, 9, 26, and 40) were separated along the two factors based on antibiotic resistance, color, and lipid hydrolysis. A PCA plot of factors 1 versus 3 showed two groupings and four outlying isolates (21, 23, 26, 40). Factor 1 as above accounted for 24.4% of the variability, while factor 3 accounted for 13.6%. Major groupings consisted of isolates (1, 4, 8, 9, 17, and 22; and 10, 15, 20, 29, 39, 43) (Figure 4b) and were separated based on colony consistency, citrate utilization, oxidase, and the presence of gelatinase. The PCA plot of factor 2 versus factor 3 showed one large grouping of isolates (1, 4, 8, 9, 10, 15, 17, 20, 21, 23, 30, 39, 43) and four outlying isolates (22, 26, 29, 40) (Figure 4c).

3.6: Pairwise Calculations

Pairwise distance calculation results showed a range of distance between organisms ranging from 0% to 1.4% based on differences in their 16S rDNA sequences. Isolates placed within the *D. nishinomiyaensis* clade showed pairwise differences up to 0.5%. Organisms placed within the *D. barathri* and *D. profundi* clade had pairwise difference calculations from 0-0.2% (Table 7).

3.7: Sequencher Nucleotide Substitution Results

Dermacoccus nishinomiyaensis DSM20448 showed between two and nineteen nucleotide differences between WCU isolates and this type strain (Figure 5). The average number of differences between all isolates and the *D. nishinomiyaensis* type strain was 5.8 nucleotides. Isolates placed within the *D. nishinomiyaensis* clade had approximately 4.1 nucleotide differences on average (Figure 5). The *Dermacoccus barathri* and *Dermacoccus profundi* variance report showed two nucleotide differences between isolate 29 and 1. Isolates within the *Dermacoccus nishinomiyaensis* clade had ~15.9 nucleotide differences from WCU 1 on average (Figure 6).

CHAPTER FOUR: DISCUSSION

Traditionally, the common threshold for a novel bacterial species is 97.5% homology for the 16S rDNA gene, however, for *Dermacoccus* there have been species that have been separated by only one unique nucleotide [8]. Although all 17 isolates in this study share many characteristics to support their placement within the genus *Dermacoccus*, primarily by classification using 16S rDNA, the variation in the physiological and metabolic tests cloud a clear picture of delineating these isolates into distinct species groupings. The phenotypic and metabolic tests that were performed were compared to the tests previously summarized for the described species in *Dermacoccus* as reported in the Bergey's Manual [8], and the results showed notable differences. Phenotypically, colony characteristics of isolates from this study showed ranges in color from yellow to orange, opaque to translucent and the elevation ranged from flat to convex. The Bergey's Manual states the known species are bright orange in color with an opaque/shiny consistency and a slightly convex elevation. Colony size was only listed for *D. nishinomiyaensis* at 2 mm or less, however, WCU 10 reached a diameter of 2.25 mm.

The temperature, salt, and pH ranges in this work also expanded the boundaries of reported ranges within the genus. According to the Bergey's Manual, the species within *Dermacoccus* generally have been shown to grow within the 10-37° C range but the isolates in this study could grow from 20-42°C with some reaching higher temperatures of 45-50°C. Salt tolerances for *Dermacoccus* species can be high, perhaps due to their isolation from areas such as cured meat, skin, and deep-sea trenches. The Bergey's Manual reports isolates belonging to *D. nishinomiyaensis* growing well when exposed to 0-5% NaCl. Many isolates in this study, however, could only tolerate up to 4% NaCl, in some cases, while others could tolerate up to 10% NaCl. The salt tolerances for *Dermacoccus profundi* and *D. barathri* are reported as high as 10% NaCl, however, the salt tolerances for the isolates in this study that were most closely related to these species (isolates 1 and 29) could not grow above 5% NaCl. The pH range for the genus has not been measured, however, the isolates could each grow well in moderately acidic and alkaline environments, from pH 5-9. Anaerobic testing was performed using two different techniques; a shake culture technique as well as an anaerobic chamber technique. The shake culture technique results were recorded due to inconsistent results with the anaerobic chamber technique.

Metabolic results also varied widely from Bergey's type species results. EnteroPluri tube results showed notable differences in the urea hydrolysis and citrate utilization tests. The Bergey's Manual shows that urea is hydrolyzed at least 80% of the time for *D. barathri*, *D. nishinomiyaensis* and *D. profundi*, however, 100% of the isolates in this study hydrolyzed urea [8]. This shows the ability of the organism to utilize water to break urea into ammonia and carbon dioxide, providing the organism a source of nitrogen. Citrate utilization as reported in the

literature only occurs 15% of the time for *Dermacoccus* species, while in this experiment 59% of isolates could use citrate as a sole carbon and energy source for growth.

Gelatin hydrolysis, starch hydrolysis, nitrate reduction, lipid hydrolysis, DNAse production, and oxidase production also had varying results from those reported for the genus in the Bergey's Manual [8]. Gelatin is reported to be hydrolyzed 100% of the time within *D. nishinomiyaensis*, however, only 76% of isolates showed gelatinase activity. The hydrolysis of starch has been reported for 35% of *Dermacoccus* strains, however, it was present in 47% of the isolates within this study. Nitrate reduction was shown to be variable in the Bergey's Manual, with 67% of isolates showing the ability to reduce nitrate to nitrite or N2. In this study, no isolates could reduce nitrate. Lipid hydrolysis results were similar to the 85% positive results reported in the Bergey's Manual. Fourteen of the seventeen (82%) isolates in this work tested positive for the ability to break down lipids. DNAse was stated to be absent within *Dermacoccus nishinomiyaensis* [8], however, 35% of isolates in this work could produce DNAse. Only one species showed an oxidase positive result in this study, however, this molecule is known to exist in *Dermacoccus* and may be weakly expressed [8].

Antibiotic testing was completed to test if the isolates had resistance to pharmaceutical therapies. Although Bergey's Manual only reports data for one of the seven antibiotics tested, the knowledge of these isolates' resistance to common antibiotics can be beneficial to understanding their pathogenicity and biology, especially due to the ability of at least one strain of *D. barathri* to cause infection [8]. Isolates showed resistance to nalidixic acid and colistin in 82% of the cultures (14/17 isolates). The bacteria showed susceptibility to the other five antibiotics; clindamycin, chloramphenicol, tetracycline, ampicillin, and nitrofurantoin. Clindamycin and chloramphenicol inhibited growth for 65% of the isolates, tetracycline inhibited growth for 70%,

ampicillin controlled growth for 53%, and nitrofurantoin was effective against 59% of the isolates, respectively. *Dermacoccus* spp. are not known to be pathogenic species; however, antibiotic resistance of the isolates in this study, as common bacterial members of the human microflora, would have a relatively simple path to taking advantage of its host in the event that conditions allowed for it. There is also growing evidence showing bacteria evolving resistance to antibiotics in environments that are not exposed to commercial versions of these drugs and this opens basic science questions about why this occurs.

The mode of attack of these antibiotics targets necessary cellular function, inhibiting the cells ability to continue the course of infection. Ampicillin attacks the cell wall synthesis of the organism [29] while colistin attacks the outer cell membrane of Gram negative bacteria [30] [31]. Clindamycin attacks the 50S large ribosomal subunit [32], tetracycline targets the 30S small ribosomal subunit [33], and chloramphenicol targets the 70S ribosome of the bacterial cell [34]. Nalidixic acid regulates binary fission by blocking DNA replication [35]. Nitrofurantoin damages the bacterial cell by interacting with nitrofurantoin reductase flavoproteins that attack the ribosomal DNA, respiration, and pyruvate metabolism [36]. Understanding the mode of killing within these antibiotics can be useful if a *Dermacoccus* isolate became opportunistically pathogenic. When the mode of killing is known it is much easier to determine what antibiotic is most useful against the Gram positive structure of *Dermacoccus* cells.

Phylogenetic trees from this study show evolutionary distance between species based on base pair changes within the 16S rDNA (Figure 2) [37] [38]. Bootstrap analyses were done to assess the accuracy of the placement of species within the tree and values were generally above 80%; however, some bootstrap values were lower, likely due to incomplete 16S rDNA

sequences of approximately 1,049 base pairs -1,212 base pairs of isolates within the *D. nishinomiyaensis* clade. The bootstrap estimates within the maximum likelihood tree show two distinct clades separated by a 99% bootstrap value (Figure 2). This separation between clades is also indicative of a novel species in each clade supported by a pairwise distance calculation of 1.4% from clade one, and 0.5% from clade two (Table 7). A bootstrap estimate was also used to separate WCU 9 from clade one, suggesting a possible third species that could be distinguished from phylogeny alone. WCU 9 is an outlier of the *D. nishinomiyaensis* clade, however the isolate differs from *D. nishinomiyaensis* by 0.70% and differs from the other three *Dermacoccus* spp. by 1.2%. Branch length values were also assigned to observe the amount of evolutionary change the organisms have undergone while allowing the tree to be easily interpreted. Branch length values were minimal, with 0.00 length values assigned to all *Dermacoccus* isolates. This shows all the organisms have minimal genetic change over time for the 16S rDNA gene.

The annotated phylogenetic tree with a heatmap shows a clearer picture of how similar organisms within this genus can be from a DNA standpoint yet how different their metabolic and phenotypic measurements may be. Closely related organisms within the *D. nishinomiyaensis* clade vary on all characteristics besides cell shape, colony shape, edge, Gram stain, urea hydrolysis, and catalase. Some characteristics vary widely, such as pH and color, denoted by the color range displayed in Figure 3. The principal components analysis showed WCU 29 and 22 to be outliers in both figures (Figures 4a and 4c), signifying those isolates are the most phenotypically diverse from the other isolates.

Isolate 29 was also separated on the basis of DNA base pairwise comparison, showing that it differed from *D. abyssi*, *D. barathri*, and *D. profundi* by 0.2%-0.3% according to the pairwise distance calculation in Table 7. Pairwise calculations show WCU 29 differs from WCU 1 by 0.2% and 10 phenotypic tests. These organisms are close in 16S rDNA sequences, however, nearly half of tests performed with these organisms gave different results, arguing against placement within the same *Dermacoccus* species.

The Bergey's Manual states there is some variation between organisms that are classified within the same species, however, no definitive threshold is expressed to determine how many phenotypic or other differences might define a novel species. Within the *Dermacoccus* genus, the maximum sequence difference between species described in the literature is 1.4% according to the pairwise calculation, yet there are four distinct species represented. The threshold to determine a novel species is 97.5% similarity. Using this pairwise calculation, organisms *Dermacoccus barathri* and *Dermacoccus profundi* would have a 16S rDNA homology of approximately 98.6%, well above the threshold, yet they have been separated as two individual species. An example of a 16S rDNA homology threshold deficiency was discovered while separating species within the *Mycobacterium* genus. According to Woo et al., five *Mycobacterium* species could not be separated from one another using 16S rDNA sequencing alone, instead requiring additional phenotypic tests to separate them alongside other gene targets such as hsp65 and rpoB [39]. This combination of phenotypic tests and metabolic capability testing alongside 16S rDNA should be more heavily relied on than a specific 97.5% 16S rDNA threshold. Other genes are often also looked at. If 16S rDNA alone was relied on for the *Dermacoccus* species with only 1.4% nucleotide difference between two distinct species, it could be argued that a 98.6% threshold could determine

novel species, distinguishing 13 of 17 isolates as novel. However, with the combination of 16S rDNA sequencing, phylogeny, phenotypic and metabolic data the possibility of 13 novel species being found in this study is possible. Conservatively, it is more likely to propose three novel species from this work based on the phylogenetic data. The isolates displaying the highest amount of novelty are isolates 9, 23, and

Whole genome sequencing of WCU 9, 23, and 29 could help to further delineate these species based on all of their genes. This would also allow for clues to further phenotypes that could be tested such as gene products used in metabolism (e.g., enzymes for carbon use, cycling of other elements), pathogenicity traits, and other characteristics that could lead to hypothesis testing in the wet lab and in genomic studies with other species. Intra species variation can occur from horizontal gene transfer of plasmids as well as mutations within the genome. According to NCBIs nucleotide database, *Dermacoccus nishinomiyaensis* contains plasmids within the whole genome sequence, which could explain the variation between isolates [40]. Plasmids allow for bacteria to pick up resistance genes from other organisms by conjugation. This phenomenon has allowed organisms to become resistant to certain pharmaceutical therapies, leading to the current antibiotic conundrum. Alongside correct use of antibiotic prescriptions, tests such as cell membrane fatty acids, cytochromes, and peptidoglycan chemistry can be used to identify and treat organisms.

CONCLUSION

The genus *Dermacoccus* was first described in 1995 when *Micrococcus nishinomiyaensis* was emended into this genus, and since the genus was discovered there have only been three

additions. The genus is relatively small and understudied. Through this experiment, many additional phenotypic and metabolic characteristics were tested, and results could be instrumental in further understanding this genus. Though *Dermacoccus* is largely nonpathogenic, there has been a documented case of opportunistic pathogenesis giving more reason to further study this group. The results of this study could be beneficial to the growing knowledge and study of the human microbiome. Classification of novel species will only expand the ability of other microbiologists to determine what niches this genus fills when found within the normal microflora of the skin and elsewhere. The organisms deemed to be novel (WCU 9, 23, 29) are given the provisional names *Dermacoccus cullowheensis*, *Dermacoccus os*, and *Dermacoccus computatri*, respectively.

REFERENCES

1. Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M., & Stahl, D. A. (2019). Brock Biology of Microorganisms (15th ed.). New York, NY: Pearson.

2. Baron, S. (1996). Medical Microbiology (4th ed.). Galveston, TX: University of Texas Medical Branch at Galveston.

3. O'Connell, S. (2017). Laboratory Manual. Cullowhee, NC: Western Carolina University.

4. Gao, Z., Tseng, C., Pei, Z., Blaser, M. (2007). Molecular Analysis of Human Forearm Superficial Skin Bacterial Biota. Proceedings of the National Academy of Sciences, 104: 2927–2932.

5. Pray, L., Pillsbury, L., Tomayko, E. Institute of Medicine. (2013). The Human Microbiome, Diet, and Health: Workshop Summary. Washington, DC: The National Academies Press.

6. NIH Human Microbiome Project Defines Normal Bacterial Makeup of the Body. (2015). https://www.nih.gov/news-events/news-releases/nih-human-microbiome-project-definesnormal-bacterial-makeup-body.

7. Genomics and Precision Health. (2017). https://blogs.cdc.gov/genomics/2017/06/07/precision-public-health/.

8. Stackebrandt, E. (2015). Dermacoccus. In Bergey's Manual of Systematics of Archaea and Bacteria (eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, B. Hedlund and S. Dedysh). doi:10.1002/9781118960608.gbm00071.

9. Szczerba, I. (2003). Occurrence and Number of Bacteria from the Micrococcus, Kocuria, Nesterenkonia, Kytococcus and Dermacoccus Genera on Skin and Mucous Membranes in Humans. Medycyna Doswiadczalna I Mikrobiologia, 55:67-74.

10. Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. PorrasAlfaro, C. R. Kuske, and J. M. Tiedje. (2014). Ribosomal Database Project: Data and Tools for High Throughput rRNA Analysis. Nucleic Acids Research, 42:D633-D642.

11. Sender R, Fuchs S, Milo R (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biology. https://doi.org/10.1371/journal.pbio.1002533.

12. Fung, T., Artis, D., & Sonnenberg, G. (2014). Anatomical Localization of Commensal Bacteria in Immune Cell Homeostasis and Disease. Immunological Review, 260:35-49.

13. Takahashi, N., Shinjoh, M.,Tomita, H., Fujino, A., Sugita, K., Katohno, Y., Kuroda, T., Kikuchi, K. (2015). Catheter-Related Blood Stream Infection Caused by Dermacoccus barathri, Representing the First Case of Dermacoccus Infection in Humans. Journal of Infection and Chemotherapy, 21:613–616.

14. Grice, E., Kong, H., Conlan, S., Deming, C., Davis, J., Young, A., Bouffard, G., Blakesley, R., Murray, P., Green, E., Turner, M., Segre, J. (2009). Topographical and Temporal Diversity of the Human Skin Microbiome. Science, 324:1190–1192.

15. Scharschmidt, T., Fischbach, M. (2003). What Lives on Our Skin: Ecology, Genomics and Therapeutic Opportunities of the Skin Microbiome. Drug Discovery Today: Disease Mechanisms, 10:e83-e89.

16. Janda, J., Abbott, S. (2007). 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. Journal of Clinical Microbiology, 45:2761-2764.

17. Donelli, G., Vuotto, C., & Mastromarino, P. (2013). Phenotyping and Genotyping are both Essential to Identify and Classify a Probiotic Microorganism. Microbial Ecology in Health and Disease, doi:10.3402/mehd.v24i0.20105.

18. Marzo (2005). EnteroPluri-Test Archivio Liofilchem. legacy.bd.com/ds/technicalCenter/inserts/L010569(02).pdf.

19. Cruz, T., Torres J. (2012). Gelatin Hydrolysis Test Protocol. ASMScience, American Society of Microbiology. http://www.asmscience.org/content/education/protocol/protocol.3776.

20. Aryal, S. (2019). Nitrate Reduction Test - Principle, Procedure, Uses and Interpretation. https://microbiologyinfo.com/nitrate-reduction-test/.

21. Mole-Blast. In Mole-Blast: Phylogenetic Tree Widget https://blast.ncbi.nlm.nih.gov/moleblast/moleblast.cgi.

22. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology, 215:403-410.

23. Kumar,S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution, 33:1870-1874.

24. Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating the Human-Ape Split by a Molecular Clock of Mitochondrial DNA. Journal of Molecular Evolution, 22:160-174.

25. Wang, Q, Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology, 73:5261-5267.

26. Letunic, I., Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent Updates and New Developments. Nucleic Acids Research. doi.org/10.1093/nar/gkz239.

27. SYSTAT Software (2002). SYSTAT 10 for Windows. SYSTAT Software, Richmond, WA.

28. Sequencher® version 5.4.6 DNA Sequence Analysis Software, Gene Codes Corporation, Ann Arbor, MI.

29. Parr, T. R., Bryan, L. E. (1984). Mechanism of Resistance of an Ampicillin-Resistant, BetaLactamase-Negative Clinical Isolate of Haemophilus influenzae Type B to Beta-Lactam Antibiotics. Antimicrobial Agents and Chemotherapy, 25:747–753.

30. Nikolaidis, I., Favini-Stabile, S., & Dessen, A. (2014). Resistance to Antibiotics Targeted to the Bacterial Cell Wall. Protein Science, 23:243-259.

31. Sun, J., Zeng, X., Li, X., Liao, X., Liu, Y., Lin, J. (2017). Plasmid-Mediated Colistin Resistance in Animals: Current Status and Future Directions. Animal Health Research Reviews, 18:136-152.

32. Long, S. S., Pickering, L. K., & Prober, C. G. (2012). Principles and Practice of Pediatric Infectious Disease (4th ed.). Edinburgh: Churchill Livingstone.

33. Ammor, M. S., Gueimonde, M., Danielsen, M., Zagorec, M., Van Hoek, A. H., De Los Reyes-Gavilán, C. G., Margolles, A. (2008). Two Different Tetracycline Resistance Mechanisms, Plasmid-Carried Tet(L) and Chromosomally Located TransposonAssociated Tet(M), Coexist in Lactobacillus sakei Rits 9. Applied and Environmental Microbiology, 74:1394-1401.

34. Cohen, J., Powderly, W. G., Opal, S. M. (2017). Infectious diseases. Amsterdam: Elsevier.

35. Tang, Y., Sussman, M., Liu, D., Poxton, I., & Schwartzman, J. (2015). Molecular Medical Microbiology (2nd ed.). Amsterdam: Academic Press.

36. Sekyere, J. (2018). Genomic Insights into Nitrofurantoin Resistance Mechanisms and Epidemiology in Clinical Enterobacteriaceae. Future Science. doi: 10.4155/fsoa-2017-0156.

37. Saitou, N., Nei, M. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. Molecular Biology and Evolution, 4:406-425.

38. 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 (USA), 101:11030-11035.

39. Woo, P., Lau, S., Teng, J., Yuen, K. (2008). Then and Now: Use of 16S rDNA Gene Sequencing for Bacterial Identification and Discovery of Novel Bacteria in Clinical Microbiology Laboratories. Clinical Microbiology and Infection, 14:908-934.

40. Hong, K., Chan, K. (2017). Dermacoccus nishinomiyaensis strain M25, Complete Genome. National Center for Biotechnology Information, U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/nucco.

Table 1. Sources of samples for the 17 *Dermacoccus* isolates obtained in this study and length of 16S rDNA sequences used in SeqMatch analyses using the Ribosomal Database Project. The

percentage (%) match indicates the relationship of each isolate to one of four *Dermacoccus* species found in the database.

Isolate	Sequence length	% match	Sample source
WCU1	1204	100% *D. barathri/profundi*	Headphones
WCU4	1094	97.8% *D. nishinomiyaensis*	Pull up bar
WCU8	1184	98.4% *D. nishinomiyaensis*	Cell phone
WCU9	1097	96.6% *D. nishinomiyaensis*	Mouth
WCU10	1057	96.2% *D. nishinomiyaensis*	Watch
WCU15	1184	97.8% *D. nishinomiyaensis*	Cell phone
WCU17	1175	98.9% *D. nishinomiyaensis*	Trash can
WCU20	1077	98.0% *D. nishinomiyaensis*	Fingernail
WCU21	1049	97.6% *D. nishinomiyaensis*	Cell phone
WCU22	1123	98.3% *D. nishinomiyaensis*	Ring
WCU23	1212	97.9% *D. nishinomiyaensis*	Calculator
WCU26	1053	97.4% *D. nishinomiyaensis*	Cell phone
WCU29	1173	98.9% *D. barathri/profundi*	Cell phone
WCU30	1126	98.3% *D. nishinomiyaensis*	Cell phone case
WCU39	1087	98.3% *D. nishinomiyaensis*	Cell phone case
	1204	99.0% *D. nishinomiyaensis*	Earring
	1104	98.3% *D. nishinomiyaensis*	Earring

WCU40

WCU43

Table 2. GenBank identification numbers of 31 isolates provided by Mole-Blast software used in phylogenetic analyses in Figures 1-3.

Culture Name	Gene bank/NCBI Accession number
WCU 1	MK880152
WCU 4	MK880153
WCU 8	MK880154
WCU 9	MK880155
WCU 10	MK880156
WCU 15	MK880157
WCU 17	MK880158
WCU 20	MK880159
WCU 21	MK880160
WCU 22	MK880161
WCU 23	MK880162
WCU 26	MK880163
WCU 29	MK880164
WCU 30	MK880165
WCU 39	MK880166
WCU 40	MK880167
WCU 43	MK880168
***Dermacoccus nishinomiyaensis* DSM 20448**	NR_044872
***Dermacoccus abyssi* MT1.1**	NR_043260
***Dermacoccus barathri* MT2.1**	NR_043261
***Dermacoccus profundi* MT2.2**	NR_043262
***Austwickia chelonae* NBRC 105200**	NR_113145
***Austwickia chelonae* W16**	NR_028904
***Calidifontibacter indicus* PC IW02**	NR_115977
***Calidifontibacter terrae* R161**	NR_156958
***Kineosphaera limosa* NBRC 100340**	NR_036896
***Luteipulveratus halotolerans* C296001**	NR_144592
***Luteipulveratus mongoliensis* MN07-A0370**	NR_112830
***Rudaeicoccus suwonensis* HOR6-4**	NR_108544
***Yimella lutea* YIM45900**	NR_116716
***Yimella radicis* py1292**	NR_152030

Table 3. Colony and cell characteristics of 17 WCU isolates used to describe colony size, cell size, cell shape, colony shape, color, consistency, edge, elevation, and Gram stain.

WCU #	Colony size	Cell size	Cell shape	Colony shape	Color	Consistency	Edge	Elevation	Gram stain
1	0.5mm	1µm	cocci	circular	yellow	opaque/shiny	entire	flat	positive
4	1.5mm	1-2µm	cocci	circular	yellow/orange	opaque/shiny	entire	raised	positive
8	0.75mm	1µm	cocci	circular	yellow/orange	opaque/shiny	entire	raised	positive
9	1mm	1µm	cocci	circular	yellow	opaque/shiny	entire	flat	positive
10	2.25mm	1.5µm	cocci	circular	orange	opaque/shiny	entire	raised	positive
15	1mm	1.5µm	cocci	circular	orange	opaque/shiny	entire	raised	positive
17	1mm	1.5µm	cocci	circular	orange	opaque/shiny	entire	flat	positive
20	1-2mm	1µm	cocci	circular	yellow	opaque/shiny	entire	convex	positive
21	2mm	1.25µm	cocci	circular	orange	opaque/shiny	entire	raised	positive
22	1mm	1µm	cocci	circular	yellow	opaque/shiny	entire	flat	positive
23	0.5-1mm	1.25µm	cocci	circular	orange	opaque/shiny	entire	raised	positive
26	2mm	1.25µm	cocci	circular	yellow/orange	translucent/shiny	entire	flat	positive
29	0.5-1mm	1µm	cocci	circular	yellow	translucent/shiny	entire	raised	positive
30	1.5mm	1µm	cocci	circular	orange	opaque/shiny	entire	raised	positive
39	2mm	1.25µm	cocci	circular	orange	opaque/shiny	entire	raised	positive
40	1.5-2mm	1µm	cocci	circular	orange	translucent/shiny	entire	raised	positive
43	0.5-1mm	1.25µm	cocci	circular	orange	opaque/shiny	entire	raised	positive

Table 4. Temperature, salt and pH tolerances for the 17 *Dermacoccus* isolates.

WCU #	Temperature (°C)	Salt	pH
1	20-42°	0-5%	5-10
4	20-45°	0-5%	7-10
8	20-50°	0-5%	7-10
9	20-42°	0-10%	7-10
10	20-42°	0-5%	7-10
15	20-42°	0-5%	5-9
17	20-42°	0-5%	5-9
20	20-42°	0-4%	5-9
21	20-42°	0-4%	7-9
22	20-42°	0-5%	7-10
23	20-42°	0-10%	7-9
26	20-42°	0-10%	7-9
29	20-42°	0-4%	7-9
30	20-45°	0-5%	7-9
39	20-42°	0-5%	7-9
40	20-42°	0-4%	5-9
43	20-42°		

Table 5

. Enteropluri results for 17 *Dermacoccus* isolates for sugar and sugar alcohol fermentations, decarboxylations, deaminations, hydrolysis of urea, and utilization of different carbon and energy sources.

Characteristics	1	4	8	9	10	15	17	20	21	22	23	26	29	30	39	40	43
Glucose Fermentation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gas Production	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lysine Decarboxylation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ornithine Decarboxylation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sulfate Reduction	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Indole Production	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Adonitol Fermentation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lactose Fermentation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Arabinose Fermentation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sorbitol Fermentation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Acetoin Production	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dulcitol Fermentation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Phenylalanine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Urea	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 6

Hydrolysis															
Citrate	+	+	-	+	+	+	+	+	-	+	-	-	-	+	+

. Metabolic capabilities of *Dermacoccus* isolates to determine if organisms can hydrolyze starch, lipids, and gelatin, breakdown DNA and Hydrogen peroxide, reduce nitrate and utilize cytochrome oxidase c.

WCU #	**Gelatin Hydrolysis**	**Starch -**	**SIM**	**Nitrate Reduction**	**Lipid Hydrolysis**	**DNAse**	**Catalase**	**Oxidase**
1	-		N/A	-	+	+	+	-
4	+	+	N/A	-	-	-	+	-
8	+	+	N/A	-	+	+	+	-
9	+	+	N/A	-	+	-	+	-
10	+	+	N/A	-	+	+	+	-
15	+	-	N/A	-	+	+	+	-
17	+	-	N/A	-	+	+	+/-	-
20	+	-	N/A	-	+	-	+	-
21	+	-	N/A	-	+	-	+	-
22	+	-	-	-	-	-	+	-
23	+	-	N/A	-	+	-	+	-
26	-	+	N/A	-	+	-	+	-
29	-	-	N/A	-	-	-	+	-
30	+	-	N/A	-	+	-	+	-
39	-	+	N/A	-	+	-	+	-
40	-	-	N/A	-	+	-	+	+

Table 7

43	+	+	N/A	-	+	+	+	-

. Antibiotic resistance or susceptibility of 17 *Dermacoccus* isolates. Resistance refers to no zone of inhibition surrounding the 30µg antibiotic disc, while susceptibility refers to a zone of inhibition surrounding the 30µg antibiotic disc.

WCU #	Nalidixic acid	Clindamycin	Colistin	Chloramphenicol	Tetracycline	Ampicillin	Nitrofurantoin
1	R	S	R	S	S	S	S
4	R	S	R	S	S	S	S
8	R	S	R	S	S	S	S
9	R	S	R	S	S	S	R
10	R	R	R	R	R	R	S
15	R	R	R	R	R	R	R
17	S	S	S	S	S	S	S
20	R	R	R	R	S	R	R
21	R	S	R	S	S	S	R
22	R	S	R	S	S	R	R
23	S	S	S	S	S	S	S
26	R	S	R	S	S	R	S

Table 8

29	R	R	R	R	R	R	S
30	R	S	S	S	S	S	S
39	R	R	R	R	R	R	R
40	S	S	R	S	S	S	S
43	R	R	R	R	R	R	R

Table 9

. Numeric assignments for parameters that signified possible novel organisms such as colony and morphological characteristics, cell characteristics, and metabolic capabilities of the 17 *Dermacoccus* isolates for analyses shown in Figures 3-6.

Parameter	Measurement	Numeric Assignment
Colony Size	0-1mm	0
	1-2mm	1
	>2mm	2
Cell Size	0-1µm	0
	>1µm	1
Colony Color	Yellow	0
	Yellow/Orange	1
	Orange	2
Colony Shape	Cocci	1
Cell Shape	Circular	1
Consistency	Opaque/Shiny	0
	Translucent/Shiny	1
Edge	Entire	1
Gram Stain	Positive	1
Elevation	Flat	0
	Raised	1
	Convex	2
Temperature	20-42°C	0
	20-45°C	1
	20-50°C	2
Salt	0-4%	0
	0-5%	1
	0-10%	2
pH	5-9pH	1
	5-10pH	2
	7-9pH	3
	7-10pH	4
Antibiotics	Susceptible	0
	Resistant	1
All Other Tests	Negative	0
	Positive	1

Table 10

. Component loading values from principal components analysis (PCA) in Figure 4a, b, c with highly weighted characteristics that drove culture isolation generally being > 0.5 or < -0.5.

	1	2	3
Colony Size Size	0.168	0.164	0.412
Cell Size	0.237	0.562	0.138
Colony color	-0.013	0.777	0.382
Colony Consistency	-0.141	-0.525	0.754
Colony Elevation	0.437	0.199	0.234
Temperature tolerance	-0.329	-0.083	-0.262
Salt tolerance	-0.573	0.103	0.011
pH tolerance	-0.32	-0.233	0.243
Citrate utilization	0.39	0.209	-0.545
Oxidase	-0.319	-0.087	0.667
Nalidixic Acid	0.56	-0.35	-0.376
Clindamycin	0.93	0.074	0.216
Colistin	0.495	-0.518	0.119
Chloramphenicol	0.93	0.074	0.216
Tetracycline	0.811	0.07	0.278
Ampicillin	0.833	-0.185	0.154
Nitrofurantoin	0.591	0.136	-0.358
Gelatinase presence	0.029	0.456	-0.656

Table 11

Amylase activity	0.187	0.009	0.012
Lipase presence	-0.095	0.752	0.245

Isolates	DSM 20448	10	15	17	20	21	22	23	26	30	39	4	40	43	8	9	MT 1.1	1	29	MT 2.1
DSM 20448																				
10	0.5																			
15	0.2	0.3																		
17	0.2																			
20	0.2																			
21	0.2																			
22	0.3					0.1														
23	0.4					0.2	0.2													
MT2.1	1.4	1.4				1.2	1.2	1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.2					
MT2.2	1.4	1.4	1.3	1.2	1.3	1.2	1.2	1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	0.2	0.0	0.2	0.0

Table 10.

Percentage pairwise nucleotide distance values. Analyses were conducted using the Maximum Composite Likelihood model.[29] This analysis involved 21 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. Pairwise deletion option was chosen. There were a total of 1157 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. [31] DSM 20448 is *D.nishinomiyaensis*, MT1.1 is *D.abyssi*, MT2.1 *D. barathri*, and MT2.2 is *D.profundi.*

		0.3	0.0					
		0.3	0.0	0.0				
		0.2	0.0	0.0	0.0			
		0.2	0.0	0.1	0.1			
		0.2	0.0	0.3	0.3			
26	0.2	0.2	0.0	0.0	0.0	0.0		
30	0.3	0.2	0.1	0.1	0.1	0.1		
39	0.3	0.2	0.1	0.1	0.1	0.1		
4	0.3	0.2	0.1	0.1	0.1	0.1		
40	0.2	0.3	0.0	0.0	0.0	0.0	0.1	0.2
43	0.3	0.2	0.1	0.1	0.1	0.1		
8	0.3	0.2	0.1	0.1	0.1	0.1		
9	0.7	0.5	0.5	0.5	0.5	0.5		
MT1.1	1.5	1.5	1.4	1.3	1.4	1.3	0.0	0.0
1	1.4	1.4	1.3	1.2	1.3	1.2	0.0	0.1
29	1.4	1.2	1.3	1.2	1.3	1.1	0.0	0.0
			1.3	1.2	1.3		0.4	1.2
						1.2		
						1.2		

0.2	0.1										
0.2	0.1	0.0									
0.2	0.1	0.0	0.0								
0.3	0.0	0.1	0.1	0.1							
0.2	0.1	0.0	0.0	0.0	0.1						
0.2	0.1	0.0	0.0	0.0	0.1	0.0					
0.3	0.5	0.4	0.4	0.4	0.5	0.4	0.4				
1.2	1.4	1.2	1.2	1.2	1.3	1.2	1.2	1.2			
1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	0.2		
1.1	1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.3	0.3	0.2	
								1.2	0.2	0.0	0.2

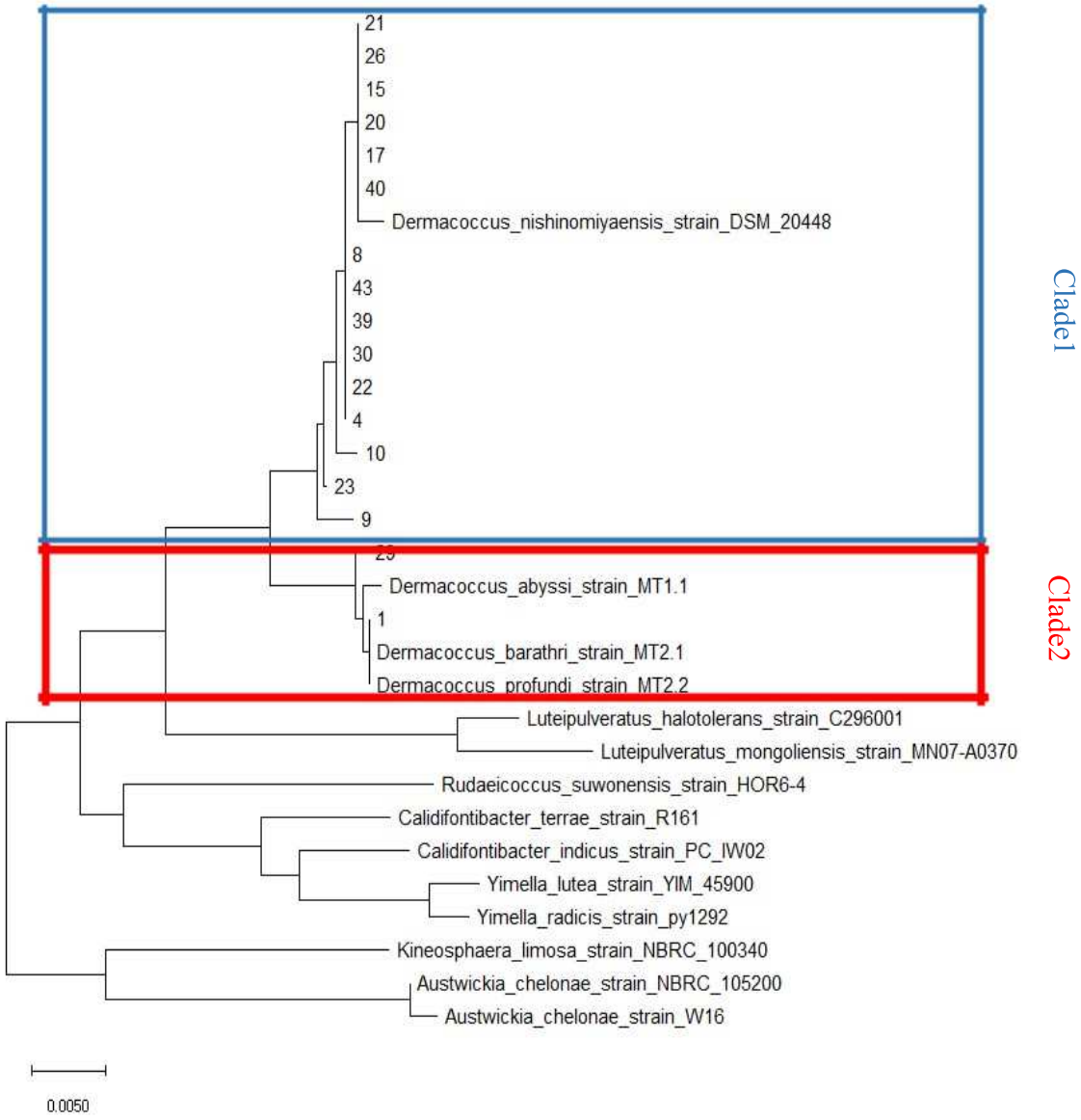
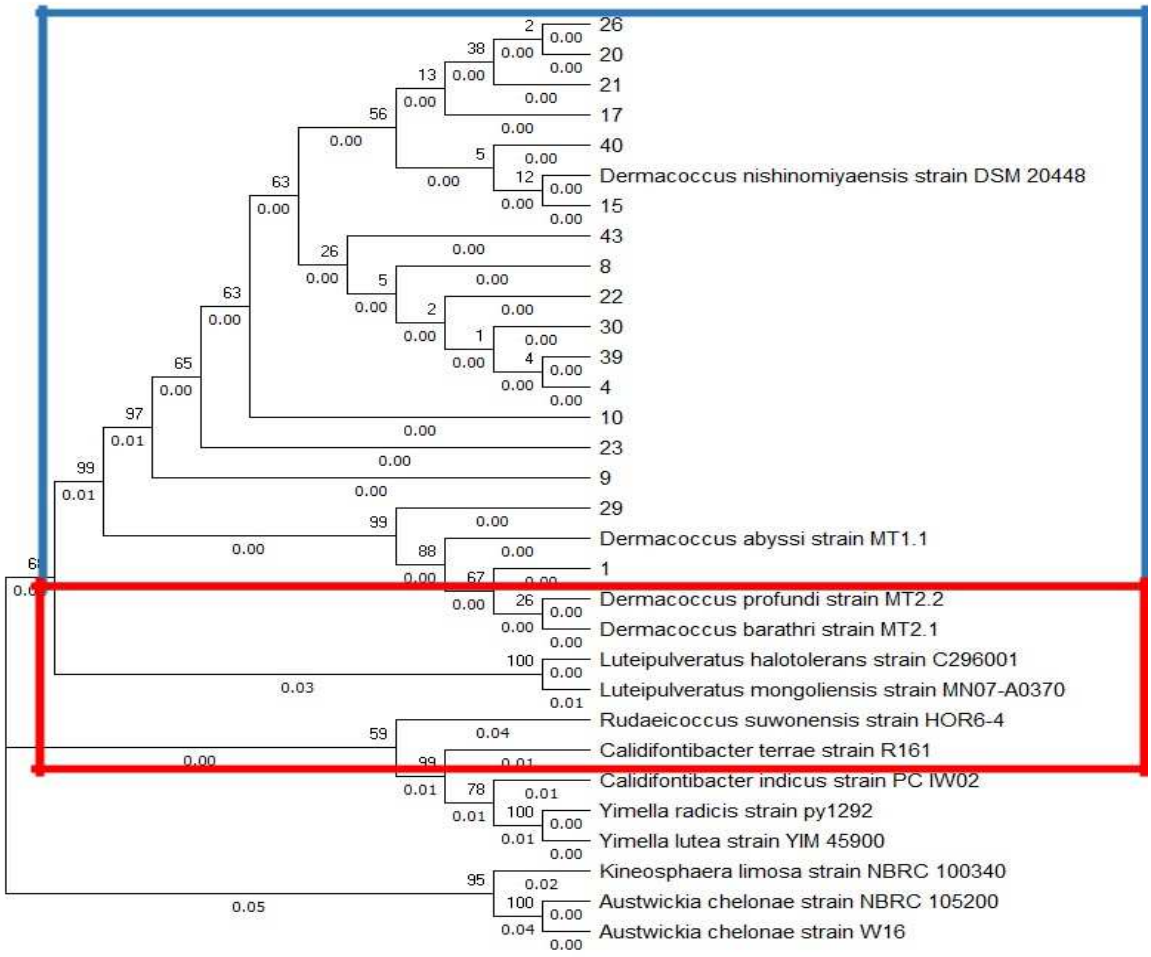


Figure 1. The evolutionary history was inferred using the Neighbor-Joining method [39]. The optimal tree with the sum of branch length = 0.08099484 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [40] and are in the units of the number of base substitutions per site. This analysis involved 31 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1271 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [23].



40

Figure 2. Molecular Phylogenetic analysis by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano model [24]. The tree with the highest log likelihood (-2903.76) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6321)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 78.04% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 31 nucleotide sequences. There were a total of 1156 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [23].

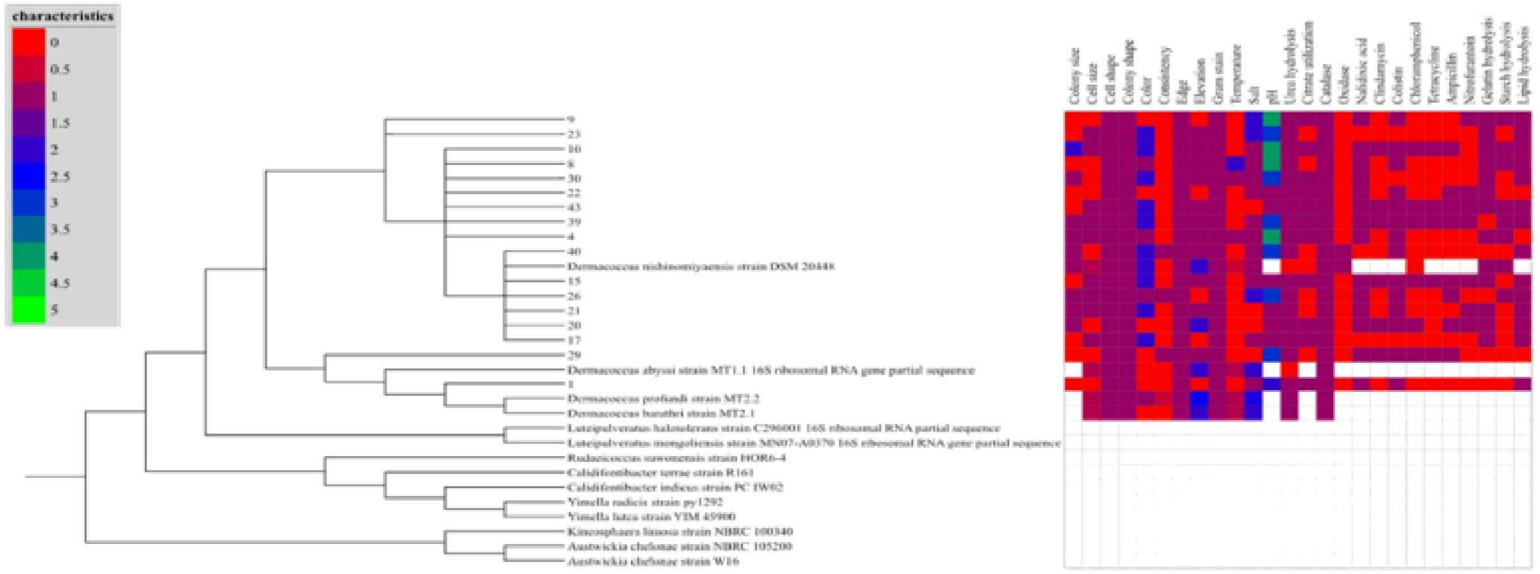


Figure 3. Annotated phylogenetic tree using a heat map to show differences within each characteristic. Colony size: <1-Red, >2-Blue. Cell size: <1-Red,>1.1 Maroon. Cell shape: Cocci-Maroon. Colony shape: Circular-Maroon. Colony color: yellow-Red, Orange-Blue.

Consistency: opaque/shiny -Red, translucent/shiny-Maroon. Edge: entire-Maroon. Elevation: flat-Red, convex-Blue. Gram stain: positive -Maroon. Temperature:20-42-Red, 20-50-Blue. Salt:0-4-Red, 0-10-Blue. pH: 3-9-Red, 7-10-Green. Urea hydrolysis, Citrate utilization, Catalase, Oxidase, Gelatin hydrolysis, Starch hydrolysis, Lipid hydrolysis, and DNAse: negative-Red, positive-Maroon Nalidixic acid, Clindamycin, Colistin, Chloramphenicol, Tetracycline, Ampicillin, and Nitrofurantoin: Susceptible-Red, ResistantMaroon.

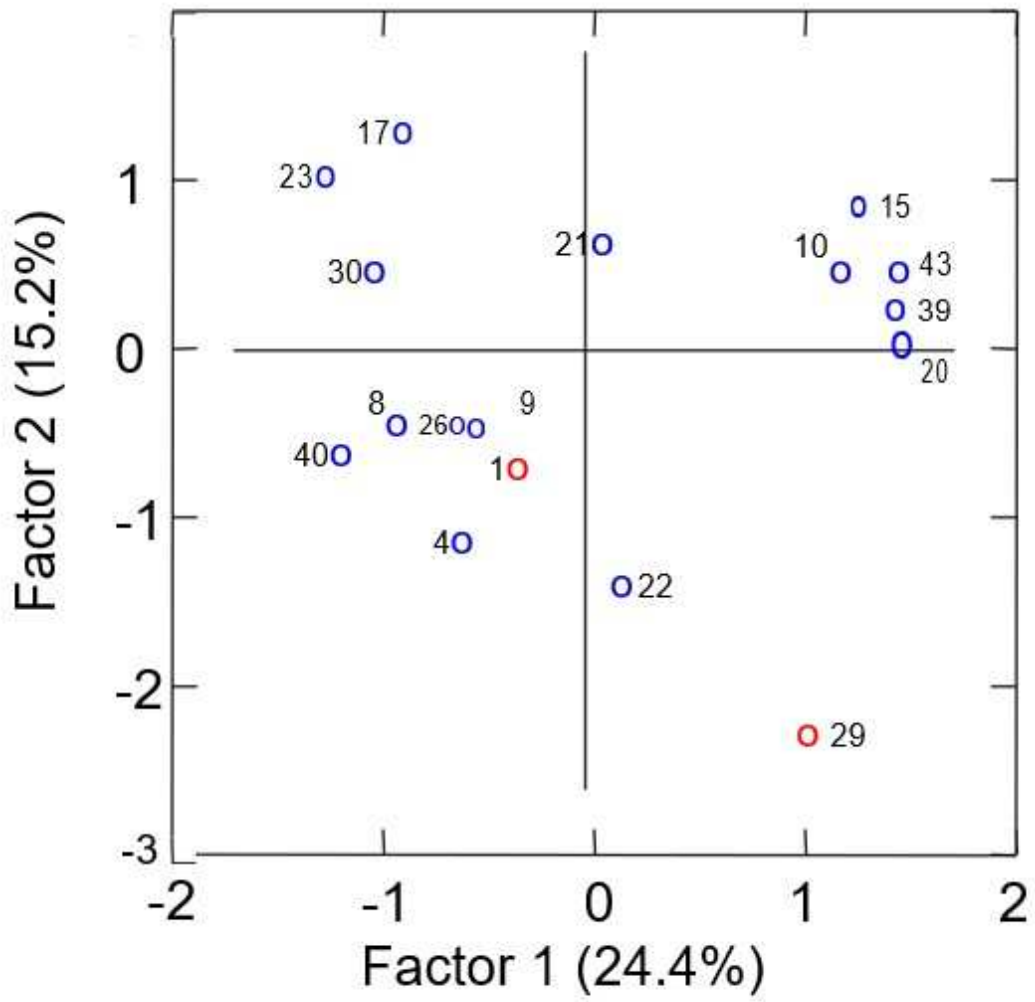


Figure 4a. Principal components analysis of metabolic and phenotypic characteristics for 17 isolates. Three distinct clusters of isolates are present (10, 15, 20, 39, and 43; 17, 23, and 30; and 1, 4, 8, 9, 26, and 40). Three outliers were identified as isolates 21, 22, and 29. Blue plotted isolates signify close 16S rDNA relationship to *D. nishinomiyaensis*, while red plotted isolates are more closely related to *D. barathri*/*D. profundi.*

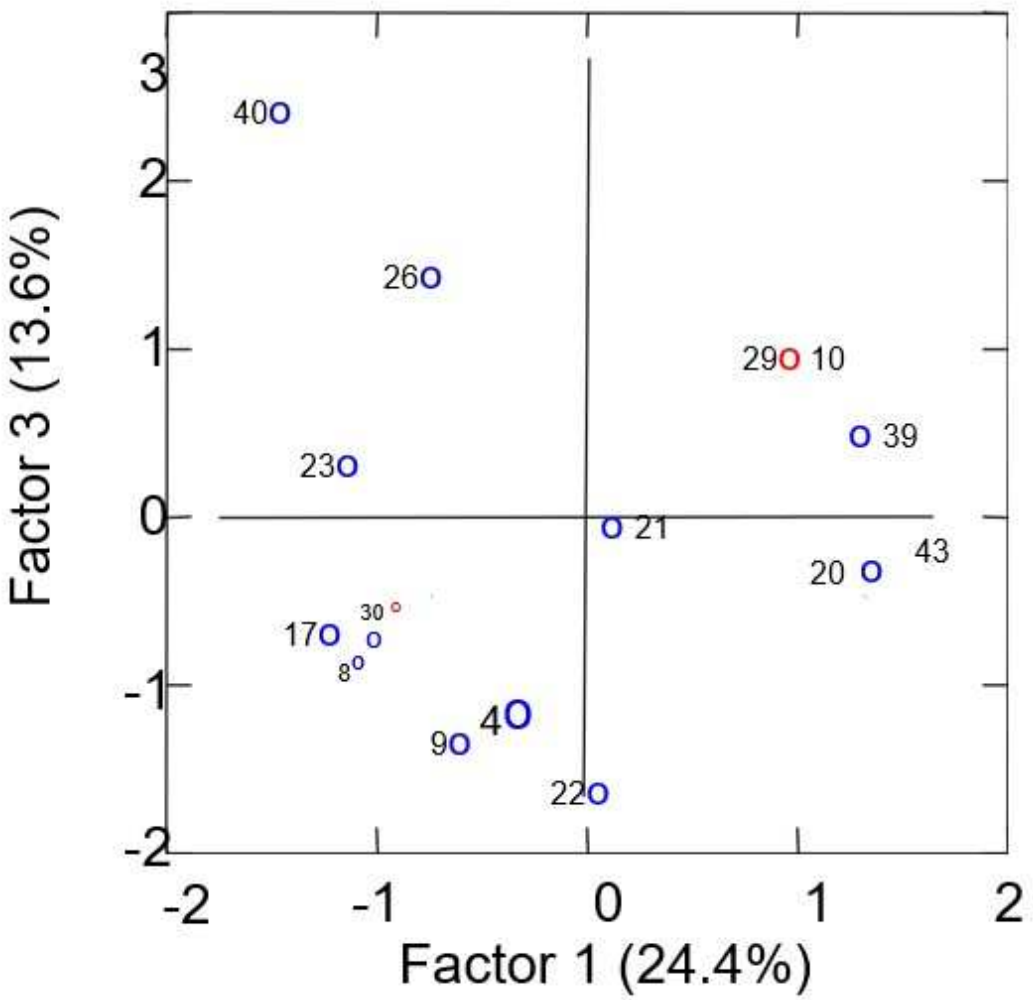


Figure 4b. Principal components analysis of metabolic and phenotypic characteristics depicting four outliers (21, 23, 26, 40) and two clusters (1, 4, 8, 9, 17, and 30 & 10, 15, 20, 39, and 43) of isolates. Blue plotted isolates signify close 16S rDNA relationship to *D. nishinomiyaensis*, while red plotted isolates are more closely related to *D. barathri*/*D. profundi.*

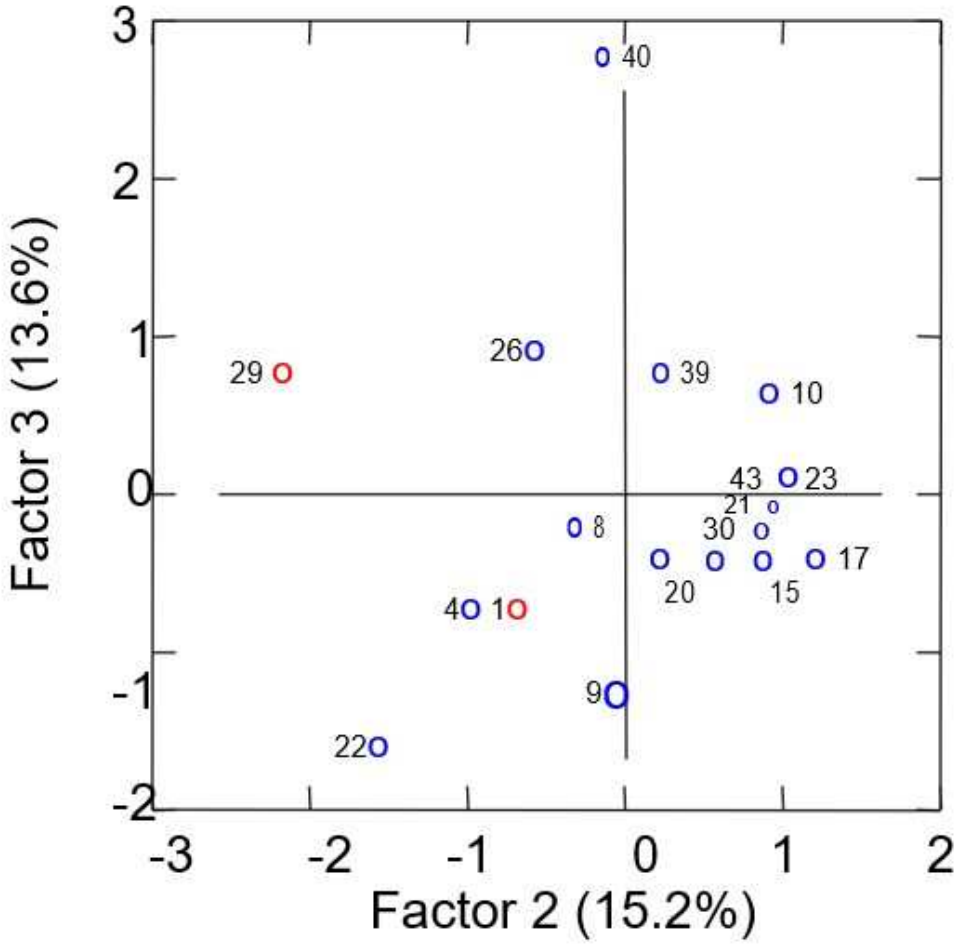


Figure 4c. Principal components analysis of metabolic and phenotypic characteristics depicting four outliers (22, 26, 29, 40). Two clusters of isolates were also present (1, 4, 8, 9, 17, 21, 23, and 30 & 10, 15, 20, 39, and 43). Blue plotted isolates signify close 16S rDNA relationship to *D. nishinomiyaensis*, while red plotted isolates are more closely related to *D. barathri*/*D. profundi.*

Reference		1	29	4	8	9	10	15	17	20	21	22	23	26	30	39	40	43	Total
101	T	C	C	X		X				X						X		X	2
102	C	T	T	X		X				X						X		X	2
144	C	T	T																2
149	G	A	A																2
152	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	17
153	T	A	A																2
155	C	T	T																2
162	A	T	T																2
177	T	A	A																2
182	G	T	T																2
240	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	17
502.1	:									G									1
534	C	T	T	T	T	T	T					T	T		T	T		T	11
557.1	:									C									1
602	G	A	A																2
781	A	G	G																2
892.1	:			A															1
970	C					A													1
973	C					T													1
1,078	T	G	G			G							G						4
1,080	C	G				A													2
1,080.1	:	A	A			A	A				N		A						6
1,081	G	T	T			A	A				N		A	N					7
1,082	G													N					1
1,084	T						G												1
1,089.1	:										N								1
1,090	C						N				N								2
1,190	T		C				X				X			X					1
1,208	C	T	T				X				X			X					2
1,322	C	X	X	X	X	X	X	X	T	X	X	X		X	X	X	X	X	1
Total		19	19	4	3	9	7	2	3	4	6	3	6	4	3	3	2	3	100

Figure 5. Sequencher variance report showing differences between isolates 16S rDNA when compared to type strain *Dermacoccus nishinomiyaensis* strain DSM20448. Nucleotide differences range from two to nineteen differences, with major differences at loci 152, 240, and 534.

Reference		1	29	4	8	9	10	15	17	20	21	22	23	26	30	39	40	43	Total
49	C			X	T	X	T	T	T	X	T	T	T	T	T	X	T	X	10
50	T			X	C	X	C	C	C	X	C	C	C	C	C	X	C	X	10
92	T			C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	15
97	A			G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	15
101	A			T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	15
103	T			C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	15
110	T			A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	15
125	A			T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	15
130	T			G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	15
450.1	:									G									1
482	T							C	C	C	C			C			C		6
505.1	:									C									1
550	A			G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	15
729	G			A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	15
840.1	:			A															1
918	C					A													1
921	C					T													1
1,026	G			T	T		T	T	T	T	T	T		T	T	T	T	T	13
1,028	G		C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	C	16
1,029	A			:	:			:	:	:	N	:		:	:	:	:	:	12
1,030	T			G	G	A	A	G	G	G	N	G	A	N	G	G	G	G	15
1,031	G													N					1
1,033	T						G												1
1,038.1	:										N								1
1,039	C						N				N								2
1,139	T		C				X				X			X					1
1,157	T			C	C	C	X	C	C	C	X	C	C	X	C	C	C	C	12
Total		0	2	15	16	14	16	17	17	17	18	16	14	17	16	14	17	14	240

Figure 6. Sequencher variance report showing differences between isolates 16S rDNA when compared to WCU1. Nucleotide differences range from two to eighteen differences, with major differences at loci 49, 50, 92, 97, 101, 103, 110, 125, 130, 550, 729, 1,026, 1,028, 1,030, and 1,157.

Appendix A: *Dermacoccus* sequences used in the analyses in this study.

>1

GTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCCTACTCTG
G
GATAAGCCTGGGAAACTGGGTCTAATACTGGATATGACCAATCACTGCATGGTGTGT
TGGTGGAAAGATTTTTTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGG

GTAATGGCCTACCAAGGCGACGACGGGTAGCGCCTGAGAGGGCGACCGGCCACACT
GGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA
ATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGT
AAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACCG
GCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAAT
TATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGC
TTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTATGGTAGGGGAGACTGG
AATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAG
GCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAGG
ATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGGCTCATT
CCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCC
GCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGC
GGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAATC
ATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCAG
CTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGT
TGCCAGCACGTGATGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAA
GGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTAC
AATGGCCGGTACAGAGGGTTGCGAAACTGTGAGGTGGAGCTAATCCCAAAAAACCG
GTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAA
TCGCAGATCAGCAACGCTGCGGT

>4

GGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGAG
TGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGT
GGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGC
CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT
TGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCG
GGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAA
GCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTC
CGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGAC
CGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGA
GACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATG
GCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCG
AACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGG
ACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGT
ACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGA
AGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACAC
CGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTG
TCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGT

>8

GGGTGAGTAACACGTGAGTAACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTG
GGTCTAATACTGGATACGACCGATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTG
GTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGG
CGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCT
GATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCA
GGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCA
GCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGA
GCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTG
CAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGC
GGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCC
ATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGT
AGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGC
CGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTC
AAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCA
ACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGC
GTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATG
TTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGT
GGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCA
A
ATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAGAGG
GTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAAAACCGGTCTCAGTTCGGATTGG
GGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACG
CTGCGGTG

>9

CTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGG
AGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTG
GTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCG
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTT
CGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAG
AAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG
TCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAG
ACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGG
GAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGA
TGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAG
CGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTG

GGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGA
GTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCG
GAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACA
CCGGAATCATGCAGAGATGTGTGAGTTTTCGGACTGGTGTACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
TTCCATGTTGCCAGCACGTAAAGGTGGGGACTCATGGGAGACTGCCGGGGTCAACT
CGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACG
CATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCA
AAAAACCGGTCTCAGTTCGGATTGGG

>10

CTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAGCACC
AGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCC
CTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCG
CATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCT
TGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCG
ACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGG
GAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGG
CCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAG
AAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCG
TTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGA
AAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTA
GGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACAC
CGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGG
GAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGT
GTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGG
GGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
GCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACAT
ACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCAT
G
GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
TCGTTCCATGTTGCCAGCACTTCAAGGGGGGGAN

>15

CTGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAGC
ACCAGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTG
CCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCT

CCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCA
GCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGG
GCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT
GGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATG
A
CGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTG
GAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGA
GCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCTG
TGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTG
GTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAA
CACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGG
GGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGT
GGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAG
TACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGC
ATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAAT
CATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCAGC
TCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCC
AGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGG
ATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGT
ACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAAAAC

>17

GAGTAACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATA
CGACCGATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCG
CGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCG
GCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCG
TGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGT
G
ACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACG
TAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTC
GCGTCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGA
CTAGAGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATA
TCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAA
GCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACG
TTGGGCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAA
GCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGG
CCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACC
AAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTAC
AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACT

GCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATG
TCTTGGGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGG
T
GGAGCTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCC
ATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAAT

>20

GGGTCTAATACTGGATACGACCGATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGT
GGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAG
GCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCC
TGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACC
AGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGGAATTATTGGGCGTAAA
GAGCTTGTAGGCGGTCTGTCGCGTCTGCTGTGAAAGACCCGGGGCTTAACTCCGGTTCT
GCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCG
GTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATT
ACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
ATGCCGTAAACGTTGGGCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCT
AACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATT
GACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAAC
CTTACCAAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGT
GTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGC
AACGAGCGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACT
GCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCT
TGGGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGC
TAATCCCAAAAAACCGGTCTCAGTTCGGAT

>21

GGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAGCA
CCAGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGC
CCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTC
CGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAG
CTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGG
CGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTG
GGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGAC
GGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGG
AGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAG
CGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCTGT

GAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGG
TAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAAC
ACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATG
GGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAG
GTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCT
GGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAG
CGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGAC
ATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGC
A
TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
CCTCGTTCCATGTTGCCAGCACTTCNNGGTGGGGANNTC

>22

ACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACC
GATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCC
TATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTG
AGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
AGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACG
GTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAG
GGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCG
TCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTA
G
AGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAG
GAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGA
AAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGG
GCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGC
CCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGG
CTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAG
CGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCG
GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTG
GGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAG
CTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACC

>23

TGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCTCACTCTGG
G
ATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGAGTGTT
GGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGGG
TAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCAC
AATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTG
TAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACC
GGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAA
TTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGG
CTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGAGACTG
GAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAA
GGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAG
GATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGACTCAT
TCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGC
CGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATG
CGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAAT
CATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATG
TTGCCAGCACGTCAAGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGA
AGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTA
CAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAAAACC
GGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTA
ATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGG

>26

TCTGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAG
CACCAGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCT
GCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATC
TCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATC
AGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAG
GGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAG
TGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATG
ACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCT
GGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCG
AGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCT
GTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGT
GGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGA
ACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCA
TGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCT
AGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGC

CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACA
AGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTG
ACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGT
G
CATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCA
ACCCTCGTTCCATGTTGCCAGCACTTCNNGTGGGGA

>29

GAGTAACCTGCCCCCTACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATA
TGACCAATCACTGCATGGTGTGTTGGTGGAAAGATTTTTTGGTGGGGGATGGACTCG
CGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCG
GCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCG
TGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGT
G
ACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACG
TAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTC
GCGTCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGA
CTAGAGTATGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATA
TCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAA
GCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACG
TTGGGCGCTAGGTGTGGGGCTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAA
GCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGG
CCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACC
AAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTAC
AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTCGTTCCATGTTGCCAGCACGTCATGGTGGGGACTCATGGGAGA
CTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTA
TGTCTTGGGCTTCACGCATGCTACAATGGCCGGCACAGAGGGTTGCGAAACTGTGA
G
GTGGAGCTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCC
CATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGC

>30

ACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACC
GATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCC
TATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTG

AGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
AGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACG
GTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAG
GGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCG
TCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTA
G
AGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAG
GAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGA
AAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGG
GCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGC
CCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGG
CTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAG
CGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCG
GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTG
GGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAG
CTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCA

>39
CTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGG
AGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTG
GTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCG
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTT
CGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAG
AAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG
TCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAG
ACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGG
GAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGA
TGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAG
CGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTG
GGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGA
GTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCG
GAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACA
CCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG

TTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCG
GAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCA
TGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAA
AAACCGGTCTCAGTTCGGATTGGG

>40

TGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCTCACTCT
GGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGAG
TGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGT
GGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGC
CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT
TGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCG
GGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAA
GCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTC
CGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCTGTGAAAGAC
CGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGA
GACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATG
GCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCG
AACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGG
ACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGT
ACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGA
GCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACACC
GGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGT
CGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTT
CCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGG
AGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCAT
GCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAA
AACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCT
AGTAATCGCAGATCAGCAACGCTGCGGTGAA

>43

TGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGA
GTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGG
TGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGG
CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTT
CGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAG

AAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG
TCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAA
GACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGG
GGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACC
GATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGG
GAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGT
GTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTG
GGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACA
TACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCA
TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAA
CCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGT
CAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTC
ACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCC
AAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAA

Appendix B: GenBank sequences with accession numbers for the isolates from this study.

>MK880152

GTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCCTACTCTG
G
GATAAGCCTGGGAAACTGGGTCTAATACTGGATATGACCAATCACTGCATGGTGTGT
TGGTGGAAAGATTTTTTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGG
GTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
CAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTT
GTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCAC
CGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA
ATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGG
GCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTATGGTAGGGGAGACT
GGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGA
AGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACA
GGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGGCTCA
TTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGG
C
CGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATG
CGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAAT
CATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATG
TTGCCAGCACGTGATGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGA
AGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTA
CAATGGCCGGTACAGAGGGTTGCGAAACTGTGAGGTGGAGCTAATCCCAAAAAACC

GGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTA
ATCGCAGATCAGCAACGCTGCGGTGA

> MK880153

GGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGAG
TGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGT
GGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGC
CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATA
TTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTC
GGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGA
AGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGT
CCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAG
ACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGG
GAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGA
TGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAG
CGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTG
GGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGG
AGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGC
GGAAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATA
CACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATG
GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
TCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAA
CTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCA
CGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCC
CAAAA AACCGGTCTCAGTTCGGATTGGGG

> MK880154

GGGTGAGTAACACGTGAGTAACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTG
GGTCTAATACTGGATACGACCGATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTG
GTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGG
CGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCT
GATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCA
GGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCA
GCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGA
GCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTG
CAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGC
GGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCC
ATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGT
AGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGC

CGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTC
AAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCA
ACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGC
GTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATG
TTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGT
GGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCA
A
ATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAGAGG
GTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAAAACCGGTCTCAGTTCGGATTGG
GGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACG
CTGCGGTGAA

> MK880155

CTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGG
AGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTG
GTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCG
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTT
CGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAG
AAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG
TCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAG
ACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGG
GAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGA
TGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAG
CGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTG
GGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGG
AGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGC
GGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATAC
ACCGGAATCATGCAGAGATGTGTGAGTTTTCGGACTGGTGTACAGGTGGTGCATGG
TTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
CGTTCCATGTTGCCAGCACGTAAAGGTGGGGACTCATGGGAGACTGCCGGGGTCAA
CTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCA
CGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCC
CAAAAAACCGGTCTCAGTTCGGATTGGGGT

> MK880156

CTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAGCACC
AGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCC
CTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCG
CATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCT

TGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCG
ACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGG
GAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGG
CCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAG
AAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCG
TTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGA
AAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTA
GGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACAC
CGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGG
GAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGT
GTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGG
GGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
GCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACAT
ACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCAT
G
GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
TCGTTCCATGTTGCCAGCACTTCAAGGGGGGGA

> MK880157

CTGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAGC
ACCAGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTG
CCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCT
CCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCA
GCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGG
GCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT
GGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATG
ACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCT
GGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCG
AGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCT
GTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGT
GGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGA
ACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCA
TGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCT
AGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGC
CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACA
AGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTG
ACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGT
GCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGC
AACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGG
GTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGG
GCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGC
TAATCCCAAAAAACCG

> MK880158

GAGTAACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATA
CGACCGATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCG
CGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCG
GCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCG
TGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGT
G
ACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACG
TAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTC
GCGTCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGA
CTAGAGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATA
TCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAA
GCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACG
TTGGGCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAA
GCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGG
CCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACC
AAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTAC
AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACT
GCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATG
TCTTGGGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGG
T
GGAGCTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCC
ATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATAT

> MK880159
GGGTCTAATACTGGATACGACCGATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGT
GGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAG
GCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCC
TGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACC
AGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGGAATTATTGGGCGTAAA
GAGCTTGTAGGCGGTCTGTCGCGTCTGCTGTGAAAGACCCGGGGCTTAACTCCGGTT
CTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGAGACTGGAATTCCTGGTGT
AGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTG
GGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCC
TGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGACTCATTCCACGAGTTCC
GTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAA

ACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGA
TGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAATCATGCAGAGATGT
GTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGA
GATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACTTC
GGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGAC
GTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTAC
AGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAAAACCGGTCTCAGTTCGG
ATTG

> MK880160

GGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAGCA
CCAGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGC
CCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTC
CGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAG
CTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGG
CGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTG
GGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGAC
GGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGG
AGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAG
CGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCTGT
GAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGG
TAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAAC
ACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATG
GGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAG
GTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCT
GGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAG
CGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGAC
ATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGC
A
TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
CCTCGTTCCATGTTGCCAGCACTTC

> MK880161

ACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACC
GATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCC
TATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTG
AGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
AGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACG
GTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAG
GGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCG

TCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTA
G
GTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAG
GAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGA
AAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGG
GCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGC
CCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGG
CTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAG
CGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCG
GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTG
GGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAG
CTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCC

> MK880162

TGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCTCACTCTGG
G
ATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGAGTGTT
GGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGGTGGGG
TAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCAC
AATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTG
TAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGAAGCACC
GGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAA
TTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGG
CTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGGAGACTG
GAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAA
GGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAG
GATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGGGACTCAT
TCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGC
CGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATG
CGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAAT
CATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCCATG
TTGCCAGCACGTCAAGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCGGAGGA
AGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCATGCTA
CAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAAAAACC
GGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTA
ATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGC

> MK880163

TCTGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGATGAAG
CACCAGCTTGCTGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCT
GCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATC
TCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATC
AGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAG
GGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAG
TGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATG
ACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCT
GGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCG
AGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCT
GTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGT
GGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGA
ACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCA
TGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCT
AGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGC
CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACA
AGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTG
ACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGT
G
CATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCA
ACCCTCGTTCCATGTTGCCAGCACTTC

>MK880164

GAGTAACCTGCCCCCTACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATA
TGACCAATCACTGCATGGTGTGTTGGTGGAAAGATTTTTTGGTGGGGGATGGACTCG
CGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCG
GCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCG
TGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGT
G
ACGGTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACG
TAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTC
GCGTCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGA
CTAGAGTATGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATA
TCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAA
GCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACG
TTGGGCGCTAGGTGTGGGGCTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAA
GCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGG
CCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACC

AAGGCTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTAC
AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTCGTTCCATGTTGCCAGCACGTCATGGTGGGGACTCATGGGAGA
CTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTA
TGTCTTGGGCTTCACGCATGCTACAATGGCCGGCACAGAGGGTTGCGAAACTGTGA
G
GTGGAGCTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCC
CATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGG

> MK880165

ACCTGCCCCTCACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACC
GATCTCCGCATGGAGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCC
TATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTG
AGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
AGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACG
GTACCTGGAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAG
GGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCG
TCTGCTGTGAAAGACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTA
GGTGTGGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAG
GAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGA
AAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGG
GCGCTAGGTGTGGGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGC
CCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGG
CTTGACATACACCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAG
CGCAACCCTCGTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCG
GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTT
GGCTTCACGCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAG
CTAATCCCAAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATG

> MK880166

CTGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGG
AGTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTG
GTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCG
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTT
CGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAG
AAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG

TCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAG
ACCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGG
GAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGA
TGGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAG
CGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTG
GGACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGG
ATACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCG
GAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACA
CCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
TTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCG
GAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCA
TGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAA
AAACCGGTCTCAGTTCGGATTGGGGT

> MK880167

TGGTGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCCTCACTC
TGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGA
GTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGG
TGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGG
CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTT
CGGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAG
AAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG
TCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTCTGTCGCGTCTGCTGTGAAAG
ACGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGG
AGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGAT
GCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGC
GAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGG
GACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGA
GTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCG
GAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACA
CCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTC
GTTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAACT
CGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCAC
GCATGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCC
AAAAAACCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAG
TCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATA

> MK880168

TGGGATAAGCCTGGGAAACTGGGTCTAATACTGGATACGACCGATCTCCGCATGGA
GTGTTGGTGGAAAGTTTTTGTGGTGGGGGATGGACTCGCGGCCTATCAGCTTGTTGG
TGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGG
CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATA
TTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTC
GGGTTGTAAACCTCTTTCACCAGGGACGAAGCTAACGTGACGGTACCTGGAGAAGA
AGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGT
CCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGA
CCGGGGCTTAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGTGGTAGGGG
AGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGAT
GGCGAAGGCAGGTCTCTGGGCCATTACTGACGCTGAGAAGCGAAAGCATGGGGAGC
GAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCGCTAGGTGTGG
GACTCATTCCACGAGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGA
GTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCG
GAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATACA
CCGGAATCATGCAGAGATGTGTGCGTCTTCGGACTGGTGTACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
TTCCATGTTGCCAGCACTTCGGGTGGGGACTCATGGGAGACTGCCGGGGTCAACTCG
GAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTTGGGCTTCACGCA
TGCTACAATGGCCGGTACAGAGGGTTGCGAAACCGTGAGGTGGAGCTAATCCCAAA
AAACCGGTCTCAGTTCGGATTGGGGTCTGCAACT