BIOFERTILIZERS FOR SUSTAINABLE AGRICULTURE: ISOLATION AND GENOMIC CHARACTERIZATION OF NITROGEN-FIXING BACTERIA FROM SUGARCANE

A Dissertation
Presented to
The Academic Faculty

by

Luz Karime Medina Cordoba

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biology in the
School of Biological Sciences

Georgia Institute of Technology
May 2020

BIOFERTILIZERS FOR SUSTAINABLE AGRICULTURE: ISOLATION AND GENOMIC CHARACTERIZATION OF NITROGEN-FIXING BACTERIA FROM SUGARCANE

Approved by:

Dr. Joel E. Kostka, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. I. King Jordan, Co-advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Frank Stewart
School of Biological Sciences
Georgia Institute of Technology

Dr. Jung H. Choi
School of Biological Sciences
Georgia Institute of Technology

Dr. Leonard W. Mayer
School of Medicine
Emory University

Date Approved: December 10, 2019

To my family and friends

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my advisor, Joel Kostka, for the opportunity to complete my PhD under his guidance. I am extremely grateful for your contributions to my research, and the experiences in your laboratory, where I learned principles that have made me a better microbiologist.

To my co-advisor, King Jordan, thank you for believing in me from the start of the PhD program. Your guidance through each part of this five-year journey has contributed to my growth as a scientist. I will always be grateful for your patience and dedicated effort in educating me on bioinformatics. Your love and dedication to science and your students have made a lasting impression on me.

I want to thank my committee members, Dr. Frank Stewart, Dr. Leonard Mayer, and Dr. Jung H. Choi; your support, comments, and suggestions have been fundamental for my research and have helped me to become a better scientist.

It would have been impossible to complete my research without my colleges and friends from the lab. Aroon Chande and Dr. Lavanya Rishishwar were extraordinary instructors during the last years of my studies. To Shashwat Deepali Nagar and Héctor Espitia, thank you for your friendship and all your help during my time in the laboratory.

I would like to thank my country, Colombia, for providing resources to help further my education. I hope to one day make a difference in the lives of every Colombian through microbiology. Finally, thank you to my family, who have been the catalyst for my motivation to grow as a person and as a professional. To my parents, Dr. Miguel Medina and Professor Mabel Cordoba, thank you for your unconditional love and for teaching me early on to love science, and more importantly, to respect others. To my brother Miguel David Medina for being my best friend and always giving me good advice. And to Kai Parham for always being there when I needed encouragement and the strength to continue.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ACC	1-aminocyclopropane-1-carboxylic acid
ANI	Average Nucleotide Identity
AMR	Antimicrobial Resistance
BLAST	Basic Local Alignment Search Tool
CGView	Circular Genome Viewer
IAA	Indole Acetic Acid
NF	Nitrogen Fixation
NCBI	National Center for Biotechnology Information
PAIs	Pathogenicity Islands
OM	Organic Matter
PCR	Polymerase Chain Reaction
PGAP	Prokaryotic Genome Annotation Pipeline
PGP	Plant Growth Promoting
RAST	Rapid Annotations Using Subsystems Technology
RISA	Ribosomal Intergenic Spacer Analysis
VFDB	The Virulence Factors Database
VF	Virulence Factor

SUMMARY

The rapid increase of the world's population has intensified the global demand for food production. In order to meet the food needs of the growing human population, farmers worldwide have been enhancing crop production via the intensive application of chemical fertilizers, which pose a severe threat to both the natural environment and the health of human populations. Accordingly, there is a pressing need to research and develop more environmentally sustainable agricultural practices that mitigate the environmental damage caused by indiscriminate use of chemical fertilizers. The goal of my thesis research was to explore the use of bacterial biofertilizers, as an alternative or complementary approach to chemical fertilizers, in support of more sustainable agricultural practices.

Biofertilizers are made up of living microorganisms that promote plant growth by increasing the availability of essential nutrients to the plants with which they are associated. Bacterial biofertilizers promote plant growth by providing bioavailable nitrogen and phosphorus, among other nutrients, via the processes of nitrogen fixation and phosphorus solubilization. My work is focused on the discovery and characterization of native nitrogen-fixing that are associated with sugarcane crops cultivated in the Cauca Valley of Colombia. I hypothesized that native nitrogen-fixing bacteria, found in association with local sugarcane crops, could serve as potent biofertilizers with the potential to simultaneously increase crop yield while reducing the reliance of chemical fertilizers. I evaluated this hypothesis by isolating nitrogen-fixing bacteria from Colombian sugarcane fields and characterizing their plant growth-promoting (PGP) potential using an integrated computational genomic and experimental approach. I focused on native bacterial

biofertilizers for several reasons: (1) they are less likely than non-native, and potentially invasive, species to pose a threat to the local environment, (2) they are more likely to effectively promote plant growth of the sugarcane crops with which they are already associated, and (3) they serve as a potentially renewable resource, in terms of bacteria can be locally sourced, cultivated and then re-applied to crops.

My research on nitrogen-fixing biofertilizers was focused on four specific aims: 1. Isolation of nitrogen-fixing bacteria from Colombian sugarcane fields, 2. Genomic characterization and computational phenotyping of the nitrogen-fixing bacterial isolates, 3. Experimental validation of the computationally predicted PGP properties of the nitrogen-fixing bacteria, and 4. Development of a nitrogenase (*nifH*) gene database to facilitate the taxonomy assignment and analysis of nitrogen-fixing bacteria. Here, I briefly summarize the results for each of my research aims.

The first aim of my research was to isolate putative plant growth-promoting bacterial biofertilizers from Colombian sugarcane fields and to characterize their genome sequences. Samples of leaves, rhizosphere soil, stem, and roots were collected from sugarcane fields in the Cauca Valley of Colombia as described in Chapter 2. A high-throughput cultivation approach was developed and applied for the enrichment and isolation of nitrogen-fixing bacteria from these environmental samples. Pure cultures of nitrogen-fixing bacteria were isolated and tested for diazotrophic potential by PCR amplification of *nifH* genes, a common molecular marker for nitrogen fixing capacity, and twenty-two distinct *nifH* positive isolates were selected for genome sequencing and analysis. Isolate genome sequences were characterized using next generation sequencing, and the resulting genome sequence assemblies and their annotations were made available in the NCBI Genbank database (BioProject PRJNA418312).

The second aim of my research was to develop a genome-enabled approach for the prioritization of native bacterial isolates with the potential to serve as biofertilizers for sugarcane fields in Colombia. I refer to the approach that I developed as computational phenotyping, which entails the prediction of specific organismal phenotypes, or biochemical abilities, based on the analysis of functionally annotated whole-genome sequences. Genome sequence analysis confirmed the presence of intact *nifH* genes and operons in the genomes of 18 of the isolates, as described in Chapter 3. Isolate genomes were also found to encode operons for phosphate solubilization, siderophore production, and other PGP phenotypes. I characterized 14 of the 22 nitrogen-fixing isolates as distinct strains of *Klebsiella pneumoniae*, and four others were members of genera that are closely related to *Klebsiella*. The quantitative approach to computational phenotyping that I developed and applied to sugarcane bacterial isolates facilitates the screening for strains that have a high potential for nitrogen fixation and other PGP phenotypes while showing minimal risk for virulence and antibiotic resistance.

The third aim of my research was to experimentally evaluate the computationally predicted PGP properties of the prioritized bacterial biofertilizer isolates. Bioinformatic predictions of the presence of PGP traits were validated with a series of experimental laboratory assays. The top six prioritized strains were evaluated for the following PGP phenotypes: nitrogen fixation, phosphate solubilization, and the production of siderophores, gibberellic acid, and indole acetic acid, as described in Chapter 3. Results from the biochemical assays were consistent with the computational phenotype predictions for these isolates, in support of their PGP potential.

The fourth aim of my research was to develop a *nifH* gene database to facilitate the taxonomy assignment and diversity analysis of environmental isolates of nitrogen-fixing bacteria. The *nifH* gene codes for a critical, and highly conserved, component of the nitrogenase enzyme complex that catalyzes the conversion of atmospheric nitrogen to bioavailable ammonium during nitrogen fixation. Accordingly, *nifH* is a molecular marker that is widely used to characterize the biochemical capacity and taxonomic composition of diazotrophic microorganisms in environmental samples. As described in Chapter 4, I developed and validated an automated method to curate and continuously update a *nifH* reference sequence database that can be used to support the characterization of metagenomic or environmental amplicon sequences. This approach yielded a far smaller, but much more reliable, set of ‘gold-standard’ *nifH* sequences, against which users can compare their metagenomic or amplicon sequence data for accurate functional prediction and taxonomic characterization of diazotrophs in environmental samples.

CHAPTER 1. INTRODUCTION: BIOFERTILIZERS FOR SUSTAINABLE AGRICULTURE

1.1 Sustainable agriculture

Sustainable agriculture is the implementation of environmentally friendly technologies for crop production (1). The dominant approach in contemporary agriculture entails the intensive use of chemical fertilizers to meet the food demands of the growing human population. These chemical fertilizers provide critical macronutrients to crops – such as nitrogen (N), phosphorus (P), potassium (K), and sulfur (S) – and are widely used to maximize agricultural yield (2). The application of chemical fertilizers represents a high cost for agricultural companies and also contributes to environmental damage, such as air pollution through the formation of microparticles, soil depletion, and water pollution via run-off (3). The main goal of sustainable agriculture is the development of environmentally friendly cropping practices that can support increased crop production while minimizing damage to the environment and threats to human populations. The use of beneficial bacteria that stimulate plant growth, *i.e.* biofertilizers, represents a particularly promising approach to sustainable agriculture (1). It has already been shown that biofertilizers are superior to chemical fertilizers for a variety of crops (4, 5).

1.2 Biological fertilizers (biofertilizers)

Biological fertilizers (biofertilizers) are made up of microbial inoculants containing beneficial bacteria that promote plant growth, thereby representing an alternative or complementary approach for increasing crop yield that is more sustainable and environmentally friendly (5). Biofertilizers augment plant growth by facilitating nutrient acquisition and stimulating hormone production, and they can prevent pathogen attacks by competitive exclusion; they are also a potentially affordable source of nutrients, since they can be locally sourced from the environment (6).

1.3 Plant growth-promoting bacteria (biofertilizers) and their role in sustainable agriculture

Soil microbiomes comprise one of the most significant reservoirs of biological (biochemical) diversity on the planet, and a subset of these microbial communities are intimately associated with plants (7). The rhizosphere is a micro-ecological zone in soils that is strongly influenced by plant roots, containing up to 10^{11} microbial cells per gram of root tissue (8), and more than 30,000 prokaryotic species (9, 10). The rhizosphere functions as an interface that facilitates intense plant-soil-microorganism interactions (11). With respect to sustainable agriculture, plant-microbiome interactions in the rhizosphere, termed the rhizobiome, represent an essential set of processes that substantially impact plant health and growth by transformation, mobilization, and solubilization of nutrients that are indispensable for plant development (12).

Plant growth-promoting (PGP) bacteria live near plant roots, on root and leaf surfaces, and within plant tissues as leaf or root endophytes (13). These microbes are also free-living in the soil, where they support the uptake of mineral nutrients essential to crop production (14). In agricultural soils, macronutrients are often present in insufficient amounts or are unavailable to plants. PGP bacteria make nutrients available to plants through nitrogen fixation and phosphate solubilization. PGP bacteria also support plant growth through production of phytohormones or antimicrobial compounds as well as induction of systemic resistance (15).

1.3.1 Mechanisms of bacterial plant growth promotion

1.3.1.1 Biological nitrogen fixation

Biological nitrogen fixation (BNF) is a microbially mediated process whereby atmospheric nitrogen (N_2) is reduced into ammonia (NH_3) in the presence of the nitrogenase enzyme (16). Nitrogen is a fundamental, limiting nutrient for plant growth that is often found in biochemically inaccessible forms in the environment. Nitrogen forms part of the nitrogenous bases and amino acids that are the building blocks of nucleic acids and proteins and is also part of the chlorophyll molecules that produce energy through photosynthesis (17). BNF is carried out by a broad diversity of nitrogen-fixing bacteria referred to as diazotrophs. Some diazotrophs can fix N_2 in the free-living state, while others perform this process in tight physical association with plants, in true symbiotic associations (14). Additional details on BNF are given in Chapter 4.

Nitrogen fixation is an exclusively microbial process carried out by diverse phylogenetic groups of prokaryotes. Diazotrophs represent a vast range of metabolic life styles, including aerobic (for example, *Azotobacter*), facultatively anaerobic (*Klebsiella*), or anaerobic (*Clostridium*) heterotrophs, anoxygenic (*Rhodobacter*) or oxygenic (*Anabaena*) phototrophs, and chemolithotrophs (*Leptospirillum ferrooxidans*) (18).

1.3.1.2 Phosphate solubilization

Phosphorous (P) is a required macronutrient that is frequently limited in the soil. For example, phosphate is a crucial part of the nucleic acid structure in plants, participates in the regulation of protein synthesis (19), and is required for cell division and energy transformation. Importantly, phosphate also helps plants convert other nutrients into functional building blocks for growth. Even though most agricultural soils contain a vast amount of inorganic and organic P, most of this nutrient is chemically unavailable for plants. Even after adding P to agricultural soils, plants may not be able to have access to P, since it is immobilized and insoluble, and therefore unavailable to plants (20).

Phosphate solubilizing bacteria such as *Pseudomonas*, *Bacillus*, *Azotobacter*, and *Bradyrhizobium* are capable of solubilizing inorganic phosphorus from insoluble compounds in the soil (21). *Actinomycetes* have the metabolic capacity to solubilize P-metal complexes to release bioavailable P in the form of orthophosphate through a variety of mechanisms involving organic acids, siderophore production, and phosphatase enzymes that play essential roles in hydrolyzing organic P forms (22).

1.3.1.3 Siderophore production

Siderophores are iron-chelating compounds that aid in iron assimilation. Most agricultural soils contain sufficient iron for plant growth (23). However, this iron is largely present in an insoluble form as Fe (III) precipitates. In the absence of assimilable iron, microorganisms developed low molecular weight siderophores as iron-chelating or complexing agents (23). Siderophores have a high affinity for ferric ion (24). Therefore, siderophores produced in PGP bacteria can bind the insoluble iron and mobilize it within the soil, facilitating iron uptake through root systems (23).

1.3.1.4 Phytohormone production

Auxins, cytokinins, and gibberellins are phytohormone signaling molecules generated within the plant that are involved in the regulation of all aspects of plant growth, development, and response to biotic stress. Indole-3-acetic acid (IAA) is a common phytohormone from the auxin group (25). This hormone directly contributes to plant physiological processes such as cell enlargement and division, tissue differentiation, and responses to light and gravity (26). Bacteria such as *Pseudomonas*, *Rhizobium*, *Bradyrhizobium*, *Agrobacterium*, *Enterobacte*r, and *Klebsiella* can produce IAA.

Gibberellic Acid (GA3) is a phytohormone that regulates a number of different developmental processes in plants. GA3 activates the cells responsible for seed germination in grains to produce mRNA molecules that code for hydrolytic enzymes (27). GA3 breaks seed dormancy, enhances stem growth, and is a secondary metabolite synthesized by bacteria as a signaling factor to the host plant. Some GA3 producing bacteria include *Bacillus*, *Azotobacte*r, and *Rhizobium* (27).

1.4 Sugarcane as a global crop for the study of plant growth-promoting bacteria

Sugarcane is a tall, perennial grass cultivated in tropical and warm temperate regions around the world, which is capable of producing high concentrations of sugar (sucrose) along with a diverse set of organic byproducts (28, 29). Currently, sugarcane is one of the most economically profitable crops for the production of bio-ethanol used as for automotive fuel (replacing gasoline) (30). Along with corn, rice, and wheat, sugarcane is one of the intensively cultivated crops around the world. Brazil, India, and China are the primary producers of sugarcane worldwide, with Brazil being the top sugarcane producer in the world (31). Sugarcane production is crucial to the economy of Colombia, and the Cauca Valley is the main sugarcane producing region in the country.

Sugarcane crops are made up of several species of the genus *Saccharum*, including *S. officinarum*, *S. robustum*, *S. sinense*, *S. edule*, *S. spontaneum* (28). Previous studies suggested an optimal hybridization of two species to yield the most productive and stable sugarcane crops: *S. officinarum* and *S. spontaneum* (32). Commercial sugarcane is a hybrid that comes from the selective breeding process between *S. officinarum*, which has high amounts of sucrose in the stem but has weak disease resistance, and *S. spontaneum*, which has resistance to diseases. (33, 34).

Sugarcane cultivation differs from other crops because sugarcane is propagated vegetatively through cane setts (stalks or shorter stem segments), which are planted when new production is required or for re-establishing an existing sugarcane field when productivity is low (35, 36). Sugarcane harvests take place nine to eighteen months after planting the cane setts. Then, sugarcane can be continually harvested several times, as new stalks, ratoons, continuously grow from the stubble (*i.e.* the part of the plants left underground after the harvest) (37). Sugarcane is a perennial crop that yields substantial biomass but requires large amounts of moisture, nutrients, and light for optimal growth efficiency (38, 39). The fertilization process is a crucial aspect of sugarcane nutrient management and production, being responsible for up to 50% yield intensification (38). Nitrogen, phosphorus, and potassium are the essential limiting nutrients for yield, with nitrogen as the primary nutrient that influences sugarcane quality (40).

1.4.1 Plant growth-promoting (PGP) bacteria isolated from sugarcane

Previous studies have shown that the sugarcane microbiome harbors diverse PGP microorganisms, especially nitrogen-fixing bacteria, which have the potential to serve as biofertilizers (41). Known nitrogen-fixing bacterial genera such as *Azospirillum*, *Azotobacter*, *Beijerinckia*, *Enterobacter*, *Klebsiella*, and *Herbaspirillum* have been isolated and characterized from sugarcane in different areas in the world (42, 43). The diversity of PGP bacteria associated with sugarcane varies depending on factors such as location, sugarcane variety, agricultural practices, and the techniques used to detect their metabolic activity (43-45). The most common methods for the detection of plant growth-promoting bacteria are described below.

1.4.2 Cultivation based studies

The isolation, cultivation, and experimental characterization of representative plant-associated microbes are crucial to understanding the physiological mechanisms that underlie growth promotion (46-48). Previous studies reported on the contributions of beneficial bacteria and microbial communities in plant development when performing culture-based experiments under controlled laboratory conditions (49, 50). Cultures of PGP bacteria isolated from sugarcane crops and soils have previously been shown to possess beneficial properties such as nitrogen fixation, phosphate solubilization, and phytohormone production (51-53). Nevertheless, it has been proposed that traditional methods of isolation of plant-associated bacteria using culture-based methods may be of limited use in the detection of novel species as well as the discovery of new metabolic pathways involved in plant growth (54). This is due to the fact that the majority of microorganisms in the environment remain uncultivated. The research reported here is focused on novel discoveries through advances facilitated by genome-enabled technologies, which have been shown to uncover unknown dimensions of plant microbiome interactions (45).

1.4.3 Microscopy based studies

When studying the mechanisms employed by PGP bacteria to colonize and promote the growth of plants, microscopy plays a significant role. Fluorescence microscopy, electron microscopy, bright field microscopy, using staining methods such us FISH (fluorescence *in situ* hybridization) and GFP (green fluorescence protein), are traditional methods used to study endophytic bacteria in sugarcane (55-57). For instance, previous microscopy studies have uncovered physical interactions between *Enterobacter* sp. UYSO10 and *Shinella* sp. UYSO24 in commercial sugarcane crops from Uruguay. This approach showed that both free-living isolates are endophytes serve to promote plant growth in sugarcane (58).

1.4.4 Genomic based studies

The majority of previous cultivation-independent research employed PCR amplification and sequencing of taxonomic genes, SSU rRNA genes, to investigate the microbial communities that are associated with sugarcane crops (39, 59, 60). Current genomic techniques such as whole-genome sequencing, metagenomics, and transcriptomics build on this previous work to elucidate bacterial gene function in the study of plant microbiome interactions (45). For example, previous research (61) elucidated the molecular mechanisms and the changes in gene expression profiles of *Burkholderia* Q208 associated with sugarcane. My research focused on the use of genome-enabled technologies to make predictions regarding the suitability of sugarcane bacterial isolates as potential biofertilizers. The motivation behind this approach is to use genomic technologies to save time, effort, and cost for downstream experimental and field studies by prioritizing the most promising biofertilizer strains. The genomic approach is also distinguished by its ability to make negative predictions, which can allow researchers to avoid the study, manipulation and environmental introduction of strains that encode antimicrobial resistance, virulence factors, or other undesirable characteristics.

1.5 The most common strategies for the application of biofertilizers in sugarcane, and the impact of these strategies on plant growth and crop yield

Even though genome sequence based technologies can be used to predict the PGP potential, understand gene expression, and predict the metabolic activity of PGP bacteria, there is still much work that remains to be done once studies of this kind are completed to ensure the maximum impact of biofertilizers that will be applied to crops. For example, finding the right formulation, *i.e.* the specific mix of PGP bacteria and other components, for biofertilizers will be essential for the application of cultivable microorganisms to crops (62). The isolation and biochemical characterization of potential PGP bacteria isolated from natural environments are vital for the classification and formulation of biofertilizer inocula (63). Some crops need higher or different concentrations among the required nutrients to meet the plant mineral nutrition needs. For instance, soil fertilization is the most critical part of sugarcane cultivation because it generates up to 50% yield increase (38). Sugarcane needs suitable quantities of nutrients such as N, P, and K to produce taller sugarcane and higher sugar levels, but nitrogen is the principal nutrient responsible for sugarcane yield and quality (38).

The success of biofertilizers depends on multiples factors such as the specific species of cultivar, soil physicochemical properties, and pathogenic microbiota present in the soil (6). In sugarcane, numerous techniques have been developed for the application of biofertilizers. These include direct contact of bioinocula with the seed coat, seedlings, or applying biofertilizers directly to the soil (64). Previous studies in Brazil reported an increase of up to 50% in sugarcane root dry mass when cane setts were inoculated with a consortium of 5 nitrogen-fixing bacteria under greenhouse conditions (44). Recent studies have also reported an increase in sugarcane productivity by spraying biofertilizer directly onto ratoons in sugarcane fields in Brazil (65).

1.6 Challenges to the use of biofertilizers

PGP bacteria are already widely employed to formulate biofertilizers; nevertheless, their application in agriculture systems is still challenged by a variety of factors, such as unpredictable results after biofertilizer application, difficulties in tracking microorganisms in the field, a lack of understanding of plant microbiome interactions, and deficiencies in agriculture practices (66). In the past ten years, more and more studies have reported on the evaluation of PGP bacteria intended to be used as biofertilizers (67). However, these studies usually do not use tracking methods to study the actual biochemical activities of biofertilizers in their natural environment (68). The use of tracking methods for biofertilizer studies is limited by the availability of few effective, low-cost, and rapid methods for evaluating the activity of inoculated PGP bacteria. For this reason, the outcomes of biofertilizer in field studies are usually assessed by simply measuring plant biomass (69). Thus, it is possible that the PGP properties of biofertilizers may not be directly related to their predicted biochemical capacity, *e.g.* their ability to provide fixed nitrogen or solubilized phosphate to plants.

1.6.1 Biofertilizers for sustainable sugarcane agriculture in Colombia

INCAUCA is a Colombian agribusiness located in the Cauca Valley, which is dedicated to developing a wide variety of products and services derived from sugarcane. This company has 54 years of experience and is a major driver of the regional economy, making significant contributions to the development of the country by supporting the production of food, energy, and fuel (ethanol) along with a variety of organic by-products. INCAUCA uses chemical fertilizers such as urea to replenish nitrogen in farm fields. We are collaborating with INCAUCA to help them increase the production of sugarcane in terms of biomass and reduce the use of damaging and costly chemical fertilizers.

Previous studies have shown that sugarcane harbors diverse PGP microorganisms, including nitrogen-fixing bacteria, which can serve as biofertilizers. The primary goal of my thesis research was the isolation and characterization of native, nitrogen-fixing bacteria from the INCAUCA sugarcane fields, with the long term goal of deploying them as PGP biofertilizers. My research entails a novel, integrated field, computational, and experimental approach to the isolation and characterization of native plant-associated soil bacteria, along with the prediction and prioritization of their functional potential as PGP biofertilizers. The discovery of novel, more ecologically friendly biofertilizers is a vital strategy in support of more sustainable agricultural practices.

CHAPTER 2. ISOLATION AND INITIAL CHARACTERIZATION OF NITROGEN-FIXING BACTERIA FROM COLOMBIAN SUGARCANE FIELDS

2.1 Abstract

A primary aim of this research was to isolate and characterize native, nitrogen-fixing bacteria from sugarcane fields in the Cauca Valley of Colombia. Samples of bulk soil, rhizosphere soil, roots, stems, and leaves were collected from plants in the INCAUCA fields of the San Fernando farm. Samples were transported in a unique 4^{o}C shipping container, consisting of refrigerated panels to preserve the samples during transport. A high-throughput cultivation approach was developed and applied to facilitate the enrichment and isolation of nitrogen-fixing bacteria from the INCAUCA samples. The initial isolation phase was followed by a molecular screening procedure to validate the nitrogen fixation potential of each isolate. A total of 22 distinct *nifH* PCR+ isolates was obtained from the INCAUCA samples and selected for genome sequence analysis. Isolate genomes were characterized using next-generation sequencing followed by sequence quality control, assembly, and an initial round of annotation.

2.2 Introduction

2.2.1 Colombian sugarcane company INCAUCA

INCAUCA is a Colombian sugarcane company located in the Cauca River Valley in the southwest region of the country between the western and central mountain ranges. INCUACA is the largest and most productive of 13 sugarcane companies in the area, with ~5,700 employees. In 2015, INCAUCA controlled almost 45,000 hectares of cultivated sugarcane fields in the Cauca Valley, and in 2017 the INCAUCA mill processed ~4.3 metric tons of sugarcane. INCAUCA has an electric plant and a large-scale fermentation facility capable of producing more than 100 million liters of ethanol per year.

Our group is collaborating with INCAUCA to help them develop cropping practices that are both effective and sustainable. The long-term goals of this collaboration are to simultaneously (i) increase crop yield, and (ii) decrease the reliance on chemical fertilizers via the discovery, characterization, and application of endemic (native) biofertilizers to INCAUCA sugarcane fields (Figure 1). Biofertilizers, in this case, are bacteria that encode the biochemical capacity to fix nitrogen and promote plant growth through other pathways, such as phosphate solubilization and the production of growth-inducing hormones. INCAUCA, like other sugarcane companies in the region, currently uses commercially available bacterial biofertilizers, which originated in other countries (primarily Brazil), with limited success. We reasoned that native bacteria should be better adapted to the local environment and thereby serve as more effective biofertilizers for Colombian sugarcane. The use of indigenous bacteria as biofertilizers should also mitigate potential threats to the environment posed by non-native and potentially invasive species of bacteria. Finally, native bacteria represent a renewable resource that the company should be able to develop through isolation and cultivation of local strains for years to come.

Figure 1. INCAUCA sugarcane fields.

Georgia Tech professor King Jordan inspects sugarcane fields at INCAUCA along with Dr. Lina Valderrama-Aguirre, the director of INCAUCA's field microbiology laboratory.

2.3 Materials and Methods

2.3.1 Sample site selection, design, and execution

Sample collection was conducted in INCAUCA company sugarcane fields located in the Cauca Valley of Colombia (Figure 2). The sample sites were carefully selected to control for variability among these critical factors: agro-ecological zone, soil type, crop age, and the number of harvests since planting.

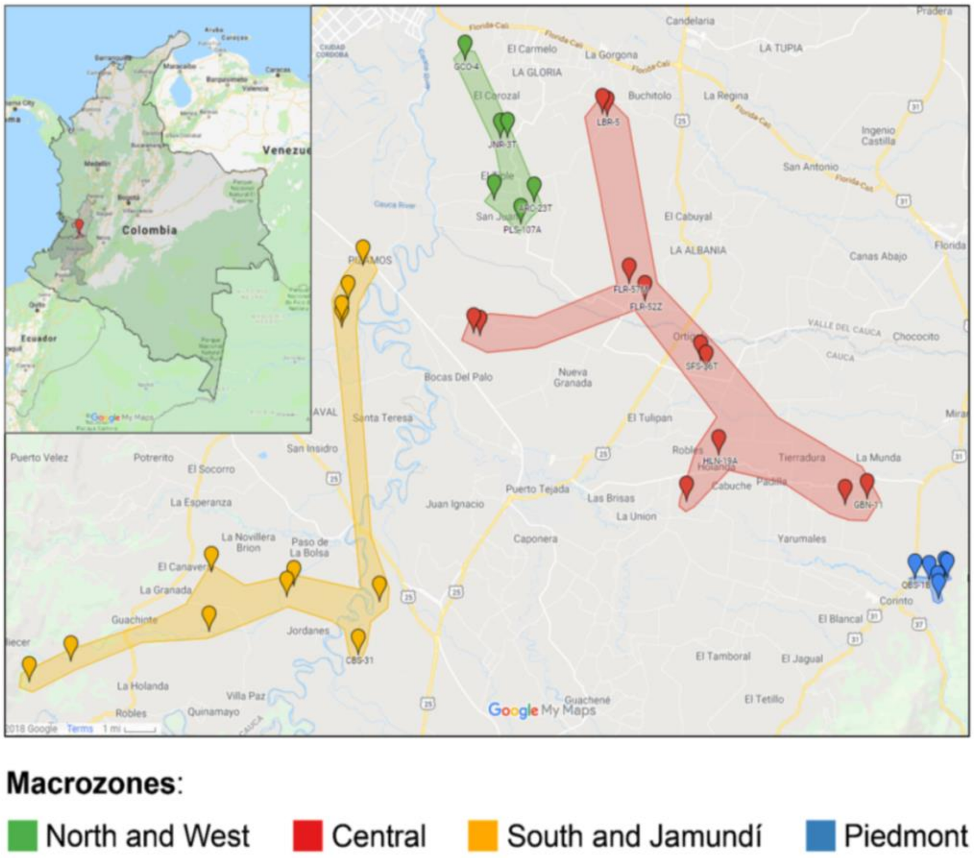


Figure 2. INCAUCA sugarcane growing zones and farms.

The location of INCAUCA fields in the southwest region of Colombia is shown in the inset. Four principal farming macrozones are shown along with the locations of individual farms within each macrozone. Sampling was conducted for representative farms and representative entry points for each sampled field.

First, the INCAUCA Chronologic Database of field information was used to select fields that have similar characteristics. Second, the Cenicaña Map Server was used to ensure that distinct sugarcane varieties that were sampled were from the same soil type. Third, a detailed map of the INAUCA fields was provided using GIS data to find access points and assess crop row orientation. All of this information was combined to select 37 sample sites in fields from within the 4 main growing macrozones where INCAUCA harvests sugarcane, the Northwest, the Southern and Jamundi, Piedmont, and Central regions (Figure 2).

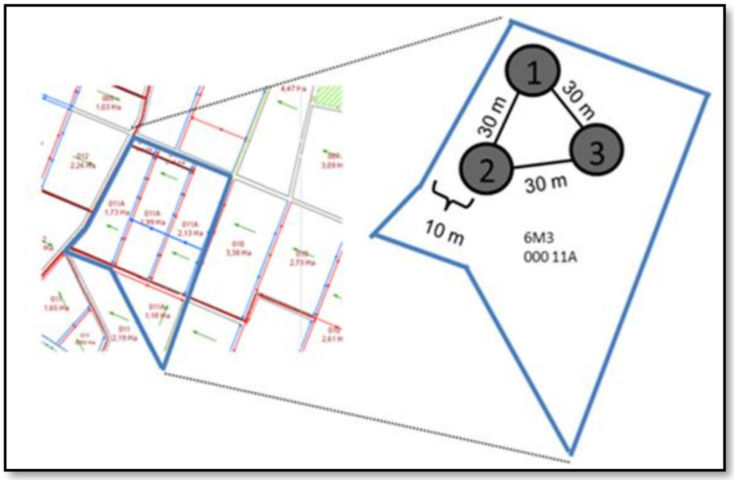


Figure 3. Triangular field sampling design.

An example of an individual field (outlined in blue) is shown, with the expansion illustrating the spacing and dimensions of the triangular sampling design.

At each of the selected fields and for each of the field-specific sampling sites, a controlled triangular sampling design was used to eliminate potential spatial effects (Figure 3). For this design, samples were taken at 3 equidistant points along an equilateral triangle having 30-meter sides. All sampling triangles were set 10 meters from the margins of the fields to eliminate field edge effects. Thirty-seven sites in sugarcane fields were sampled with 3 replicates each, as described above, for a total of 111 replicates. A total of 5 sampled compartments were taken from each individual sampling site replicate: 2 soil (bulk soil & rhizosphere soil) and 3 plant tissues (root, stem & leaf) to account for the expectation that each compartment will exhibit differences with respect to the diversity and abundance of nitrogen fixers (Figure 4). This procedure resulted in a total of 555 microbiome samples from INCAUCA fields. The soil physicochemical characterization of INCAUCA fields are shown in Table 1.

Figure 4. Sugarcane sampling.

Georgia Tech researcher Chris Gaby, postdoc in the Kostka lab, is shown sampling sugarcane along with a member of the INCAUCA microbiology laboratory. Resulting samples of sugarcane stem and leaf are shown.

Table 1. Physicochemical characterization of INCAUCA sugarcane soil fields.

Field	Last harvest	Sampling Date	Sand %	Loam %	Clay %	Moist per %	Moist prop	pH	OM[1] %
0	09-11-13	06-12-14	40.60	20.67	38.73	35.26	0.35	5.1	4.38
1	09-11-13	06-03-14	12.60	56.67	30.73	18.11	0.18	5.3	1.63
2	08-16-13	06-09-14	13.93	30.67	55.40	25.50	0.25	5.3	2.50
2c	10-17-13	06-18-14	41.93	33.33	24.73	20.22	0.20	7.1	3.97
3a	08-03-13	06-16-14	33.93	24.67	41.40	19.26	0.19	6.9	2.12
3t	10-19-13	06-16-14	37.27	25.33	37.40	14.64	0.15	6.6	1.95
4	09-19-13	06-16-14	29.27	26.67	44.07	17.71	0.18	6.4	2.45
5	11-16-13	05-30-14	53.93	22.00	24.07	6.55	0.07	6.4	1.75
5b	11-10-13	06-03-14	59.93	17.42	22.65	9.77	0.10	6.4	1.67
6	12-21-13	06-09-14	19.93	28.00	52.07	26.39	0.26	5.6	3.04
7	10-13-13	06-09-14	19.27	33.33	47.40	25.31	0.25	5.6	3.15
7a	10-16-13	06-19-14	23.93	44.33	31.73	18.10	0.18	6.8	2.88
8	08-29-13	06-18-14	39.27	28.67	32.07	18.98	0.19	6.3	3.00
8a	10-24-13	06-24-14	12.60	22.67	64.73	21.78	0.22	4.8	2.12
9t	10-17-13	06-19-14	20.60	36.67	42.73	16.82	0.17	6.3	2.18
9z	07-28-13	06-12-14	23.27	36.67	40.07	23.39	0.23	5.1	2.86
11	10-16-13	06-17-14	29.27	26.00	44.73	19.03	0.19	6.3	2.72
11a	10-22-13	06-17-14	25.93	26.00	48.07	19.87	0.20	6.4	2.71
13a	08-05-13	06-12-14	15.93	28.00	56.07	27.44	0.27	5.4	2.58
14a	10-02-13	06-19-14	45.93	25.33	28.73	12.29	0.12	5.9	1.40
19	10-17-13	06-13-14	14.60	30.67	54.73	29.95	0.30	6.0	3.20
19a	09-27-13	06-05-14	27.93	41.33	30.73	9.26	0.09	5.9	1.50
19b	10-15-13	06-12-14	15.27	32.00	52.73	27.56	0.28	5.8	2.80
22	10-18-13	06-12-14	13.27	23.33	63.40	34.65	0.35	5.2	3.44
23t	09-02-13	06-10-14	41.93	26.00	32.07	10.38	0.10	6.8	2.75
31	09-18-13	06-24-14	9.27	16.67	74.07	24.23	0.24	5.4	2.39
32t	08-15-13	06-05-14	55.29	22.65	22.07	12.22	0.12	6.5	1.87
36t	09-27-13	06-05-14	27.27	37.33	35.40	18.79	0.19	6.4	2.45
52Z	12-02-13	06-17-14	43.93	33.33	22.73	16.15	0.16	5.8	1.44
57M	11-16-13	06-17-14	29.33	33.27	37.40	18.37	0.18	6.9	2.46
107a	09-27-13	06-10-14	25.95	27.31	46.73	20.58	0.21	6.8	2.28
117	11-20-13	06-10-14	39.27	30.00	30.73	13.53	0.14	6.6	2.12

Table 1 (continued).

Field	P BrayII^2	P fixed	S mg.kg^{-1}	Ca mg.kg^{-1}	Mg mg.kg^{-1}	K mg.kg^{-1}	Na mg.kg^{-1}	CEC^3 cmol.kg^{-1}	CaMg ratio	Ca+Mg K ratio
0	8.47	40.09	23.53	12.27	11.65	0.35	0.42	43.89	0.97	68.42
1	53.00	none	none	4.31	1.99	0.29	0.10	14.02	none	none
2	2.88	33.10	30.37	22.11	7.13	0.38	0.18	28.75	3.11	77.04
2c	338.87	15.44	71.93	22.06	6.74	1.03	0.38	23.90	3.27	28.78
3a	18.53	15.87	Traces	39.25	7.16	0.45	0.36	5.49	102	102.72
3t	16.01	12.31	4.45	40.00	5.82	0.43	0.32	25.90	6.88	53.42
4	23.25	12.88	4.57	44.17	8.30	0.50	0.31	27.13	5.33	105.79
5	79.10	none	none	7.81	2.56	0.45	0.26	16.35	none	none
5b	60.99	none	none	6.92	2.18	0.40	0.25	16.28	none	none
6	2.65	43.50	45.90	28.87	7.97	0.31	0.19	27.52	3.67	126.41
7	2.66	42.93	36.73	28.11	8.45	0.30	0.17	30.95	3.33	56.89
7a	178.88	12.30	31.60	11.99	4.89	0.80	0.23	22.34	2.51	22.43
8	6.51	20.71	18.93	21.69	9.98	0.44	0.18	30.48	2.17	44.41
8a	7.39	36.52	81.13	3.17	3.78	0.23	0.13	33.89	0.86	31.50
9t	18.80	11.29	12.20	8.16	7.63	0.35	0.24	23.37	1.10	30.09
9z	14.16	29.40	14.70	11.57	1.12	0.23	0.04	16.79	10.38	55.14
11	11.65	22.56	12.03	13.30	10.53	0.55	0.23	32.72	1.26	44.87
11a	16.23	23.56	17.43	11.09	7.35	0.42	0.32	29.76	1.53	29.95
13a	3.27	44.36	23.00	16.63	4.72	0.30	0.10	21.81	3.70	32.37
14a	4.61	15.44	12.60	3.23	2.06	0.16	0.11	16.25	1.58	16.23
19	2.54	41.94	4.53	34.80	11.88	0.38	0.16	34.01	2.93	123.94
19a	60.32	9.29	7.03	18.04	2.74	0.33	0.21	15.47	6.58	63.04
19b	2.93	39.23	6.30	25.15	10.95	0.31	0.16	31.67	2.30	116.79
22	4.22	40.94	33.33	17.04	9.43	0.37	0.15	27.49	1.81	42.89
23t	39.93	13.44	none	18.29	7.60	0.40	0.31	25.84	2.36	65.42
31	3.27	29.54	59.23	6.77	7.44	0.29	0.13	30.85	0.91	33.35
32t	260.14	14.02	none	9.62	3.83	0.52	0.25	none	2.47	26.54
36t	137.27	14.16	none	16.40	6.46	0.64	0.50	none	2.54	26.59
52Z	60.04	9.46	22.07	4.55	2.67	0.22	0.13	7.57	1.70	31.82
57M	137.99	12.45	18.77	6.13	3.59	0.26	0.13	10.11	1.67	20.06
107a	118.99	18.43	5.10	23.01	9.49	0.49	0.64	18.73	2.42	65.86
117	172.69	13.73	10.20	13.81	5.18	0.48	0.51	37.13	2.67	25.29

Table 1 (continued).

Field	CaK ratio	Sat Ca	Sat Mg	Sat K	Sat Na	B mg.kg^{-1}	Cu mg.kg^{-1}	Fe mg.kg^{-1}	Mn mg.kg^{-1}	Zn mg.kg^{-1}
0	35.15	49.27	46.76	1.40	1.71	0.27	157.48	157.48	26.20	3.78
1	none	none	none	None	none	0.53	4.14	88.76	14.43	0.97
2	58.28	72.97	23.59	1.28	0.58	0.30	6.01	126.20	30.49	2.10
2c	22.12	72.96	22.35	3.44	1.26	0.24	4.38	71.99	9.00	3.23
3a	86.89	83.13	15.15	0.96	0.18	0.18	48.64	13.87	0.62	33.93
3t	92.36	85.88	12.50	0.93	0.68	0.21	2.94	67.35	15.64	0.37
4	89.12	82.88	15.59	0.94	0.59	0.29	3.68	70.37	17.03	1.21
5	none	none	none	None	none	none	2.12	67.03	9.61	1.05
5b	none	none	none	None	none	none	1.60	60.35	7.75	0.99
6	99.98	76.42	21.29	0.84	0.50	0.16	8.69	101.80	59.02	1.70
7	96.19	75.69	22.73	0.79	0.46	0.15	7.72	112.91	46.46	1.84
7a	15.70	66.60	27.71	4.40	1.29	0.30	4.23	2593.72	13.28	1.68
8	49.15	67.04	31.03	1.37	0.56	0.15	7.20	66.98	32.55	3.64
8a	14.44	28.24	33.33	2.08	1.12	0.22	6.20	85.72	16.94	2.14
9t	24.09	50.21	46.14	2.10	1.55	0.15	2.55	53.39	26.05	0.90
9z	50.29	81.43	7.86	1.62	0.26	0.28	3.47	89.54	10.33	1.65
11	25.08	53.89	42.90	2.27	0.94	0.20	6.83	69.77	12.35	2.06
11a	28.96	57.96	38.24	2.15	1.65	0.15	6.67	60.86	12.13	2.87
13a	57.99	75.30	21.03	1.34	0.44	0.17	5.70	63.88	19.30	2.19
14a	20.28	58.13	37.04	2.90	1.93	0.14	2.28	58.41	24.89	0.60
19	92.32	73.66	25.21	0.80	0.33	0.21	20.85	71.15	35.13	1.50
19a	54.71	84.60	12.86	1.55	0.98	0.32	5.12	127.86	6.54	0.11
19b	81.40	68.75	29.95	0.85	0.45	0.25	8.69	95.95	48.58	2.29
22	46.65	60.74	33.61	1.31	0.55	0.18	10.39	137.41	27.92	1.31
23t	46.57	66.79	30.51	1.62	1.08	0.22	3.63	64.16	13.56	0.71
31	24.49	44.96	49.45	1.93	0.89	0.19	6.93	310.34	16.06	2.27
32t	19.00	67.08	27.24	3.95	1.73	0.41	4.26	99.73	9.19	2.46
36t	25.95	68.32	26.91	2.66	2.11	0.28	7.34	136.14	13.57	1.60
52Z	20.05	59.36	35.52	3.21	1.91	0.23	3.19	73.68	18.16	0.88
57M	23.69	58.97	35.31	4.07	1.64	0.24	6.25	82.14	18.30	2.08
107a	46.56	68.29	28.29	1.47	1.95	0.22	6.57	92.52	10.98	1.60
117	29.24	68.63	26.36	2.38	2.63	0.26	2.58	62.97	17.76	1.59

[1] OM = organic matter.

[2] BrayII = soil test used to determine phosphorus bioavailability in crops and soils, usually used in tropical acid soils.

[3] CEC= Cation exchange capacity

2.3.2 *Sample acquisition*

Bulk soil, rhizophere, roots, stem, and leaf were sampled as described above and processed by the team at INCAUCA (Figure 5). The samples were shipped in a unique 4 °C shipping container consisting of refrigerated panels to preserve the samples during transport (Credo Cube, Minnesota Thermal Sciences). The samples arrived in good condition, and the refrigerated panels were confirmed to be cold and still solid upon arrival, thereby confirming that proper pre-conditioning of the panels was conducted and that the samples remained refrigerated during transit.

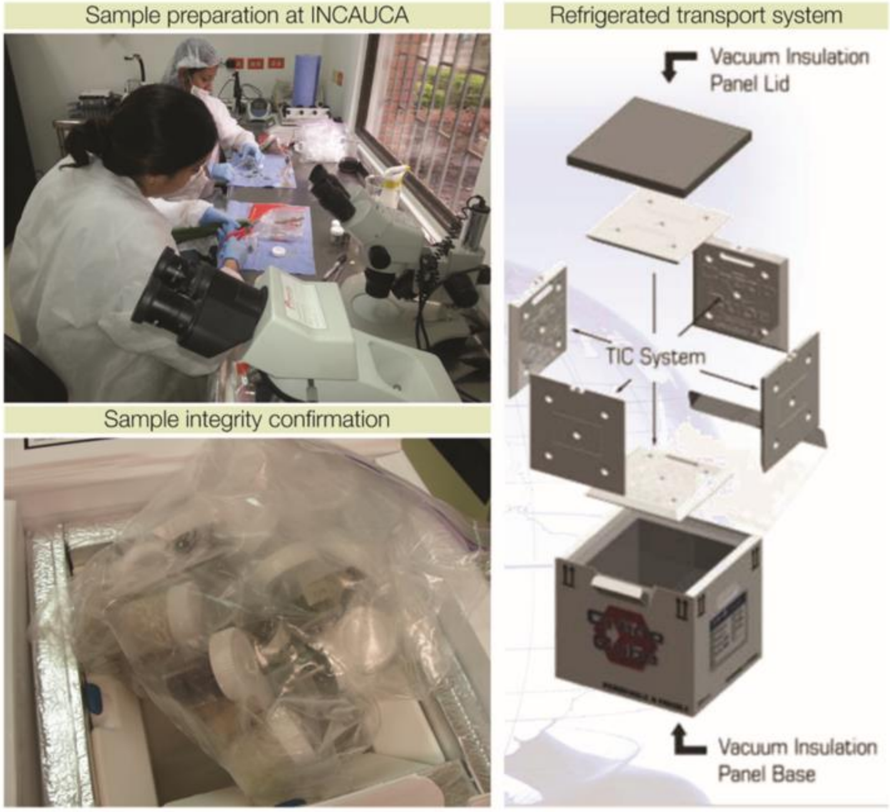


Figure 5. Sample preparation and shipping system.

2.3.3 Cultivation of putative nitrogen-fixing bacteria from sugarcane

A high-throughput enrichment approach was developed to enable the cultivation of multiple strains of putative nitrogen-fixing bacteria from sugarcane field samples (Figure 6). A 96-well plate system was employed with modified Jensen's nitrogen-free, selective media that allows for the enrichment of putative nitrogen-fixing bacteria that can grow in the absence of nitrogen. A range of carbon sources (glucose, sucrose, citric acid, maleic acid, lactate, cellulose, and xylose) and inoculum types (leaves, rhizosphere soil, stem, and roots) were investigated using the nitrogen-free media to enrich for a variety nitrogen fixers from distinct plant-associated communities, which may grow under different conditions. Enrichment plates were incubated under normal atmospheric conditions, or in a hypoxic atmosphere of 2% oxygen, a condition which lowers cellular oxygen and stimulates the oxygen-sensitive nitrogenase enzyme (Figure 6).

Droplets from individual wells of this high-throughput enrichment system were streaked for isolation with an inoculating loop onto Petri plates made of the same nitrogen-free media, containing the same carbon sources to isolate individual colonies. The isolation plates were incubated under normal or hypoxic atmospheric conditions. All isolates of nitrogen-fixing bacteria were frozen as glycerol stocks at -80°C for long-term preservation (Figure 7).

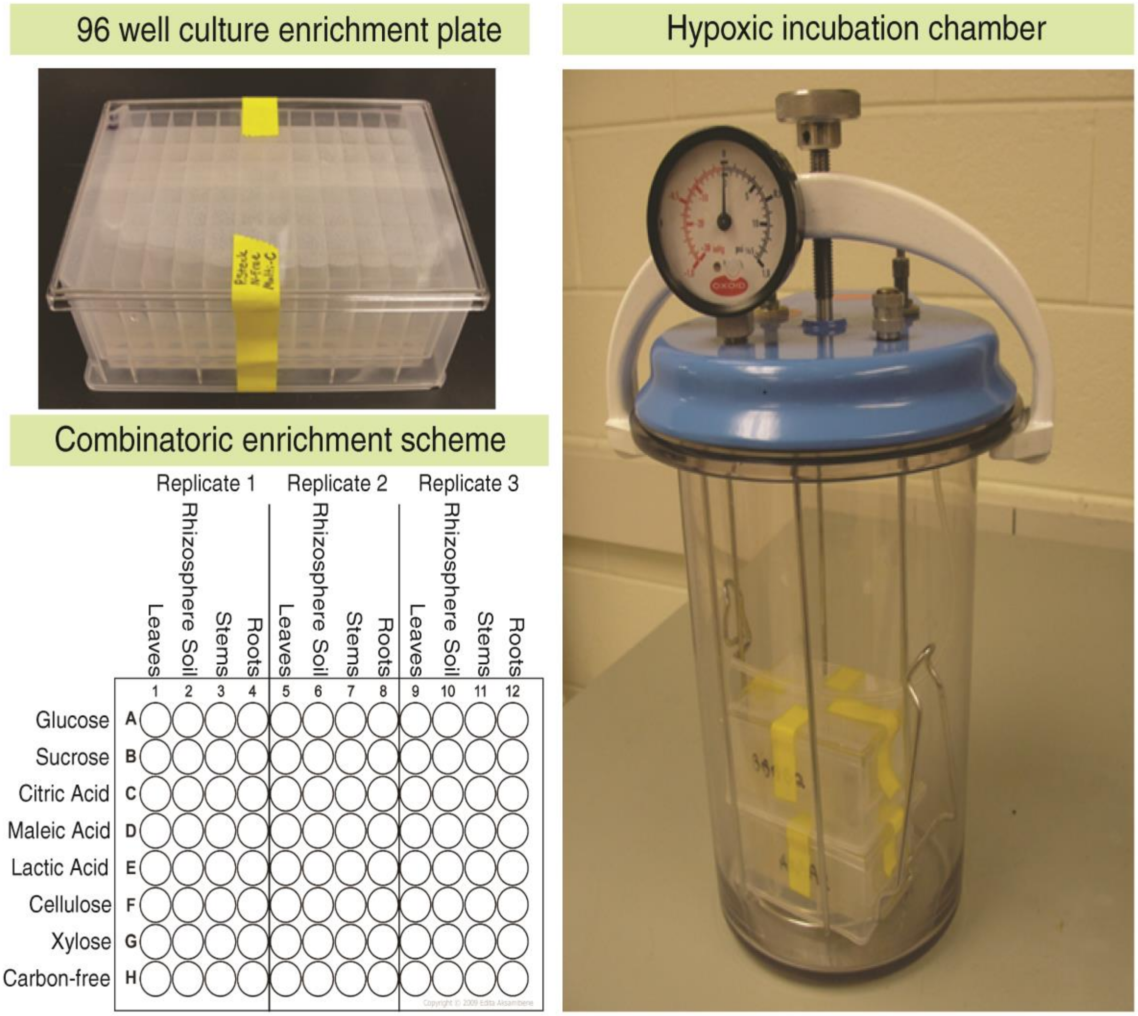


Figure 6. Platform for high-throughput enrichment of nitrogen-fixing bacteria.

A 96-well plate system was used to cultivate isolates from different sugarcane compartments (leaves, rhizosphere soil, stems, and roots) with nitrogen-free media based on multiple carbon sources (glucose, sucrose, citric acid etc.). Isolates were cultivated under atmospheric oxygen conditions and in a hypoxic atmosphere of 2% oxygen, using the incubation chamber shown.

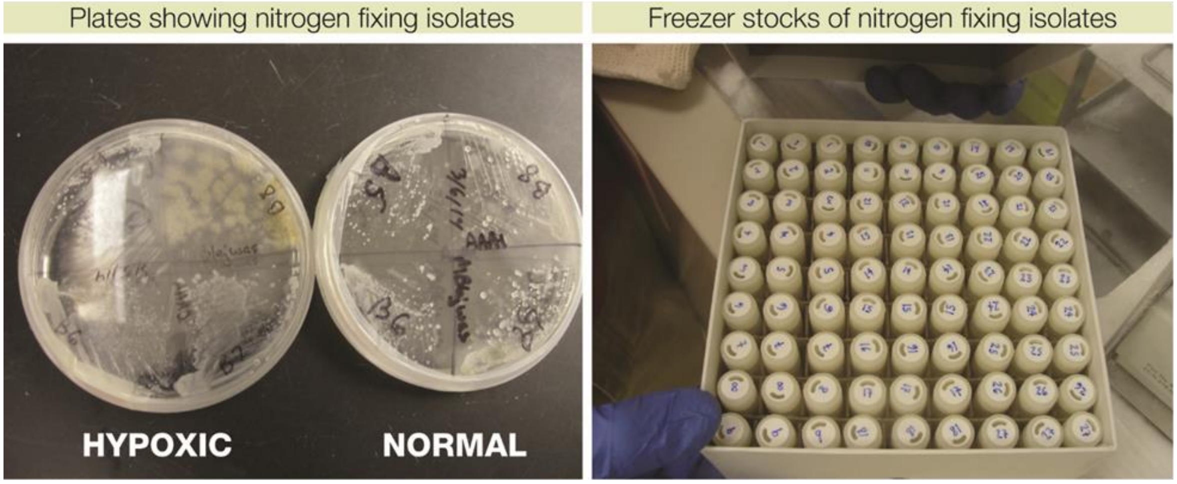


Figure 7. Cultivation and storage of nitrogen-fixing bacterial isolates.

2.3.4 Molecular screening and identification of nitrogen fixers

Plates of putative nitrogen-fixing bacteria were visually inspected for colonies that showed distinct morphologies and were thereby more likely to represent unique bacterial strains. Colonies with unique morphologies were selected and screened for their nitrogen fixation potential by PCR amplification and sequencing of a nitrogenase gene, *nifH*, a commonly used molecular marker for nitrogen fixation using previously described methods (70). Once a set of putative nitrogen-fixing isolates was obtained, an additional PCR-based technique, Ribosomal Intergenic Spacer Analysis (RISA) (71) was used to identify genetically unique strains by PCR amplification with the primers S-D-Bact-1522-b-S-20 (5'-TGCGGCTGGATCCCCTCCTT-3') and L-D-Bact-132-a-A-18 (5'-CCGGGTTTCCCCATTCGG-3') and subsequent band pattern visualization by agarose gel electrophoresis (Figure 8).

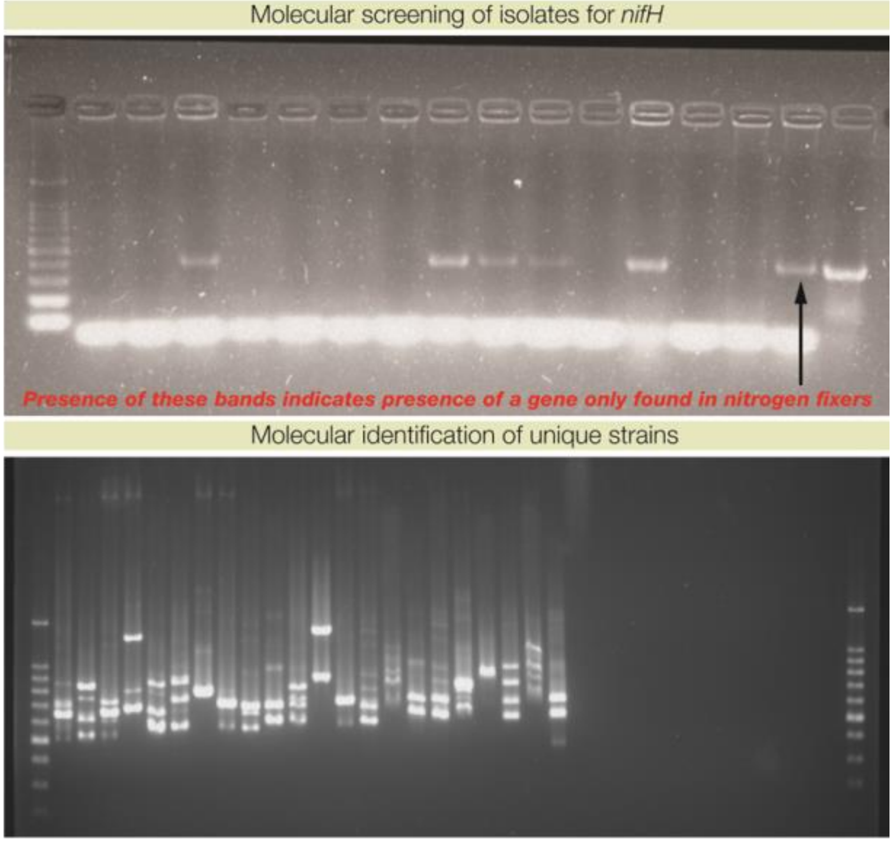


Figure 8. Molecular screening of putative nitrogen-fixing bacterial isolates.

The top panel shows the results of nifH PCR amplification with bands indicating gene presence. The bottom panel shows the results of RISA with distinct banding patterns indicating unique bacterial strains.

2.3.5 Initial genome characterization of putative nitrogen-fixing bacteria

A total of 22 distinct *nifH* PCR+ isolates from the initial cultivation and screening steps were grown overnight in LB medium (Difco) at 37°C. Genomic DNA was isolated using the E.Z.N.A. bacterial DNA kit (Omega Bio-tek), and paired-end fragment libraries were constructed using the Nextera XT DNA library preparation kit (Illumina), with a fragment length of 1,000 bp. Libraries were sequenced on an Illumina MiSeq platform using V3 chemistry, yielding approximately 400,000 paired-end 300-bp sequence reads per sample. Sequence read quality control was performed using the program FastQC version0.11.5 (72). Adapter/primer sequences and low-quality bases and reads ($Q < 20$) were removed using Trimmomatic (v.0.35) (73). The 22 isolate genomes were assembled using the de novo assembler SPAdes (v.3.6) (74). Isolate genome sequences were annotated using the Rapid Annotations using Subsystems Technology (RAST) Web server (75, 76) (Figure 9).

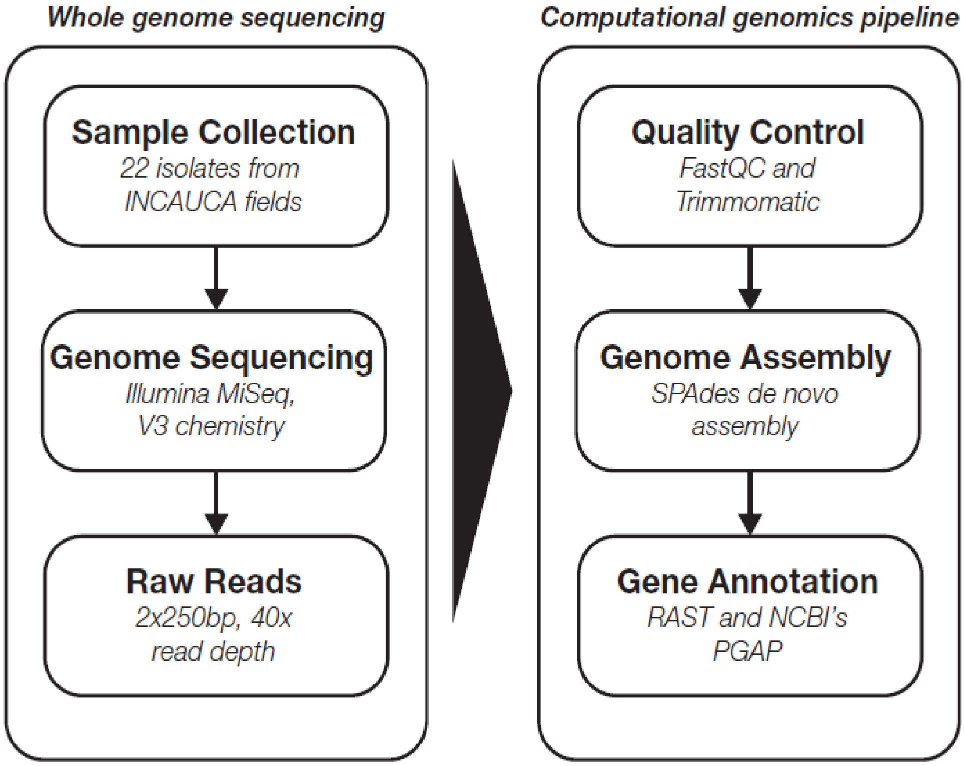


Figure 9. Intial whole genome sequencing and analysis pipeline.

2.4 Results

2.4.1 Cultivation of putative nitrogen-fixing bacteria from sugarcane

A systematic cultivation approach, incorporating seven carbon substrates in nitrogen-free media, was employed to isolate putative nitrogen-fixing bacteria from four different sugarcane plant compartments. These isolates were screened for nitrogen fixation potential through PCR amplification of *nifH* genes. This initial screening procedure yielded several hundred clonal isolates of putative nitrogen-fixing bacteria, and Ribosomal Intergenic Spacer Analysis (RISA) was subsequently used to identify the (presumably) genetically unique strains from the broader set of clonal isolates. A total of 22 potentially unique strains of putative nitrogen-fixing bacteria were isolated in this way and selected for genome sequence analysis.

2.4.2 Initial genome characterization of putative nitrogen-fixing bacteria

Genome sequencing, assembly, and annotation was conducted for the 22 putative nitrogen-fixing isolates (Table 2). Isolate genomes were sequenced to an average of 67x coverage (range: 50x – 88x) and genome sizes range from 4.0 Mb to 6.1Mb. The GC content varied from 41.82% – 66.79%, with a distinct mode at ~57%. The genome assemblies are robust with a range of 24 – 294 contigs ≥500bp in length and averages of N50=310,166bp and L50=8.4. Genome sequence assemblies and their annotations are available in the NCBI Genbank database (BioProject PRJNA418312) and 15 *Klebsiella* isolates are described in a Genome Announcement (77) (Figure 10). Extensive methodological details and additional results of the comprehensive comparative genome analysis of these isolates are presented in Chapter 3.

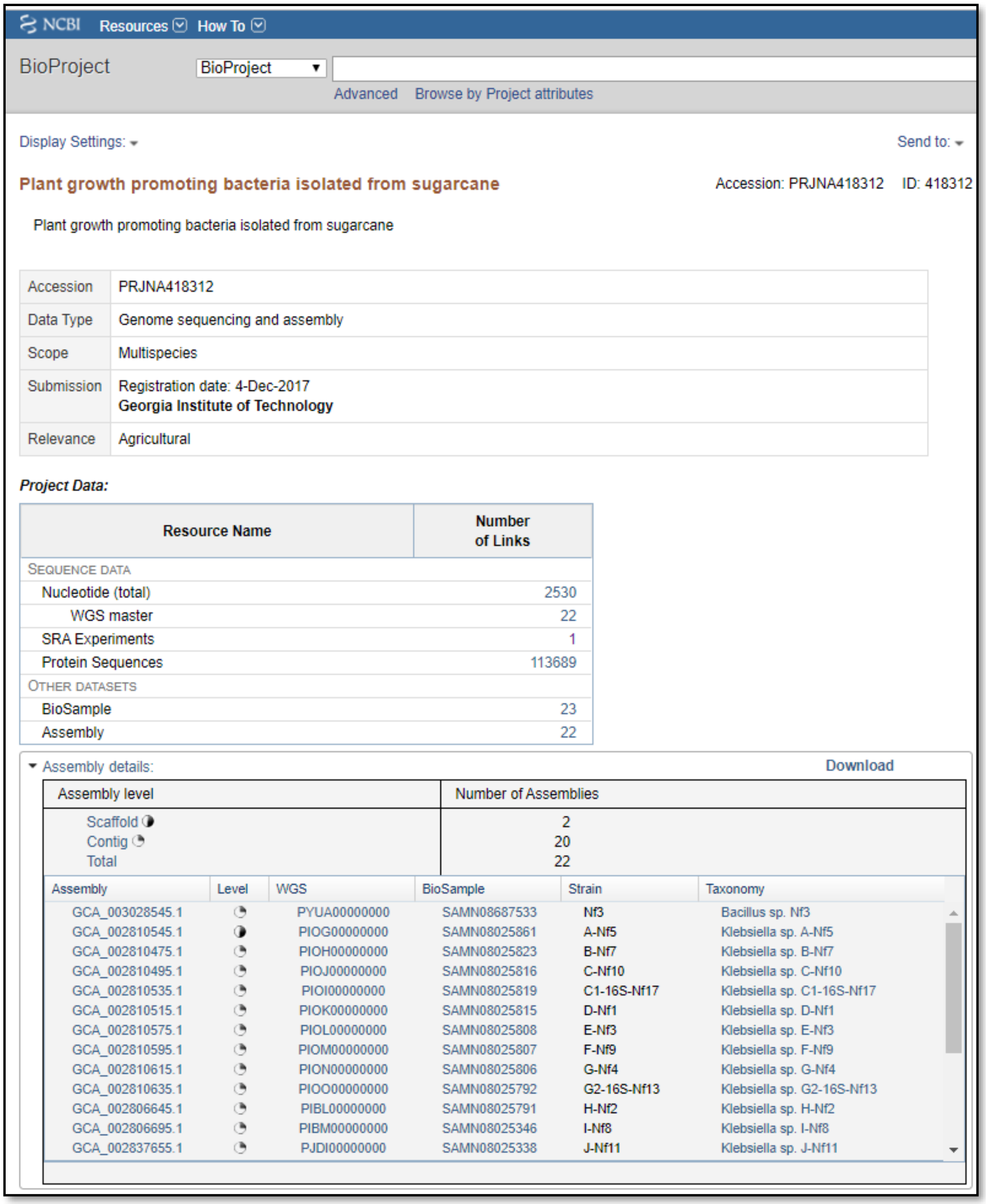
Table 2. Genome assembly statistics for the isolates characterized here.

Sample ID	Genome Length (bp)	N50[a]	L50[b]	GC(%)	# of Contigsc
SCK1	4,522,541	402,304	4	66.79	24
SCK2	5,231,439	417,927	5	59.33	53
SCK3	3,824,428	670,745	3	41.82	150
SCK4	4,511,030	223,239	8	66.79	55
SCK5	5,774,634	162,673	13	53.1	98
SCK6	6,094,823	117,689	15	56.73	294
SCK7	5,693,007	282,996	7	57.03	50
SCK8	5,695,902	281,292	9	57.03	50
SCK9	5,579,618	311,650	6	57.03	42
SCK10	5,591,472	614,324	3	57.03	34
SCK11	5,696,136	382,597	5	57.15	268
SCK12	5,817,089	176,655	10	57.02	79
SCK13	5,476,221	358,490	5	57.34	33
SCK14	5,465,811	300,899	5	57.34	41
SCK15	5,564,330	330,579	5	57.15	43
SCK16	5,795,921	478,592	3	54.06	84
SCK17	5,475,984	358,490	4	57.34	35
SCK18	5,476,135	422,400	3	57.34	32
SCK19	5,688,396	270,585	7	57.09	56
SCK20	5,500,801	82,111	20	57.45	165
SCK21	5,324,920	112,078	15	55.26	100
SCK22	5,847,607	65,329	29	57.02	181

[a] When the contigs of an assembly are arranged from largest to smallest, N50 is the length of the contig that makes up at least 50% of the genome

[b] L50 is the number of contigs equal to or longer than N50. L50 is the minimal number of contigs that cover half the assembly

[c] Number of contigs ≥500bp in length



Plant growth promoting bacteria isolated from sugarcane

Accession: PRJNA418312 ID: 418312

Plant growth promoting bacteria isolated from sugarcane

Accession	PRJNA418312
Data Type	Genome sequencing and assembly
Scope	Multispecies
Submission	Registration date: 4-Dec-2017 **Georgia Institute of Technology**
Relevance	Agricultural

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (total)	2530
WGS master	22
SRA Experiments	1
Protein Sequences	113689
OTHER DATASETS	
BioSample	23
Assembly	22

Assembly details:

Download

Assembly level	Number of Assemblies
Scaffold ◑	2
Contig ◔	20
Total	22

Assembly	Level	WGS	BioSample	Strain	Taxonomy
GCA_003028545.1	◔	PYUA00000000	SAMN08687533	Nf3	Bacillus sp. Nf3
GCA_002810545.1	◑	PIOG00000000	SAMN08025861	A-Nf5	Klebsiella sp. A-Nf5
GCA_002810475.1	◔	PIOH00000000	SAMN08025823	B-Nf7	Klebsiella sp. B-Nf7
GCA_002810495.1	◔	PIOJ00000000	SAMN08025816	C-Nf10	Klebsiella sp. C-Nf10
GCA_002810535.1	◔	PIOI00000000	SAMN08025819	C1-16S-Nf17	Klebsiella sp. C1-16S-Nf17
GCA_002810515.1	◔	PIOK00000000	SAMN08025815	D-Nf1	Klebsiella sp. D-Nf1
GCA_002810575.1	◔	PIOL00000000	SAMN08025808	E-Nf3	Klebsiella sp. E-Nf3
GCA_002810595.1	◔	PIOM00000000	SAMN08025807	F-Nf9	Klebsiella sp. F-Nf9
GCA_002810615.1	◔	PION00000000	SAMN08025806	G-Nf4	Klebsiella sp. G-Nf4
GCA_002810635.1	◔	PIOO00000000	SAMN08025792	G2-16S-Nf13	Klebsiella sp. G2-16S-Nf13
GCA_002806645.1	◔	PIBL00000000	SAMN08025791	H-Nf2	Klebsiella sp. H-Nf2
GCA_002806695.1	◔	PIBM00000000	SAMN08025346	I-Nf8	Klebsiella sp. I-Nf8
GCA_002837655.1	◔	PJDI00000000	SAMN08025338	J-Nf11	Klebsiella sp. J-Nf11

Figure 10. Screenshot of the BioProject PRJNA418312 available in the NCBI.

CHAPTER 3. GENOMIC CHARACTERIZATION AND COMPUTATIONAL PHENOTYPING OF NITROGEN-FIXING BACTERIA ISOLATED FROM COLOMBIAN SUGARCANE FIELDS

3.1 Abstract

Previous studies have shown that the sugarcane microbiome harbors diverse plant growth-promoting (PGP) microorganisms, including nitrogen-fixing bacteria, and the objective of this study was to design a genome-enabled approach to prioritize sugarcane associated nitrogen-fixing bacteria according to their potential as biofertilizers. Using a systematic high throughput approach, 22 pure cultures of nitrogen-fixing bacteria were isolated and tested for diazotrophic potential by PCR amplification of nitrogenase (*nifH*) genes, common molecular markers for nitrogen fixation capacity. Genome sequencing confirmed the presence of intact nitrogenase *nifH* genes and operons in the genomes of 18 of the isolates. Isolate genomes also encoded operons for phosphate solubilization, siderophore production operons, and other PGP phenotypes. *Klebsiella pneumoniae* strains comprised 14 of the 22 nitrogen-fixing isolates, and four others were members of closely related genera to *Klebsiella*. A computational phenotyping approach was developed to rapidly screen for strains that have a high potential for nitrogen fixation and other PGP phenotypes while showing low risk for virulence and antibiotic resistance. The majority of sugarcane isolates were below a genotypic and phenotypic threshold, showing uniformly low predicted virulence and antibiotic resistance compared to clinical isolates. Six prioritized strains were experimentally evaluated for PGP phenotypes:

nitrogen fixation, phosphate solubilization, and the production of siderophores, gibberellic acid and indole acetic acid. Results from the biochemical assays were consistent with the computational phenotype predictions for these isolates. Our results indicate that computational phenotyping is a promising tool for the assessment of benefits and risks associated with bacteria commonly detected in agricultural ecosystems.

3.2 Introduction

The human population is expected to double in size within the next 50 years, which will in turn lead to a massive increase in the global demand for food (78). Given the scarcity of arable land worldwide, an increase in agricultural production of this magnitude will require vast increases in cropping intensity and yield (22). It has been estimated that as much as 90% of the increase in global crop production will need to come from increased yield alone (79). At the same time, climate change and other environmental challenges will necessitate the development of agricultural practices that are more ecologically friendly and sustainable.

Chemical fertilizers that provide critical macronutrients to crops – such as nitrogen (N), phosphorus (P), potassium (K), and sulfur (S) – are widely used to maximize agricultural yield (2). The application of chemical fertilizers represents a major cost for agricultural companies and also contributes to environmental damage, in the form of eutrophication, hypoxia, harmful algal blooms, and air pollution through the formation of microparticles (3). Biological fertilizers (biofertilizers) are comprised of microbial inoculants that promote plant growth, thereby representing an alternative or complementary approach for increasing crop yield, which is more sustainable and environmentally friendly. Biofertilizers augment plant growth through nutrient acquisition, hormone production, and by boosting immunity to pathogens (6).

Sugarcane is a tall, perennial grass cultivated in tropical and warm temperate regions around the world, which is capable of producing high concentrations of sugar (sucrose) and diverse byproducts (28). Sugarcane is consistently ranked as one of the top ten planted crops in the world (31). Sugarcane agriculture plays a vital role in the economy of Colombia by supporting the production of food products and biofuel (ethanol). The long-term goals of this work are to develop more effective and sustainable sugarcane cropping practices in Colombia by simultaneously (i) increasing crop yield, and (ii) decreasing the reliance on chemical fertilizers via the discovery, characterization, and application of endemic (native) biofertilizers to Colombian sugarcane fields.

Most sugarcane companies in Colombia currently use commercially available biofertilizers, consisting primarily of nitrogen-fixing bacteria, which were discovered and isolated from other countries (primarily Brazil), with limited success. We hypothesized that indigenous bacteria should be better adapted to the local environment and thereby serve as more effective biofertilizers for Colombian sugarcane. The use of indigenous bacteria as biofertilizers should also mitigate potential threats to the environment posed by non-native, and potentially invasive, species of bacteria. Finally, indigenous bacteria represent a renewable resource that agronomists can continually develop through isolation and cultivation of local strains.

The advent of next-generation sequencing technologies has catalyzed the development of genome-enabled approaches to harness plant microbiomes in sustainable agriculture (39, 43). The objective of this study was to use genome analysis to predict the local bacterial isolates that have the greatest potential for plant growth promotion while representing the lowest risk for virulence and antibiotic resistance. Putative biofertilizer strains were isolated and cultivated from Colombian sugarcane fields, and computational phenotyping was employed to predict their potential utility as biofertilizers. We then performed a laboratory evaluation of the predicted plant growth promoting properties of the prioritized bacterial biofertilizer isolates, with the aim of validating our computational phenotyping approach.

3.3 Materials and methods

3.3.1 Genome sequencing, assembly, and annotation

A total of 22 distinct *nifH* PCR+ isolates that passed the initial cultivation and screening steps were grown in LB medium (Difco) at 37°C for subsequent genomic DNA extraction as a previously described in Chapter 2. Isolate genomic DNA libraries were sequenced on the Illumina MiSeq platform using V3 chemistry, yielding approximately 400,000 paired-end 300bp sequence reads per sample. A list of all genome sequence analysis programs that were used for this study is provided Table 3. Sequence read quality control and trimming were performed using the programs FastQC version0.11.5 (72) and Trimmomatic (v.0.35) (73). *De novo* sequence assembly was performed using the program SPAdes (v.3.6) (74). Assembled genome sequences were annotated using the Rapid Annotations using Subsystems Technology (RAST) Web server (75, 76) and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (80). The 15 *Klebsiella* isolates characterized in this way were briefly described in a Genome Announcement (77), and the analysis here includes 7 additional non-*Klebsiella* isolates.

Table 3. List of genome sequence analysis software used in this study.

Software	Version	URL	Function
Barrnap	0.9	https://github.com/tseemann/barrnap	16S rRNA prediction
BCFtools	1.7	http://www.htslib.org/	Calling SNPs and consensus sequences from alignment
BLAST	2.2.28	https://blast.ncbi.nlm.nih.gov/Blast.cgi	Sequence database searches and alignment
BWA	0.7.15	http://bio-bwa.sourceforge.net/	Read mapping
CGView	1.0	http://wishart.biology.ualberta.ca/cgview/	Bacterial genome visualization and browsing
FastQC	0.11.5	https://www.bioinformatics.babraham.ac.uk/projects/fastqc	FASTQ quality control
MUSCLE	3.8.425	https://www.ebi.ac.uk/Tools/msa/muscle/	Multiple Sequence Alignment
Genome Workbench	2.12.0	https://www.ncbi.nlm.nih.gov/tools/gbench/	Compering and analyzing genetic data
MEGA	7.1	https://www.megasoftware.net	DNA analysis and protein sequence data from species and populations
PATRIC3/ mic_prediction	0b5494a	https://github.com/PATRIC3/ mic_prediction	Minimum inhibitor concentration prediction for *Klebsiella* isolates
Mummer	3.23	http://bcb.gmu.edu/software/mummer/manual/	DNA alignment and comparison
PATRIC	3.5.20	https://www.patricbrc.org/	Genome and sequence analysis database
RAST	N/A	http://rast.nmpdr.org/	Rapid Annotation of microbial genome
Samtools	1.7	http://www.htslib.org/	Manipulating NGS alignments
SPAdes	3.6	http://cab.spbu.ru/software/spades/	*de novo* assemble of 22 isolate genomes
TaxaDiva	0.11.3	https://github.com/lavanyarishishwar/taxadiva	Analysis of *nifH* amplicon sequences
Trimmomatic	0.35	http://www.usadellab.org/cms/?page=trimmomatic	Remove low quality reads/regions
VSEARCH	2.7.1	https://github.com/torognes/vsearch	Clustering and searching nucleotides sequences

3.3.2 Comparative genomic analysis

Average Nucleotide Identity (ANI) was employed to assign the taxonomy of the bacterial isolates characterized here (81, 82). ANI values were calculated using the program MuMmer3 (v.3.23) (83), with the standard settings of DNA fragments showing >75% identity over >75% of their length used for ANI calculations. ANI cutoff values of 97% for species assignment and 85% for genus assignment were used. Isolate genome sequences were initially compared to a collection of ~7,000 NCBI RefSeq complete bacterial genome sequences for taxonomic assignment (84). Isolates with >97% ANI to a single reference genome were assigned the same taxonomy. Isolates with >95% ANI to multiple reference genomes from the same genus were subject to an additional round of majority-rule based taxonomic assignment. For this procedure: 1) all completed or scaffold assemblies available for each genus was retrieved from NCBI Genbank, 2) ANIs for ambiguous isolates were computed against the corresponding genus-specific reference sets, and 3) majority-rule taxonomic assignments were made for the ambiguous isolates based on the most commonly observed species showing ANI values >95% for each genus. Pairwise ANI values for the bacterial isolates characterized here and their most closely related reference strains were converted to p-distances and used to reconstruct a neighbor-joining phylogenetic tree (85) using the program Molecular Evolutionary Genetics Analysis (MEGA, v.7.1) (86).

In support to ANI analysis, taxonomic assignment was also conducted by targeting small subunit ribosomal RNA (SSU rRNA) gene sequences. SSU rRNA sequences were extracted from the isolate genome sequences using the program Barrnap (v.0.9) (87). For isolates where only partial or no 16S rRNA sequences were able to be extracted with

Barrnap, full length 16S rRNA sequences were reconstructed by read mapping to the most closely related reference genome using BWA (v.0.7.15-r1140) (88) followed by consensus 16S rRNA sequence calling using SAMtools (v.1.7) and BCFtools (v.1.7) (89). When multiple, full-length 16S rRNA sequences were present in a single isolate genome sequence, VSEARCH (v.2.7.1) (90) was used to cluster the sequences and pick a single representative sequence centroid. The resulting full-length 16S rRNA sequences, along with 16S rRNA reference sequences from the NCBI 16S rRNA RefSeq database (84), were aligned using MUSCLE (v.3.8.425) (91) and a neighbor-joining phylogenetic tree (85), based on pairwise p-distances, was constructed using MEGA.

Nitrogenase enzyme encoding *nifH* gene sequences were extracted from isolate genome sequences, clustered, and taxonomically assigned using the TaxaDiva (v.0.11.3) method developed by our group (70). For this approach, isolate genome sequence reads in FASTQ format were subjected to quality control and trimming followed by alignment to a *nifH* reference sequence database to produce *nifH* sequences and taxonomic assignments for each isolate. Details of this approach and the specific parameters used for genome-based *nifH* gene reconstruction and taxonomic assignment were previously described (70).

Whole genome sequence comparisons between bacterial isolates characterized here and the *K. pneumoniae* type strain 342 were performed using BLAST+ (v.2.2.28) (92) and visualized with the program CGView (v.1.0) (93). Isolate contig sequences were searched against the *K. pneumoniae* 342 genome sequence, and BLAST best hits whit >70% nucleotide identity were retained. The resulting BLAST best hits were parsed using a custom Perl script and mapped along the *K. pneumoniae* 342 genome sequence.

3.3.3 Computational phenotyping

Computational phenotyping was performed by searching the bacterial isolate genome sequences characterized here for the presence/absence of genes or features related to four functional classes of interest, with respect to their potential as biofertilizers: (i) nitrogen fixation (NF), (ii) plant growth-promotion (PGP), (iii) virulence factors (iv), and (4) antimicrobial resistance (AMR). Gene panels were manually curated by searching the literature (NCBI PubMed) for genes implicated in nitrogen fixation and plant growth promotion. For nitrogen fixing (NF), all *n*=21 known bacterial nitrogen fixation (*nif*) genes were collected from the literature and gene (protein) annotation databases (NCBI and RAST, (75, 94). For plant growth promotion (PGP), a total of *n*=25 genes from six distinct functional subcategories were collected in the same way. The resulting custom gene panels are shown in Table 4 (NF genes) and Table 5 PGP genes). A gene panel containing subset of *n*=44 virulence factor (VF) genes that can be found among any of the bacterial isolates characterized here were taken from the Virulence Factors Database (VFDB) (95). All of the gene (nucleotide) sequences in VFDB were compared against the bacterial isolate genome sequences characterized here using BLAST (92), and VF genes that showed any hits >75% sequence identity over 75% of their length were retained for the gene panel. Gene (nucleotide) sequences from all three gene panels were used as queries to search against the bacterial isolate genome sequence assemblies using BLAST+, requiring the same criteria of >75% sequence identity over >75% of their length for a gene to be considered present in a genome. The presence of multiple gene copies was considered here, with each unique locus counting towards an isolate's score. Finally, VF genes were organized into operons to control for the genomic complexity of some VF phenotypes. For example, the siderophore aerobacterin is synthesized by 4 genes (96) which results in production of a single siderophore – when all four genes are present and co-located, a +1 is added to the Siderophore production phenotype score. Similarly, exotoxin is encoded by a single gene and the presence of this gene adds a +1 to the exotoxin score.

Table 4. Custom gene panel for nitrogen fixation genes.

The 21 bacterial nitrogen fixation genes (nif) used for computational phenotyping are shown along with brief descriptions of their function and their associated TIGRRFAM protein family accession numbers (taken from the RAST annotation system) (97).

Gene name	Gene description	Function	Protein Family
nifH	Nitrogenase iron protein	Structural: dinitrogenase reductase (Fe protein); α2 dimer, Fe4S4 center. Multiple roles: catalysis, FeMo-co synthesis and insertion	TIGR01287
nifD	Nitrogenase molybdenum-iron protein alpha chain	Structural: part of the nitrogenase complex that catalyzes the key enzymatic reactions in nitrogen fixation	TIGR01282
nifJ	Structural- pyruvate: ferredoxin (flavodoxin) oxidoreductase	Structural: pyruvate; flavodoxin oxidoreductase; α2 dimer; FeS centers. Donates electrons to flavodoxin	TIGR01752
nifF	Flavodoxin, NifF	Structural: flavodoxin; donates electrons to flavodoxin	TIGR02176
nifA	Nif-specific regulatory protein	Regulatory: positive regulatory role	TIGR01817
nifL	Nitrogen fixation negative regulator NifL	Regulatory: negative regulator role. Required for the inhibition of NifA activity in response to oxygen and low level of fixed nitrogen (https://www.uniprot.org/uniprot/P30663)	TIGR02938
nifE	Nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	Biosynthesis of the prosthetic group of nitrogenase (FeMo cofactor) (https://www.uniprot.org/uniprot/P10996)	TIGR01283
nifQ	Iron-molybdenum cofactor biosynthesis protein NifQ	Biosynthesis: Involved in molybdenum processing, early role in Mo processing for FeMo-co; not obligate at normal Mo	pfam04891
nifX	Nitrogen fixation protein NifX	Biosynthesis: involved in MoFe cofactor biosynthesis	TIGR02663
nifB	Nitrogenase cofactor biosynthesis protein NifB	Biosynthesis: FeMo-co Biosynthesis. Oxygen-labile FeS protein. Membrane - bound in *K. pneumonie*	TIGR01290

Table 4 (Continued).

nifV	Homocitrate synthase	Biosynthesis: this protein is a Fe-Mo-cofactor biosynthetic component	TIGR02660
nifS	Cysteine desulfurase NifS	Biosynthesis: catalyzes the removal of elemental sulfur atoms from cysteine to produce alanine	TIGR03402
nifM	Nitrogen fixation protein NifM	Processing: required for processing NifH A	TIGR02933
nifY	Protein NifY	Processing: attaches to apodinitrogenase to allow FeMo-co insertion	TIGR02663
nifK	Nitrogenase molybdenum-iron protein subunit beta	Structural: part of the nitrogenase complex that catalyzes the key enzymatic reactions in nitrogen fixation	TIGR01286
nifN	Nitrogenase iron-molybdenum cofactor biosynthesis protein NifN	Play a role in the biosynthesis of the prosthetic group of nitrogenase	TIGR01285
nifU	Iron-sulfur cluster assembly scaffold protein NifU	May be involved in the formation or repair of [Fe-S] clusters present in iron-sulfur proteins	TIGR02000
nifW	Nitrogenase-stabilizing/protective protein NifW	May protect the nitrogenase Fe-Mo protein from oxidative damage	pfam03206
nifO	Nitrogenase-associated protein NifO	Nitrogenase-associated protein	TIGR01616
nifT	Putative nitrogen fixation protein NifT	Nitrogen fixation global process	TIGR02934
nifZ	Putative NifZ	Nitrogenase activity	pfam04319

Table 5. Custom gene panels for plant growth promotion genes.

The 25 plant growth promotion genes used for computational phenotyping are shown along with brief descriptions of their function and their associated Genbank protein family accession numbers. The 25 genes are organized into six functional subcategories (97).

Plant growth promotion funtional subcategories	Gene name	Gene description	Function	Protein accession
Phosphate solubilization	*pstA*	Phosphate ABC transporter, permease protein PstA	Inorganic phosphate transmembrane transporter activity	WP_007508606.1
	pstB	Phosphate transport ATP-binding protein	Part of the ABC transporter complex PstSACB involved in phosphate import. Responsible for energy coupling to the transport system	YP_006895224.1
	pstC	Phosphate ABC transporter permease subunit PstC	Part of the binding-protein-dependent transport system for phosphate	WP_004107261.1
	pstS	Phosphate transporter subunit	ATPase-coupled phosphate ion transmembrane transporter activity	YP_005229600.1
Indole-3-Acetic Acid (IAA)	*ipdC*	Indole-3-pyruvate decarboxylase	Carboxy-lyase activity, magnesium ion binding and thiamine pyrophosphate binding	YP_001005545.1
	pvdO	Pyoverdine biosynthesis-like protein	Kinase activity: catalysis of the transfer of a phosphate group, usually from ATP, to a substrate molecule	NP_746332.1
	pvdN	Pyoverdine biosynthesis-like protein	Catalytic activity	NP_746331.1
	pvdP	Pyoverdine biosynthesis-like protein	Tyrosinase that drives maturation of the pyoverdine 2 chromophore	NP_746329.1
	pvdE	Pyoverdine ABC transporter ATP-binding protein/permease	ATPase activity and peptide transporter activity	NP_746333.1
Siderophore production	*pvdT*	Pyoverdine ABC transporter ATP-binding protein/permease	Hydroxymethyl-, formyl- and related transferase activity	NP_746327.1
	fpvA	Ferripyoverdine receptor	Iron ion binding and receptor activity	NP_251088.1

Table 6 (continued).

	mbtH	Balhimycin biosynthetic protein MbtH	Essential for the adenylation of amino acid in non-ribosomal peptide synthesis.	YP_036475.1
	dhbB	Isochorismatase (siderophore specific)	Catalytic activity	NP_391077.1
	dhbE	2,3-dihydroxybenzoate-AMP ligase	Involved in the biosynthesis of the catecholic siderophore	NP_391078.1
	fhu	Fhu operon transcription regulator	Essential for the utilization of ferric siderophores of the hydroxamate type and for the uptake of the antibiotic albomycin (98)	WP_008842386.1
ACC deaminase activity	*acdS*	Butyryl-CoA dehydrogenase AcdS	Catalyzes a cyclopropane ring-opening reaction, the irreversible conversion of 1-aminocyclopropane-1-carboxylate (ACC) to ammonia and alpha-ketobutyrate. Allows growth on ACC as a nitrogen source	NP_717985.1
	rimM	Ribosome maturation factor	Ribosome binding and rRNA binding	NP_417099.4
	dcyD	D-cysteine desulfhydrase, PLP-dependent	Catalyzes the alpha, beta-elimination reaction of D-cysteine and of several D-cysteine derivatives. It could be a defense mechanism against D-cysteine. Can also catalyze the degradation of 3-chloro-D-alanine	NP_416429.4
Acetoin & butanediol synthesis	*budA*	Alpha-acetolactate decarboxylase	Acetolactate decarboxylase activity	YP_001008160.1
	budC	Acetoin reductase	Catalyzes the reversible reduction of (S)-acetoin to 2,3-butanediol in the presence of NADH	NP_721691.1
	poxB	Pyruvate dehydrogenase, thiamine triphosphate-binding, FAD-binding	Catalytic activity	NP_415392.1
Peroxidases	*osmC*	OsmC family peroxiredoxin	Peroxidase activity, hyperosmotic response, and response to oxidative stress	WP_047723870.1
	oxyR	DNA-binding transcriptional regulator OxyR	Hydrogen peroxide sensor. Activates the expression of a regulon of hydrogen peroxide-inducible genes	YP_005224403.1

Table 6 (continued).

	efeB	Deferrochelatase/peroxidase EfeB	Involved in the recovery of exogenous heme iron. Extracts iron from heme while preserving the tetrapyrrol ring intact	YP_002411956.1
	tpx	Thiol peroxidase	Thiol-specific peroxidase that catalyzes the reduction of hydrogen peroxide and organic hydroperoxides to water and alcohols. Involved in cell protection against oxidative stress by detoxifying peroxides.	YP_009359303.1

For antimicrobial resistance (AMR), the PATRIC3/mic prediction tool (commit: 0b5494a) (99) was used to predict minimum inhibitory concentrations (MIC) for *n*=20 antibiotic classes using the isolate genome sequences characterized here. Performance standards for antimicrobial susceptibility testing were taken from the Clinical & Laboratory Standards Institute (CLSI) (100) and used to convert the predicted MIC levels to sensitive, intermediate, and resistant phenotypes.

A composite score was developed to characterize each bacterial isolate genome sequence with respect to the presence/absence of genes from the NF, PGP, and VF gene panels along with the predicted AMR levels. For each bacterial isolate i, $Score_i = S.NF_i + S.PGP_i - S.VF_i - S.AMR_i$. $S.NF_i = \sum_{j=1}^{21} P_j$, where $P_j \in \{0,1\}$ corresponding to absence and presence of individual *nif* genes based on the results of BLAST+ searches against bacterial isolate genome assemblies. $S.PGP_i = \sum_{j=1}^{6} \frac{\sum_{k=1}^{k} count_k \times w_k}{\max(count_{j,i_{1\ldots22}})}$, where $count_k$ is the number of times any given PGP gene occurs in the bacterial isolate i , $j \in$ $\left\{\begin{matrix} Phospate\ solubilization, IAA\ production, Siderophore\ production, ACC\ deaminase\ activity, \\ Acetoin\ butanediol\ synthesis, Peroxiade \end{matrix}\right\}$ $, w_j$ is a category-specific weight, and $\max(count_{j,i_{1\ldots22}})$ is the maximum observed count of genes for each category of genes j across all 22 genomes. $S.VF_i = \sum_{j=1}^{6} \frac{\sum_{k=1}^{k} count_k \times w_j}{\max(count_{j,i_{1\ldots22}})}$, where $count_k$ is the number of times any given VF gene or operon occurs in the bacterial isolate genome from category j, where $j \in$ $\{Adherence, Invasion, Capsules, Endotoxins, Exotoxins, Siderophores\}$, w_k is a category-specific weight and $\max(count_{j,i_{1\ldots22}})$ is the maximum observed count of genes for each category of genes j across all 22 genomes. Category-specific weights for the

PGP and VF gene panels are $w_i = 1$ for all genes except siderophore genes, which are present in both the PGP and VF panels. Their relative contributions to each phenotype were approximated to yield gene-panel specific weights. PGP siderophore genes are weighted as $w = 0.7$, and VF siderophore genes are weighted as $w = 0.3$, in light of their known relative contributions to plant growth promotion versus virulence. $S.AMR_i = \sum_{j=1}^{20} A_j$, where $A_j \in \{0,0.5,1\}$, corresponding to predicted sensitive, intermediate, and resistant AMR phenotypes. Each of the four sets of functional category-specific scores were normalized by the maximum score such that each category contributed equally to the final genome score.

An additional set of VF genome scores, computed using the same gene panel and formula described above, were compared for the environmental bacterial isolates characterized here together with a set of genome sequences from 29 clinically-associated isolates from closely related *Klebsiella* species taken from the NCBI RefSeq database (Table 6).

Table 6. Genome sequences for clinically-associated isolates from closely related *Klebsiella* species.

Isolate identifiers are Genbank accession numbers are shown for the 29 clinically-associated Klebsiella isolates used for virulence profiling.

Clinical Klebsiella isoaltes	**GenBank Accession number**
Klebsiella pneumoniae subsp. pneumoniae KpO3210	GCA_000300735.3
Klebsiella pneumoniae UCI 62	GCA_000694875.1
Klebsiella pneumoniae strain KPPR1	GCA_000755605.1
Klebsiella pneumoniae strain 4640	GCA_000788025.1
Klebsiella pneumoniae strain 50606221	GCA_001463105.1
Klebsiella pneumoniae strain KPOII-1	GCA_001596935.1
Klebsiella pneumoniae strain CCBH6984	GCA_001720345.1
Klebsiella quasipneumoniae strain B8095	GCA_001729665.1
Klebsiella pneumoniae strain KL064	GCA_002120065.1
Klebsiella pneumoniae strain CNTD94	GCA_002150545.1
Klebsiella pneumoniae strain G749	GCA_002187025.1
Klebsiella pneumoniae subsp. strain KPCTRSRTH02	GCA_002248045.1
Klebsiella pneumoniae strain D7	GCA_002271825.1
Klebsiella pneumoniae strain BA2005	GCA_002287665.1
Klebsiella pneumoniae isolate UBA7514	GCA_002478325.1
Klebsiella pneumoniae strain 18ES	GCA_002738115.1
Klebsiella pneumoniae WCHKP015091	GCA_002852395.1
Klebsiella quasipneumoniae strain D52	GCA_002855865.1
Klebsiella pneumoniae strain CST_2_1	GCA_002910655.1
Klebsiella pneumoniae strain KPNIH26	GCA_002920135.1
Klebsiella oxytoca strain 2880STDY5682542	GCA_900083965.1
Klebsiella pneumoniae isolate PEI-B-P-08	GCA_900108765.2
Klebsiella pneumoniae strain VRES1170	GCA_900172725.1
Klebsiella pneumoniae strain KP06M	GCA_900177035.1
Klebsiella pneumoniae subsp. strain 3189STDY5864809	GCA_900181985.1
Klebsiella pneumoniae 573	GCA_000822285.1
Klebsiella aerogenes UCI 46	GCA_000534115.1
Klebsiella quasipneumoniae subsp. similipneumoniae	GCA_900181145.1
Klebsiella pneumoniae 342	GCA_000019565.1

3.3.4 Experimental validation

Predictions made by computational phenotyping were validated using five distinct experimental assays: (1) Acetylene reduction assay for nitrogen fixation activity, (2) Phosphate solubilization assay, (3) Siderophore production assay, (4) Gibberellic acid production assay, and (5) Indole acetic acid production assay.

3.3.4.1 Acetylene reduction assay.

Potential nitrogen fixation activity was quantified using the acetylene reduction assay (101). Strains were cultivated in pressure tubes containing 10 ml of nitrogen-free media Jensen’s broth with Phytagel and sealed with a rubber stopper. Cultures were incubated for 120 hours at room temperature in a shaker (at 1800 rpm), and growth was monitored by optical density at 600 nm wavelength. Acetylene was added to the headspace (2.5 ml of acetylene was added to a 25 ml headspace), and 100 μl of headspace was sampled for quantification of ethylene production via gas chromatography after 120 hours. Rates (nmol C_2H_4/ml/h) were determined by linear regression of ethylene concentrations over time after comparison to a standard curve.

3.3.4.2 Phosphate solubilization assay

Plant growth promotion phenotypes were verified using a series of cultivation-based assays. Phosphate solubilization activity was evaluated using Pikovskaya's agar medium as previously described (102). Isolate cultures from LB agar medium were resuspended in LB medium and diluted to an OD600 of 0.5, approximate cell density of 4 $\times 10^8$ cells/ml. 100 µl of this suspension was pelleted at 15,000 g at room temperature and buffer exchanged into 5 ml of 0.1% Peptone Salt solution (pH 8). 50 µl of the 0.1% Peptone Salt cell suspension was inoculated onto a sterile filter paper disk. Inoculated filters were cultivated on Pikovskaya's agar medium for 48 hours. After 48 hours of growth, the potential to solubilize phosphate was determined by the formation a clear zone around each colony. A phosphate solubilization index was calculated by measuring the colony and halo zone using the following formula: $(Colony\ diameter + Halo\ zone\ diameter)$ / $(Colony\ diameter)$. According to (103), phosphate solubilization efficacy can be approximated with values of 1 to 2 considered as low solubilizers, values from 2 to 3 considered as medium solubilizers, and values above 3 considered as high solubilizers.

3.3.4.3 Siderophore production assay

Siderophore production was assessed using the chrome azurol S (CAS) media as a previously described (104, 105). CAS is a blue indicator dye sensitive to the oxidation state of iron and iron (III) and LB media. Isolates were cultivated on CAS media by adding 50 µl of each suspension (same culture density and media used for inoculum for Phosphate solubilization assay) of isolates onto a sterile paper disc. Isolates exhibiting a yellowish-orange halo after 5 days of incubation at 28 ± 2 °C were considered positive for siderophore production. A siderophore production index was calculated by measuring the colony and halo diameter using the following formula $(Colony\ diameter + Halo\ zone\ diameter)$ / $(Colony\ diameter)$.

3.3.4.4 Gibberellic acid production assay

Gibberellic acid (GA) production was assessed using the enzyme-linked immunosorbent assay kit MyBiosource (MBS2000244) for Gibberellic acid following the manufacturer protocol recommendation (106). Briefly, GA production was calculated by averaging triplicate reading (optical density) for each standard, control, and samples using a standard curve for linear regression.

3.3.4.5 Indole acetic acid production assay

Indole acetic acid (IAA) production was evaluated using a colorimetric method according to (107). Each isolate was cultivated in tryptophan broth and incubated for 48 hours. After the incubation, 10 ml of culture was centrifuged, and 2 ml of the supernatant was combined with 2ml of Salkowski reagent (0.5M FeCl3 and 70% perchloric acid). A standard curve was prepared of IAA (Sigma-Aldrich) and optical density was measured at 540 nm.

3.4 Results

3.4.1 Genome characterization of putative nitrogen-fixing bacteria

Genome sequencing and assembly summary statistics for the 22 isolates are shown in Table 2. Isolate genomes were sequenced to an average of 67x coverage (range: 50x – 88x) and genome sizes range from 4.0 Mb to 6.1Mb. GC content varies from 41.82% – 66.79%, with a distinct mode at ~57%. The genome assemblies are robust with a range of

24 – 294 contigs ≥500bp in length and averages of N50=310,166bp and L50=8.4. Genome sequence assemblies, along with their functional annotations, can all be found using the NCBI BioProject PRJNA418312. Individual BioSample, Genbank Accession, and Assembly Accession numbers for the 22 isolates are shown in Table 7.

Table 7. Genome sequencing assembly and functional annotations.

Isolate identifiers, NCBI BioSample identifiers, Genbank nucleotide accession numbers, and Genbank assembly accession numbers are shown for the 22 sugarcane-associated bacterial isolates characterized here.

Isolate ID	BioSample	GenBank accession	Assembly
SCK1	SAMN08687486	PYTV00000000	GCA_003028475.1
SCK2	SAMN08687489	PYTW00000000	GCA_003028435.1
SCK3	SAMN08687533	PYUA00000000	GCA_003028545.1
SCK4	SAMN08687959	PYTX00000000	GCA_003028535.1
SCK5	SAMN08687985	PYTY00000000	GCA_003028515.1
SCK6	SAMN08025861	PIOG00000000	GCA_002810545.1
SCK7	SAMN08025823	PIOH00000000	GCA_002810475.1
SCK8	SAMN08025819	PIOI00000000	GCA_002810535.1
SCK9	SAMN08025816	PIOJ00000000	GCA_002810495.1
SCK10	SAMN08025815	PIOK00000000	GCA_002810515.1
SCK11	SAMN08025808	PIOL00000000	GCA_002810575.1
SCK12	SAMN08025807	PIOM00000000	GCA_002810595.1
SCK13	SAMN08025806	PION00000000	GCA_002810615.1
SCK14	SAMN08025792	PIOO00000000	GCA_002810635.1
SCK15	SAMN08025791	PIBL00000000	GCA_002806645.1
SCK16	SAMN08025790	PYTZ00000000	GCA_003028445.1
SCK17	SAMN08025346	PIBM00000000	GCA_002806695.1
SCK18	SAMN08025338	PJDI00000000	GCA_002837655.1
SCK19	SAMN08025337	PIBN00000000	GCA_002806655.1
SCK20	SAMN08115734	PJDJ00000000	GCA_002837625.1
SCK21	SAMN08025336	PIBO00000000	GCA_002806725.1
SCK22	SAMN08115735	PJDK00000000	GCA_002837615.1

3.4.2 Comparative genomic analysis

Average nucleotide identity (ANI; Figure 11) and 16S rRNA gene sequence analysis (Figure 12) were employed in the taxonomic assignment of nitrogen-fixing isolates and the results of both approaches were highly concordant (Table 8), with ANI yielding superior resolution to 16S rRNA gene sequence analysis. A total of eight different species and seven different genera were identified among the 22 isolates characterized. Analysis of *nifH* gene sequences also gave similar results; however, four of the isolates were not found to encode *nifH* genes, despite their (apparent) ability to grow on nitrogen-free media and the positive *nifH* PCR results. This could be due to false-positives in the original PCR analysis for the presence of *nifH* genes, or to changes in the composition of (possibly mixed) bacterial cultures during subsequent growth steps after the initial isolation on nitrogen-free media.

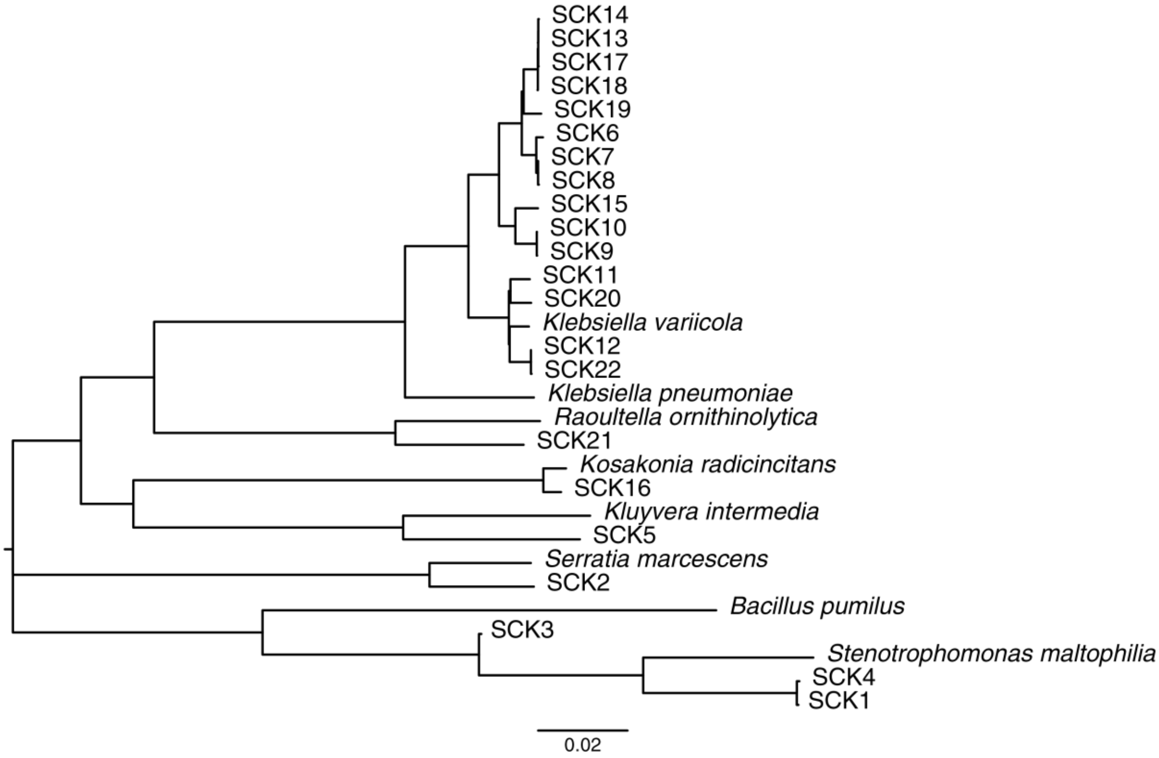


Figure 11. Phylogeny of the bacterial isolates characterized here (SCK numbers) together with their most closely related bacterial type strains.

Phylogeny of the bacterial isolates characterized here (SCK numbers) together with their most closely related bacterial type strains. The phylogeny was reconstructed using pairwise average nucleotide identities between whole genome sequence assemblies, converted to p-distances, with the neighbor-joining method. Horizontal branch lengths are scaled according the p-distances as shown.

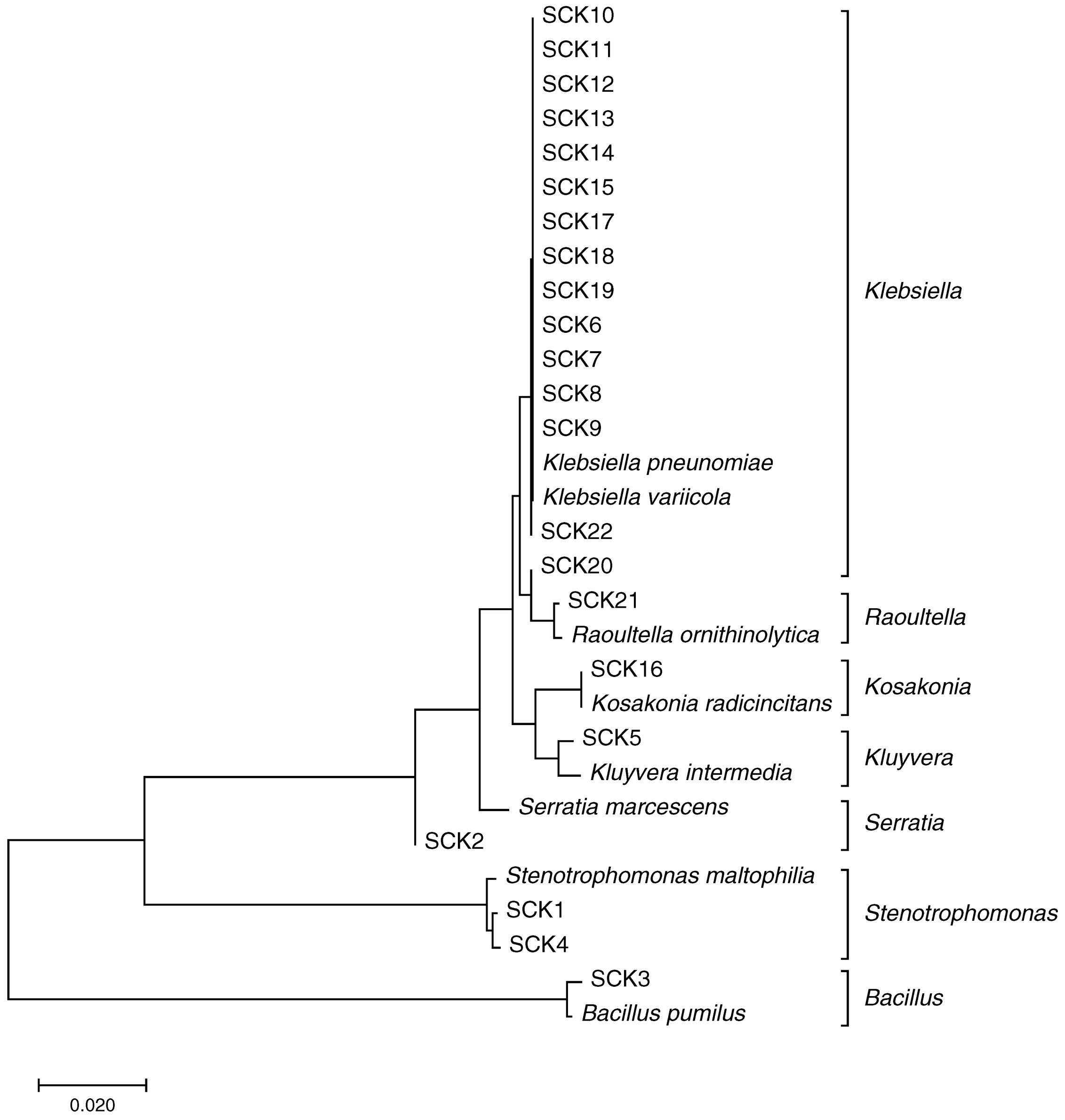


Figure 12. Phylogeny of the 16S rRNA genes for the bacterial isolates characterized here (SCK numbers) together with their most closely related bacterial type strains

The phylogeny was reconstructed using p-distances between 16S rRNA sequences with the neighbor-joining method.

Table 8. Identity of the most closely related species (genus) for the isolates characterized here.

Species (genus) identification was performed using average nucleotide identity (ANI), 16S rRNA and nifH sequence comparisons.

Strain	ANI	16S	*nifH*
SCK1	*Stenotrophomonas maltophilia*	*Stenotrophomonas*	NA
SCK2	*Serratia marcescens*	*Serratia*	NA
SCK3	*Bacillus pumilus*	*Bacillus*	NA
SCK4	*Stenotrophomonas maltophilia*	*Stenotrophomonas*	NA
SCK5	*Kluyvera intermedia*	*Kluyvera*	*Kluyvera*
SCK6	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK7	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK8	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK9	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK10	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK11	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK12	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK13	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK14	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK15	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK16	*Kosakonia radicincitans*	*Kosakonia*	*Kosakonia*
SCK17	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK18	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK19	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK20	*Klebsiella pneumoniae*	*Klebsiella*	*Klebsiella*
SCK21	*Raoultella ornithinolytica*	*Raoultella*	*Raoultella*
SCK22	*Klebsiella variicola*	*Klebsiella*	*Klebsiella*

The majority of isolates, 14 of 22, were characterized as *Klebsiella pneumoniae*, consistent with previous studies showing that *K. pneumoniae* strains are capable of fixing nitrogen (108); in fact, the canonical *nif* operons were defined in the *K. pneumoniae* type strain 342 genome sequence (70). *K. pneumoniae* is also known to be an opportunistic pathogen that can cause disease in immunocompromised human hosts (109), which raises obvious safety concerns regarding its application to crops as part of a biofertilizer inoculum. We performed a comparative sequence analysis between the endophytic nitrogen-fixing *K. pneumoniae* type strain 342, which is capable of infecting the mouse urinary tract and lung (110), and five of the isolates identified as *K. pneumoniae* here. All genomes were shown to contain the *nif* cluster, which contains five functionally related *nif* operons involved in nitrogen fixation (Figure 13). In contrast, the four most critical pathogenicity islands implicated in the virulence of *K. pneumoniae* 342 were all missing in the environmental *K. pneumoniae* isolates characterized here (PAI 1-4 in Figure 13A). The absence of pathogenicity islands in the genome of the endophytic nitrogen-fixer *K. michiganensis* Kd70 was associated with an inability to infect the urinary tract in mice (111). Our results indicate that nitrogen-fixing *K. pneumoniae* environmental isolates from Colombian sugarcane fields do not pose a health risk compared to clinical and environmental isolates that have previously been associated with pathogenicity. We explore this possibility in more detail in the following section on computational phenotyping.

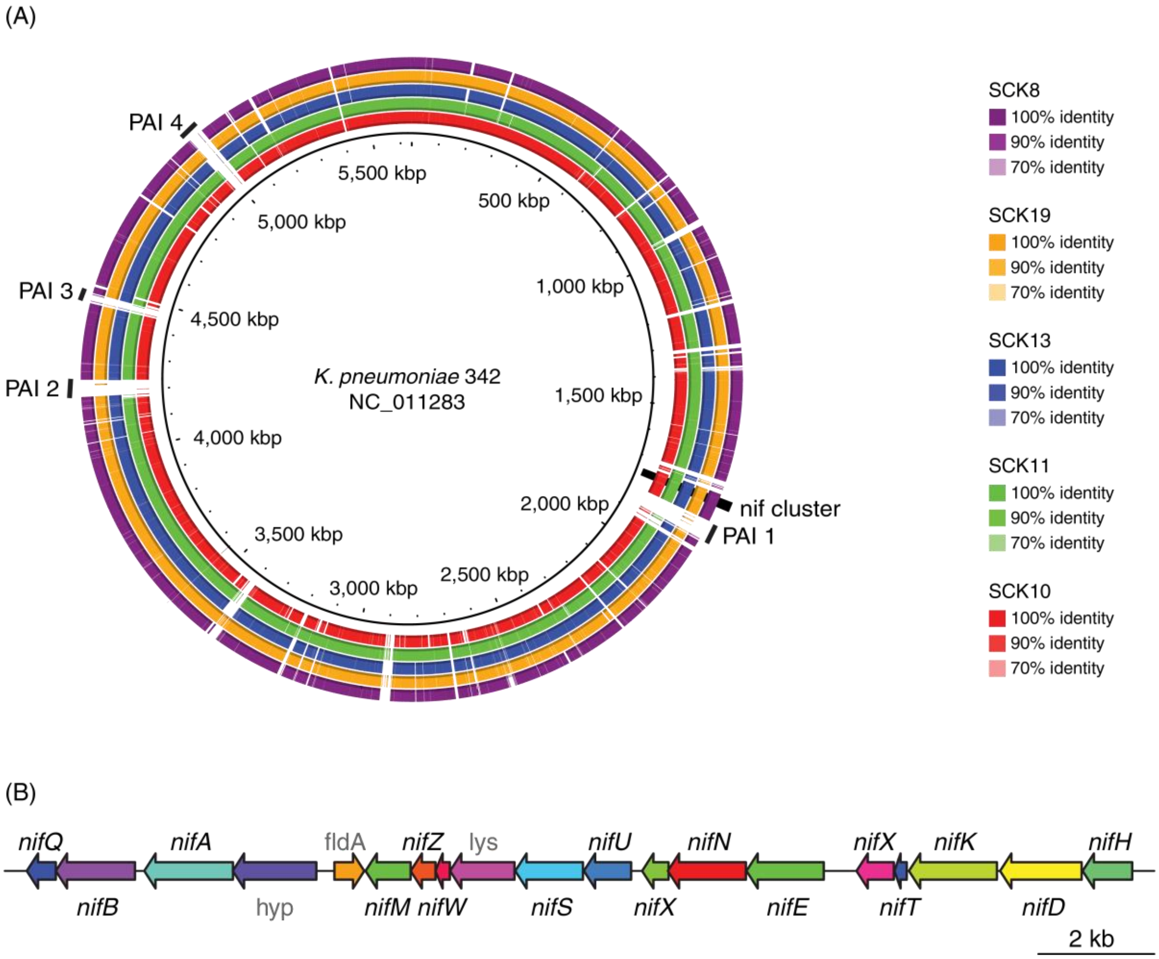


Figure 13. Comparison of the *K. pneumoniae* type strain 342 to *K. pneumoniae* sugarcane isolates characterized here.

(A) BLAST ring plot showing synteny and sequence similarity between K. pneumoniae 342 and five K. pneumoniae sugarcane isolates. The K. pneumoniae 342 genome sequence is shown as the inner ring, and syntenic regions of the five K. pneumoniae sugarcane isolates are shown as rings with strain-specific color-coding according to the percent identity between regions of K. pneumoniae 342 and the sugarcane isolates. The genomic locations of nif operon cluster along with four important pathogenicity islands (PAIs) are indicated. PAI1 – type IV secretion and aminoglycoside resistance, PAI2 hemolysin and fimbria secretion, heme scavenging, PAI3 – radical S-adenosyl-L-methionine (SAM) and antibiotic resistance pathways, PAI4 – fosfomycin resistance and hemolysin production. (B) A scheme of the nif operon cluster present in both K. pneumoniae 342 and the five K. pneumoniae sugarcane isolates.

The *nifH* genes from the *Klebsiella* isolates characterized here form two distinct phylogenetic clusters (Figure 14). This finding is consistent with previous results showing multiple clades of *nifH* among *Klebsiella* genome sequences (112-114) and underscores the potential functional diversity, with respect to nitrogen fixation, for the sugarcane isolates.

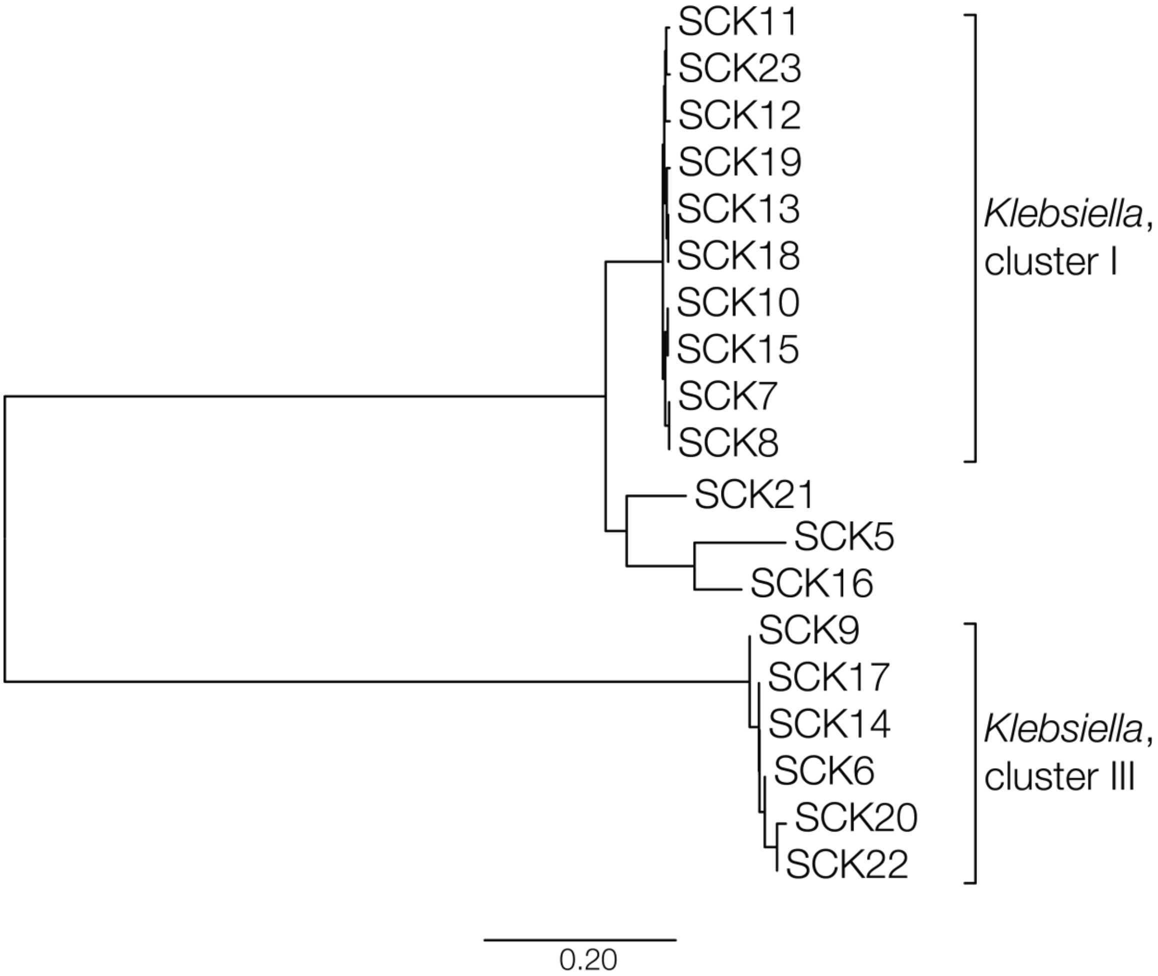


Figure 14. Phylogeny of the *nifH* genes for the *Klebsiella* bacterial isolates characterized here (SCK numbers).

The phylogeny was reconstructed using pairwise nucleotide p-distances between nifH genes recovered from the isolate genome sequences using the neighbor-joining method. Horizontal branch lengths are scaled according the p-distances as shown.

3.4.3 Computational phenotyping

Computational phenotyping, also referred to as reverse genomics, was used to evaluate the potential of the bacterial isolates characterized here to serve as biofertilizers for Colombian sugarcane fields. For the purpose of this study, computational phenotyping entails the prediction of specific organismal phenotypes, or biochemical capacities, based on the analysis of functionally annotated genome sequences (115). The goal of the computational phenotyping performed here was to identify isolates that show the highest predicted capacity for plant growth promotion while presenting the lowest risk to human populations. Accordingly, bacterial isolate genome sequences were screened for gene features that correspond to the desirable (positive) characteristics of (i) nitrogen fixation and (ii) plant growth promotion and the disadvantageous (negative) characteristics of (iii) virulence and (iv) antimicrobial resistance. Genome sequences were scored and ranked according to the combined presence or absence of these four categories of gene features as described in the Materials and Methods. To compute genome scores, the presence of nitrogenase and plant growth promoting genes contribute positive values, whereas the presence of virulence factors and predicted antibiotic resistance yield negative values. Scores for each of the four specific phenotypic categories were normalized and combined to yield a single composite score for each bacterial isolate genome. The highest scoring isolates are predicted as best candidates to be included as part of a sugarcane biofertilizer inoculum (Figure 15). The predicted biochemical capacities of the highest scoring isolates were subsequently experimentally validated.

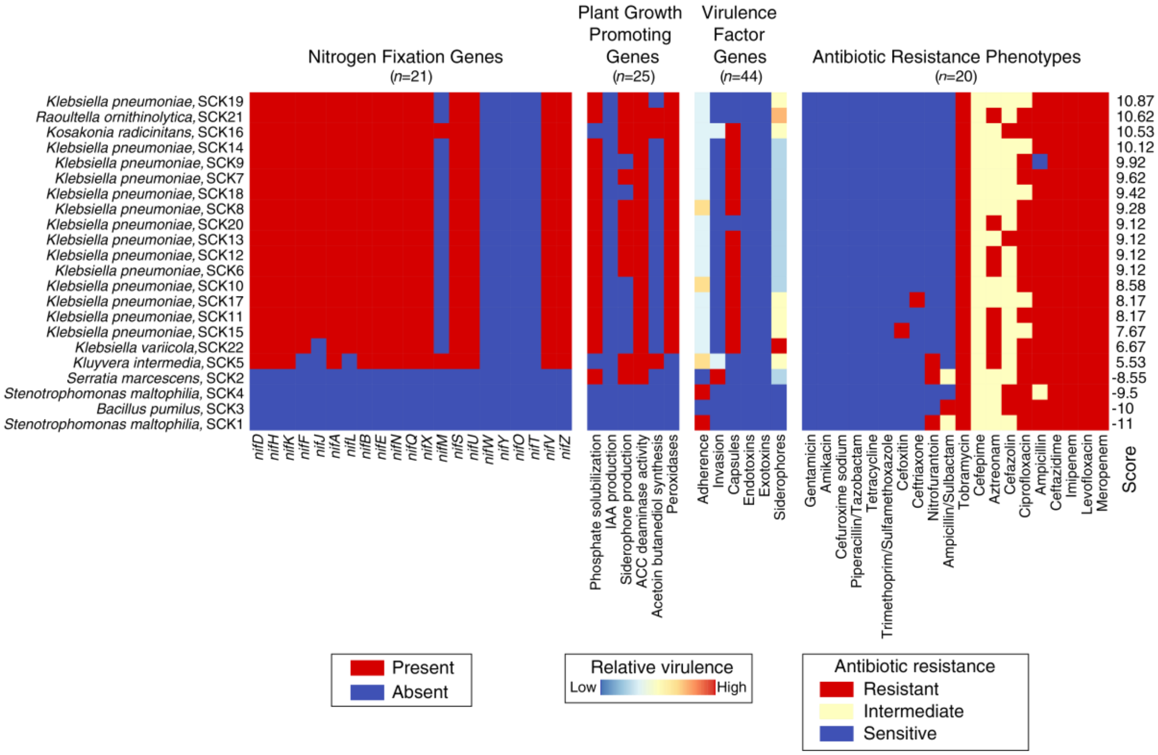


Figure 15. Computational phenotyping of the sugarcane bacterial isolates characterized here.

The presence (red) and absence (blue) profiles for nitrogen fixation genes, plant growth promoting genes, and virulence factor genes are shown for the 22 bacterial isolates. Results are shown for all n=21 nitrogen-fixing genes. Results for plant growth promoting genes (n=25) and virulence factor genes (n=44) are merged into six gene categories each. Predicted antibiotic resistance profiles are shown for n=20 antibiotic classes. Detailed results for gene presence/absence and predicted antibiotic resistance profiles are shown in Table 11-14. The results for all four phenotypic classes of interest were merged into a single priority score for each isolates (right side of plot), as described in the Materials and Methods, and used to rank the isolates with respect to their potential as biofertilizers.

Isolates are ranked according to their composite genome scores, with a value of 10.87 observed as the highest potential for biofertilizer production (Figure 15). Individual gene and phenotype scores are color coded for each genome, and the four functional-specific categories are shown separately. The *nif* gene presence/absence profiles were found to be highly similar for all but four of the bacterial isolates characterized here, those which are not members of the *Klebsiella* genus, or closely related species, and do not encode any *nif* genes. The four non-nitrogen fixing isolates represent bacterial species that are commonly found in soil (116-119), but they are not predicted to be viable biofertilizers. The *Kosakonia radicincitans* genome encodes the largest number of *nif* genes (n=17) observed for any of the Colombian sugarcane isolates. This is consistent with previous studies showing that isolates of this species are capable of fixing nitrogen (120). The 14 characterized *K. pneumoniae* genomes all contain 16 out of 21 *nif* genes, including the core *nifD* and *nifK* genes, which encode the heterotetramer core of the nitrogenase enzyme, and the *nifH* gene, which encodes the dinitrogenase reductase subunit (121). These genomes also all encode the nitrogenase master regulators *nifA* and *nifL*. The missing *nif* genes for the *K. pneumoniae* isolates correspond to accessory structural and regulatory proteins that are not critical for nitrogen fixation. Accordingly, all of *K. pneumoniae* isolate genomes are predicted to encode the capacity for nitrogen fixation, consistent with previous results (110, 122). The single *Raoultella ornithinolytica* isolate characterized here also contains the same 16 *nif* genes; *Raoultella* species have previously been isolated from sugarcane (123) and have also been demonstrated to fix nitrogen (124).

Initially, a total of 29 canonical bacterial plant growth promoting genes were mined from the literature, 25 of which were found to be present in at least one of the bacterial isolate genome sequences characterized here. These 25 plant growth promoting genes were organized into six distinct functional categories: phosphate solubilization, indolic acetic acid (IAA) production, siderophore production, 1-aminocyclopropane-1-carboxylate (ACC) deaminase, acetoin butanediol synthesis, and peroxidases. For the purposes of visualization (Figure 15), each functional category is deemed to be present in an isolate genome sequence if all required genes for that function can be found, but the weighted scoring for these categories is based on individual gene counts as described in the Materials and Methods. The *R. ornithinolytica* isolate shows the highest predicted capacity for plant growth promotion, with 5 of the 6 functional categories found to be fully present. The majority of *K. pneumoniae* isolates also show similar, but not identical, plant growth promoting gene presence/absence profiles, with 3 or 4 functional categories present. The capacity for siderophore production is predicted to vary among *K. pneumoniae* isolates. The *K. radicincitans* genome also encodes 4 functional categories of plant growth promoting genes, but differs from the *K. pneumoniae* isolates with respect to absence of phosphate solubilization genes and the presence of acetoin butanediol synthesis genes. Three of the four species found to lack *nif* genes also do not score present for any of the plant growth promoting gene categories, further underscoring their predicted lack of utility as biofertilizers.

Initially, a total of ~2,500 virulence factor genes were mined from the Virulence Factor Database (VFDB) (95), 44 of which were found to be present in at least one of the bacterial isolate genome sequences characterized here. These 44 virulence factors were organized into six distinct functional categories related to virulence and toxicity: adherence, invasion, capsules, endotoxins, exotoxins, and siderophores. The weighted scores for these categories were computed based on individual gene presence/absence patterns (Figure 15). In contrast to the *K. pneumoniae* clinical isolates which have previously been characterized as opportunistic pathogens, the *K. pneumoniae* environmental isolates showed uniformly low virulence scores. The virulence factor genes found among the *K. pneumoniae* environmental isolates correspond to adherence proteins, capsules, and siderophores. As shown in Figure 13, genomes of environmental isolates lack coding capacity for important invasion and toxin proteins, including the Type IV secretion system, which are found in clinical *K. pneumoniae* isolates. The *R. ornithinolytica* and *K. radicincitans* isolates, both of which show high scores for nitrogen fixation and plant growth promotion, gave higher virulence scores in comparison to the environmental *K. pneumoniae* isolates. Whereas *Bacillus pumilus* had the lowest virulence score for any of the isolates, the remaining three non-nitrogen fixing isolates had the highest virulence scores and were shown to encode well-known virulence factors, such as Type IV, hemolysin, and fimbria secretion systems.

The predicted antibiotic resistance phenotypes for all characterized isolates were fairly similar across the 20 classes of antimicrobial compounds for which predictions were made. The majority of the *K. pneumoniae* genomes, along with the relatively high scoring *R. ornithinolytica* and *K. radicincitans* isolate genomes, indicated predicted susceptibility to 10 of the 20 classes of antimicrobial compounds, intermediate susceptibility for 2-4, and predicted resistance to 5-8. The highest level of predicted antibiotic resistance was seen for *Serratia marcescens*, with resistance predicted for 8 compounds and intermediate susceptibility predicted for 4.

Computational phenotyping scores for the four categories were normalized and combined into a final score, with respect to their potential as biofertilizers (Figure 15). Most of the top positions are occupied by *K. pneumoniae* isolates, with the exception of the second-ranked *R. ornithinolytica* and the third-ranked *K. radicincitans*. The results of a similar analysis of four additional plant associated *Klebsiella* genomes are shown in Figure 16.

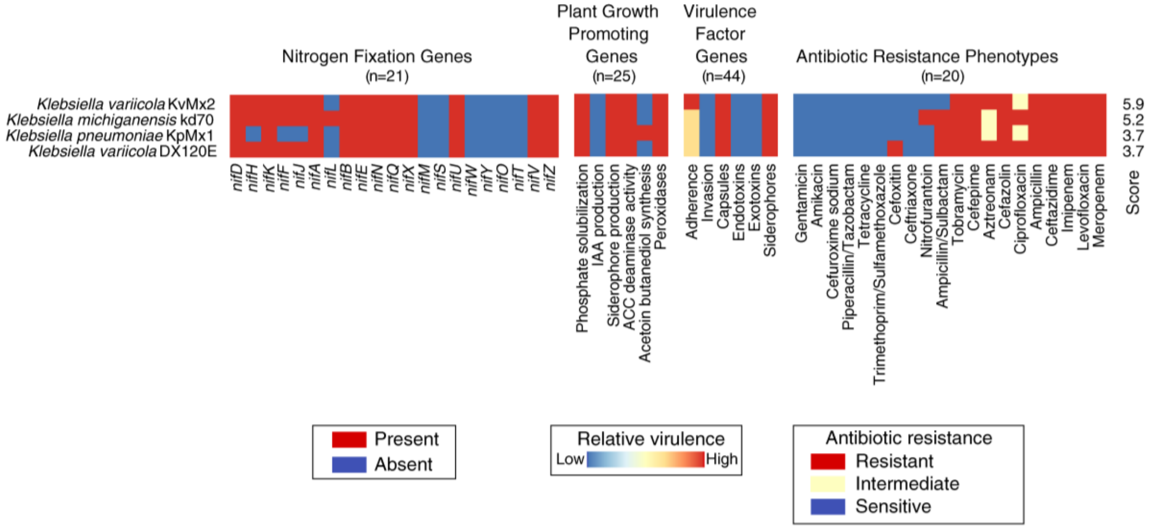


Figure 16. Implementation of the computational phenotyping approach in plant-associate *Klebsiella* from the literature review.

The computational phenotyping approach used to evaluate our 22 environmental samples was employed to analyze four genome sequences of plant-associated Klebsiella that have been previously isolated and reported in previous studies.

3.4.4 Virulence comparison

The results described in the previous section indicate that the majority of the *K. pneumoniae* strains isolated from Colombian sugarcane fields have the highest overall potential as biofertilizers, including a low predicted potential for virulence. Nevertheless, the fact that strains of *K. pneumoniae* have previously been characterized as opportunistic pathogens (125) raises concerns when considering the use of *K. pneumoniae* as part of a bioinoculum that will be applied to sugarcane fields. With this in mind, we performed a broader comparison of the predicted virulence profiles for Colombian sugarcane isolates along with a collection of 29 clinical isolates of *K. pneumoniae* and several other closely related species (See Table 6 for isolate accession numbers). For this comparison, the same virulence factor scoring scheme described in the previous section was applied to all 51 genome sequences (Figure17). Perhaps most importantly, a very clear distinction was observed in the virulence score distribution, whereby all 29 clinical strains show a substantially higher predicted virulence (from 4.45 to 2.11) in comparison to the environmental isolates (1.55 to 0.00). Furthermore, the three environmental isolates that show the highest predicted virulence correspond to species with low predicted capacity for both nitrogen fixation and plant growth promotion; as such, these isolates would not be considered as potential biofertilizers. In particular, the *K. pneumoniae* environmental isolates showed uniformly low predicted virulence compared to clinical isolates of the same species. Thus, the results support, in principle, the use of the environmental *K. pneumoniae* isolates as biofertilizers for Colombian sugarcane fields.

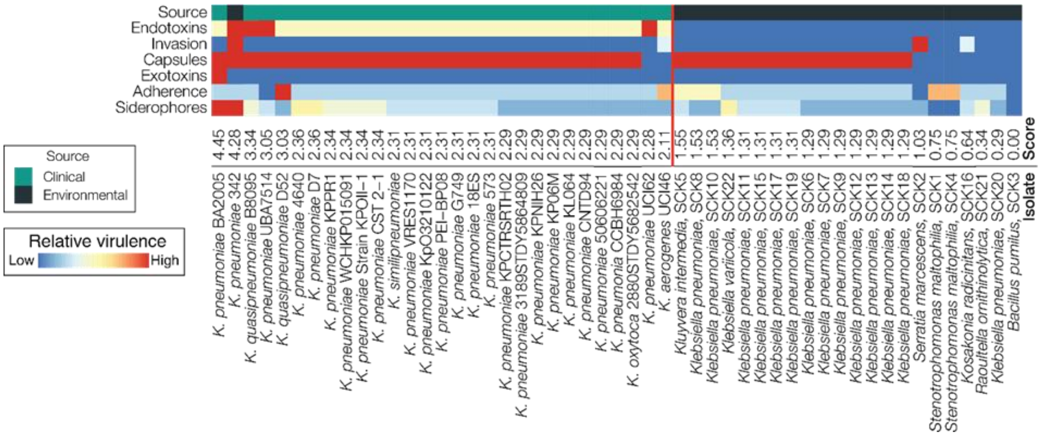


Figure 17. Comparison of predicted virulence profiles for clinical *K. pneumoniae* isolates compared to the environmental (sugarcane) bacterial isolates characterized here.

As in Figure 15, predicted virulence profiles for six classes of virulence factor genes are shown for each isolate. Isolate-specific virulence factor scores are shown for each isolate *are based on the presence/absence profiles for the n=44 virulence factor genes as described in the Materials and Methods. The virulence factor genes are used to rank the genomes from most (left) to least (right) virulent. Clinical versus environmental samples are shown to the left and right, respectively, of the red line, based on their virulence scores.*

3.4.5 Experimental validation of prioritized isolates

The top six scoring isolates from the computational phenotyping were subjected to a series of cultivation-based phenotypic assays in order to validate their predicted biochemical activities: (i) acetylene reduction (a proxy for nitrogen fixation), (ii) phosphate solubilization, (iii) siderophore production, (iv) gibberellic acid production, and (v) indole acetic acid production.

Nitrogen fixation activity, as determined by acetylene reduction to ethylene, was observed in all six isolates, three of which had higher levels in comparison to the positive control (Figure 18A). All six of the isolates showed high levels of phosphate solubilization (Figure 18B & C) and siderophore production (Figure 18D & E) compared to the respective negative controls. All six isolates showed the ability to produce gibberellic acid (Figure 18F), whereas none were able to produce indole acetic acid. The biochemical assay results are consistent with the computational phenotype predictions for these isolates.

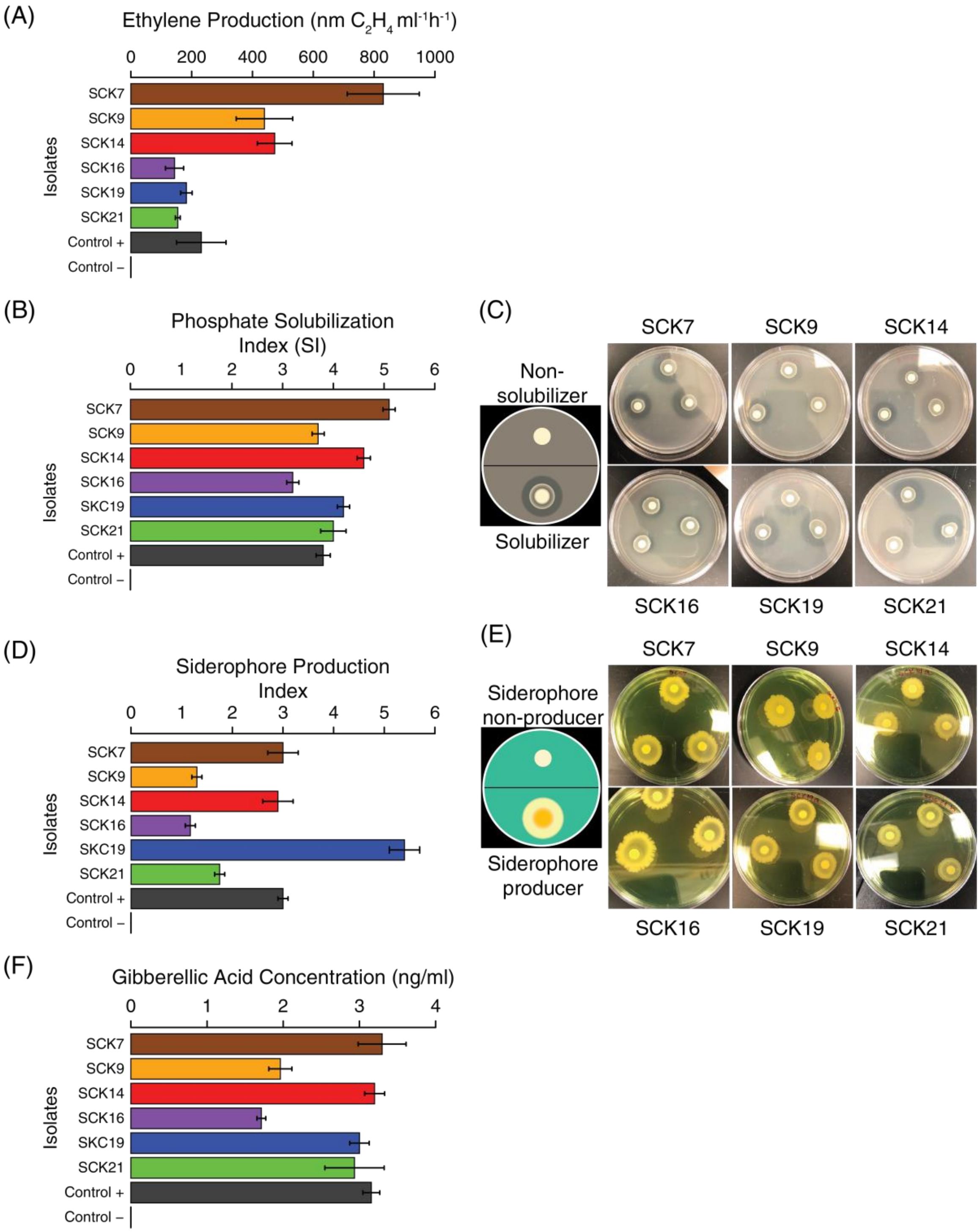


Figure 18. Experimental validation of prioritized biofertilizer isolates.

The computationally predicted plant growth promoting phenotypes for the top six isolates were experimentally validated. All six strains were capable of acetylene reduction, i.e. ethylene production (A), phosphate solubilization (B&C), siderophore production (D&E), and gibberellic Acid production (F).

3.5 Discussion

Members of the *Enterobacteriaceae* are often observed in cultivation-independent studies of sugarcane and nitrogen-fixing *Enterobacteriaceae* are often isolated from sugarcane plants worldwide (59, 126-129). The majority of isolates that were obtained in this study from Colombian sugarcane belonged to the family *Enterobacteriaceae,* with the *Klebsiella* as the most abundant genus along with *Serratia*, *Kluyvera*, *Stenotrophomonas*, and *Bacillus*. *Klebsiella* are Gram-negative, facultative anaerobic bacteria found in soils, plants, or water (130). *Klebsiella* species have been isolated from a large variety of crops worldwide, such as sugarcane, rice, wheat, and maize (130-132). *Klebsiella* species associated with plants have been shown to fix nitrogen and express other plant growth promoting traits (131, 133). Specifically, *Klebsiella* species are abundant amongst the cultivable strains of *Enterobacteriaceae* obtained from sugarcane (59). For example, a survey of sugarcane in Guangxi, China observed that *Klebsiella* was the most abundant plant-associated nitrogen-fixing bacterial group (59), and among the strains isolated, *K. variicola* was shown to colonize sugarcane and promote plant growth (131). In addition, endophytic *Klebsiella* spp. have been isolated from commercial sugarcane in Brazil, and their potential for plant growth promotion was evaluated *in vitro* (15). Finally in Pakistan, the phenotypic diversity of plant growth promoting associated with sugarcane was determined, with *Klebsiella* also appearing as one of the most abundant bacteria found (127). At the same time, *Klebsiella* and other groups of *Enterobacteriaceae* commonly detected in agricultural systems are abundant in the human microbiome and often contain closely related members that are known opportunistic pathogens (134-137). The coexistence of microbial species that contain plant beneficial traits with closely related strains that potentially cause human diseases presents a challenge for the development of sustainable agriculture. How can we effectively perform a risk-benefit analysis of bacterial strains for potential use in the agricultural biotechnology industry? Thus, the overall goal of this study was to develop high throughput methods for the isolation and screening of nitrogen-fixing bacteria for their potential as biofertilizers.

3.5.1 Computational phenotyping for the prioritization of potential biofertilizers

A computational phenotyping approach was developed for the screening of plant growth promoting bacteria for their potential to serve as biofertilizers. Computational phenotyping entails the implementation of a variety of bioinformatic and statistical methods to predict phenotypes of interest based on whole genome sequence analysis (138, 139). This approach has been used for a variety of applications in the biomedical sciences: prediction of clinically relevant phenotypes, study of infectious diseases, identification of opportunistic pathogenic bacteria in the human microbiome, and cancer treatment decisions (140, 141). To our knowledge, this study represents the first time computational phenotyping has been used for agricultural applications. To implement computational phenotyping for the prioritization of potential biofertilizers, we developed a scoring scheme based on the genome content of four functional gene categories of interest: nitrogen-fixing genes, other plant growth promoting genes, virulence factor genes, and antimicrobial resistance genes.

The results of the computational phenotyping predictions, confirmed by laboratory experiments, support the potential use of selected bacterial strains isolated from Colombian sugarcane fields as biofertilizers with minimum health risk to the human population. In particular, all isolates with higher scores (5.53 to 10.87, Figure 15) in our scheme were found to demonstrate the potential to fix nitrogen and to promote plant growth in other ways, while lacking many of the important known virulence factors and antibiotic resistance genes that can be found in clinical isolates of the same species. In general, isolates SCK7, SCK14, and SCK19 appeared to possess more potent plant growth promoting properties compared to isolates SCK9, SCK16, and SCK21 (Figure15). Our computational phenotyping scheme also has valuable negative predictive value. Isolates that contained few or none of the beneficial traits that characterize biofertilizers, *Bacillus pumilus* SCK3 and *Stenotrophomonas maltophilia* SCK1, had the lowest scores (-10 and -11 respectively). Finally, it is also worth reiterating that the computationally predicted biochemical activities related to plant growth promotion were all validated by experimental results (Figure 18).

3.5.2 Virulence profiling for the prioritization of potential biofertilizers

Opportunistic pathogens are microorganisms that usually do not cause disease in a healthy host, but rather colonize and infect an immunocompromised host (142, 143). For example, *Klebsiella spp*. including *Klebsiella pneumoniae*, *Klebsiella oxytoca*, and *Klebsiella granulomatis* were associated with nosocomial diseases (144) and other hospital-acquired infections, primarily in immunocompromised persons (145). The potential for virulence, along with the presence of antimicrobial resistance genes, is an obvious concern when proposing to use *Klebsiella* spp. as biofertilizers. Importantly, we found that the environmental *Klebsiella* isolates did not contain pathogenicity islands associated with many virulence factor genes usually found in clinical isolates of *Klebsiella* spp. (Figure 13). Our results are corroborated by a previous study of *Klebsiella michiganensis* Kd70 isolated from the intestine of larvae of *Diatraea saccharalis*, for which the genome was shown to contain multiple genes associated with plant growth promotion and root colonization, but lacked pathogenicity islands in its genome (111). In order to shed further light on this problem, we extended our study of environmental isolates from Colombian sugarcane to comparisons with genomes of *Klebsiella* clinical isolates associated with opportunistic infections in humans along with a number other environmental isolates with available genome sequences (Figure 17). The virulence factor profiles for all of the environmental isolates were clearly distinct from the clinical strains, which show uniformly higher virulence profile scores, underscoring the relative safety of *Klebsiella* environmental isolates for use as biofertilizers.

3.5.3 Potential for the use of computational phenotyping in other microbiology applications

The results obtained from the computational phenotyping approach developed in this study serve as a proof of principle in support of genomic guided approaches to sustainable agriculture. In particular, computational phenotyping can serve to substantially narrow the search space for potential plant growth promoting bacterial isolates, which can be further interrogated via experimental methods. Computational phenotyping can be used to simultaneously identify beneficial properties of plant associated bacterial isolates while avoiding potentially negative characteristics. In principle, this approach can be applied to a broad range of potential plant growth promoting isolates, or even assembled metagenomes, from managed agricultural ecosystems.

We can also envision a number of other potential applications for computational phenotyping of microbial genomes. The computational phenotyping methodology developed here has broad potential including diverse applications in agriculture, plant and animal breeding, food safety, water quality microbiology along with other industrial microbiology applications such as bioenergy, quality control/quality assurance, and fermentation microbiology as well as human health applications such as pathogen antibiotic resistance, virulence predictions, and microbiome characterization. For instance, computational phenotyping could be useful in food safety related to vegetable crop production. Vegetables harbor a diverse bacterial community dominated by the family *Enterobacteriaceae*, Gram-negative bacteria that include a huge diversity of plant growth promoting bacteria and enteric pathogens (146). Vegetables such as lettuce, spinach, and carrots are usually consumed raw, which increases the concern of bacterial infections or human disease outbreaks associated with consumption of vegetables (142).

Increasing antibiotic resistance, generated by the abuse of antibiotics in agriculture as well as medicine, is another major threat to human health (147), and the food supply chain creates a direct connection between the environmental habitat of bacteria and human consumers (148). Our computational phenotyping approach could provide for an additional food safety solution, which could be used to prevent the spread of antibiotic resistance pathogens genes present in the food chain.

CHAPTER 4. DEVELOPMENT OF A *nifH* DATABASE TO FACILITATE THE TAXONOMIC ASSIGNMENT AND DIVERSITY ANALYSIS OF NITROGEN-FIXING BACTERIA

4.1 Nitrogen fixation

Nitrogen fixation is the reduction of atmospheric dinitrogen gas N_2 to ammonium NH_3. When this biochemical reaction is carried out by microorganisms, it is referred to as biological nitrogen fixation (149, 150). Biological nitrogen fixation entails a series of protonation and reduction steps that includes the co-formation of one H_2 equivalent and the hydrolysis of 16 ATP equivalents. The biological nitrogen fixation reaction is written as: $N_2 + 8H^+ + 8e^- + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16P_i$. Biological nitrogen fixation is central to life because it allows abundant nitrogen from the air, which is relatively inert and non-reactive, to be converted into a form that can be readily metabolized by almost all organisms (16). To put it another way, biological nitrogen fixation allows molecular nitrogen to become ‘bioavailable’. Life as we know it simply would not exist without this critical biochemical process.

Diazotrophs are bacteria and archaea that encode the nitrogenase enzyme complex that catalyzes nitrogen fixation (113). The structure of the nitrogenase enzyme complex is illustrated in Figure 19A. The left half of the complex is shown in the ‘cartoon’ style of the Protein Data Bank (PDB https://www.rcsb.org/), and the right half of the complex is shown in the ‘space filling’ style. However, both halves of the complex are essentially symmetrical. The nitrogenase complex has two main components as shown in the figure:

component I is the MoFe protein, or the dinitrogenase, and component II is the Fe protein, also referred to as the nitrogenase iron protein or dinitrogenase reductase. The left half of the nitrogenase complex also shows the three metalloclusters – labeled F, P, and M – that participate in the reduction of dinitrogen (N_2) to ammonium (NH_3). The M metallocluster (FeMo-co) serves as the catalytic core for nitrogen reduction, and the other two clusters, F (4Fe-4SD) and P (8Fe-4S), facilitate electron transfer from the Fe protein to the M catalytic site (151).

The MoFe protein consists of a heterodimer of two proteins, shown in the blue and green colors, which are encoded by the *nifD* and *nifK* genes. The Fe protein consists of homodimer, each unit of which is encoded by the *nifH* (152). These crucial catalytic *nif* genes are organized into operons, along with a number of other accessory *nif* genes, as shown in Figure 19B. The *nifH*, *nifD*, and *nifK* genes are color coded following the nitrogenase enzyme complex structure in Figure 19A, and the accessory nif genes are categorized as: Fe protein maturation (*nifM* brown), M cluster biosynthesis (*nifY*, *nifE*, *nifN*, *nifX*, *nifB*, *nifQ* purple), F cluster biosynthesis (*nifU*, *nifS*, *nifZ* orange), transcriptional regulation (*nifA* magenta), and unknown function (*nifT*, *nifW*, yellow). The *nifH* gene is the most highly conserved member of the nitrogenase enzyme complex, and it is therefore widely used as both a biomarker for nitrogen fixing capacity as well as a taxonomic marker for the diversity of nitrogen fixing bacteria (153).

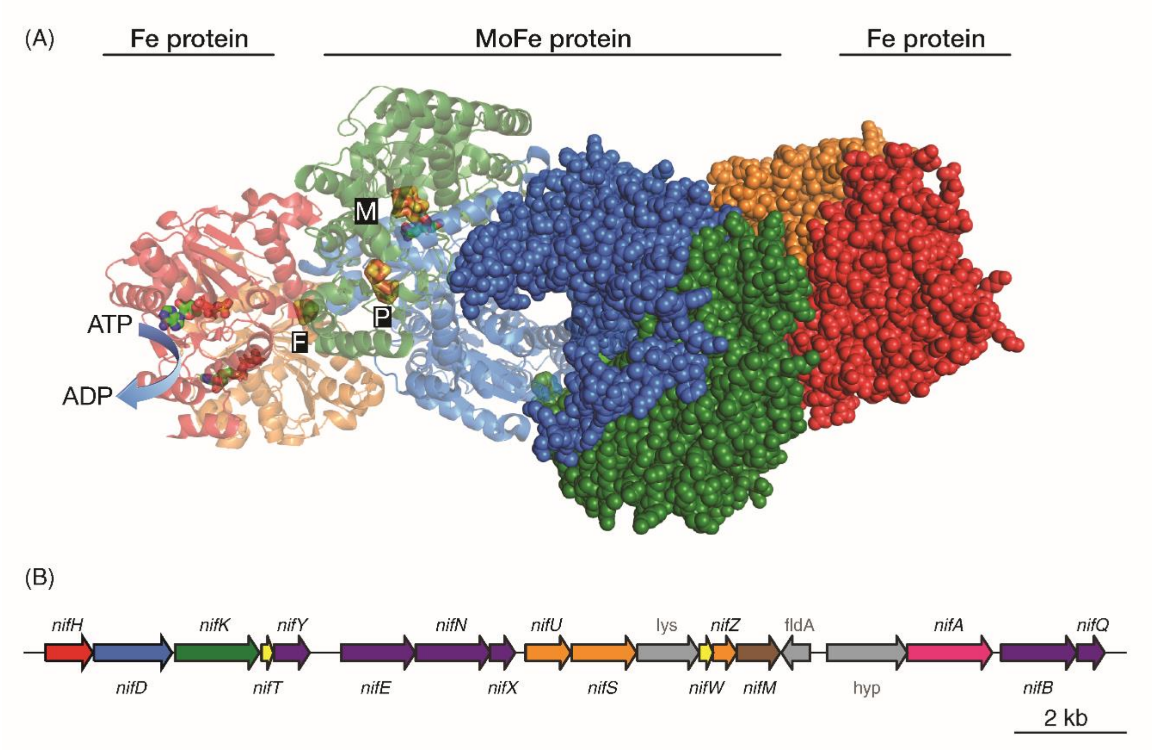


Figure 19. Nitrogenase enzyme complex and nif gene operons.

The nitrogenase enzyme complex is shown in panel A and the nif gene operons are shown in panel B. The relative lengths, locations, and orientations of the nif genes/operons shown are taken from the Klebsiella pnuemoniae genome sequence strain 342 annotations (Genbank accession NC_011283). The nitrogenase enzyme complex and the nif genes are described in detail in the text. The representation of the nitrogenase enzyme complex is adopted from Figure 1 in (151).

4.2 Need for a *nifH* gene database

The microbial ecology research community samples broadly in the natural environment – including water and soils along with plant and animal associated biomes – in order to determine the biochemical capacity and taxonomic composition of the bacterial communities that live therein. More and more, these studies rely on the use of next generation sequencing data, *i.e*. metagenomics. Metagenomic studies can entail a broad and unbiased characterization of all of the DNA in any given environment, or they can use high-throughput sequencing of PCR amplicons to focus on specific genes of interest. Amplicon sequencing studies are most often used to characterize taxonomic markers, such as SSU RNA gene sequences for bacteria and archaea. But amplicon sequencing can also be focused on specific functional markers; this is the case for *nifH* gene amplicon sequencing studies, which aim to characterize the capacity for biological nitrogen fixation found in any given environment. Amplicon sequencing runs can yield hundreds to thousands of *nifH* sequences in a given experiment, depending on the PCR amplification and sequencing technology that is used. This high volume of sequences must be characterized with respect to (1) their identity as *bona fide nifH* genes and (2) their specific taxonomic origin. This characterization is done by comparing the *nifH* amplicon sequences against a sequence database containing known *nifH* gene sequences from a wide variety of microorganisms and environments. Accordingly, the utility of the *nifH* gene as a biomarker for nitrogen fixing capacity and as a taxonomic marker for diazotroph diversity is critically dependent upon the availability of suitable gene sequence databases against which researchers can compare their own sequences of interest.

The most obvious solution to this challenge would be for investigators to compare their *nifH* amplicon sequences to the most comprehensive sequence database that exists – the NCBI Genbank database. However, this is not a practical or reliable solution for a number of reasons. First and foremost, the massive size of the Genbank database mitigates the practicality of direct searches against it. Searches of this kind waste a massive amount of computational resources sifting through sequences that are not directly relevant to the researchers' questions. This not only wastes time and resources for the investigators of any given amplicon sequence study, it can also waste valuable resources of the entire research community when the work is done on a shared compute environment, which is becoming more and more standard these days. For example, Georgia Tech's PACE high-performance compute environment is often bogged down by metagenomics researchers who are performing needlessly slow BLAST searchers of their sequences of interest against Genbank. Searching against Genbank also limits the kind of search algorithms that can be used, and the aforementioned BLAST algorithm is particularly slow compared to more modern sequence similarity comparison algorithms. Furthermore, while it is true that Genbank will contain more *nifH* gene sequences that any other database, the reliability of the sequences deposited therein is highly variable. Curation of sequences deposited to Genbank is left up to individual researchers, and there is no quality control with respect to sequence annotation. For example, since the *nifH* gene, like many other genes, has many different designations, labels of *nifH* sequences can vary widely in the database. This can include errors whereby the *nifH* sequence is simply misnamed or errors in taxonomic assignment. Thus, even when strong, statistically significant sequence similarity is found between an investigator's *nifH* gene amplicon sequence and a database sequence, the database annotation may yield erroneous information regarding the sequence's provenance. Finally, as I will elaborate on below, the vast majority of *nifH* gene sequences found in the Genbank database are either partial gene sequences and/or sequences for which there is no taxonomic information. Sequences of this kind will also mislead investigators' who rely on BLAST-based sequence comparison against Genbank. For all of these reasons, the research community has recognized the need for the development and maintenance of sequence databases dedicated exclusively to the *nifH* gene.

4.3 Problems with previous *nifH* gene databases

Prior to the present study, two main *nifH* gene sequence databases were available to the microbial research community: (1) the curated *nifH* database from Jonathan Zehr's Marine Microbial Ecology laboratory at the University of California Santa Cruz (154) and (2) the comprehensive aligned *nifH* database developed by Chris Gaby and Daniel Buckley at Cornell University (149). The comprehensive aligned *nifH* database was recently coupled with a sequence similarity search tool and released together as the TaxaDiva high-throughput *nifH* amplicon sequencing and analysis pipeline by the laboratory of Joel Kostka at Georgia Tech. In this study, flaws are identified in each of these databases that diminish their potential utility for the research community. In light of these problems, I have developed a new approach for the automated creation and maintenance of a gold-standard *nifH* gene sequence database for the microbial research community. In this chapter, I enumerate the problems posed by the previous databases along with the solutions that my approach provides, and I describe the final *nifH* database in detail. I refer to my final database as the *nifH* gene reference sequence database.

The main problem with previously constructed *nifH* gene sequence repositories is that these custom databases are manually curated by investigators from the *nifH* research community. Thus, each has its own idiosyncrasies that limit its utility and adoption by researchers outside of the lab of origin, and more importantly, custom databases do not allow for automated updates. This latter point is the most critical flaw that is addressed in the database pipeline developed here. Sequence submissions from microorganisms are growing exponentially (Figure 20). New submissions, which will inevitably include many *nifH* sequences from whole genomes or amplicons, have the potential to greatly enhance studies of diazotrophs. However, if new sequences are not properly incorporated into the databases, they are essentially worthless to the field, particularly when you consider the aforementioned problems with direct searching of Genbank for *nifH* sequences. The last updates of the curated *nifH* gene database and the comprehensive aligned *nifH* database occurred in 2017 and 2014, respectively. Clearly, there is a need for a databases that can be automatically updated.

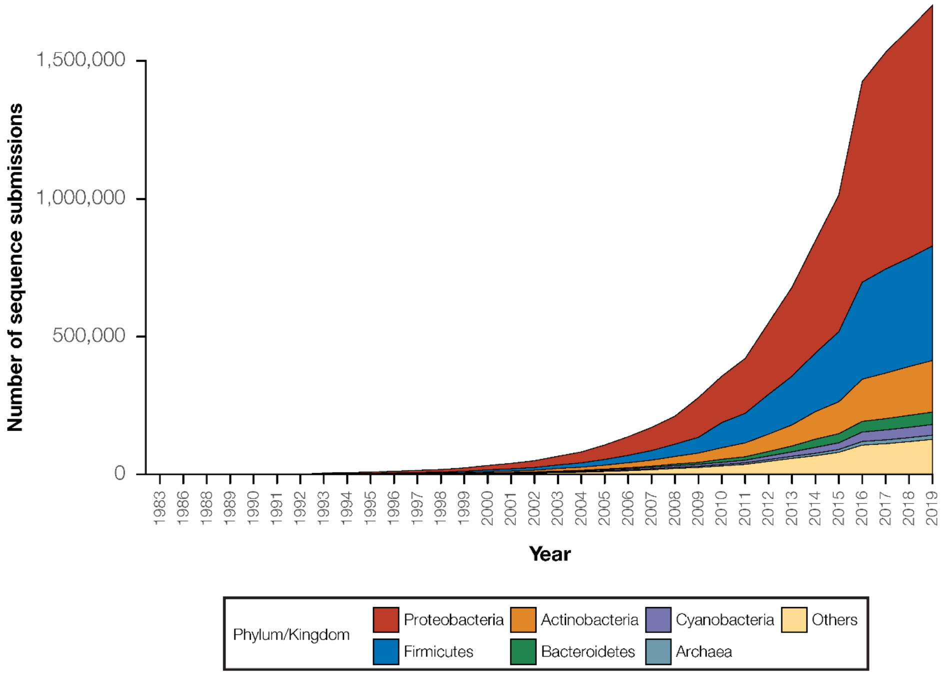


Figure 20. Increase in the number of bacterial and archaeal sequences submitted to Genbank.

The cumulative number of Genbank sequence submissions per year is shown for seven main bacterial and archaeal phyla, colored as shown in the legend.

Another pressing challenge for existing *nifH* databases relates to sequence quality and the reliability of sequence annotations. One advantage of manual, or semi-automated, curation of sequence databases is that it should, in principle, allow for the collection of a higher-quality, validated set of sequences. However, the work presented here with the curated *nifH* gene database suggests that this is far from true. The curated *nifH* database from Jonathan Zehr is maintained using the antiquated ARB file format, which is meant to support graphical user interfaces and is no longer used by the bioinformatics community. For this reason, simple entry into database and manipulation of its sequences is difficult. This is a good example of a boutique database that is intended to be used by the developers and a small group of collaborators. Best practices in bioinformatics are moving away from this kind of database development for obvious reasons.

In this study, the comprehensive aligned *nifH* database was evaluated and improved with state-of-the-art bioinformatics approaches. Surprisingly, it was shown 90% of the sequences in the database can be considered as "uncharacterized" sequences in the sense that they are not associated with any taxonomic label. Such sequences can only be used to confirm the identity of an amplicon sequence as *nifH* and little else. It was also observed that 85% of the sequences in the database are sequence fragments with less than 75% of the expected length range for a complete *nifH* gene sequence (700-1000 bp). Incomplete sequences are problematic as they can cause spuriously high sequence similarity scores with query sequences that are either a different length and/or correspond to a different, but overlapping, amplified region of the *nifH* gene. Both of these problems – uncharacterized sequences and sequence fragments – are caused by the inclusion of numerous *nifH* sequences characterized by amplicon sequencing studies. These are exactly the kinds of problems that should be overcome by manual curation, but this was not the case with this database. Thus, the use of manual curation for the comprehensive aligned *nifH* database simultaneously ensures that it cannot be regularly updated (last update in 2014) while not reducing the number of uninformative or misleading sequences in the database.

4.4 Creation and evaluation of the *nifH* gene reference sequence database

The objective of this study was to create a new and improved *nifH* database that directly addresses the challenges faced by existing databases in a highly dynamic computational environment. The new *nifH* database presented herein includes several distinguishing features including that it is automatically updated and predominantly composed of full-length sequences that are annotated with taxonomic labels. In other words, this new and improved *nifH* sequence database is comprised of gold-standard sequences with maximum utility for investigators with amplicon or metagenome sequences in need of characterization. To meet this objective, a series of automated pipelines was developed to retrieve, validate, and annotate *nifH* sequences from the NCBI Genbank databases. The pipeline was developed in four stages, each representing an improved version of the previous effort. Here, I present results of the final *nifH* database creation automated pipeline, with an emphasis on how this bioinformatics approach generates a more reliable and thereby useful set of sequences.

4.4.1 Automated database pipeline: sequence retrieval and validation

As discussed previously, manual curation of gene sequence databases should, in principle, generate the most reliable set of sequences in the database. However, this is rarely the case in practice. In addition, manual curation is plagued by the challenge to continuously manually update the database in the face of the ever increasing amount of new sequence data generated and deposited to databases (Figure 20). A compromise to these two solutions is semi-automated updates, whereby an initial automated sequence search and retrieval step is done followed by manual curation. This is the approach used by the curated *nifH* gene database from Zehr and colleagues. The semi-automated approach suffers from problems posed by both automated and manual curation approaches, rather than solving them both as it is intended to do. For this reason, a different approach to *nifH* database creation was pursued that involves a long research phase whereby multiple automated pipelines are developed and deployed and then the results of each pipeline, in terms of the characteristics of the sequences that are retrieved and stored in the database, are deeply interrogated. The new approach revealed numerous problems with the automated pipeline. However, rather than manually curating the sequences to resolve the problems, the pipeline was redesigned and rerun to more efficiently and sustainably address the problems. This approach requires more initial effort (the four stages were implemented over a year); however, the product yielded a fully automated pipeline and creation of a reliable database of gold-standard sequences with the capability to be constantly updated, with a minimum amount of additional effort from database administrators.

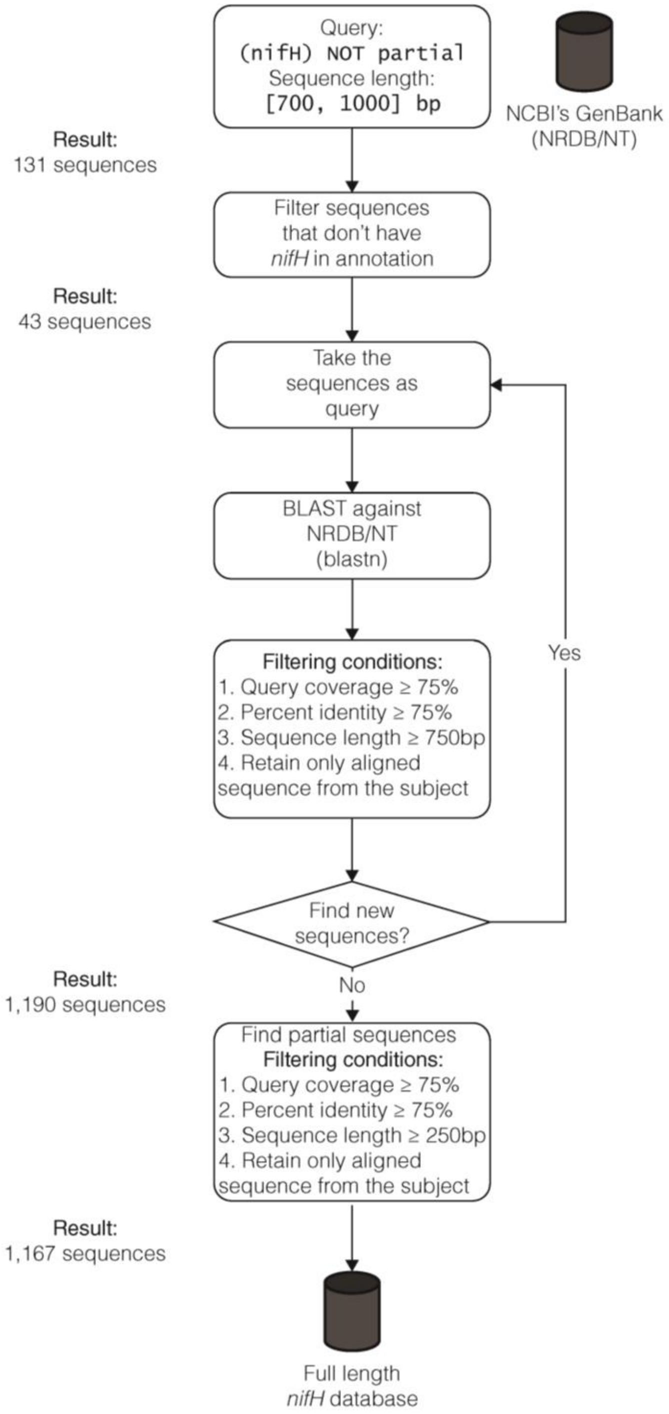


Figure 21. Scheme of the automated pipeline used for the creation of the *nifH* gene reference sequence database.

Steps in the pipeline are described in detail in the text.

The automated pipeline uses a combination of text and sequence similarity searches, together with a stringent set of filtering criteria, to ensure the retrieval of a reliable set of *nifH* gene sequences. The pipeline starts by performing a string search of the Genbank NRDB/NT database annotations for gene sequences that are annotated as *nifH* and are not annotated as partial. A sequence length range is also mandated. The resulting initial set of sequences is then used as a query for a BLASTn sequence similarity search of Genbank NRDB/NT, with a series of filtering conditions that ensure high quality full-length sequences. The process iterates until it converges when no new sequences are retrieved. The number of sequences retrieved by the initial text search and all subsequent BLASTn searches is shown in Figure 22. The database converged after 9 rounds of sequence similarity searching. A total of 1,190 full length *nifH* gene sequences were retrieved in this way. A step to retrieve a minimal set of partial sequences was added to ensure that the most comprehensive set of sequences is retained. The final set consists of 1,167 gold-standard full-length *nifH* sequences and an additional 2,028 partial sequences that are included for the purposes of taxonomic breadth.

Evaluation of several aspects of the resulting *nifH* gene reference sequence database underscored its quality compared to the previously developed comprehensive aligned *nifH* database. For comparison, several iterations of the comprehensive aligned *nifH* database were considered. The first version of this previous *nifH* database was released in 2009, followed by updates in 2012 and finally 2014. First and foremost, while the new database contains an order of magnitude fewer sequences compared to the most recent release of the comprehensive aligned *nifH* database, it contains a much higher quality sequence dataset. Surprisingly, only 134 of 32,954 sequences in the previous database (0.4%) were shown to be full-length, whereas the new database is comprised of 1,167 full-length sequences out of a total of 3,690 sequences (31.6%; Figure 23). Furthermore, the previous database was predominated by 30,989 taxonomically uncharacterized sequences (94.0%), whereas all of the sequences in the new database are taxonomically characterized. The enrichment of taxonomically informative sequences in our database is underscored by a comparison of the total amount of taxonomic diversity found in each database (Figure 24). Our *nifH* gene reference sequence database has far more bacterial/archaeal classes, genera, and species represented than the previous database despite the fact that it only has ~10% the total number of sequences. Finally, our database has a *nifH* gene length distribution that is greatly shifted towards longer, presumably full-length sequences, compared to the previous database (Figure 25). These results underscore the success of our iterated automated approach to *nifH* gene sequence database development and maintenance.

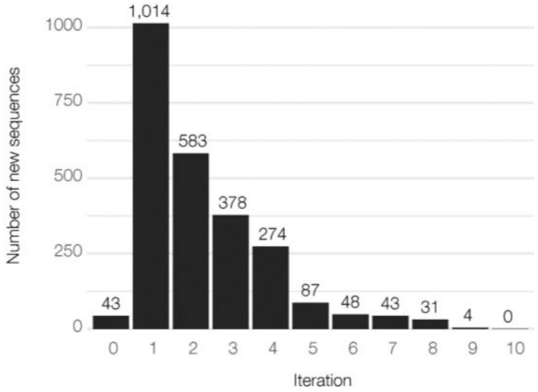


Figure 22. Numbers of new *nifH* gene sequences retrieved for each search iteration of the automated pipeline.

Iteration 0 corresponds to the text search of Genbank, and iterations 1-10 correspond to the BLASTn searches.

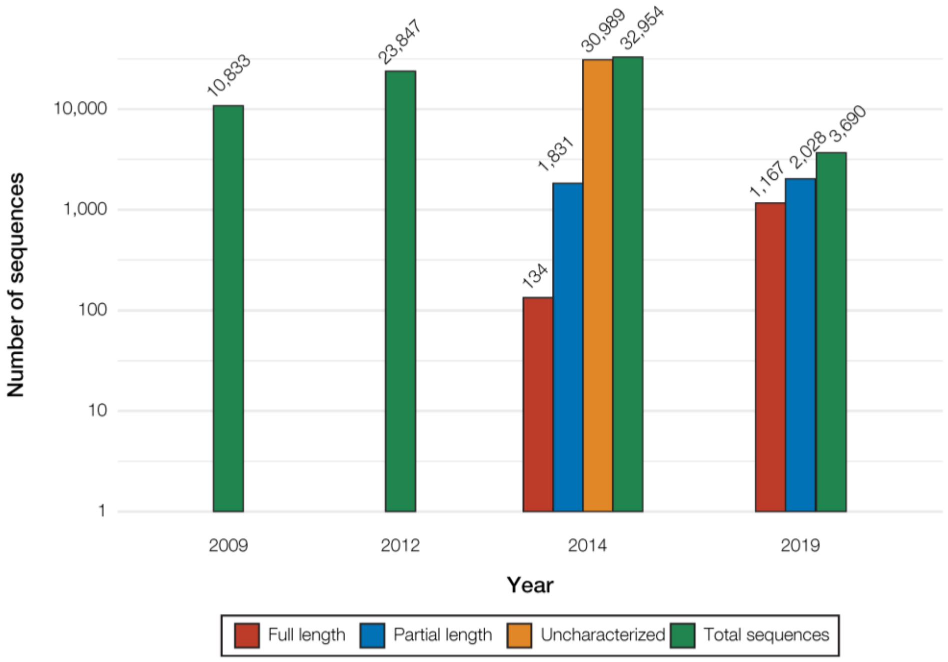


Figure 23. Numbers of sequences included in the comprehensive aligned *nifH* database compared to our new *nifH* gene reference sequence database.

Data are shown for three previous releases of the comprehensive aligned nifH database: 2009, 2012, and 2014. Total numbers of sequences in each database are shown in green along with the numbers of full-length (red), partial (blue), and taxonomically uncharacterized (orange) sequences for the most recent version of each database.

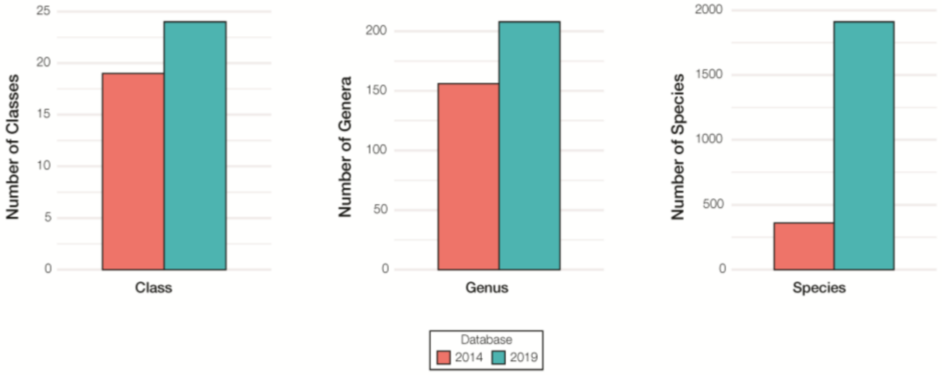


Figure 24. Comparison of the taxonomic diversity of the comprehensive aligned *nifH* database (coral) versus our new *nifH* gene reference sequence database (mint green).

For each database, the numbers of distinct bacterial/archaeal classes, genera, and species are shown.

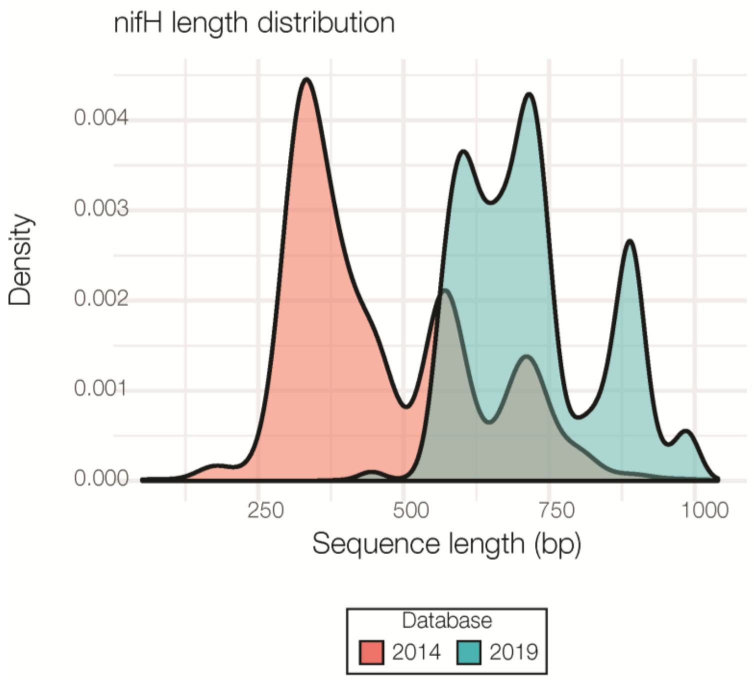


Figure 25. Comparison of the *nifH* gene sequence length distributions for the comprehensive aligned *nifH* database (coral) versus our new *nifH* gene reference sequence database (mint green)

4.5 A natural genetic classification system for *nifH*

Previously, *nifH* gene sequences have been characterized into five clusters based on their sequence diversity and taxonomic distribution: clusters I-V. I only consider clusters I-III here, as clusters IV and V are currently less well characterized. Inconsistencies in previous *nifH* gene sequence databases also suggested that the classification scheme used for *nifH* may have issues. Thus, the new database of gold-standard *nifH* gene sequences was leveraged to construct a more natural classification system for *nifH* gene sequences based explicitly on the observed *nifH* sequence diversity found in the database.

To revise the classification system, all-against-all pairwise global protein sequence alignments were computed using the Smith-Waterman dynamic programming algorithm, implemented in the water program from the EMBOSS suite, for the 1,167 full-length sequences in the database. Pairwise alignments were used to compute pairwise sequence *p*-distances, where *p* is the total number of differences between sequences divided by the alignment length. The resulting pairwise distance matrix was then projected into two dimensions using principal components analysis (PCA; Figure 26A).

The first two principal components (PCs) capture 83.7% of the variance in the pairwise distance matrix (PC1=67.4% and PC2=16.3%) indicating that the PCA is very well powered to capture and visually represent *nifH* sequence diversity. The PCA shows very clear clustering with three apparent major clusters, two of which correspond to Cluster I *nifH* gene sequences and one of which corresponds to both Cluster II and Cluster III sequences.

At first glance, Cluster II and Cluster III *nifH* sequences appeared to overlap somewhat in PC space, suggesting a dissonance between their classification and their actual sequence diversity (Figure 26A). However, this could be an artifact of applying PCA analysis on a very diverse set of sequences. To look at this more closely, PCA analysis was re-run on a pairwise distance matrix containing only Cluster II and Cluster III sequences. This subsequent PCA analysis revealed that Cluster II and Cluster III sequences form distinct clusters consistent with their initial characterization (Figure 26B).

In the cases of both PCA plots, the sequence groupings suggest the possibility of additional clusters that reflect a finer level of sequence structure. For example, there are two very clearly distinct Cluster I groups in the PCA, which are primarily distinguished by PC1, and additional structure is observed within each of these groups along PC2 (Figure 26A). Thus, Cluster I could be represented by four distinct groups, and similarly, Cluster II forms two very distinct groups in the second PCA plot (Figure 26B and Table 15).

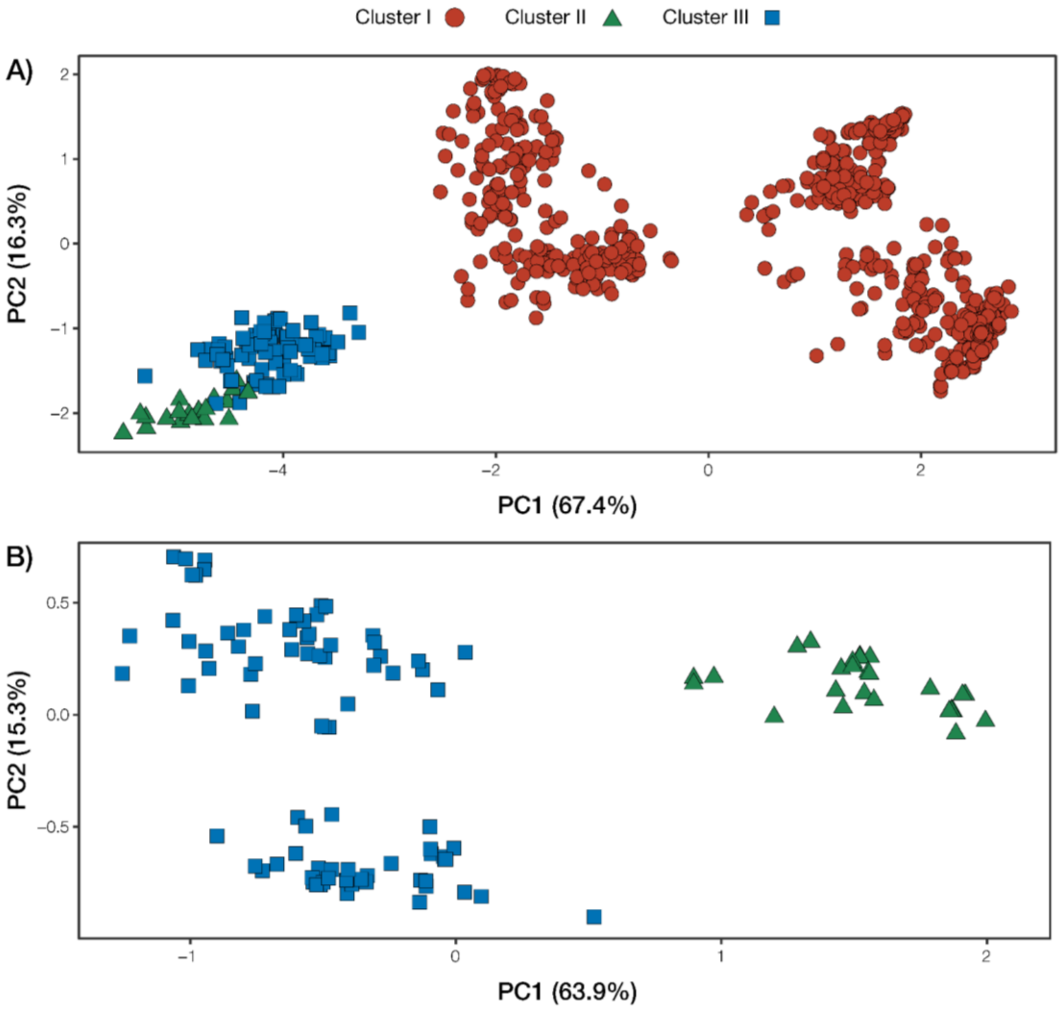


Figure 26. Sequence diversity of *nifH* genes from the *nifH* gene reference sequence database.

The sequence diversity of full-length nifH genes is visualized using principal components analysis (PCA) as described in the text. Individual nifH sequences are shown as red circles (Cluster I), green triangles (Cluster II), or blue squares (Cluster III). (A) All sequences considered together. (B) Cluster II and Cluster III sequences considered separately.

Given the observed sequence structure among the gold-standard *nifH* sequences in the updated database (Figure 26), an unsupervised clustering method was applied on all full-length sequences from the database in an effort to create a more natural genetic classification for *nifH* sequences. A density-based clustering method – HDBSCAN – was employed that operates on the three-dimensional space defined by the first three PCs to identify coherent clusters of *nifH* sequences. This approach is a modification of the DBSCAN method implemented in python (155). Importantly, the clustering algorithm does not require a pre-defined number of groups; instead, it relies entirely on the structure of the data to delineate clusters and to decide on the most optimal number of clusters. In other words, this density-based clustering method best reflects the goal to create a natural genetic classification for *nifH* that rests solely on observed sequence diversity, as opposed to any expert user input. As discussed previously in the context of the comprehensive aligned *nifH* database, and again below with respect to the taxonomic classification accuracy of the linked TaxADivA application (70), while expert user input is typically considered to be desirable for biological databases and classification, it can also introduce and propagate numerous errors along with false sense of confidence. Accordingly, the proposed classification scheme is objective rather than subjective. The new scheme does not necessarily replace the current classification system, but rather, it provides a possible alternative that investigators in the field may wish to consider to inform future efforts.

Density-based clustering of full-length *nifH* sequences using the HDBSCAN algorithm yields five distinct clusters in three-dimensional space. A two-dimensional projection of this clustering scheme is shown in Figure 27. It should be noted that the density-based clustering algorithm was not able to classify 198 out of 1,183 full-length *nifH* sequences, *i.e.* 17% of *nifH* sequences fall between the clusters identified by HBDSCAN. Interestingly, all of these sequences fall between clusters 3 and 4, and they correspond to sequences that are currently classified as cluster I (see Figure 26). This is an artifact of the clustering method and does not necessarily reflect the fact that these sequences are unclassifiable. In the future, when more sequences become available, these sequences could be grouped unambiguously into a single cluster. Nevertheless, a 17% failure rate is unacceptably high for a clustering method that aims to objectively classify sequences. Thus, the best solution for *nifH* classification may actually entail a semi-automatic method that uses an initial round of clustering followed by expert user annotation that combines visual inspection of the clustered and un-clustered sequences in PC space together with known taxonomic and functional information regarding the un-clustered sequences.

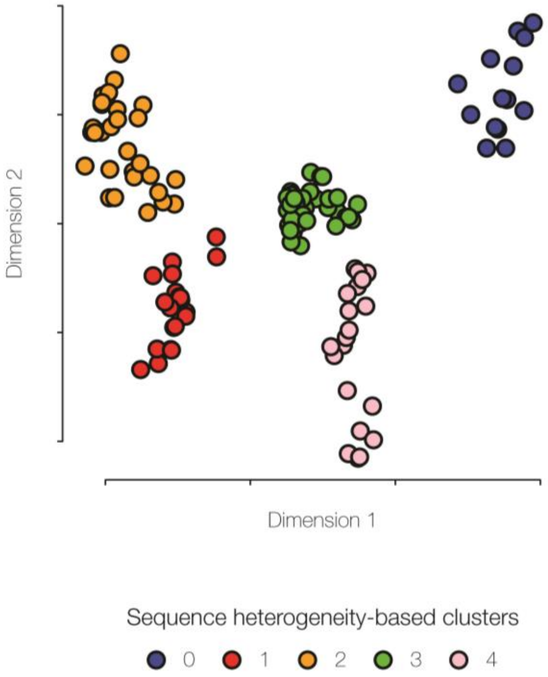


Figure 27. Density-based clustering of *nifH* genes from the *nifH* gene reference sequence database.

Details of the clustering method are described in the text. The five discrete clusters are color-coded as shown in the legend. Cluster 0 corresponds to clusters II and III in Figure 25. Clusters 1-4 correspond to cluster I in Figure 25. The distinction between the previous annotations and observed clustering underscore the need for a more natural classification system for nifH.

As can be seen by comparing the results of the initial *nifH* PCA in Figure 26 and the subsequent density-based clustering in Figure 27, most of the *nifH* sequence diversity (clusters 1-4 in Figure 27) is captured by the previously annotated cluster I. This distinction leads to the hypothesis that the natural genetic clusters of *nifH* may better capture the distribution of taxonomic diversity than the previous classification system. To test this hypothesis, taxonomic classification was reevaluated for all of the full-length sequences in the new *nifH* database using the TaxADivA program (70).

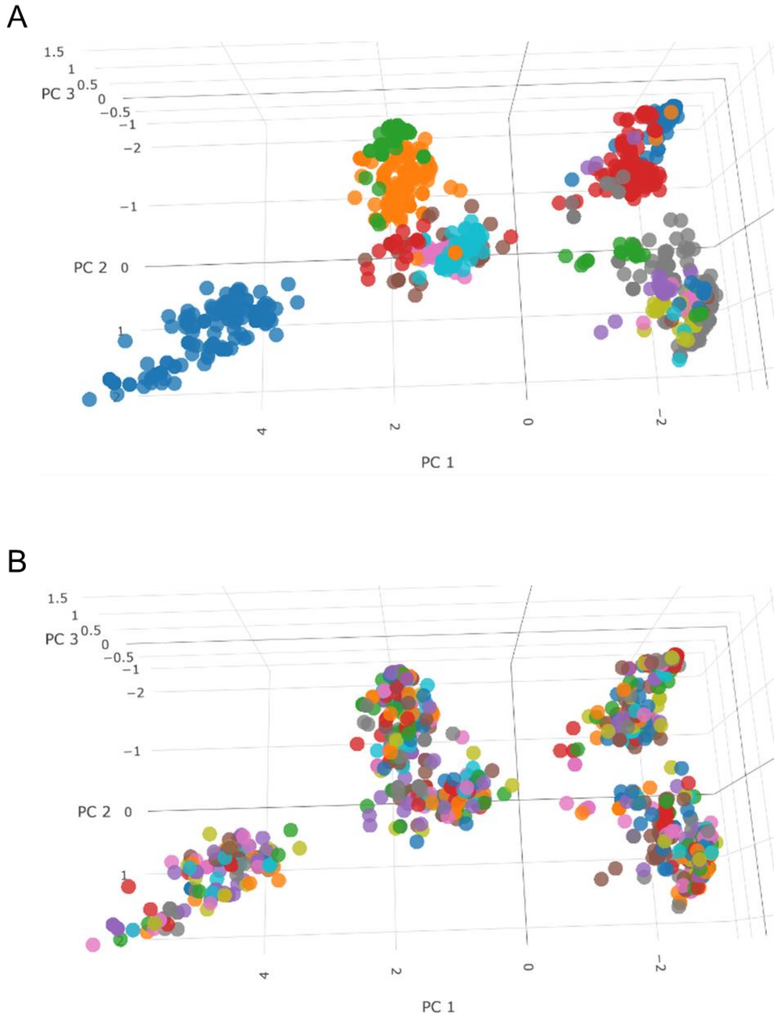


Figure 28. Taxonomic labeling of *nifH* sequence clusters.

Nodes are color-coded based on the species identity. Note that individual species names are not shown here; thus, color can be taken to represent the total amount of taxonomic diversity for each cluster and the distribution of taxonomic diversity among clusters. (A) Taxonomic labeling based on TaxADivA. (B) Taxonomic labeling based on the NCBI Genbank database. The Genbank taxonomic labels are the true labels.

The results of the taxonomic labeling analysis with TaxADivA were both unexpected and alarming. TaxADivA assigned a relatively small number of distinct species and each cluster was made up of only a few species (Figure 28A). One cluster was made up entirely of a single species: *Clostridium pasteurianum*. This cluster corresponds to the previously defined clusters II and III together, yet all of the sequences therein were somehow mapped to the same individual sequence and same species in the TaxADivA database. This seemingly makes no sense whatsoever and calls into question (1) the validity of the TaxADivA approach to sequence similarity searching and/or (2) the validity of the *nifH* sequences and taxonomic labels curated in the previously developed comprehensive *nifH* database that is coupled to TaxADivA. TaxADivA uses a standard BLASTn algorithm for comparing query to database sequences, and it selects the best hit as the hit with the highest bit score (assuming it passes a sequence similarity threshold). Of course, BLAST is the most widely used bioinformatics algorithm, and it has been extensively vetted over the years (92). BLAST should also have no problem quantifying sequence similarity and best hits at the level of sequence identity needed to classify *nifH* amplicon sequences. Accordingly, it is highly unlikely that the problem lies with the TaxADivA sequence similarity search algorithm. Therefore, it is concluded that the problem must lie with the comprehensive *nifH* database.

The comprehensive *nifH* database was previously shown to be comprised of nearly all partial sequences. However, it is unclear how a predominance of partial sequences could lead to the kind of result seen in Figure 28A. Thus, to address this inconsistency, full-length *nifH* sequences from the new database were reclassified using their Genbank accession numbers. It is important to note that these full-length *nifH* sequences were iteratively retrieved from Genbank using the pipeline illustrated in Figure 21. In other words, these sequences were recently taken directly from Genbank using a fully automated method, and therefore their species labels should therefore faithfully represent their taxonomic origins. Classification of clustered full-length *nifH* sequences in this way yields a completely different set of species labels (Figure 28B), and each cluster is highly diverse compared to what is observed for the TaxADivA taxonomic labels (Figure 28A). Taking the Genbank species labels as the ground truth for their taxonomy, yields an abysmal accuracy rate of 8.7% for taxonomic classification with TaxADivA. This is obviously an unacceptable level of accuracy for use in any amplicon based environmental study of *nifH*. Examination of the previously developed *nifH* comprehensive database suggest that it is not only dominated by partial sequences, but that the sequences in the database are wildly mislabeled.

Short term and long term solutions to this problem are proposed here. In the short term, it is proposed that the new gold-standard *nifH* reference database should replace the previously developed comprehensive *nifH* database and be coupled with the TaxADivA pipeline for taxonomic classification of *nifH* amplicon sequencesOver the long term, a new and improved sequence similarity search algorithm will be developed, based on the k-mer paradigm used for the previously published stringMLST algorithm (156), which will be coupled with the *nifH* reference database.

CHAPTER 5. CONCLUSIONS AND FUTURE PROSPECTS

Sustainable agriculture aims for the development of environmentally friendly agricultural practices to increase crop production. Biofertilizers represent one of the most promising areas of research and development in support of more sustainable agricultural practices. Biofertilizers consist of beneficial bacteria that promote plant growth, and they are an alternative or complementary approach, compared to traditionally employed chemical fertilizers, for increasing crop yield. For this thesis research, members of the Kostka and Jordan laboratories at Georgia Tech collaborated with a Colombian sugarcane company to help them increase the production of sugarcane biomass, while reducing the use of damaging and costly chemical fertilizers, by isolating and characterizing plant growth-promoting bacteria from sugarcane.

As described in Chapter 2, a variety of microbial and genomic-based techniques were employed to isolate and characterize nitrogen-fixing and plant growth-promoting (PGP) bacteria native to Colombian sugarcane fields. A high-throughput systematic cultivation approach was used to isolate pure cultures of nitrogen-fixing bacteria, corresponding to 22 unique strains that were evaluated for diazotrophic potential by PCR amplification of nitrogenase (*nifH*) genes. Most of the strains obtained in this study belong to the Enterobacteriaceae family, with the *Klebsiella* as the most abundant genus.

In addition to the culture-based techniques, a genome-enabled approach was developed to prioritize native bacterial isolates with the potential to serve as biofertilizers for sugarcane fields in Colombia. As discussed in Chapter 3, the approach was based on computational phenotyping, which entails predictions related to traits of interest, based on bioinformatic analysis of whole-genome sequences. The computational phenotyping approach rapidly screened for strains that have a high potential for nitrogen fixation and other PGP phenotypes while showing minimal risk for virulence and antibiotic resistance. In addition to their positive PGP properties, most sugarcane isolates fell below a genotypic and phenotypic threshold, showing uniformly low predicted virulence and antibiotic resistance factors compared to clinical isolates. Bioinformatic predictions of the presence of plant growth-promoting traits were confirmed in six prioritized strains using experimental assays under laboratory conditions. This research demonstrated the utility of computational phenotyping for assessing the benefits and risks posed by bacterial isolates that can be used as biofertilizers.

As discussed in Chapter 4, the development of an improved *nifH* gene database to facilitate taxonomic assignment and diversity analysis of nitrogen-fixing microbial communities is crucial for the study of diazotrophs in agricultural and natural ecosystems. The *nifH* gene is a standard molecular marker for nitrogen fixation, and the assessment of *nifH* sequences available in public databases facilitates the study of nitrogen fixing-bacteria. In this thesis, an improved *nifH* database was developed through the automated retrieval of all published full-length *nifH* gene sequences from the NCBI non-redundant database to improve the taxonomic assignment of sequences generated from metagenomic and *nifH* gene amplicon studies. This work identified a number of crucial problems with the existing approach to *nifH* taxonomic characterization and diversity analysis and proposed a number of possible solutions that can be pursued for future research.

To briefly recapitulate the main findings of my research, this thesis explored the potential of native plant growth-promoting bacteria from a Colombian sugarcane fields to serve as biofertilizers, thereby contributing to the development of sustainable agriculture practices. I reported the potential of selected bacterial strains that can fix nitrogen and promote plant growth as biofertilizers while lacking many of the critical virulence factors that are in clinical isolates of the same species; therefore, these isolates are good candidates for biofertilizers.

While this work is promising, it is only a step in the direction of using biofertilizers for sustainable agriculture. More data is needed to directly demonstrate the capability of biofertilizers to sustainably promote crop plant growth. Perhaps most importantly, follow-up greenhouse and field studies are required to validate the best candidate isolates under environmental conditions for their performance as biofertilizers. This future work will be performed by INCAUCA scientists in Colombia. Furthermore, despite significant efforts to characterize and formulate biofertilizers for agricultural settings, the actual mechanisms that support their ability to increase plant growth remain poorly understood. Most importantly, there is a lack of research and methodologies that accurately evaluate the performance of biofertilizers and their metabolic activity in the field. Additional work on tracking the activity of biofertilizers, and the macronutrients that they provide to plants, in greenhouse and field conditions are needed to better understand the mechanisms that underlie the utility of biofertilizers.

A deep understanding of the environmental and human health risks posed by biofertilizer inoculants is also essential, given that they are made from living bacteria and applied directly to crop fields that are integrated with the wider environment. The quantitative approach to computational phenotyping developed in this thesis can serve to narrow the search for potential plant growth-promoting bacterial isolates, which can be further assessed via experimental methods. It is important to reiterate that computational phenotyping can not only be used to find the beneficial properties of plant-associated bacterial strains, but it also has important negative predictive utility with respect to avoiding any potentially harmful characteristics. Here, it should be stressed that the computational phenotyping approach has potential that goes far beyond the scope of research activities, focused on biofertilizers, described in this thesis. Genome sequencing followed by computational phenotyping could be applied to a broad range of applications in environmental and industrial microbiology, food safety, water quality, and antibiotic resistance studies. The key to this approach, as validated in this thesis via the integration of computational analysis and experimental assays, is the ability to make accurate phenotypic predictions that save can downstream time, effort, and cost by focusing subsequent research and development efforts on bacterial strains that are simultaneously predicted to be most beneficial and least harmful for any given microbiology application.

APENDIX A. SUPPLEMENTARY TABLES FOR CHAPTER 2

Table 9. Agro ecological zones sampled for this project.

Agroecological Zone	Number of Fields Sampled
10H3	2
10H5	4
11H0	2
11H1	2
11H3	6
27H3	2
28H3	2
31H3	2
6H1	3
6H2	4
6H3	3
8H3	4
Guadual (control)	1
Total	**37**

Table 10. Sugar cane varieties sampled for this project.

Variety	Number of Fields Sampled
CC85-92	21
CC93-4418	7
CC84-75	4
CC01-1940	2
CC97-7170	1
SP71-6949	1
Guadual (control)	1
Total	37

APENDIX B. SUPPLEMENTARY TABLES FOR CHAPTER 3

Table 11. Presence/absence calls for nitrogen fixation genes panel.

Nitrogen fixation genes panel profiles are shown for the 22 sugarcane-associated bacterial isolates characterized here. These are the data that are visualized as a heatmap in Figure15.

Nif genes											
Sample	***nifH***	***nifD***	***nifJ***	***nifF***	***nifA***	***nifL***	***nifE***	***nifQ***	***nifB***	***nifV***	***nifS***
SCK1	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK6	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK7	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK8	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK9	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK10	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK11	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK12	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK13	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK14	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK15	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK16	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK17	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK18	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK19	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK20	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present
SCK21	Present	Present	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK22	Present	Present	Absent	Absent	Present	Present	Present	Present	Present	Present	Present

Table 11 (continued).

Nif Genes										
Sample	***nifM***	***nifY***	***nifK***	***nifN***	***nifU***	***nifW***	***nifT***	***nifZ***	***nifX***	***nifO***
SCK1	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Present	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK6	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK7	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK8	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK9	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK10	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK11	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK12	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK13	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK14	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK15	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK16	Present	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK17	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK18	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK19	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK20	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK21	Absent	Absent	Absent	Present	Present	Absent	Absent	Present	Present	Absent
SCK22	Absent	Absent	Absent	Present	Present	Absent	Present	Present	Present	Absent

Table 12. Presence/absence calls for plant growth promotion genes panel.

Plant growth promotion genes panel profiles are shown for the 22 sugarcane-associated bacterial isolates characterized here. These are the data that are visualized as a heatmap in Figure15.

	Phosphate Solubilitation				**IAA**	**Siderophere production**							
Sampl	***pstA***	***pstB***	***pstC***	***pstS***	***ipdC***	***pvdO***	***pvdN***	***pvdP***	***pvdE***	***pvdT***	***fpvA***	***mbtH***	***dhbB***
SCK1	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK6	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK7	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK8	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK9	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK10	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK11	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK12	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK13	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK14	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK15	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK16	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK17	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK18	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK19	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK20	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK21	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK22	Absent	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent

Table 12 (continued).

	ACC deaminase activity					Acetoin & butanediol			Peroxidases			
Sample	***dhbE***	***fhu***	***acdS***	***rimM***	***dcyD***	***budA***	***budC***	***poxB***	***osmC***	***oxyR***	***efeB***	***tpx***
SCK1	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Present	Absent	Absent	Present	Absent	Absent	Absent	Present	Absent	Absent	Absent	Absent
SCK6	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK7	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK8	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Absent	Absent	Absent
SCK9	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK10	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK11	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK12	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK13	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK14	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK15	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK16	Present	Absent	Absent	Present	Absent	Absent	Absent	Present	Present	Present	Absent	Absent
SCK17	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK18	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK19	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK20	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent
SCK21	Present	Absent	Absent	Present	Absent	Absent	Absent	Present	Absent	Present	Absent	Absent
SCK22	Present	Absent	Absent	Present	Absent	Absent	Absent	Present	Absent	Present	Absent	Absent

Table 13. Presence/absence calls for virulence factor genes panel.

Virulence factors genes panel profiles are shown for the 22 sugarcane-associated bacterial isolates characterized here. These are the data that are visualized as a heatmap in Figure15.

Adherence											
Sample	***csgD***	***cheW***	***fliA***	***fliH***	***flil***	***fliM***	***fliN***	***fliZ***	***flgI***	***motA***	***pilT***
SCK1	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Present
SCK2	Absent	Absent	Present	Absent	Absent	Present	Absent	Present	Absent	Present	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Present	Absent	Absent	Absent	Absent	Present	Absent	Present	Absent	Present
SCK5	Present	Absent	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent
SCK6	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK7	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK8	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK9	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK10	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK11	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK12	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK13	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK14	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK15	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK16	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK17	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK18	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK19	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK20	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK21	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK22	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent

Table 13 (continued).

Adherence											
Sample	***pilR***	***pilG***	***pilU***	***yagZ/ecpA)***	***yagY/ecpB***	***yagX/ecpC)***	***yagW/ecpD***	***yagV/ecpE***	***(ykgK/ecpR***	***entA***	***entB***
SCK1	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Absent	Present
SCK6	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
SCK7	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK8	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK9	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK10	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Present	Present
SCK11	Absent	Absent	Absent	Present	Present	Present	Present	Present	Absent	Present	Present
SCK12	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK13	Absent	Absent	Absent	Present	Absent	Present	Present	Present	Present	Present	Present
SCK14	Absent	Absent	Absent	Present	Present	Present	Absent	Present	Present	Present	Present
SCK15	Absent	Absent	Absent	Absent	Present	Present	Present	Present	Present	Absent	Absent
SCK16	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK17	Absent	Absent	Absent	Present	Absent	Absent	Present	Present	Present	Present	Present
SCK18	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK19	Absent	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present
SCK20	Absent	Absent	Absent	Present	Present	Present	Present	Present	Present	Present	Present
SCK21	Absent	Absent	Absent	Absent	Absent	Present	Absent	Absent	Absent	Absent	Absent
SCK22	Absent	Absent	Absent	Absent	Present	Absent	Present	Present	Present	Present	Present

Table 13 (continued).

Siderophores											
Sample	***entC***	***entE***	***entF***	***entS***	***fes***	***fepA***	***fepC***	***fepD***	***fepG***	***fimH***	***fyuA***
SCK1	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Absent	Present	Absent	Absent	Present	Present	Present	Present	Present	Absent	Absent
SCK6	Absent	Present	Absent	Absent	Absent	Present	Present	Present	Present	Absent	Absent
SCK7	Absent	Present	Absent	Absent	Absent	Present	Present	Present	Present	Absent	Absent
SCK8	Absent	Present	Absent	Absent	Present	Present	Present	Present	Present	Absent	Absent
SCK9	Absent	Present	Absent	Absent	Absent	Present	Present	Absent	Present	Absent	Absent
SCK10	Absent	Present	Absent	Absent	Absent	Present	Present	Absent	Present	Absent	Absent
SCK11	Absent	Present	Present	Absent	Absent	Present	Present	Absent	Present	Absent	Absent
SCK12	Present	Present	Absent	Absent	Absent	Absent	Present	Absent	Present	Absent	Absent
SCK13	Present	Present	Absent	Absent	Absent	Present	Present	Present	Present	Absent	Absent
SCK14	Present	Present	Absent	Absent	Absent	Present	Present	Present	Present	Absent	Absent
SCK15	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK16	Absent	Present	Present	Absent	Present	Absent	Present	Present	Present	Present	Absent
SCK17	Absent	Present	Absent	Absent	Absent	Present	Present	Present	Present	Absent	Absent
SCK18	Present	Absent	Absent	Absent	Absent	Present	Present	Present	Present	Absent	Absent
SCK19	Absent	Present	Absent	Absent	Absent	Present	Present	Absent	Present	Absent	Absent
SCK20	Present	Present	Present	Absent	Absent	Absent	Present	Absent	Present	Absent	Absent
SCK21	Absent	Present	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Present
SCK22	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent

Table 13 (continued).

	Siderophores									Invation	Capsules
Sample	***irp1***	***irp2***	***ybtA***	***ybtE***	***ybtP***	***ybtQ***	***ybtS***	***ybtT***	***ybtX***	***hsiC1/vipB***	***wzi***
SCK1	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK2	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent
SCK3	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK4	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SCK5	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent
SCK6	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK7	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK8	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
SCK9	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK10	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK11	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK12	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK13	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK14	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK15	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK16	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
SCK17	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK18	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK19	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK20	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
SCK21	Present	Present	Present	Present	Present	Present	Present	Present	Present	Absent	Present
SCK22	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present

Table 14. **Presence/absence calls for predicted antibiotic resistance profiles genes panel.**

Predicted antibiotic resistance profiles genes panel profiles are shown for the 22 sugarcane-associated bacterial isolates characterized here. These are the data that are visualized as a heatmap in Figure15.

Antibiotic resistance genes										
Sample	**Amikacin**	**Cefuroxime sodium**	**Gentamicin**	**Piperacillin/Tazobactam**	**Tetracycline**	**Trimethoprim/Sulfamethoxazole**	**Cefoxitin**	**Ceftriaxone**	**Nitrofurantoin**	**Ampicillin/Sulbactam**
SCK1	S	S	S	S	S	S	S	S	I	I
SCK2	S	S	S	S	S	S	S	S	I	I
SCK3	S	S	S	S	S	S	S	S	S	R
SCK4	S	S	S	S	S	S	S	S	S	S
SCK5	S	S	S	S	S	S	S	S	I	S
SCK6	S	S	S	S	S	S	S	S	S	S
SCK7	S	S	S	S	S	S	S	S	S	S
SCK8	S	S	S	S	S	S	S	S	S	S
SCK9	S	S	S	S	S	S	S	S	S	S
SCK10	S	S	S	S	S	S	S	S	S	S
SCK11	S	S	S	S	S	S	S	S	S	S
SCK12	S	S	S	S	S	S	S	S	S	S
SCK13	S	S	S	S	S	S	S	S	S	S
SCK14	S	S	S	S	S	S	S	S	S	S
SCK15	S	S	S	S	S	S	I	S	S	S
SCK16	S	S	S	S	S	S	S	S	S	S
SCK17	S	S	S	S	S	S	S	I	S	S
SCK18	S	S	S	S	S	S	S	S	S	S
SCK19	S	S	S	S	S	S	S	S	S	S
SCK20	S	S	S	S	S	S	S	S	S	S
SCK21	S	S	S	S	S	S	S	S	S	S
SCK22	S	S	S	S	S	S	S	S	S	S

Table 14 (continued).

Antibioc Resistence Genes										
Sample	**Tobramycin**	**Cefepime**	**Aztreonam**	**Cefazolin**	**Ciprofloxacin**	**Ampicillin**	**Ceftazidime**	**Imipenem**	**Levofloxacin**	**Meropenem**
SCK1	I	I	I	I	R	R	R	R	R	R
SCK2	I	I	I	I	R	R	R	R	R	R
SCK3	I	I	I	R	R	R	R	R	R	R
SCK4	I	I	I	R	R	I	R	R	R	R
SCK5	I	I	R	I	R	R	R	R	R	R
SCK6	I	I	R	I	R	R	R	R	R	R
SCK7	I	I	I	I	R	R	R	R	R	R
SCK8	I	I	I	I	R	R	R	R	R	R
SCK9	I	I	I	I	R	S	R	R	R	R
SCK10	I	I	I	I	R	R	R	R	R	R
SCK11	I	I	R	I	R	R	R	R	R	R
SCK12	I	I	R	I	R	R	R	R	R	R
SCK13	I	I	I	R	R	R	R	R	R	R
SCK14	I	I	I	I	I	R	R	R	R	R
SCK15	I	I	R	I	I	R	R	R	R	R
SCK16	I	I	I	R	R	R	R	R	R	R
SCK17	I	I	I	I	I	R	R	R	R	R
SCK18	I	I	I	I	I	R	R	R	R	R
SCK19	I	I	I	I	I	R	R	R	R	R
SCK20	I	I	R	I	R	R	R	R	R	R
SCK21	I	I	R	I	R	R	R	R	R	R
SCK22	I	I	R	I	R	R	R	R	R	R

APENDIX C. SUPPLEMENTARY TABLES FOR CHAPTER 4

Table 15. Lis of all full length sequences used for the genetic classification for *nifH*

Accession	Taxonomy	Length	Cluster
AB094963.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium elkanii*	783	cluster I
AB126254.2	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacterium;Heliobacterium gestii*	849	cluster I
AB189453.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas azotifigens*	879	cluster I
AB189641.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae;Halorhodospira;Halorhodospira halophila*	873	cluster I
AB191041.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacterium;Heliobacterium chlorum*	789	cluster I
AB191042.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacterium;Heliobacterium gestii*	786	cluster I
AB191043.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacterium;Heliobacterium modesticaldum*	756	cluster I
AB191044.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacillus;Heliobacillus mobilis*	789	cluster I
AB191045.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliorestis;Heliorestis baculata*	768	cluster I
AB191046.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliorestis;Heliorestis daurensis*	777	cluster I
AB196476.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum;Herbaspirillum sp. B501*	876	cluster I
AB196525.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacterium;Heliobacterium chlorum*	849	cluster I
AB233495.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Amorphomonas;Amorphomonas oryzae*	762	cluster I

Table 15 (continued).

AB293988.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Zehria;Zehria sp. KO68DGA*	900	cluster I
AB303392.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae;Azorhizobium;Azorhizobium caulinodans*	888	cluster I
AB303393.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae;Azorhizobium;Azorhizobium caulinodans*	888	cluster I
AB485747.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus graminis*	864	cluster I
AB489070.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus fujiensis*	861	cluster I
AB489139.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus abekawaensis*	843	cluster I
AB519450.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium loti*	714	cluster I
AB684292.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	753	cluster I
AB684295.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	753	cluster I
AB684427.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	753	cluster I
AB685466.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	882	cluster I
AB758605.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	894	cluster I
AB808482.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Leptolyngbyaceae;Leptolyngbya;Leptolyngbya boryana*	861	cluster I
AE001437.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium acetobutylicum*	813	cluster III

Table 15 (continued).

AE006469.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
AE006470.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobaculum;Chlorobaculum tepidum*	813	cluster III
AE008692.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
AE010299.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina acetivorans*	801	cluster II
AE017180.2	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter sulfurreducens*	855	cluster I
AE017282.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylococcus;Methylococcus capsulatus*	879	cluster I
AE017286.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio vulgaris*	825	cluster III
AF003336.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Crocosphaera;Crocosphaera subtropica*	981	cluster I
AF012326.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7120*	894	cluster I
AF016484.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Microcoleaceae;Trichodesmium;Trichodesmium erythraeum*	771	cluster I
AF030414.3	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Gluconacetobacter;Gluconacetobacter diazotrophicus*	843	cluster I
AF031817.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
AF167538.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Microcoleaceae;Trichodesmium;Trichodesmium erythraeum*	771	cluster I

Table 15 (continued).

AF194084.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;bacterium*	840	cluster I
AF200742.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus sp. BH72*	885	cluster I
AF218126.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium gallicum*	780	cluster I
AF266462.3	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	816	cluster III
AF547999.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferrooxidans*	873	cluster I
AH000924.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp.*	846	cluster I
AH010242.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	894	cluster I
AJ299453.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus durus*	861	cluster I
AJ302315.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia tuberum*	858	cluster I
AJ457912.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium mediterraneum*	735	cluster I
AJ457913.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	735	cluster I
AJ457914.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	735	cluster I
AJ457917.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium mediterraneum*	735	cluster I
AJ515294.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus durus*	861	cluster I

Table 15 (continued).

AJ567343.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;Burkholderia vietnamiensis*	759	cluster I
AL672114.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium japonicum*	849	cluster I
AM114193.2	*Archaea;Euryarchaeota;Methanomicrobia;Methanocellales;Methanocellaceae;Methanocella;Methanocella arvoryzae*	807	cluster III
AM236084.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	888	cluster I
AM406670.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus sp. BH72*	885	cluster I
AM889285.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Gluconacetobacter;Gluconacetobacter diazotrophicus*	843	cluster I
AP007255.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum magneticum*	879	cluster I
AP009049.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium kluyveri*	813	cluster II
AP009384.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae;Azorhizobium;Azorhizobium caulinodans*	888	cluster I
AP010904.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio magneticus*	771	cluster III
AP010946.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum lipoferum*	840	cluster I
AP012206.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium japonicum*	894	cluster I
AP012279.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium sp. S23321*	888	cluster I

Table 15 (continued).

AP012292.1	*Bacteria;Firmicutes;Negativicutes;Selenomonadales;Selenomonadaceae;Selenomonas;Selenomonas ruminantium*	765	cluster III
AP012304.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus sp. KH32C*	861	cluster I
AP012320.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;unclassified family;Rubrivivax;Rubrivivax gelatinosus*	873	cluster I
AP012342.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferrooxidans*	873	cluster I
AP012549.1	*Bacteria;Cyanobacteria;unclassified class;unclassified order;unclassified family;unclassified genus;cyanobacterium endosymbiont of Epithemia turgida*	870	cluster I
AP012557.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium loti*	849	cluster I
AP012603.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium oligotrophicum*	885	cluster I
AP014633.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Thiotrichaceae;Thioploca;Thioploca ingrica*	855	cluster I
AP014636.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio tritonius*	876	cluster I
AP014638.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Leptolyngbyaceae;Leptolyngbya;Leptolyngbya boryana*	861	cluster I
AP014642.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Leptolyngbyaceae;Leptolyngbya;Leptolyngbya boryana*	861	cluster I
AP014683.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;unclassified family;unclassified genus;Burkholderiales bacterium GJ-E10*	882	cluster I
AP014685.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium diazoefficiens*	894	cluster I

Table 15 (continued).

AP014800.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodovulum;Rhodovulum sulfidophilum*	873	cluster I
AP014854.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Blastochloris;Blastochloris viridis*	831	cluster I
AP014950.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
AP014951.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	858	cluster I
AP017295.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. NIES-3756*	888	cluster I
AP017305.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Fischerella;Fischerella sp. NIES-3754*	882	cluster I
AP017367.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Leptolyngbyaceae;Leptolyngbya;Leptolyngbya sp. O-77*	855	cluster I
AP017372.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae;Halorhodospira;Halorhodospira halochloris*	870	cluster I
AP017378.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio ferrophilus*	729	cluster III
AP017564.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Pseudanabaenaceae;Pseudanabaena;Pseudanabaena sp. ABRG5-3*	867	cluster I
AP017605.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium loti*	849	cluster I
AP017626.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae;Pleomorphomonas;Pleomorphomonas sp. SM30*	849	cluster I
AP017912.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter;Geobacter sulfurreducens*	855	cluster I

Table 15 (continued).

AP017928.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylocaldum;Methylocaldum marinum*	879	cluster I
AP018166.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Anabaena;Anabaena cylindrica*	891	cluster I
AP018172.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. NIES-2098*	885	cluster I
AP018174.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Aphanizomenonaceae;Anabaenopsis;Anabaenopsis circularis*	882	cluster I
AP018178.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. NIES-2100*	885	cluster I
AP018180.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc carneum*	885	cluster I
AP018184.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. NIES-2111*	888	cluster I
AP018194.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Scytonemataceae;Scytonema;Scytonema sp. HK-05*	900	cluster I
AP018203.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Leptolyngbyaceae;Leptolyngbya;Leptolyngbya boryana*	861	cluster I
AP018207.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix brevissima*	882	cluster I
AP018216.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus variabilis*	891	cluster I
AP018222.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc linckia*	891	cluster I
AP018223.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc linckia*	885	cluster I
AP018233.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Rivulariaceae;Microchaete;Microchaete diplosiphon*	885	cluster I
AP018248.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Tolypothrichaceae;Tolypothrix;Tolypothrix tenuis*	885	cluster I

Table 15 (continued).

AP018254.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. NIES-3974*	879	cluster I
AP018255.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. NIES-4071*	795	cluster I
AP018268.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Scytonemataceae;Scytonema;Scytonema sp. NIES-4073*	900	cluster I
AP018269.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Cylindrospermum;Cylindrospermum sp. NIES-4074*	873	cluster I
AP018280.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. NIES-4101*	858	cluster I
AP018288.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. NIES-4103*	882	cluster I
AP018290.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. NIES-4105*	795	cluster I
AP018298.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Fischerella;Fischerella sp. NIES-4106*	861	cluster I
AP018307.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Fortieaceae;Aulosira;Aulosira laxa*	885	cluster I
AP018318.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. HK-01*	882	cluster I
AP018326.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc commune*	891	cluster I
AP018327.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc commune*	876	cluster I
AP018341.1	*Bacteria;Cyanobacteria;unclassified class;unclassified order;unclassified family;unclassified genus;cyanobacterium endosymbiont of Rhopalodia gibberula*	870	cluster I
AP018449.1	*Bacteria;Firmicutes;Negativicutes;Selenomonadales;Sporomusaceae;Methylomusa;Methylomusa anaerophila*	843	cluster I
AP018724.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Thioalkalispiraceae;Sulfurivermis;Sulfurivermis fontis*	879	cluster I

Table 15 (continued).

AP018738.1	*Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Gallionellaceae;Ferriphaselus; Ferriphaselus amnicola*	867	cluster I
AP018756.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Metakosakonia;Metakosakonia sp. MRY16-398*	879	cluster I
AP018795.1	*Bacteria;Proteobacteria;Acidithiobacillia;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus;Acidithiobacillus ferridurans*	879	cluster I
AP018907.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Blastochloris;Blastochloris sp. GI*	822	cluster I
AY115490.2	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp. Mrp182*	846	cluster I
AY204398.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferrooxidans*	873	cluster I
AY242355.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
AY318755.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	843	cluster I
AY544164.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Delftia;Delftia tsuruhatensis*	861	cluster I
AY603957.3	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	819	cluster III
AY728386.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Crocosphaera;Crocosphaera subtropica*	981	cluster I
AY728387.1	*Bacteria;Cyanobacteria;unclassified class;unclassified order;unclassified family;unclassified genus;cyanobacterium endosymbiont of Rhopalodia gibba*	870	cluster I
AY912109.3	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus massiliensis*	864	cluster I
AY934869.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium vignae*	741	cluster I

Table 15 (continued).

AY934871.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium yuanmingense*	735	cluster I
AY934872.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	747	cluster I
AY934873.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. CCBAU 65255*	747	cluster I
AY934875.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. CCBAU 33220A*	780	cluster I
AY934876.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium hainanense*	777	cluster I
AY934877.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium sp. CCBAU 53044B*	750	cluster I
BA000012.4	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium japonicum*	849	cluster I
BA000019.2	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7120*	891	cluster I
BA000040.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium diazoefficiens*	894	cluster I
BX572597.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	828	cluster I
BX572607.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	837	cluster I
BX950851.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Pectobacterium;Pectobacterium atrosepticum*	879	cluster I
CP000089.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Azonexaceae;Dechloromonas;Dechloromonas aromatica*	861	cluster I

Table 15 (continued).

CP000096.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Pelodictyon;Pelodictyon luteolum*	771	cluster III
CP000099.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
CP000108.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobium;Chlorobium chlorochromatii*	771	cluster III
CP000117.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus variabilis*	891	cluster I
CP000142.2	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Pelobacter;Pelobacter carbinolicus*	792	cluster I
CP000143.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP000148.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter;Geobacter metallireducens*	855	cluster I
CP000230.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Rhodospirillum;Rhodospirillum rubrum*	828	cluster I
CP000239.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Synechococcaceae;Synechococcus;Synechococcus sp. JA-3-3Ab*	855	cluster I
CP000240.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Synechococcaceae;Synechococcus;Synechococcus sp. JA-2-3B'a(2-13)*	855	cluster I
CP000249.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia casuarinae*	861	cluster I
CP000250.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	843	cluster I
CP000271.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia xenovorans*	879	cluster I

Table 15 (continued).

CP000283.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	843	cluster I
CP000301.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	843	cluster I
CP000304.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas stutzeri*	879	cluster I
CP000393.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Microcoleaceae;Trichodesmium;Trichodesmium erythraeum*	771	cluster I
CP000463.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	843	cluster I
CP000471.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Magnetococcales;Magnetococcaceae;Magnetococcus;Magnetococcus marinus*	855	cluster I
CP000478.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae;Syntrophobacter;Syntrophobacter fumaroxidans*	774	cluster III
CP000482.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Pelobacter;Pelobacter propionicus*	858	cluster I
CP000492.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobium;Chlorobium phaeobacteroides*	816	cluster III
CP000494.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium sp. BTAi1*	858	cluster I
CP000528.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio vulgaris*	825	cluster III
CP000529.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Polaromonas;Polaromonas naphthalenivorans*	858	cluster I

Table 15 (continued).

CP000544.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae;Halorhodospira;Halorhodospira halophila*	873	cluster I
CP000577.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP000607.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobium;Chlorobium phaeovibrioides*	813	cluster III
CP000612.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfotomaculum;Desulfotomaculum reducens*	798	cluster III
CP000613.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Rhodospirillum;Rhodospirillum centenum*	819	cluster I
CP000616.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	879	cluster I
CP000661.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP000673.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium kluyveri*	813	cluster II
CP000698.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter uraniireducens*	858	cluster I
CP000721.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	819	cluster III
CP000724.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Alkaliphilus;Alkaliphilus metalliredigens*	816	cluster III
CP000740.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium medicae*	888	cluster I
CP000769.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Anaeromyxobacteraceae;Anaeromyxobacter;Anaeromyxobacter sp. Fw109-5*	855	cluster I

Table 15 (continued).

CP000780.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanoregulaceae;Methanoregula;Methanoregula boonei*	804	cluster III
CP000781.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae;Xanthobacter;Xanthobacter autotrophicus*	888	cluster I
CP000806.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Crocosphaera;Crocosphaera subtropica*	981	cluster I
CP000820.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp. EAN1pec*	867	cluster I
CP000860.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Candidatus Desulforudis;Candidatus Desulforudis audaxviator*	705	cluster II
CP000930.2	*Bacteria;Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacterium;Heliobacterium modesticaldum*	849	cluster I
CP000943.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium;Methylobacterium sp. 4-46*	834	cluster I
CP000964.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP000975.1	*Bacteria;Verrucomicrobia;Methylacidiphilae;Methylacidiphilales;Methylacidiphilaceae;Methylacidiphilum;Methylacidiphilum infernorum*	855	cluster I
CP001013.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;unclassified family;Leptothrix;Leptothrix cholodnii*	879	cluster I
CP001016.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae;Beijerinckia;Beijerinckia indica*	867	cluster I
CP001037.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc punctiforme*	891	cluster I
CP001046.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia phymatum*	879	cluster I
CP001076.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	981	cluster I

Table 15 (continued).

CP001087.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae;Desulfobacterium;Desulfobacterium autotrophicum*	768	cluster III
CP001089.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter lovleyi*	855	cluster I
CP001096.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	837	cluster I
CP001097.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobium;Chlorobium limicola*	771	cluster III
CP001099.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobaculum;Chlorobaculum parvum*	816	cluster III
CP001100.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chloroherpeton;Chloroherpeton thalassium*	816	cluster II
CP001101.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobium;Chlorobium phaeobacteroides*	774	cluster III
CP001108.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Prosthecochloris;Prosthecochloris aestuarii*	774	cluster III
CP001110.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Pelodictyon;Pelodictyon phaeoclathratiforme*	816	cluster III
CP001124.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter bemidjiensis*	858	cluster I
CP001131.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Anaeromyxobacteraceae;Anaeromyxobacter;Anaeromyxobacter sp. K*	861	cluster I
CP001132.1	*Bacteria;Proteobacteria;Acidithiobacillia;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus;Acidithiobacillus ferrooxidans*	879	cluster I
CP001150.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I

Table 15 (continued).

CP001151.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	750	cluster II
CP001157.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	870	cluster I
CP001189.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Gluconacetobacter;Gluconacetobacter diazotrophicus*	843	cluster I
CP001192.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
CP001197.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio vulgaris*	828	cluster III
CP001219.1	*Bacteria;Proteobacteria;Acidithiobacillia;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus;Acidithiobacillus ferrooxidans*	879	cluster I
CP001280.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae;Methylocella;Methylocella silvestris*	879	cluster I
CP001287.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Rippkaea;Rippkaea orientalis*	861	cluster I
CP001291.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Gloeothece;Gloeothece citriformis*	858	cluster I
CP001312.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter capsulatus*	876	cluster I
CP001322.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae;Desulfatibacillum;Desulfatibacillum aliphaticivorans*	771	cluster III
CP001338.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanoregulaceae;Methanosphaerula;Methanosphaerula palustris*	801	cluster III
CP001344.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Cyanothecaceae;Cyanothece;Cyanothece sp. PCC 7425*	855	cluster I

Table 15 (continued).

CP001349.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium;Methylobacterium nodulans*	834	cluster I
CP001390.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter;Geobacter daltonii*	858	cluster I
CP001614.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Cellvibrionaceae;Teredinibacter;Teredinibacter turnerae*	876	cluster I
CP001616.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Aeromonadales;Aeromonadaceae;Tolumonas;Tolumonas auensis*	855	cluster I
CP001623.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
CP001629.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae;Desulfomicrobium;Desulfomicrobium baculatum*	783	cluster III
CP001649.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio salexigens*	825	cluster III
CP001654.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya;Dickeya paradisiaca*	870	cluster I
CP001661.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter;Geobacter sp. M21*	855	cluster I
CP001666.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium ljungdahlii*	816	cluster III
CP001698.1	*Bacteria;Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae;Spirochaeta;Spirochaeta thermophila*	801	cluster III
CP001701.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Rippkaea;Rippkaea orientalis*	861	cluster I
CP001715.1	*Bacteria;Proteobacteria;Betaproteobacteria;unclassified order;unclassified family;Candidatus Accumulibacter;Candidatus Accumulibacter phosphatis*	858	cluster I

Table 15 (continued).

CP001720.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfofarcimen;Desulfofarcimen acetoxidans*	786	cluster III
CP001722.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP001831.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP001842.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Candidatus Atelocyanobacterium;Candidatus Atelocyanobacterium thalassa*	858	cluster I
CP001891.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP001896.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;Allochromatium; Allochromatium vinosum*	882	cluster I
CP001940.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfurivibrio;Desulfurivibrio alkaliphilus*	825	cluster III
CP001965.1	*Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Gallionellaceae;Sideroxydans; Sideroxydans lithotrophicus*	876	cluster I
CP001998.1	*Bacteria;Verrucomicrobia;Opitutae;Puniceicoccales;Puniceicoccaceae;Coraliomargarita;Coraliomargarita akajimensis*	780	cluster III
CP002016.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia sp. CCGE1002*	879	cluster I
CP002028.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Thermincola;Thermincola potens*	798	cluster III
CP002031.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter sulfurreducens*	855	cluster I
CP002038.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya dadantii*	849	cluster I

Table 15 (continued).

CP002039.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum;Herbaspirillum seropedicae*	876	cluster I
CP002048.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Syntrophomonadaceae;Syntrophothermus;Syntrophothermus lipocalidus*	750	cluster II
CP002059.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus azollae*	888	cluster I
CP002085.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfarculales;Desulfarculaceae;Desulfarculus; Desulfarculus baarsii*	780	cluster III
CP002116.1	*Bacteria;Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae;Sediminispirochaeta;Sediminispirochaeta smaragdinae*	798	cluster III
CP002117.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanomicrobiaceae;Methanolacinia;Methanolacinia petrolearia*	804	cluster III
CP002118.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium acetobutylicum*	813	cluster III
CP002160.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium cellulovorans*	810	cluster III
CP002171.1	*Bacteria;Firmicutes;Clostridia;Thermoanaerobacterales;Thermoanaerobacterales Family III. Incertae Sedis;Thermoanaerobacterium;Thermoanaerobacterium thermosaccharolyticum*	816	cluster III
CP002198.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Gloeothece;Gloeothece verrucosa*	858	cluster I
CP002279.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium opportunistum*	888	cluster I
CP002281.1	*Bacteria;Fusobacteria;Fusobacteriia;Fusobacteriales;Fusobacteriaceae;Ilyobacter;Ilyobacter polytropus*	759	cluster III
CP002292.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Rhodomicrobium;Rhodomicrobium vannielii*	873	cluster I

Table 15 (continued).

CP002298.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio vulgaris*	825	cluster III
CP002345.1	*Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Paludibacteraceae;Paludibacter;Paludibacter propionicigenes*	816	cluster II
CP002364.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfobulbus;Desulfobulbus propionicus*	825	cluster III
CP002400.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ethanoligenens;Ethanoligenens harbinense*	801	cluster II
CP002418.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	837	cluster I
CP002431.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Pseudodesulfovibrio;Pseudodesulfovibrio aespoeensis*	825	cluster III
CP002432.1	*Bacteria;Chrysiogenetes;Chrysiogenetes;Chrysiogenales;Chrysiogenaceae;Desulfurispirillum; Desulfurispirillum indicum*	825	cluster I
CP002436.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Erwiniaceae;Pantoea;Pantoea sp. At-9b*	876	cluster I
CP002447.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	888	cluster I
CP002479.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter sp. M18*	855	cluster I
CP002547.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Syntrophobotulus;Syntrophobotulus glycolicus*	843	cluster I
CP002582.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Cellulosilyticum;Cellulosilyticum lentocellum*	813	cluster III
CP002622.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas stutzeri*	879	cluster I

Table 15 (continued).

CP002629.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae;Desulfobacca;Desulfobacca acetoxidans*	768	cluster III
CP002660.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium acetobutylicum*	813	cluster III
CP002736.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfotomaculum;Desulfotomaculum nigrificans*	798	cluster III
CP002738.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylomonas;Methylomonas methanica*	879	cluster I
CP002741.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP002780.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfotomaculum;Desulfotomaculum ruminis*	798	cluster III
CP002783.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP002801.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia symbiont of Datisca glomerata*	870	cluster I
CP002850.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP002865.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP002985.1	*Bacteria;Proteobacteria;Acidithiobacillia;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus;Acidithiobacillus ferrivorans*	876	cluster I
CP002987.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;Acetobacterium;Acetobacterium woodii*	819	cluster III
CP003046.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Rhodospirillum;Rhodospirillum rubrum*	828	cluster I

Table 15 (continued).

CP003051.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;Thioflavicoccus;Thioflavicoccus mobilis*	879	cluster I
CP003065.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Hungateiclostridiaceae;Hungateiclostridium;Hungateiclostridium clariflavum*	816	cluster III
CP003066.1	*Bacteria;Firmicutes;Clostridia;Thermoanaerobacterales;Thermoanaerobacterales Family III. Incertae Sedis;Thermoanaerobacterium;Thermoanaerobacterium thermosaccharolyticum*	816	cluster III
CP003107.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus terrae*	864	cluster I
CP003153.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Azospira;Azospira oryzae*	858	cluster I
CP003154.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;Thiocystis;Thiocystis violascens*	876	cluster I
CP003184.1	*Bacteria;Firmicutes;Clostridia;Thermoanaerobacterales;Thermoanaerobacterales Family III. Incertae Sedis;Thermoanaerobacterium;Thermoanaerobacterium saccharolyticum*	816	cluster III
CP003218.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP003220.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio desulfuricans*	825	cluster III
CP003221.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfocurvibacter;Desulfocurvibacter africanus*	813	cluster III
CP003242.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio sp. EJY3*	879	cluster I
CP003243.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanocellales;Methanocellaceae;Methanocella;Methanocella conradii*	804	cluster III
CP003246.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Yersiniaceae;Rahnella;Rahnella aquatilis*	867	cluster I
CP003259.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium sp. BNL1100*	819	cluster III

Table 15 (continued).

CP003261.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	819	cluster III
CP003273.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfallas;Desulfallas gibsoniae*	798	cluster III
CP003358.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium australicum*	888	cluster I
CP003406.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Yersiniaceae;Rahnella;Rahnella aquatilis*	867	cluster I
CP003516.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. KJ006*	879	cluster I
CP003548.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7107*	882	cluster I
CP003549.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Rivulariaceae;Rivularia;Rivularia sp. PCC 7116*	789	cluster I
CP003552.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7524*	882	cluster I
CP003565.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	846	cluster I
CP003590.1	*Bacteria;Cyanobacteria;unclassified class;Pleurocapsales;Hyellaceae;Pleurocapsa;Pleurocapsa minor*	762	cluster I
CP003597.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcidiopsidales;Chroococcidiopsidaceae;Chroococcidiopsis;Chroococcidiopsis thermalis*	876	cluster I
CP003630.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Microcoleaceae;Microcoleus;Microcoleus sp. PCC 7113*	885	cluster I
CP003642.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Cylindrospermum;Cylindrospermum stagnale*	888	cluster I
CP003659.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Anabaena;Anabaena cylindrica*	891	cluster I

Table 15 (continued).

CP003683.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP003709.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP003715.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP003732.1	*Bacteria;Firmicutes;Clostridia;Thermoanaerobacterales;Thermoanaerobacteraceae;Thermacetogenium;Thermacetogenium phaeum*	750	cluster II
CP003865.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia phenoliruptrix*	879	cluster I
CP003869.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Dehalobacter;Dehalobacter sp. DCA*	843	cluster I
CP003870.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Dehalobacter;Dehalobacter sp. CF*	843	cluster I
CP003936.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP003943.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. PCC 7507*	903	cluster I
CP003945.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Aphanothecaceae;Halothece;Halothece sp. PCC 7418*	855	cluster I
CP003985.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfocapsa;Desulfocapsa sulfexigens*	828	cluster III
CP004017.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium tropici*	915	cluster I
CP004078.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sabinae*	861	cluster I
CP004121.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium saccharoperbutylacetonicum*	816	cluster III

Table 15 (continued).

CP004138.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP004887.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP005094.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	870	cluster I
CP005095.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	870	cluster I
CP005955.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	930	cluster I
CP005991.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Enterobacter;Enterobacter sp. R4-368*	879	cluster I
CP006583.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium huakuii*	891	cluster I
CP006585.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio gigas*	777	cluster III
CP006763.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium autoethanogenum*	816	cluster III
CP006777.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	819	cluster III
CP006818.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP006879.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium gallicum*	978	cluster I
CP006941.2	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I

Table 15 (continued).

CP006988.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. IE4771*	981	cluster I
CP007031.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;Marichromatium ;Marichromatium purpuratum*	879	cluster I
CP007048.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
CP007053.1	*Bacteria;Verrucomicrobia;Opitutae;Opitutales;Opitutaceae;unclassified genus;Opitutaceae bacterium TAV5*	819	cluster III
CP007070.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
CP007142.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Saccharospirillaceae;Gynuella;Gynuella sunshinyii*	876	cluster I
CP007203.2	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Aphanizomenonaceae;Nodularia;Nodularia spumigena*	894	cluster I
CP007215.3	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia sacchari*	879	cluster I
CP007243.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferriphilum*	873	cluster I
CP007268.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae;Ectothiorhodospira;Ectothiorhodospira haloalkaliphila*	882	cluster I
CP007569.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	894	cluster I
CP007643.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	981	cluster I
CP007744.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Pectobacterium;Pectobacterium atrosepticum*	879	cluster I

Table 15 (continued).

CP007793.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
CP008700.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP008746.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
CP008762.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia xenovorans*	879	cluster I
CP008788.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP008841.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP008852.1	*Bacteria;Firmicutes;Negativicutes;Selenomonadales;Sporomusaceae;Pelosinus;Pelosinus sp. UFO1*	825	cluster II
CP009125.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Pectobacterium;Pectobacterium atrosepticum*	879	cluster I
CP009145.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP009241.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sp. FSL H7-0357*	864	cluster I
CP009267.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	825	cluster II
CP009268.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	825	cluster II
CP009274.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I

Table 15 (continued).

CP009280.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sp. FSL P4-0081*	864	cluster I
CP009282.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sp. FSL R5-0912*	864	cluster I
CP009283.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sp. FSL R7-0273*	864	cluster I
CP009284.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sp. FSL R7-0331*	864	cluster I
CP009285.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus borealis*	864	cluster I
CP009286.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus stellifer*	861	cluster I
CP009287.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus graminis*	864	cluster I
CP009288.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus durus*	861	cluster I
CP009428.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus odorifer*	864	cluster I
CP009460.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya fangzhongdai*	849	cluster I
CP009506.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina siciliae*	801	cluster II
CP009507.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina siciliae*	801	cluster II
CP009508.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina siciliae*	801	cluster II
CP009517.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
CP009520.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina vacuolata*	801	cluster II

Table 15 (continued).

CP009524.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina sp. Kolksee*	801	cluster II
CP009526.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
CP009528.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
CP009530.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
CP009632.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	879	cluster I
CP009687.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium aceticum*	807	cluster III
CP009788.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter pickeringii*	855	cluster I
CP009863.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
CP009933.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium scatologenes*	807	cluster III
CP009978.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio natriegens*	879	cluster I
CP010086.2	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	819	cluster III
CP010268.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I
CP010313.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	894	cluster I
CP010415.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter chroococcum*	870	cluster I

Table 15 (continued).

CP010430.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter sulfurreducens*	855	cluster I
CP010523.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP010557.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Raoultella;Raoultella ornithinolytica*	858	cluster I
CP010802.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Desulfuromonas;Desulfuromonas soudanensis*	867	cluster I
CP010803.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Aurantimonadaceae;Martelella;Martelella endophytica*	876	cluster I
CP010836.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingomonas;Sphingomonas hengshuiensis*	864	cluster I
CP010869.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Confluentimicrobium;Confluentimicrobium sp. EMB200-NS6*	825	cluster I
CP010978.1	*Bacteria;Firmicutes;Negativicutes;Selenomonadales;Sporomusaceae;Pelosinus;Pelosinus fermentans*	825	cluster II
CP011072.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus sp. CIB*	858	cluster I
CP011114.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus durus*	861	cluster I
CP011360.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium diazoefficiens*	894	cluster I
CP011382.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Calotrichaceae;Calothrix;Calothrix sp. 336/3*	888	cluster I
CP011412.1	*Bacteria;Proteobacteria;Gammaproteobacteria;unclassified order;unclassified family;Sedimenticola;Sedimenticola thiotaurini*	870	cluster I
CP011420.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I

Table 15 (continued).

CP011512.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus peoriae*	864	cluster I
CP011602.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;unclassified family;Phytobacter;Phytobacter ursingii*	879	cluster I
CP011636.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP011835.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter chroococcum*	870	cluster I
CP011930.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum;Herbaspirillum seropedicae*	876	cluster I
CP011966.3	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	819	cluster III
CP011994.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae;Ectothiorhodospira;Ectothiorhodospira sp. BSL-9*	882	cluster I
CP012036.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc piscinale*	882	cluster I
CP012373.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Thiotrichaceae;Beggiatoa;Beggiatoa leptomitoformis*	876	cluster I
CP012395.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium autoethanogenum*	816	cluster III
CP012401.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum thiophilum*	840	cluster I
CP012487.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Enterobacter;Enterobacter sp. FY-07*	879	cluster I
CP012661.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Defluviimonas;Defluviimonas alba*	882	cluster I
CP012899.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. CCGE1001*	879	cluster I

Table 15 (continued).

CP012914.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
CP012946.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Blastochloris;Blastochloris viridis*	831	cluster I
CP012960.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP013018.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	825	cluster II
CP013019.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	825	cluster II
CP013021.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae;Agarivorans;Agarivorans gilvus*	876	cluster I
CP013053.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium americanum*	978	cluster I
CP013109.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium americanum*	915	cluster I
CP013127.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium diazoefficiens*	894	cluster I
CP013238.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium butyricum*	816	cluster III
CP013353.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium butyricum*	816	cluster III
CP013395.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;Burkholderia vietnamiensis*	879	cluster I
CP013421.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;Burkholderia ubonensis*	879	cluster I
CP013434.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;Burkholderia vietnamiensis*	879	cluster I

Table 15 (continued).

CP013441.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	879	cluster I
CP013456.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	879	cluster I
CP013489.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium butyricum*	816	cluster III
CP013492.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N6212*	981	cluster I
CP013497.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N621*	981	cluster I
CP013502.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium esperanzae*	981	cluster I
CP013508.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N1341*	981	cluster I
CP013514.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N1314*	981	cluster I
CP013519.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N113*	981	cluster I
CP013530.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013535.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013540.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013544.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013550.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I

Table 15 (continued).

CP013555.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013561.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013566.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013572.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013578.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013583.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013587.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP013592.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N871*	981	cluster I
CP013598.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N741*	981	cluster I
CP013604.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N731*	981	cluster I
CP013633.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N324*	981	cluster I
CP013639.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N541*	981	cluster I
CP013646.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. N941*	981	cluster I
CP013737.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum;Herbaspirillum rubrisubalbicans*	876	cluster I

Table 15 (continued).

CP013949.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium sp. CCGE-LA001*	882	cluster I
CP013985.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP014007.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia oryzae*	879	cluster I
CP014071.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP014137.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Brenneria;Brenneria goodwinii*	882	cluster I
CP014170.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium tyrobutyricum*	816	cluster III
CP014206.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Pseudodesulfovibrio;Pseudodesulfovibrio indicus*	825	cluster III
CP014331.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium sp. MF28*	819	cluster III
CP014360.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylomonas;Methylomonas sp. DH-1*	873	cluster I
CP014476.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylomonas;Methylomonas denitrificans*	882	cluster I
CP014595.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Yangia;Yangia sp. CCB-MM3*	870	cluster I
CP014671.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Immundisolibacterales;Immundisolibacteraceae;Immundisolibacter;Immundisolibacter cernigliae*	861	cluster I
CP014696.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP014705.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium butyricum*	816	cluster III

Table 15 (continued).

CP015039.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodovulum;Rhodovulum sp. P5*	876	cluster I
CP015062.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	891	cluster I
CP015064.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	870	cluster I
CP015080.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Desulfuromonas;Desulfuromonas sp. DDH964*	870	cluster I
CP015113.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia radicincitans*	879	cluster I
CP015137.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya solani*	849	cluster I
CP015210.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP015285.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum humicireducens*	840	cluster I
CP015287.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP015322.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium amorphae*	891	cluster I
CP015418.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodovulum;Rhodovulum sulfidophilum*	873	cluster I
CP015421.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodovulum;Rhodovulum sulfidophilum*	873	cluster I
CP015423.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I

Table 15 (continued).

CP015455.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Pelobacter;Pelobacter acetylenicus*	693	cluster I
CP015518.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Pelobacter;Pelobacter acetylenicus*	813	cluster I
CP015519.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae;Pelobacter;Pelobacter sp. SFB93*	852	cluster I
CP015756.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium estertheticum*	816	cluster III
CP015848.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum sp. ME-1*	879	cluster I
CP016027.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Halothiobacillaceae;Halothiobacillus;Halothiobacillus sp. LS2*	864	cluster I
CP016079.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium loti*	849	cluster I
CP016087.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium saccharoperbutylacetonicum*	819	cluster III
CP016090.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	819	cluster III
CP016210.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus olearius*	885	cluster I
CP016280.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium tyrobutyricum*	816	cluster III
CP016290.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	888	cluster I
CP016333.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium butyricum*	816	cluster III
CP016337.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia sacchari*	879	cluster I

Table 15 (continued).

CP016344.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP016346.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio natriegens*	879	cluster I
CP016348.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio natriegens*	879	cluster I
CP016350.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio natriegens*	879	cluster I
CP016352.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio natriegens*	879	cluster I
CP016428.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium icense*	882	cluster I
CP016432.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Prosthecochloris;Prosthecochloris sp. CIB 2401*	720	cluster III
CP016448.1	*Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Sterolibacteriaceae;Methyloversatilis;Methyloversatilis sp. RAC08*	876	cluster I
CP016811.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
CP016928.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya solani*	849	cluster I
CP017104.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium gallicum*	978	cluster I
CP017237.1	*Bacteria;Firmicutes;Clostridia;Thermoanaerobacterales;Thermoanaerobacteraceae;Moorella; Moorella thermoacetica*	750	cluster II
CP017243.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	981	cluster I
CP017269.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Geosporobacter;Geosporobacter ferrireducens*	816	cluster III

Table 15 (continued).

CP017284.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP017289.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP017305.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Chlorobaculum;Chlorobaculum limnaeum*	816	cluster III
CP017448.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae;Acidihalobacter;Acidihalobacter prosperus*	855	cluster I
CP017450.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella sp. LTGPAF-6F*	879	cluster I
CP017453.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya solani*	849	cluster I
CP017454.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya solani*	849	cluster I
CP017481.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Pectobacterium;Pectobacterium polaris*	879	cluster I
CP017482.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Pectobacterium;Pectobacterium polaris*	879	cluster I
CP017562.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia sprentiae*	879	cluster I
CP017603.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium formicaceticum*	807	cluster III
CP017637.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	894	cluster I
CP017638.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya dianthicola*	870	cluster I
CP017849.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I

Table 15 (continued).

CP017967.3	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I
CP017968.3	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I
CP018016.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia radicincitans*	879	cluster I
CP018229.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	888	cluster I
CP018307.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP018335.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium kluyveri*	816	cluster II
CP018362.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP018889.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Thiotrichaceae;Beggiatoa;Beggiatoa leptomitoformis*	876	cluster I
CP019240.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Rhodoferax;Rhodoferax antarcticus*	876	cluster I
CP019483.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP019486.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP019509.1	*Bacteria;Proteobacteria;Betaproteobacteria;Neisseriales;Chromobacteriaceae;Aquaspirillum;Aquaspirillum sp. LM1*	885	cluster I
CP019585.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP019636.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;unclassified family;unclassified genus;Nostocales cyanobacterium HT-58-2*	882	cluster I
CP019698.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfotomaculum;Desulfotomaculum ferrireducens*	798	cluster III

Table 15 (continued).

CP019936.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;unclassified genus;Chromatiaceae bacterium 2141T.STBD.0c.01a*	879	cluster I
CP019948.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae;Methylocystis;Methylocystis bryophila*	882	cluster I
CP019966.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	831	cluster I
CP019967.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas;Rhodopseudomonas palustris*	831	cluster I
CP020026.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;unclassified genus;Rhodocyclaceae bacterium*	879	cluster I
CP020028.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus kribbensis*	864	cluster I
CP020114.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Aphanizomenonaceae;Nodularia;Nodularia spumigena*	894	cluster I
CP020358.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP020370.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;Thiodictyon;Candidatus Thiodictyon syntrophicum*	873	cluster I
CP020384.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodovulum;Rhodovulum sp. MB263*	873	cluster I
CP020395.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	879	cluster I
CP020470.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter blasticus*	843	cluster I
CP020559.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium formicaceticum*	807	cluster III

Table 15 (continued).

CP020657.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella sp. M5al*	858	cluster I
CP020847.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP020872.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya;Dickeya fangzhongdai*	849	cluster I
CP020898.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
CP020910.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	930	cluster I
CP020950.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. CIAT894*	981	cluster I
CP020953.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium drakei*	807	cluster III
CP021027.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. TAL182*	981	cluster I
CP021031.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. NXC14*	930	cluster I
CP021053.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP021070.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. WSM1497*	888	cluster I
CP021125.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. Kim5*	981	cluster I
CP021138.1	*Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Gallionellaceae;Sulfuriferula;Sulfuriferula sp. AH1*	888	cluster I
CP021217.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I

Table 15 (continued).

CP021375.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. ACO-34A*	846	cluster I
CP021731.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus sp. DN11*	843	cluster I
CP021780.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus donghaensis*	864	cluster I
CP021794.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021798.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021801.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021805.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021809.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021813.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021824.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021827.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021830.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
CP021850.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Hungateiclostridiaceae;Pseudoclostridium;Pseudoclostridium thermosuccinogenes*	804	cluster III
CP021913.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Sagittula;Sagittula sp. P11*	870	cluster I

Table 15 (continued).

CP021965.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus odorifer*	864	cluster I
CP022110.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Nitrospirillum;Nitrospirillum amazonense*	858	cluster I
CP022129.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylovulum;Methylovulum psychrotolerans*	858	cluster I
CP022187.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus communis*	852	cluster I
CP022188.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus communis*	852	cluster I
CP022253.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
CP022348.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP022364.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum sp. TSH58*	867	cluster I
CP022423.1	*Bacteria;Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Vitreoscilla;Vitreoscilla filiformis*	891	cluster I
CP022571.1	*Bacteria;Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae;Prosthecochloris;Prosthecochloris sp. GSB1*	816	cluster III
CP022669.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	876	cluster I
CP022823.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasivariicola*	879	cluster I
CP022991.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia aromaticivorans*	879	cluster I
CP023065.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium sp. CCBAU 05631*	846	cluster I

Table 15 (continued).

CP023069.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sojae*	846	cluster I
CP023072.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	846	cluster I
CP023185.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP023467.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya dadantii*	870	cluster I
CP023484.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya sp. Secpp 1600*	849	cluster I
CP023548.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sp. CZR27*	840	cluster I
CP023669.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylomonas;Methylomonas koyamae*	873	cluster I
CP023677.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP023682.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP023715.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas;Zymomonas mobilis*	834	cluster I
CP023737.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae;Methylosinus;Methylosinus trichosporium*	885	cluster I
CP024202.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae;Methylomicrobium;Methylomicrobium sp. wino1*	879	cluster I
CP024309.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	915	cluster I

Table 15 (continued).

CP024313.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. NXC24*	915	cluster I
CP024710.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya solani*	849	cluster I
CP024711.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya solani*	849	cluster I
CP024785.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc flagelliforme*	891	cluster I
CP024956.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Pectobacterium;Pectobacterium atrosepticum*	879	cluster I
CP024996.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum;Herbaspirillum rubrisubalbicans*	876	cluster I
CP025003.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya fangzhongdai*	849	cluster I
CP025015.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	888	cluster I
CP025084.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Yersiniaceae;Serratia;Serratia sp. ATCC 39006*	879	cluster I
CP025085.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Yersiniaceae;Serratia;Serratia sp. ATCC 39006*	879	cluster I
CP025149.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas stutzeri*	879	cluster I
CP025286.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ethanoligenens;Ethanoligenens harbinense*	801	cluster II
CP025287.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ethanoligenens;Ethanoligenens harbinense*	801	cluster II
CP025288.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ethanoligenens;Ethanoligenens harbinense*	801	cluster II

Table 15 (continued).

CP025505.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	888	cluster I
CP025696.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sp. lzh-N1*	864	cluster I
CP026192.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;unclassified genus;Enterobacteriaceae bacterium ENNIH2*	879	cluster I
CP026193.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;unclassified genus;Enterobacteriaceae bacterium ENNIH1*	879	cluster I
CP026197.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;unclassified genus;Enterobacteriaceae bacterium ENNIH3*	879	cluster I
CP026269.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP026275.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP026285.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP026681.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. 'Peltigera membranacea cyanobiont' N6*	888	cluster I
CP026692.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. 'Lobaria pulmonaria (5183) cyanobiont'*	888	cluster I
CP026715.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella oxytoca*	879	cluster I
CP027064.2	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP027225.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;unclassified family;Phytobacter;Phytobacter sp. SCO41*	879	cluster I

Table 15 (continued).

CP027526.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum gryphiswaldense*	876	cluster I
CP027527.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum gryphiswaldense*	876	cluster I
CP027663.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Acidiferrobacterales;Acidiferrobacteraceae;Acidiferrobacter;Acidiferrobacter sp. SPIII_3*	882	cluster I
CP028425.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;unclassified genus;Zoogloeaceae bacteirum Par-f-2*	879	cluster I
CP028555.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP028944.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Vibrio;Vibrio sp. dhg*	879	cluster I
CP028953.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
CP028960.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;unclassified family;Methyloceanibacter;Methyloceanibacter sp. wino2*	855	cluster I
CP028990.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
CP029141.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP029233.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	846	cluster I
CP029331.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Thauera;Thauera hydrothermalis*	879	cluster I
CP029353.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum sp. CFH 70021*	831	cluster I

Table 15 (continued).

CP029425.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium ottawaense*	894	cluster I
CP029426.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium sp. 2 39S1MB*	888	cluster I
CP029432.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP029437.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP029443.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP029453.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	846	cluster I
CP029597.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP029603.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium diazoefficiens*	894	cluster I
CP029722.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
CP029770.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella michiganensis*	879	cluster I
CP029829.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum sp. M2T2B2*	840	cluster I
CP030018.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium acetobutylicum*	813	cluster III

Table 15 (continued).

CP030040.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;Acetobacterium;Acetobacterium sp. KB-1*	819	cluster III
CP030171.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella quasipneumoniae*	879	cluster I
CP030173.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
CP030271.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP030762.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
CP031560.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pectobacteriaceae;Dickeya; Dickeya dianthicola*	870	cluster I
CP031750.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP031842.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Azonexaceae;Dechloromonas;Dechloromonas sp. HYN0024*	879	cluster I
CP031941.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sphaeroides*	891	cluster I
CP032687.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. CCGE531*	915	cluster I
CP032692.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. CCGE532*	915	cluster I
CP032696.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium jaguaris*	978	cluster I
CP033434.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I

Table 15 (continued).

CP033442.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP033446.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP033450.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sphaeroides*	840	cluster I
CP033631.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella sp. P1CD1*	879	cluster I
CP033824.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella sp. FDAARGOS_511*	858	cluster I
CP033901.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	879	cluster I
CP034086.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae;Methylocystis;Methylocystis rosea*	885	cluster I
CP034443.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. M9A.F.Ca.ET.002.03.1.2*	891	cluster I
CP034444.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. M1D.F.Ca.ET.043.01.1.1*	891	cluster I
CP034445.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. M2A.F.Ca.ET.043.02.1.1*	891	cluster I
CP034446.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. M2A.F.Ca.ET.043.05.1.1*	891	cluster I
CP034447.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. M1E.F.Ca.ET.045.02.1.1*	717	cluster I

Table 15 (continued).

CP034448.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M1B.F.Ca.ET.045.04.1.1*	891	cluster I
CP034449.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M2A.F.Ca.ET.046.03.2.1*	891	cluster I
CP034450.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M4B.F.Ca.ET.058.02.1.1*	891	cluster I
CP034451.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M3A.F.Ca.ET.080.04.2.1*	891	cluster I
CP034452.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M6A.T.Cr.TU.016.01.1.1*	891	cluster I
CP034453.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M7D.F.Ca.US.005.01.1.1*	891	cluster I
CP034454.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M8A.F.Ca.ET.057.01.1.1*	891	cluster I
CP034455.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium;Mesorhizobium sp. M1A.F.Ca.IN.022.06.1.1*	891	cluster I
CP034675.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Cellulosilyticum;Cellulosilyticum sp. WCF-2*	813	cluster III
CT573213.2	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia alni*	858	cluster I
CU234118.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium sp. ORS 278*	885	cluster I
CU633751.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
D00666.2	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Leptolyngbyaceae;Leptolyngbya;Leptolyngbya boryana*	858	cluster I

Table 15 (continued).

DQ058415.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	981	cluster I
DQ078751.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Oscillatoriaceae;Limnoraphis;Limnoraphis robusta*	855	cluster I
DQ176436.2	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Cronbergia;Cronbergia siamensis*	879	cluster I
DQ284983.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. NCHA22*	795	cluster I
DQ315787.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus variabilis*	882	cluster I
DQ385910.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Mastigocladus;Mastigocladus laminosus*	801	cluster I
DQ385913.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Mastigocladus;Mastigocladus laminosus*	801	cluster I
DQ385920.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Mastigocladus;Mastigocladus laminosus*	801	cluster I
DQ485702.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium elkanii*	780	cluster I
DQ485714.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	780	cluster I
DQ485715.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	780	cluster I
EF070142.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. CCBAU 33062*	795	cluster I
EF070143.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. CCBAU 33037*	780	cluster I

Table 15 (continued).

EF158799.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia ferrariae*	852	cluster I
EF158800.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia tropica*	852	cluster I
EF158801.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia tropica*	834	cluster I
EF158802.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia unamae*	849	cluster I
EF158803.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia unamae*	879	cluster I
EF158804.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia unamae*	849	cluster I
EF158805.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia xenovorans*	852	cluster I
EF158806.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Trinickia;Trinickia caryophylli*	849	cluster I
EF158807.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia silvatlantica*	840	cluster I
EF158808.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia silvatlantica*	852	cluster I
EF158809.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	852	cluster I
EF158810.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	831	cluster I

Table 15 (continued).

EF158811.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	849	cluster I
EF165526.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
EF165535.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
EU130410.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium temperatum*	786	cluster I
EU130411.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	786	cluster I
EU130412.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11185*	738	cluster I
EU130413.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11196*	738	cluster I
EU130414.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	741	cluster I
EU130415.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11214*	741	cluster I
EU130416.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11217*	741	cluster I
EU130417.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11226*	738	cluster I
EU130419.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	786	cluster I

Table 15 (continued).

EU130422.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium caraganae*	741	cluster I
EU130423.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11300*	738	cluster I
EU130424.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11231*	756	cluster I
EU177582.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium elkanii*	783	cluster I
EU177583.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	762	cluster I
EU177584.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	756	cluster I
EU177585.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium elkanii*	750	cluster I
EU177586.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	711	cluster I
EU177587.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium huakuii*	762	cluster I
EU177588.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
EU177589.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium yuanmingense*	771	cluster I
EU177591.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	720	cluster I
EU177593.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium liaoningense*	711	cluster I

Table 15 (continued).

EU177594.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium gallicum*	753	cluster I
EU177595.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
EU177597.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
EU252602.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	741	cluster I
EU252603.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium tarimense*	744	cluster I
EU252605.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium gobiense*	744	cluster I
EU252606.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium gobiense*	744	cluster I
EU252607.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium tarimense*	741	cluster I
EU252608.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium tarimense*	744	cluster I
EU379969.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	732	cluster I
EU514542.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. SCAU2*	801	cluster I
EU514543.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. SCAU5*	768	cluster I
EU514544.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sangaii*	801	cluster I

Table 15 (continued).

EU514545.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. SCAU9*	801	cluster I
EU514546.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. SCAU11*	786	cluster I
EU563935.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	792	cluster I
EU748941.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. BD56*	750	cluster I
EU748943.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. ML108*	747	cluster I
EU770974.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	738	cluster I
EU770975.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	780	cluster I
EU770976.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium temperatum*	738	cluster I
EU770977.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium loessense*	738	cluster I
EU770978.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	780	cluster I
EU770979.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	780	cluster I
EU770980.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	780	cluster I
EU770981.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium mongolense*	780	cluster I

Table 15 (continued).

EU770982.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	780	cluster I
FJ348670.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. RSA104*	747	cluster I
FJ514071.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium yuanmingense*	747	cluster I
FJ648713.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus sp. SWF66166*	798	cluster I
FJ745292.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium plurifarium*	804	cluster I
FJ745298.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. ORS 3404*	807	cluster I
FM202369.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. NZP2309*	780	cluster I
FM202371.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. 9LBC*	747	cluster I
FO203427.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Pseudodesulfovibrio;Pseudodesulfovibrio piezophilus*	822	cluster III
FQ311868.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum ;Azospirillum lipoferum*	840	cluster I
FQ859181.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Hyphomicrobium;Hyphomicrobium sp. MC1*	882	cluster I
FR850692.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
FR850696.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I

Table 15 (continued).

FR850699.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
FR850717.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
FR850727.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
FR850731.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
FR850740.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
FR850775.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	735	cluster I
GQ167279.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11226*	774	cluster I
GQ167280.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium mediterraneum*	825	cluster I
GQ167281.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11185*	777	cluster I
GQ167282.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium tianshanense*	780	cluster I
GQ167283.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	825	cluster I
GQ167284.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	819	cluster I
GQ167285.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium temperatum*	819	cluster I

Table 15 (continued).

GQ167286.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11196*	774	cluster I
GQ167287.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11206*	777	cluster I
GQ167288.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	777	cluster I
GQ167289.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11214*	777	cluster I
GQ167290.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11217*	780	cluster I
GQ167291.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11231*	804	cluster I
GQ167292.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium amorphae*	777	cluster I
GQ167293.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium septentrionale*	819	cluster I
GQ167294.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCBAU 11257*	777	cluster I
GQ167295.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium huakuii*	780	cluster I
GQ167296.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium caraganae*	741	cluster I
GQ167297.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium plurifarium*	825	cluster I

Table 15 (continued).

GQ226014.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sphaerophysae*	735	cluster I
GQ342604.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Erwiniaceae;Pantoea;Pantoea sp. At-9b*	876	cluster I
GQ892931.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia phymatum*	852	cluster I
GU299762.1	*Bacteria;Verrucomicrobia;Methylacidiphilae;Methylacidiphilales;Methylacidiphilaceae;Methylacidiphilum;Methylacidiphilum fumariolicum*	855	cluster I
HE577327.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
HE616895.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	846	cluster I
HE663493.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Pararhodospirillum;Pararhodospirillum photometricum*	873	cluster I
HE956757.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae;Methylocystis;Methylocystis sp. SC2*	885	cluster I
HE995407.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
HG518324.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. IRBG74*	891	cluster I
HG794546.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum gryphiswaldense*	876	cluster I
HG917868.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium bornimense*	819	cluster III
HG917869.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium bornimense*	819	cluster III

Table 15 (continued).

HG938354.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Neorhizobium;Neorhizobium galegae*	873	cluster I
HG938357.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Neorhizobium;Neorhizobium galegae*	873	cluster I
HM019529.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia phymatum*	852	cluster I
HM019531.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia phymatum*	852	cluster I
HM019532.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia phymatum*	852	cluster I
HM063961.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. DG*	747	cluster I
HM146187.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I
HM583798.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sabinae*	861	cluster I
HM583799.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sabinae*	849	cluster I
HM583800.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus sabinae*	843	cluster I
HQ115045.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-4*	852	cluster I
HQ115046.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-28*	849	cluster I
HQ115047.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-40*	852	cluster I
HQ115048.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-90*	852	cluster I

Table 15 (continued).

HQ115050.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-299*	852	cluster I
HQ115051.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-328*	852	cluster I
HQ115052.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CATux-331*	852	cluster I
HQ115053.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-01*	852	cluster I
HQ115054.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-09*	849	cluster I
HQ115055.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-11*	852	cluster I
HQ115056.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-34*	852	cluster I
HQ115057.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-50*	852	cluster I
HQ115058.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-70*	852	cluster I
HQ115059.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia sp. enrichment culture clone CACua-88*	852	cluster I
HQ132362.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium helanshanense*	780	cluster I
HQ670655.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. CCGE 510*	891	cluster I

Table 15 (continued).

J01740.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	858	cluster I
J05111.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7120*	885	cluster I
JF461491.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. CCBAU 23010*	768	cluster I
JF461492.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. CCBAU 41134*	771	cluster I
JF461493.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. CCBAU 51317*	771	cluster I
JF738072.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium undicola*	759	cluster I
JF789586.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. GZL13-3*	777	cluster I
JF907684.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium amorphae*	765	cluster I
JN021929.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium mesoamericanum*	891	cluster I
JN021930.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium grahamii*	891	cluster I
JN390678.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferriphilum*	873	cluster I
JN558693.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium laguerreae*	741	cluster I
JN558695.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. FB1206*	750	cluster I

Table 15 (continued).

JN558696.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. FB2405*	741	cluster I
JN558698.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. FB9071*	747	cluster I
JN571917.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium sp. JNVU TV1*	753	cluster I
JN585763.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Acaryochloridaceae;Acaryochloris;Acaryochloris sp. HICR111A*	858	cluster I
JN593338.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium sp. JNVU TW8*	780	cluster I
JN593339.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium yuanmingense*	747	cluster I
JN624730.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	840	cluster I
JQ231105.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferrooxidans*	873	cluster I
JQ231108.1	*Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Leptospirillum;Leptospirillum ferrodiazotrophum*	873	cluster I
JQ903614.1	*unclassified kingdom;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;Cloning vector SBa_000534*	879	cluster I
JX074064.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;Cyanobacteriaceae;Cyanobacterium;Cyanobacterium sp. NBRC 102756*	855	cluster I
JX863572.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	915	cluster I
JX863573.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium tropici*	915	cluster I

Table 15 (continued).

K00487.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. ANU289*	882	cluster I
K00490.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	858	cluster I
K01620.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	894	cluster I
KC428648.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium sp. JNVU RM18*	759	cluster I
KC428649.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium sp. JNVU VT29*	741	cluster I
KC989924.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Enterobacter;Enterobacter sp. R4-368*	879	cluster I
KF483559.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium canariense*	747	cluster I
KF483561.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium canariense*	744	cluster I
KF483562.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium canariense*	738	cluster I
KF483563.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. GV135*	762	cluster I
KF483564.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. GV137*	774	cluster I
KF483565.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium canariense*	756	cluster I
KF636517.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	840	cluster I

Table 15 (continued).

KF636525.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	819	cluster I
KF679800.1	*unclassified kingdom;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;Cloning vector pCV_cistron14*	879	cluster I
KF679801.1	*unclassified kingdom;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;Cloning vector pCV_cistron34*	879	cluster I
KF732646.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	879	cluster I
KF749193.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. SDD22-4*	759	cluster I
KF749194.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium liaoningense*	762	cluster I
KF749203.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	735	cluster I
KF749207.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	774	cluster I
KF749210.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	774	cluster I
KF806460.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium retamae*	882	cluster I
KF806461.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium valentinum*	882	cluster I
KF828820.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. BR 3262*	741	cluster I
KF828825.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. BR 3267*	747	cluster I

Table 15 (continued).

KF859874.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA0203*	768	cluster I
KF859875.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA0206*	768	cluster I
KF859879.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA0601*	768	cluster I
KF859884.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA1704*	768	cluster I
KF859886.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA1601*	768	cluster I
KF859890.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA0403*	768	cluster I
KF859894.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA0506*	768	cluster I
KF859898.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ISA2507*	768	cluster I
KJ018175.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. JNVU AJ32*	771	cluster I
KJ018178.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. JNVU AJ14*	744	cluster I
KJ021871.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Oceanospirillaceae;Marinobacterium;Marinobacterium mangrovicola*	816	cluster I
KJ021873.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Oceanospirillaceae;Marinobacterium;Marinobacterium lutimaris*	786	cluster I
KJ790194.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium tarimense*	762	cluster I

Table 15 (continued).

KJ790195.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium gallicum*	765	cluster I
KJ814970.1	*Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Zoogloeaceae;Azoarcus;Azoarcus sp. CIB*	867	cluster I
KJ943307.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Cellvibrionaceae;unclassified genus;Cellvibrionaceae bacterium Bs08*	885	cluster I
KM052348.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; uncultured Bradyrhizobium sp.*	780	cluster I
KM052351.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; uncultured Bradyrhizobium sp.*	780	cluster I
KM052355.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; uncultured Bradyrhizobium sp.*	768	cluster I
KM052357.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; uncultured Bradyrhizobium sp.*	780	cluster I
KM052359.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; uncultured Bradyrhizobium sp.*	780	cluster I
KM076744.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. JNVU AL39*	744	cluster I
KM103912.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Pseudacidovorax;Pseudacidovorax intermedius*	879	cluster I
KM192241.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. ATQ1*	756	cluster I
KM192242.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp. ATQ2*	759	cluster I
KM192243.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. ATQ3*	756	cluster I

Table 15 (continued).

KM192244.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. STA1*	759	cluster I
KM192245.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. STA2*	756	cluster I
KM192246.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. CHU1*	756	cluster I
KM192247.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. CHU2*	753	cluster I
KM192248.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. CEQ1*	756	cluster I
KM361895.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Neorhizobium;Neorhizobium galegae*	753	cluster I
KM361896.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	756	cluster I
KM361897.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium giardinii*	750	cluster I
KM361898.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	759	cluster I
KP069485.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	738	cluster I
KP069496.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	738	cluster I
KP069508.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	738	cluster I
KP069509.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	738	cluster I

Table 15 (continued).

KP219173.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. SCAUd6*	735	cluster I
KP234518.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium embrapense*	882	cluster I
KP428823.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	876	cluster I
KP433913.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	888	cluster I
KP445528.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	891	cluster I
KP886488.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum moscoviense*	882	cluster I
KP901096.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus sp. UYMAM01A*	747	cluster I
KR071032.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. JNVU CP30*	750	cluster I
KR071039.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. JNVU TW7*	747	cluster I
KR075961.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium loti*	801	cluster I
KR075962.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium robiniae*	792	cluster I
KR075968.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer adhaerens*	885	cluster I
KR075981.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia sacchari*	870	cluster I

Table 15 (continued).

KR075982.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella variicola*	861	cluster I
KR149137.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium viridifuturi*	744	cluster I
KR149138.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium viridifuturi*	747	cluster I
KR149139.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium viridifuturi*	744	cluster I
KR259165.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. JNVU RM47*	744	cluster I
KR491970.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. TUTMFJ1A1*	741	cluster I
KT803836.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Ensifer;Ensifer sp. JNVU TW10*	759	cluster I
KT820417.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	843	cluster I
KT944070.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	888	cluster I
KU144965.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium UPO35*	879	cluster I
KU509347.1	*Bacteria;Verrucomicrobia;Methylacidiphilae;Methylacidiphilales;Methylacidiphilaceae;Methylacidiphilum;Methylacidiphilum sp. RTK17.1*	855	cluster I
KU724157.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium stylosanthis*	747	cluster I
KU886163.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc flagelliforme*	891	cluster I

Table 15 (continued).

KX035104.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Mastigocladus;Mastigocladus sp. CHP1*	882	cluster I
KX394363.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp.*	747	cluster I
KX489622.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium leguminosarum*	891	cluster I
KX622571.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Paraburkholderia;Paraburkholderia tropica*	801	cluster I
KX622575.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; Burkholderia vietnamiensis*	876	cluster I
KX650081.1	*unclassified kingdom;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;Cloning vector pMON261406*	861	cluster I
KX758645.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I
KX758646.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus polymyxa*	864	cluster I
KX820332.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Aphanizomenonaceae;Dolichospermum;Dolichospermum flos-aquae*	840	cluster I
KY246982.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246983.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246984.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	753	cluster I
KY246985.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246986.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I

Table 15 (continued).

KY246988.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246989.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I
KY246990.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246991.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246992.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I
KY246993.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246994.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246995.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246996.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246997.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246998.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY246999.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247000.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247001.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I

Table 15 (continued).

KY247002.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247003.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247004.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247005.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I
KY247006.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247007.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247008.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247009.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247010.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247011.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247012.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247013.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247014.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247015.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I

Table 15 (continued).

KY247016.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247017.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247018.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247019.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247020.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247021.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I
KY247022.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247023.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247024.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY247025.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	744	cluster I
KY748058.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium esperanzae*	747	cluster I
KY941327.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	756	cluster I
KY941329.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	753	cluster I
KY941337.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	756	cluster I

Table 15 (continued).

KY941340.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I
KY941342.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	747	cluster I
KY941343.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp.*	693	cluster I
L04499.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Anabaena;Anabaena sp. L-31*	879	cluster I
L23514.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc commune*	891	cluster I
L34879.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus azollae*	882	cluster I
L41344.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia alni*	858	cluster I
LC107542.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	750	cluster I
LC107543.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LC107544.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LC107545.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium lupini*	744	cluster I
LC107546.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium lupini*	750	cluster I
LC107547.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LC107548.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium tropici*	750	cluster I
LC107549.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	750	cluster I

Table 15 (continued).

LC107550.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium alamii*	744	cluster I
LC107551.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium pachyrhizi*	750	cluster I
LC107552.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	750	cluster I
LC107553.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LC107554.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	750	cluster I
LC107555.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium tropici*	750	cluster I
LC107556.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium alamii*	750	cluster I
LC107557.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium japonicum*	750	cluster I
LC107558.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LC107559.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LC107560.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium alamii*	750	cluster I
LC107561.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium pusense*	750	cluster I
LM997411.1	*Bacteria;Verrucomicrobia;Methylacidiphilae;Methylacidiphilales;Methylacidiphilaceae;Methyl acidiphilum;Methylacidiphilum fumariolicum*	855	cluster I

Table 15 (continued).

LN713501.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium australicum*	747	cluster I
LN713502.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium opportunistum*	738	cluster I
LN713503.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP11*	747	cluster I
LN713504.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	750	cluster I
LN713507.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP33*	738	cluster I
LN713508.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP34*	747	cluster I
LN713511.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	744	cluster I
LN713513.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP58*	750	cluster I
LN713515.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP63*	744	cluster I
LN713518.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	756	cluster I
LN713520.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP87*	741	cluster I
LN713523.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium opportunistum*	744	cluster I

Table 15 (continued).

LN713524.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP117*	750	cluster I
LN713525.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium tamadayense*	753	cluster I
LN713526.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP130*	741	cluster I
LN713527.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. CCANP133*	747	cluster I
LN831776.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus riograndensis*	864	cluster I
LN907867.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Blastochloris;Blastochloris viridis*	831	cluster I
LN908213.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium beijerinckii*	816	cluster III
LN997848.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum;Magnetospirillum sp. XM-1*	879	cluster I
LO017727.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;magneto-ovoid bacterium MO-1*	876	cluster I
LR130778.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Vallitaleaceae;Petrocella;Petrocella atlantisensis*	810	cluster III
LR131271.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Raoultella;Raoultella terrigena*	858	cluster I
LR134235.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella aerogenes*	879	cluster I
LR134253.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella aerogenes*	858	cluster I

Table 15 (continued).

LS398110.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ORS 3257*	882	cluster I
LT607803.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Variovorax;Variovorax sp. HW608*	879	cluster I
LT614670.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	747	cluster I
LT614672.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. LLAN64*	747	cluster I
LT614673.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. LLAN66*	744	cluster I
LT614674.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium sp. LLAN67*	747	cluster I
LT629751.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas oryzae*	879	cluster I
LT799040.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Kosakonia;Kosakonia oryzae*	879	cluster I
LT841305.1	*Bacteria;Proteobacteria;Acidithiobacillia;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus;Acidithiobacillus ferrivorans*	879	cluster I
LT859959.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; Bradyrhizobium sp. ORS 285*	885	cluster I
LT896716.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter; Geobacter sp. DSM 9736*	855	cluster I
LT906662.1	*Bacteria;Firmicutes;Clostridia;Thermoanaerobacterales;Thermoanaerobacterales Family III. Incertae Sedis;Thermoanaerobacterium;Thermoanaerobacterium sp. RBIITD*	813	cluster II

Table 15 (continued).

LT907844.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Cohaesibacteraceae;Cohaesibacter; Cohaesibacter sp. ES.047*	867	cluster I
LT907975.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Pseudodesulfovibrio;Pseudodesulfovibrio profundus*	822	cluster III
LT960614.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;unclassified family;Hartmannibacter;Hartmannibacter diazotrophicus*	846	cluster I
LT976979.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT976980.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT976981.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT976994.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT976998.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT977072.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT984799.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT984805.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT984808.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT984809.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
LT984815.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I

Table 15 (continued).

LT991978.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;Cupriavidus taiwanensis*	879	cluster I
M10587.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
M11579.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	867	cluster I
M15238.1	*Bacteria;Proteobacteria;Acidithiobacillia;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus;Acidithiobacillus ferrooxidans*	876	cluster I
M15270.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter capsulatus*	876	cluster I
M15942.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium phaseoli*	981	cluster I
M16709.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp.*	888	cluster I
M16710.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium sp.*	888	cluster I
M20568.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	867	cluster I
M21132.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp.*	852	cluster I
M23528.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	750	cluster II
M32371.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	870	cluster I
M33774.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Rhodospirillum;Rhodospirillum rubrum*	828	cluster I

Table 15 (continued).

M64344.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
M73020.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter chroococcum*	873	cluster I
MF374509.1	*Bacteria;Cyanobacteria;unclassified class;Chroococcales;unclassified family;unclassified genus;uncultured Chroococcales cyanobacterium*	801	cluster I
MG594028.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Aphanizomenonaceae;Aphanizomenon;Aphanizomenon flos-aquae*	882	cluster I
MG604339.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	735	cluster I
MH234567.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter sp.*	840	cluster I
MH339856.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium;Bradyrhizobium sp.*	744	cluster I
MH551311.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	801	cluster I
MH551317.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	882	cluster I
MH551326.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	864	cluster I
MH551333.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	873	cluster I
MH551458.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	876	cluster I
MH551467.1	*Bacteria;unclassified phylum;unclassified class;unclassified order;unclassified family;unclassified genus;uncultured bacterium*	855	cluster I

Table 15 (continued).

MH555146.1	*Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacillus beijingensis*	864	cluster I
MK226192.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; Mesorhizobium ciceri*	891	cluster I
U00090.2	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium fredii*	846	cluster I
U22146.1	*Bacteria;Cyanobacteria;unclassified class;Synechococcales;Synechococcaceae;Synechococcus;Synechococcus sp.*	861	cluster I
U23507.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Microcoleaceae;Trichodesmium;Trichodesmium thiebautii*	846	cluster I
U49514.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Mastigocladus;Mastigocladus laminosus*	882	cluster I
U49515.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Hapalosiphonaceae;Mastigocladus;Mastigocladus laminosus*	861	cluster I
U49859.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus variabilis*	858	cluster I
U53362.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp. EuIK1*	861	cluster I
U53363.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp.*	861	cluster I
U68183.1	*Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;Desulfovibrio gigas*	777	cluster III
U80928.5	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;Rhizobium etli*	981	cluster I
U89346.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Trichormus;Trichormus variabilis*	882	cluster I
U90952.1	*Bacteria;Cyanobacteria;unclassified class;Oscillatoriales;Microcoleaceae;Trichodesmium;Trichodesmium erythraeum*	771	cluster I

Table 15 (continued).

V00001.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7120*	885	cluster I
V00631.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	858	cluster I
V01215.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium;Sinorhizobium meliloti*	888	cluster I
V01482.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 7120*	885	cluster I
X03916.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter chroococcum*	870	cluster I
X07472.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	819	cluster III
X07473.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	816	cluster III
X07474.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	825	cluster II
X07475.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	819	cluster III
X07476.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	819	cluster III
X07477.1	*Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium;Clostridium pasteurianum*	816	cluster III
X07866.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter capsulatus*	876	cluster I
X13303.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella;Klebsiella pneumoniae*	858	cluster I
X13519.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter vinelandii*	870	cluster I

Table 15 (continued).

X17522.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia alni*	867	cluster I
X51500.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
X51609.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Azospirillum;Azospirillum brasilense*	867	cluster I
X51756.1	*Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Azotobacter;Azotobacter chroococcum*	870	cluster I
X56072.1	*Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosarcinaceae;Methanosarcina;Methanosarcina barkeri*	801	cluster II
X57006.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp.*	858	cluster I
X63352.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter capsulatus*	876	cluster I
X70033.1	*Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter;Rhodobacter capsulatus*	750	cluster II
X73983.1	*Bacteria;Actinobacteria;Actinobacteria;Frankiales;Frankiaceae;Frankia;Frankia sp.*	861	cluster I
X96609.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae;Alcaligenes;Alcaligenes faecalis*	876	cluster I
Z31716.1	*Bacteria;Cyanobacteria;unclassified class;Nostocales;Nostocaceae;Nostoc;Nostoc sp. PCC 6720*	882	cluster I
Z54207.1	*Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum;Herbaspirillum seropedicae*	876	cluster I

PUBLICATIONS

1. **Medina-Cordoba**, L.K., Chande, A.T., Rishishwar, L., Mayer, L.W., Marino-Ramirez, L., Valderrama-Aguirre, L.C., Valderrama-Aguirre, A., Kostka, J.E., Jordan, I.K. (2018). Genome Sequences of 15 Klebsiella sp. Isolates from Sugarcane Fields in Colombia's Cauca Valley. Genome Announcements. Vol 6.
2. **Medina-Cordoba**, L.K., Chande, A.T., Rishishwar, L., Mayer, L.W., Valderrama-Aguirre, L.C., Valderrama-Aguirre, A., Gaby, J.G., Kostka, J.E, Jordan, I.K. (2019). Genomic characterization and computational phenotyping of nitrogen-fixing bacteria isolated from Colombian sugarcane fields. bioRxiv 780809.

REFERENCES

1. Gangwar M, Saini P, Nikhanj P, Kaur S. 2017. Plant Growth-Promoting Microbes (PGPM) as Potential Microbial Bio-Agents for Eco-Friendly Agriculture. Advances in Soil Microbiology: Recent Trends and Future Prospects, Vol 2: Soil-Microbe-Plant Interaction 4:37-55.

2. Stewart WM, Dibb DW, Johnston AE, Smyth TJ. 2005. The contribution of commercial fertilizer nutrients to food production. Agronomy Journal 97:1-6.

3. Savci S. 2012. Investigation of effect of chemical fertilizers on environment. International Conference on Environmental Science and Development 1:287-292.

4. Macilwain C. 2004. Organic: Is it the future of farming? Nature 428:792-793.

5. Mahanty T, Bhattacharjee S, Goswami M, Bhattacharyya P, Das B, Ghosh A, Tribedi P. 2017. Biofertilizers: a potential approach for sustainable agriculture development. Environ Sci Pollut Res Int 24:3315-3335.

6. Bhardwaj D, Ansari MW, Sahoo RK, Tuteja N. 2014. Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. Microbial Cell Factories 13.

7. Curtis TP, Sloan WT, Scannell JW. 2002. Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci U S A 99:10494-9.

8. Egamberdieva D, Kamilova F, Validov S, Gafurova L, Kucharova Z, Lugtenberg B. 2008. High incidence of plant growth-stimulating bacteria associated with the rhizosphere of wheat grown on salinated soil in Uzbekistan. Environ Microbiol 10:1-9.

9. Berendsen RL, Pieterse CM, Bakker PA. 2012. The rhizosphere microbiome and plant health. Trends Plant Sci 17:478-86.

10. Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JH, Piceno YM, DeSantis TZ, Andersen GL, Bakker PA, Raaijmakers JM. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science 332:1097-100.

11. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat Rev Microbiol 11:789-99.

12. Hayat R, Ali S, Amara U, Khalid R, Ahmed I. 2010. Soil beneficial bacteria and their role in plant growth promotion: a review. Annals of Microbiology 60:579-598.

13. Turner TR, James EK, Poole PS. 2013. The plant microbiome. Genome Biol 14:209.

14. Unkovich M. 2008. Measuring plant-associated nitrogen fixation in agricultural systems. Australian Centre for International Agricultural Research, Canberra, ACT.

15. Beneduzi A, Moreira F, Costa PB, Vargas LK, Lisboa BB, Favreto R, Baldani JI, Passaglia LMP. 2013. Diversity and plant growth promoting evaluation abilities of bacteria isolated from sugarcane cultivated in the South of Brazil. Applied Soil Ecology 63:94-104.

16. Howard JB, Rees DC. 1996. Structural basis of biological nitrogen fixation. Chemical Reviews 96:2965-2982.

17. Wagner SC. 2011. Biological Nitrogen Fixation. Nature Education Knowledge 3:1.

18. Dixon R, Kahn D. 2004. Genetic regulation of biological nitrogen fixation. Nature Reviews Microbiology 2:621-631.

19. Illmer P, Schinner F. 1992. Solubilization of inorganic phosphates by microorganisms isolated from forest soils. Soil Biology & Biochemistry 24:389-395.

20. Hameeda B, Harini G, Rupela OP, Wani SP, Reddy G. 2008. Growth promotion of maize by phosphate-solubilizing bacteria isolated from composts and macrofauna. Microbiol Res 163:234-42.

21. Chen YP, Rekha PD, Arun AB, Shen FT, Lai WA, Young CC. 2006. Phosphate solubilizing bacteria from subtropical soil and their tricalcium phosphate solubilizing abilities. Applied Soil Ecology 34:33-41.

22. Bargaz A, Lyamlouli K, Chtouki M, Zeroual Y, Dhiba D. 2018. Soil microbial resources for improving fertilizers efficiency in an integrated plant nutrient management system. Front Microbiol 9:1606.

23. Neilands JB. 1995. Siderophores: structure and function of microbial iron transport compounds. J Biol Chem 270:26723-6.

24. Sepúlveda Cisternas I, Salazar JC, García-Angulo VA. 2018. Overview on the Bacterial Iron-Riboflavin Metabolic Axis. Frontiers in Microbiology 9.

25. Simon S, Petrasek J. 2011. Why plants need more than one type of auxin. Plant Sci 180:454-60.

26. Leveau JHJ, Lindow SE. 2005. Utilization of the plant hormone indole-3-acetic acid for growth by *Pseudomonas putida* strain 1290. Applied and Environmental Microbiology 71:2365-2371.

27. Bottini R, Cassan F, Piccoli P. 2004. Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. Appl Microbiol Biotechnol 65:497-503.

28. Cherubin MR, Karlen DL, Cerri CE, Franco AL, Tormena CA, Davies CA, Cerri CC. 2016. Soil quality indexing strategies for evaluating sugarcane expansion in Brazil. PLoS One 11:e0150860.

29. Cheavegatti-Gianotto A, de Abreu HMC, Arruda P, Bespalhok Filho JC, Burnquist WL, Creste S, di Ciero L, Ferro JA, de Oliveira Figueira AV, de Sousa Filgueiras T, Grossi-de-Sá MdF, Guzzo EC, Hoffmann HP, de Andrade Landell MG, Macedo N, Matsuoka S, de Castro Reinach F, Romano E, da Silva WJ, de Castro Silva Filho M, César Ulian E. 2011. Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. Tropical Plant Biology 4:62-89.

30. Meyer PM, Rodrigues PHM, Millen DD. 2013. Impact of biofuel production in Brazil on the economy, agriculture, and the environment. Animal Frontiers 3:28-37.

31. Selman-Housein G, Lopez MA, Ramos O, Carmona ER, Arencibia AD, Menendez E, Miranda F. 2000. Towards the improvement of sugarcane bagasse as raw material for the production of paper pulp and animal feed. Plant Genetic Engineering: Towards the Third Millennium 5:189-193.

32. Irvine JE. 1999. Saccharum species as horticultural classes. Theoretical and Applied Genetics 98:186-194.

33. Lakshmanan P, Geijskes RJ, Aitken KS, Grof CLP, Bonnett GD, Smith GR. 2005. Sugarcane Biotechnology: The Challenges and Opportunities. In Vitro Cellular and Developmental Biology - Plant 41:345-363, 19.

34. Lazenby A, Matheson EM. 1975. Australian Field Crops: Tropical cereals, oilseeds, grain legumes and other crops. Angus & Robertson.

35. Singh J, Singh AK, Sharma MP, Singh PR, Srivastava AC. 2011. Mechanization of Sugarcane Cultivation in India. Sugar Tech 13:310-314.

36. OECD. 2016. Sugarcane (*Saccharum spp.)* Safety Assessment of Transgenic Organisms in the Environment, Volume 6: OECD Consensus Documents doi:doi:https://doi.org/10.1787/9789264253421-en. OECD Publishing, Paris.

37. Ohyama T, Momose A, Ohtake N, Sueyoshi K, Sato T, Nakanishi Y, Asis CA, Ruamsungsri S, Ando S. 2014. Nitrogen fixation in sugarcane doi:10.5772/56993. IntechOpen Limited, London.

38. Gopalasundaram P, Bhaskaran A, Rakkiyappan P. 2012. Integrated nutrient management in sugarcane. Sugar Tech 14:3-20.

39. Li HB, Singh RK, Singh P, Song QQ, Xing YX, Yang LT, Li YR. 2017. Genetic diversity of nitrogen-fixing and plant growth promoting *Pseudomonas* species isolated from sugarcane rhizosphere. Front Microbiol 8:1268.

40. Bordonal RdO, Carvalho JLN, Lal R, de Figueiredo EB, de Oliveira BG, La Scala N. 2018. Sustainability of sugarcane production in Brazil. A review. Agronomy for Sustainable Development 38:13.

41. Rodrigues AA, Forzani MV, Soares RdS, Sibov ST, Vieira JDG. 2016. Isolation and selection of plant growth-promoting bacteria associated with sugarcane. Pesquisa Agropecuária Tropical 46:149-158.

42. Yeoh YK, Paungfoo-Lonhienne C, Dennis PG, Robinson N, Ragan MA, Schmidt S, Hugenholtz P. 2016. The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application. Environ Microbiol 18:1338-51.

43. Dong M, Yang Z, Cheng G, Peng L, Xu Q, Xu J. 2018. Diversity of the bacterial microbiome in the roots of four *Saccharum* species: *S. spontaneum, S. robustum, S. barberi,* and *S. officinarum*. Front Microbiol 9:267.

44. dos Santos SG, Chaves VA, Ribeiro FD, Alves GC, Reis VM. 2019. Rooting and growth of pre-germinated sugarcane seedlings inoculated with diazotrophic bacteria. Applied Soil Ecology 133:12-23.

45. Levy A, Conway JM, Dangl JL, Woyke T. 2018. Elucidating Bacterial Gene Functions in the Plant Microbiome. Cell Host Microbe 24:475-485.

46. Armanhi JSL, de Souza RSC, Damasceno NdB, de Araújo LM, Imperial J, Arruda P. 2018. A Community-Based Culture Collection for Targeting Novel Plant Growth-Promoting Bacteria from the Sugarcane Microbiome. Frontiers in Plant Science 8.

47. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M. 2002. Cultivating the uncultured. Proc Natl Acad Sci U S A 99:15681-6.

48. Vartoukian SR, Palmer RM, Wade WG. 2010. Strategies for culture of 'unculturable' bacteria. FEMS Microbiol Lett 309:1-7.

49. Castrillo G, Teixeira PJ, Paredes SH, Law TF, de Lorenzo L, Feltcher ME, Finkel OM, Breakfield NW, Mieczkowski P, Jones CD, Paz-Ares J, Dangl JL. 2017. Root microbiota drive direct integration of phosphate stress and immunity. Nature 543:513-518.

50. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Glavina del Rio T, Jones CD, Tringe SG, Dangl JL. 2015. PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349:860-4.

51. Dobereiner J. 1961. Nitrogen-fixing bacteria of the genusBeijerinckia Derx in the rhizosphere of sugar cane. Plant and Soil 15:211-216.

52. Bastián F, Cohen A, Piccoli P, Luna V, Bottini* R, Baraldi R, Bottini R. 1998. Production of indole-3-acetic acid and gibberellins A1 and A3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. Plant Growth Regulation 24:7-11.

53. Olivares FL, Baldani VLD, Reis VM, Baldani JI, Döbereiner J. 1996. Occurrence of the endophytic diazotrophs *Herbaspirillum* spp. in roots, stems, and leaves, predominantly of Gramineae. Biology and Fertility of Soils 21:197-200.

54. Schloss PD, Handelsman J. 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol 6:229.

55. Anonymous. 2017. Bacterial Endophyte Colonization and Distribution within Plants. Microorganisms 5.

56. Iniguez AL, Dong Y, Triplett EW. 2004. Nitrogen fixation in wheat provided by *Klebsiella pneumoniae* 342. Mol Plant Microbe Interact 17:1078-85.

57. Castanheira NL, Dourado AC, Pais I, Semedo J, Scotti-Campos P, Borges N, Carvalho G, Barreto Crespo MT, Fareleira P. 2017. Colonization and beneficial effects on annual ryegrass by mixed inoculation with plant growth promoting bacteria. Microbiol Res 198:47-55.

58. Taulé C, Castillo A, Villar S, Olivares F, Battistoni F. 2016. Endophytic colonization of sugarcane (*Saccharum officinarum*) by the novel diazotrophs *Shinella* sp. UYSO24 and *Enterobacter* sp. UYSO10. Plant and Soil 403:403-418.

59. Lin L, Li Z, Hu C, Zhang X, Chang S, Yang L, Li Y, An Q. 2012. Plant growth-promoting nitrogen-fixing enterobacteria are in association with sugarcane plants growing in Guangxi, China. Microbes Environ 27:391-8.

60. Solanki MK, Wang Z, Wang F-Y, Li C-N, Lan T-J, Singh RK, Singh P, Yang L-T, Li Y-R. 2017. Intercropping in Sugarcane Cultivation Influenced the Soil Properties and Enhanced the Diversity of Vital Diazotrophic Bacteria. Sugar Tech 19:136-147.

61. Paungfoo-Lonhienne C, Lonhienne TGA, Yeoh YK, Donose BC, Webb RI, Parsons J, Liao W, Sagulenko E, Lakshmanan P, Hugenholtz P, Schmidt S, Ragan MA. 2016. Crosstalk between sugarcane and a plant-growth promoting Burkholderia species. Scientific Reports 6.

62. Suman A, Yadav AN, Verma P. 2016. Endophytic Microbes in Crops: Diversity and Beneficial Impact for Sustainable Agriculture, p 117-143. *In* Singh DP, Singh HB, Prabha R (ed), Microbial Inoculants in Sustainable Agricultural Productivity:

Vol 1: Research Perspectives doi:10.1007/978-81-322-2647-5_7. Springer India, New Delhi.

63. Glick BR. 2012. Plant growth-promoting bacteria: mechanisms and applications. Scientifica (Cairo) 2012:963401.

64. Bashan Y, de-Bashan LE, Prabhu SR, Hernandez J-P. 2014. Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). Plant and Soil 378:1-33.

65. Pereira W, Sousa JS, Schultz N, Reis VM. 2019. Sugarcane Productivity as a Function of Nitrogen Fertilization and Inoculation with Diazotrophic Plant Growth-Promoting Bacteria. Sugar Tech 21:71-82.

66. Malusà E, Pinzari F, Canfora L. 2016. Efficacy of Biofertilizers: Challenges to Improve Crop Production, p 17-40. *In* Singh DP, Singh HB, Prabha R (ed), Microbial Inoculants in Sustainable Agricultural Productivity: Vol 2: Functional Applications doi:10.1007/978-81-322-2644-4_2. Springer India, New Delhi.

67. Timmusk S, Behers L, Muthoni J, Muraya A, Aronsson A-C. 2017. Perspectives and Challenges of Microbial Application for Crop Improvement. Frontiers in Plant Science 8.

68. Timmusk S, Behers L, Muthoni J, Muraya A, Aronsson AC. 2017. Perspectives and Challenges of Microbial Application for Crop Improvement. Front Plant Sci 8:49.

69. Lesueur D, Deaker R, Herrmann L, Bräu L, Jansa J. 2016. The Production and Potential of Biofertilizers to Improve Crop Yields, p 71-92. *In* Arora NK, Mehnaz S, Balestrini R (ed), Bioformulations: for Sustainable Agriculture doi:10.1007/978-81-322-2779-3_4. Springer India, New Delhi.

70. Gaby JC, Rishishwar L, Valderrama-Aguirre LC, Green SJ, Valderrama-Aguirre A, Jordan IK, Kostka JE. 2018. Diazotroph community characterization via a high-throughput *nifH* amplicon sequencing and analysis pipeline. Appl Environ Microbiol 84.

71. Cardinale M, Brusetti L, Quatrini P, Borin S, Puglia AM, Rizzi A, Zanardini E, Sorlini C, Corselli C, Daffonchio D. 2004. Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. Applied and Environmental Microbiology 70:6147-6156.

72. Andrews S. FastQC a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 07/31/2017.

73. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

74. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19:455-477.

75. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. 2008. The RAST Server: rapid annotations using subsystems technology. BMC genomics 9:75.

76. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R. 2013. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Research 42:D581-D591.

77. Medina-Cordoba LK, Chande AT, Rishishwar L, Mayer LW, Marino-Ramirez L, Valderrama-Aguirre LC, Valderrama-Aguirre A, Kostka JE, Jordan IK. 2018. Genome sequences of 15 *Klebsiella* sp. isolates from sugarcane fields in Colombia's Cauca Valley. Genome Announc 6.

78. Fess TL, Kotcon JB, Benedito VA. 2011. Crop breeding for low input agriculture: A sustainable response to feed a growing world population. Sustainability 3:1742-1772.

79. Tilman D, Balzer C, Hill J, Befort BL. 2011. Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci U S A 108:20260-4.

80. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614-24.

81. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences of the United States of America 102:2567-2572.

82. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. International Journal of Systematic and Evolutionary Microbiology 57:81-91.

83. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12.

84. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H,

Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-45.

85. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-25.

86. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547-1549.

87. Seemann T. 2018. barrnap: BAsic Rapid Ribosomal RNA Predictor, v0.9. https://github.com/tseemann/barrnap.

88. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

89. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.

90. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

91. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

92. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

93. Grant JR, Arantes AS, Stothard P. 2012. Comparing thousands of circular genomes using the CGView comparison tool. BMC Genomics 13:202.

94. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2015. GenBank. Nucleic Acids Res 43:D30-5.

95. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. Hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res 44:D694-697.

96. Bailey DC, Alexander E, Rice MR, Drake EJ, Mydy LS, Aldrich CC, Gulick AM. 2018. Structural and functional delineation of aerobactin biosynthesis in hypervirulent *Klebsiella pneumoniae*. J Biol Chem 293:7841-7852.

97. Consortium TU. 2018. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47:D506-D515.

98. Mademidis A, Köster W. 1998. Transport activity of FhuA, FhuC, FhuD, and FhuB derivatives in a system free of polar effects, and stoichiometry of components involved in ferrichrome uptake. Molecular and General Genetics MGG 258:156-165.

99. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, Shukla M, Stevens RL, Xia F, Yoo H, Davis JJ. 2018. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. Sci Rep 8:421.

100. Wayne PA. 2015. CLSI. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fifth Informational Supplement.

101. Rice WA, Paul EA. 1971. The acetylene reduction assay for measuring nitrogen fixation in waterlogged soil. Can J Microbiol 17:1049-56.

102. Chakdar H, Dastager SG, Khire JM, Rane D, Dharne MS. 2018. Characterization of mineral phosphate solubilizing and plant growth promoting bacteria from termite soil of arid region. 3 Biotech 8:463.

103. Silva Filho GN, Vidor C. 2000. Solubilização de fostatos por microrganismos na presença de fontes de carbono. Revista Brasileira de Ciência do Solo 24:311-319.

104. Schwyn B, Neilands JB. 1987. Universal chemical assay for the detection and determination of siderophores. Anal Biochem 160:47-56.

105. Lakshmanan V, Shantharaj D, Li G, Seyfferth AL, Sherrier DJ, Bais HP. 2015. A natural rice rhizospheric bacterium abates arsenic accumulation in rice (*Oryza sativa* L.). Planta 242:1037-1050.

106. de Salamone IEG, Hynes RK, Nelson LM. 2001. Cytokinin production by plant growth promoting rhizobacteria and selected mutants. Canadian Journal of Microbiology 47:404-411.

107. Gordon SA, Weber RP. 1951. Colorimetric estimation of indoleacetic acid. Plant Physiol 26:192-195.

108. Postgate JR. 1982. Biological nitrogen fixation: fundamentals. Philos Trans R Soc Lond B Biol Sci 296:375-385.

109. Li B, Zhao Y, Liu C, Chen Z, Zhou D. 2014. Molecular pathogenesis of *Klebsiella pneumoniae*. Future Microbiol 9:1071-81.

110. Fouts DE, Tyler HL, DeBoy RT, Daugherty S, Ren Q, Badger JH, Durkin AS, Huot H, Shrivastava S, Kothari S, Dodson RJ, Mohamoud Y, Khouri H, Roesch LF, Krogfelt KA, Struve C, Triplett EW, Methe BA. 2008. Complete genome sequence of the N_2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. PLoS Genet 4:e1000141.

111. Dantur KI, Chalfoun NR, Claps MP, Tortora ML, Silva C, Jure A, Porcel N, Bianco MI, Vojnov A, Castagnaro AP, Welin B. 2018. The endophytic strain *Klebsiella michiganensis* Kd70 lacks pathogenic island-like regions in its genome and is incapable of infecting the urinary tract in mice. Front Microbiol 9:1548.

112. Rosenblueth M, Martinez L, Silva J, Martinez-Romero E. 2004. *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. Systematic and Applied Microbiology 27:27-35.

113. Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. Mol Biol Evol 21:541-54.

114. Zehr JP, Jenkins BD, Short SM, Steward GF. 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environ Microbiol 5:539-54.

115. Weimann A, Mooren K, Frank J, Pope PB, Bremges A, McHardy AC. 2016. From genomes to phenotypes: Traitar, the microbial trait analyzer. mSystems 1:e00101-16.

116. Deredjian A, Alliot N, Blanchard L, Brothier E, Anane M, Cambier P, Jolivet C, Khelil MN, Nazaret S, Saby N, Thioulouse J, Favre-Bonte S. 2016. Occurrence of *Stenotrophomonas maltophilia* in agricultural soils and antibiotic resistance properties. Research in Microbiology 167:313-324.

117. Caulier S, Gillis A, Colau G, Licciardi F, Liepin M, Desoignies N, Modrie P, Legreve A, Mahillon J, Bragard C. 2018. Versatile antagonistic activities of soil-borne *Bacillus* spp. and *Pseudomonas* spp. against *Phytophthora infestans* and other potato pathogens. Front Microbiol 9:143.

118. Badran S, Morales N, Schick P, Jacoby B, Villella W, Lorenz T. 2018. Complete genome sequence of the *Bacillus pumilus* phage Leo2. Genome Announc 6.

119. Pavan ME, Franco RJ, Rodriguez JM, Gadaleta P, Abbott SL, Janda JM, Zorzopulos J. 2005. Phylogenetic relationships of the genus *Kluyvera*: transfer of *Enterobacter intermedius* Izard et al. 1980 to the genus *Kluyvera* as *Kluyvera intermedia* comb. nov. and reclassification of *Kluyvera cochleae* as a later synonym of *K. intermedia*. Int J Syst Evol Microbiol 55:437-42.

120. Berger B, Wiesner M, Brock AK, Schreiner M, Ruppel S. 2015. *K. radicincitans*, a beneficial bacteria that promotes radish growth under field conditions. Agronomy for Sustainable Development 35:1521-1528.

121. Stacey G, Burris RH, Evans HJ. 1992. Biological Nitrogen Fixation. Chapman and Hall, New York.

122. Scott KF, Rolfe BG, Shine J. 1981. Biological nitrogen fixation: primary structure of the *Klebsiella pneumoniae nifH* and *nifD* genes. J Mol Appl Genet 1:71-81.

123. Luo T, Ou-Yang XQ, Yang LT, Li YR, Song XP, Zhang GM, Gao YJ, Duan WX, An Q. 2016. *Raoultella* sp. strain L03 fixes N_2 in association with micropropagated sugarcane plants. J Basic Microbiol 56:934-40.

124. Schicklberger M, Shapiro N, Loque D, Woyke T, Chakraborty R. 2015. Draft genome sequence of *Raoultella terrigena* R1Gly, a diazotrophic endophyte. Genome Announc 3.

125. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao HW, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen KV, Nguyen TV, Dao TT, Mensinke M, Minh VL, Nhu NTK, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR. 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proceedings of the National Academy of Sciences of the United States of America 112:E3574-E3581.

126. Magnani GS, Didonet CM, Cruz LM, Picheth CF, Pedrosa FO, Souza EM. 2010. Diversity of endophytic bacteria in Brazilian sugarcane. Genet Mol Res 9:250-8.

127. Mehnaz S, Baig DN, Lazarovits G. 2010. Genetic and phenotypic diversity of plant growth promoting rhizobacteria isolated from sugarcane plants growing in pakistan. J Microbiol Biotechnol 20:1614-23.

128. Ferrara FID, Oliveira ZM, Gonzales HHS, Floh EIS, Barbosa HR. 2012. Endophytic and rhizospheric enterobacteria isolated from sugar cane have different potentials for producing plant growth-promoting substances. Plant and Soil 353:409-417.

129. Taule C, Mareque C, Barlocco C, Hackembruch F, Reis VM, Sicardi M, Battistoni F. 2012. The contribution of nitrogen fixation to sugarcane (*Saccharum officinarum* L.), and the identification and characterization of part of the associated diazotrophic bacterial community. Plant and Soil 356:35-49.

130. Bagley ST. 1985. Habitat association of *Klebsiella* species. Infect Control 6:52-8.

131. Wei CY, Lin L, Luo LJ, Xing YX, Hu CJ, Yang LT, Li YR, An QL. 2014. Endophytic nitrogen-fixing *Klebsiella variicola* strain DX120E promotes sugarcane growth. Biology and Fertility of Soils 50:657-666.

132. Ji SH, Gururani MA, Chun SC. 2014. Isolation and characterization of plant growth promoting endophytic diazotrophic bacteria from Korean rice cultivars. Microbiol Res 169:83-98.

133. Lin L, Wei C, Chen M, Wang H, Li Y, Li Y, Yang L, An Q. 2015. Complete genome sequence of endophytic nitrogen-fixing *Klebsiella variicola* strain DX120E. Stand Genomic Sci 10:22.

134. Denton M, Kerr KG. 1998. Microbiological and clinical aspects of infection associated with *Stenotrophomonas maltophilia*. Clin Microbiol Rev 11:57-80.

135. Downing KJ, Leslie G, Thomson JA. 2000. Biocontrol of the sugarcane borer Eldana saccharina by expression of the *Bacillus thuringiensis* cry1Ac7 and *Serratia marcescens* chiA genes in sugarcane-associated bacteria. Appl Environ Microbiol 66:2804-10.

136. Ribeiro VB, Zavascki AP, Rozales FP, Pagano M, Magagnin CM, Nodari CS, da Silva RC, Dalarosa MG, Falci DR, Barth AL. 2014. Detection of bla(GES-5) in carbapenem-resistant *Kluyvera intermedia* isolates recovered from the hospital environment. Antimicrob Agents Chemother 58:622-3.

137. Juhnke ME, des Jardin E. 1989. Selective medium for isolation of *Xanthomonas maltophilia* from soil and rhizosphere environments. Appl Environ Microbiol 55:747-50.

138. Richesson RL, Sun JM, Pathak J, Kho AN, Denny JC. 2016. Clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. Artificial Intelligence in Medicine 71:57-61.

139. Drouin A, Giguere S, Deraspe M, Marchand M, Tyers M, Loo VG, Bourgault AM, Laviolette F, Corbeil J. 2016. Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. Bmc Genomics 17.

140. Berger AH, Brooks AN, Wu X, Shrestha Y, Chouinard C, Piccioni F, Bagul M, Kamburov A, Innielinski M, Hogstrom L, Zhu C, Yang X, Pantel S, Sakai R, Kaplan N, Root D, Narayan R, Natoli T, Lahr D, Tirosh I, Tamayo P, Getz G, Wong B, Doench J, Subramanian A, Golub TR, Meyerson M, Boehm JS. 2016. High-throughput phenotyping of lung cancer somatic mutations. Cancer Research 76.

141. Bone WP, Washington NL, Buske OJ, Adams DR, Davis J, Draper D, Flynn ED, Girdea M, Godfrey R, Golas G, Groden C, Jacobsen J, Kohler S, Lee EMJ, Links AE, Markello TC, Mungall CJ, Nehrebecky M, Robinson PN, Sincan M, Soldatos AG, Tifft CJ, Toro C, Trang H, Valkanas E, Vasilevsky N, Wahl C, Wolfe LA, Boerkoel CF, Brudno M, Haendel MA, Gahl WA, Smedley D. 2016. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. Genetics in Medicine 18:608-617.

142. Berg G, Erlacher A, Smalla K, Krause R. 2014. Vegetable microbiomes: is there a connection among opportunistic infections, human health and our 'gut feeling'? Microb Biotechnol 7:487-95.

143. Fishman JA. 2013. Opportunistic infections--coming to the limits of immunosuppression? Cold Spring Harb Perspect Med 3:a015669.

144. Rosenblueth M, Martinez L, Silva J, Martinez-Romero E. 2004. *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. Syst Appl Microbiol 27:27-35.

145. Podschun R, Ullmann U. 1998. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin Microbiol Rev 11:589-603.

146. Osterblad M, Pensala O, Peterzens M, Heleniusc H, Huovinen P. 1999. Antimicrobial susceptibility of Enterobacteriaceae isolated from vegetables. J Antimicrob Chemother 43:503-9.

147. Canica M, Manageiro V, Abriouel H, Moran-Gilad J, Franz CMAP. 2019. Antibiotic resistance in foodborne bacteria. Trends in Food Science & Technology 84:41-44.

148. Bengtsson-Palme J. 2017. Antibiotic resistance in the food supply chain: where can sequencing and metagenomics aid risk assessment? Current Opinion in Food Science 14:66-71.

149. Gaby JC, Buckley DH. 2014. A comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. Database (Oxford) 2014:bau001.

150. Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, Paramasivan P, Ryu M-H, Oldroyd GED, Poole PS, Udvardi MK, Voigt CA, Ané J-M, Peters JW. 2016. Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. Applied and Environmental Microbiology 82:3698-3710.

151. Oldroyd GED, Dixon R. 2014. Biotechnological solutions to the nitrogen problem. Current Opinion in Biotechnology 26:19-24.

152. Hu Y, Ribbe MW. 2015. Nitrogenase and homologs. J Biol Inorg Chem 20:435-45.

153. Zehr JP, Capone DG. 1996. Problems and promises of assaying the genetic potential for nitrogen fixation in the marine environment. Microbial Ecology 32:263-281.

154. Heller P, Tripp HJ, Turk-Kubo K, Zehr JP. 2014. ARBitrator: a software pipeline for on-demand retrieval of auto-curated *nifH* sequences from GenBank. Bioinformatics 30:2883-2890.

155. Campello RJGB, Moulavi D, Sander J. Density-Based Clustering Based on Hierarchical Density Estimates, p 160-172. *In* (ed), Springer Berlin Heidelberg,

156. Gupta A, Jordan IK, Rishishwar L. 2017. stringMLST: a fast k-mer based tool for multilocus sequence typing. Bioinformatics 33:119-121.