THE ANALYSIS OF THE DIVERSITY AND DISTRIBUTION OF LEAF ENDOPHYTIC BACTERIAL COMMUNITIES

By

TAO DING

Bachelor of Science in Cell and Molecular Biology
University of Science and Technology of China
Hefei, China
2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2012

THE ANALYSIS OF THE DIVERSITY AND

DISTRIBUTION OF LEAF ENDOPHYTIC BACTERIAL

COMMUNITIES

Thesis Approved:

Dr. Ulrich K. Melcher

Thesis Adviser

Dr. Udaya DeSilva

Dr. Ramamurthy Mahalingam

Dr. Mostafa Elshahed

Outside Committee Member

Dr. Sheryl A. Tucker

Dean of the Graduate College

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

CHAPTER II

CHAPTER III

Table Page

LIST OF FIGURES

Figure Page

CHAPTER II

Figure Page

CHAPTER III

CHAPTER I

REVIEW OF LITERATURE

1 Endophytic bacteria and leaf endophytic bacteria

1.1 Endophytic bacteria

Plants and animals, including human beings are normally associated with diverse microorganisms, especially bacteria. *Escherichia coli* is best known for its role in animal gut mutualistic bacteria, contributing 0.1% to the complete gut flora (24). Gut flora greatly stimulate immunity, nutrient absorption, mucosal barrier fortification and xenobiotic metabolism of the host (37). Similarly, plants are also associated with diverse bacteria which contribute to host plants in a way similar to how they do so to host animals. Bacteria are associated with plants in many ways. Rhizosphere bacteria are associated with the roots of plants while those associated with leaves are designated phyllosphere bacteria. Besides those rhizosphere and phyllosphere bacteria which are loosely attached to the plant surface, called epiphytes, bacteria can also live inside the plants, the endophytic bacteria.

Although plants can be infected by pathogenic bacteria and cause disease symptoms,

bacteria are also found living inside plants without causing any symptoms, suggesting plants may require associated bacteria for their growth. The conceptions of endophytic bacteria or bacterial endophytes were introduced to describe bacteria which live inside healthy plants (61). Hallmann proposed a commonly accepted definition of endophytic bacteria: endophytic bacteiria are those bacteria that "do not visibly harm the host plant but can be isolated from the surface-disinfested plant tissue or the inner parts of plants" (33). This definition emphasizes that endophytic bacteria are not pathogenic or causing symptoms; the possible mutualistic interaction between endophytic bacteria and host plants might be a result of positive mutual selection (85). Hallmann's definition also indicated that endophytic bacteria are distinguishable from rhizosphere and phyllosphere bacteria, and only isolated from surface-disinfested plants.

1.2 Function of endophytic bacteria and related research

Endophytic bacteria have been found in almost all known plant species (70), and compared to rhizosphere bacteria or pathogenic bacteria, endophytic bacteria live at lower population densities (33). Before the application of cultivation-independent molecular techniques, bacterial isolation was the major research method and endophytic bacteria have been successfully isolated from a large diversity of plants (81). Many examining methods have also been established to detect endophytic bacteria, for example: Bell *et al*. (5) introduced the media screening method in their study of endophytic bacteria in grapevine. Dong *et al* (22) employed electron microscopy in the research of sugarcane endophytic bacteria.

Since endophytic bacteria are not pathogenic and causing no symptom to host plants, their roles are of great interest to discover. Scientists have shown that endophytic bacteria play several important roles in host plants, either beneficial to the host or potentially dangerous to it. Reports indicated that endophytic bacteria have biocontrol capacity against plant pathogens, insects and nematodes (33). Duijff *et al*. discovered the ability of the O-antigenic side chain of the outer membrane lipopolysaccharides of endophytic *Pseudomonas fluorescens* to induce resistance against fusarium wilt disease in tomatoes (23). Azevedo *et al*. summarized the role played by endophytic bacteria in biocontrol against pest and disease affecting cultivated plants (3). Most of plant disease preventions by endophytic bacteria have been shown to be based on novel compounds synthesized by endophytic bacteria. As a result, the metabolites of endophytic bacteria are studied to identify new medications for human, animal and plant diseases (79).

Chanway *et al*. found that endophytic bacteria accelerate the seedling emergence in their experiment inoculating plant growth-promoting soil bacteria in tree roots (14). In another experiment, Bent and Chanway (6) found the growth-promoting effects of endophytic Bacillus polymyxa on lodgepole pine. Many bacterial endophytes have a natural capacity to degrade xenobiotics (30) and are resistant to heavy metals (72) or antimicrobials (72). These discoveries indicated that endophytic bacteria can play important roles in phytoremediation (30, 56, 77) .

Besides the beneficial effects discussed above, endophytic bacteria may also be potentially dangerous. Although endophytic bacteria are not pathogenic to their host plants and cause no symptoms, they still could be pathogenic to other plants, the cattle that consume them and even human beings, leading endophytic bacteria to be an

important issue in agricultural production and food safety. Both common and opportunistic human pathogenic bacteria have been found living as endophytes in plants especially cultivated plants. Cooley *et al*. found that two enteric pathogens, *Salmonella enterica* and *Escherichia coli* O157:H7 can live in *Arabidopsis thaliana* as endophytes for at least 21 days causing no symptom to host plants (19). These pathogens have also been detected in fresh produce, posing a risk for human beings. Guo *et al*. found *Salmonella* as endophytes in tomatoes (31). Islam *et al* found *S. enteric* serovar Typhimurium as endophytes in carrots and radishes which were treated with contaminated water (40). In an experiment using fresh bovine manure in planting lettuce, radish and carrot, the prevalence of *E. coli*, which are often used as an indicator of potential contamination with fecal pathogens, was significantly higher in lettuce than radish and carrot (39), leading us to pay more attention to the safety of lettuce. Common human pathogens of the genus of *Mycobacterium* have been detected as endophytes in Coon *et al*.'s research of wheat roots using terminal restriction fragment length polymorphism (T-RFLP) (18) . Pirttilä *et al*. also found *Mycobacterium* as endophytes in buds of Scots pine (65). Some opportunistic human pathogens were also found in cultivated plants. The most important discovery is that several opportunistic pathogenic bacteria have been found living as endophytes in rice, an important crop plant, including *Sphingomonas paucimobilis* (27), *Chromobacterium violaceum* (63), *Serratia marcescens* (32) and *Sphingobacterium sp*. (63).

1.3 Potential applications of endophytic bacteria

Endophytic bacteria have many important functions that benefit the host plants and the environment, leading scientists to explore ways to use endophytic bacteria. The existence

of endophytic bacteria can promote host plant growth by many mechanisms. Verma *et al*. (90) identified *Pantoea agglomerans*, living in the seeds of rice, as a potential plant growth promoting endophytic diazotroph for deep water rice because of the bacteria's ability to solubilize mineral phosphates, enhancing the availability of phosphate for microbial and plant growth (90). Lee *et al*. found that *Gluconacetobacter diazotrophicus*, an endophyte of sugarcane can produce indole-3-acetic acid (IAA), a well known plant growth-promoting substance besides its nitrogen-fixation role, to promote sugarcane growth especially when the fertilizer input is low (47). Endophytic bacteria may also contribute essential vitamins to host plants. Pirttila *et al*. found that *Methylobacterium extorquens*, an endophyte of Scots pine, produced adenine derivatives which can be used in cytokinin biosynthesis (64).

Similar to systemic-acquired resistance (SAR), certain endophytic bacteria induce a phenomenon called induced systemic resistance (ISR) (71). Since endophytic bacteria can help protect host plants against pathogenic bacteria and fungi, viruses, insects and nematodes, endophytic bacteria have been used to inoculate plants for biocontrol purposes. Kerry summarized the endophytic bacterial agents for biocontrol of plant-parasitic nematodes (43). Berg and Hallman's book chapter discussed the use of endophytic bacteria for the biocontrol of pathogenic fungi (8).

Endophytic bacteria also have the potential to improve phytobioremediation since they can degrade xenobiotics (30). Van Aken *et al*. isolated a strain of Methylobacterium from hybrid poplar trees which can degrade nitro-aromatic compounds such as 2,4,6-trinitrotoluene (88). Endophytic bacteria can also be engineered to have some biodegradation capacity for phytoremediation purposes. Bract *et al*. reported that a

natural endophyte of yellow lupine *Burkholderia cepacia* that was genetically modified by introduction of pTOM toluene-degradation plasmid, significantly degrades toluene (4). This strategy can significantly improve the phytoremediation efficiency. Newman and Reynolds summarized the advantages of using endophytic bacteria to improve xenobiotic remediation (58).

Many endophytic bacteria like *Pseudomonas*, *Burkholderia* and *Bacillus* are well known for their secondary metabolites which can be used as antibiotics, anticancer drugs, antifungal, antiviral and immunosuppressant agents (51). This leads scientists to develop new drugs and novel treatment from the natural products of endophytic bacteria. Besides medication purposes, natural products can also inspire new materials. Catalan *et al*. have shown that *Herbaspirillum seropedicae*, a diazotrophic endophyte, accumulates significant levels of poly-3-hydroxybutyrate, a bioplastic (13).

2 Current research of endophytic bacterial communities

2.1 The development of techniques for microbial community research

Microbiology started from the cultivation of human pathogenic bacteria in laboratories. Since then, the classical methods to study microbial communities are mainly based on isolation and media-culture. As the knowledge of microorganisms expanded greatly in the 20th century, scientists realized that most bacteria especially those from natural environments, are unculturable on laboratory media. Therefore the estimation of the diversity of certain microbial communities is biased and much lower than the reality. As a result, many culture-independent methods were developed. Most techniques were based on the hypervariable regions of specific genes to identify microorganisms; 16S ribosomal

RNA genes were the most popular genes since they contain both hypervariable regions for species identification and highly conserved regions which can be amplified using universal primers for all prokaryotes (35). Phylogenetic analysis of the 16S rDNA has been used widely in taxonomic research. Tiedje's group has demonstrated that the analyses based on 16S rDNA are congruent with those based on genomic approaches, and has established and maintained the Ribosomal Database Project (RDP) (17).

Since Sanger sequencing was introduced, sequencing of the cloned polymerase chain reaction (PCR) products of rDNA has been the dominant method to provide phylogentic information to characterize bacterial communities (35). However, for complex bacterial communities, cloning and Sanger sequencing are time-consuming and labor-intensive, so many rapid profiling fingerprint techniques were developed based on 16S rDNA, including denaturing gradient gel electrophoresis (DGGE) (57), temperature gradient gel electrophoresis (TGGE) (36), and single-strand conformation polymorphism (SSCP) (46), length heterogeneity-PCR (LH-PCR) (83), and terminal restriction fragment length polymorphism (T-RFLP) (49). These techniques are very useful for comparison of the differences among bacterial communities from diverse environments.

Since 454 pyrosequencing was introduced in 2005 as the first successfully commercialized next-generation sequencing (NGS) technique (53), the NGS, especially 454 pyrosequencing has been widely used in profiling of microbial communities. Hamady *et al*. (34) designed an error-barcode set and made tag pyrosequencing of hundreds of microbial community DNA samples simultaneously available, greatly improve the sequencing efficiency. Roche 454 pyrosequencing generates long sequencing

reads (400-500 bp), which avoid the problem of artificial recombinants due to the highly conserved regions of 16S rDNA, making the sequencing results more reliable.

2.2 Research of endophytic bacterial communities

Since endophytic bacteria have many unknown potentials, the structures of the endophytic bacterial communities need to be characterized to understand how the endophytic bacteria interact with the host plants. Although traditional cultivation-based methods can not reveal the real diversity of endophytic bacterial communities, Elvira-Recuenco *et al*. isolated endophytic bacteria from pea cultivars under field conditions and found that endophytic bacterial populations decreased from the lower to the upper part of the stem (26). This result supported the theory that endophytic bacteria originated from soil and rhizosphere bacteria.

The introduction of cultivation-independent methods greatly improved the microbial community research, and fingerprint techniques based on 16S rDNA and pyrosequencing have been widely applied. Since agricultural production and food safety are the biggest concerns triggering research on endophytes, most endophytic bacterial community research was focused on cultivated crop plants, including rice, maize and potatoes. In Chelius *et al*.'s research of endophytic bacteria in maize roots, they found direct 16S rDNA PCR identified many more endophytes than culture methods, with *Alphaproteobacteria* as the dominant group (15). In that research, Chelius also proposed a pair of 16S rDNA primer, 799F and 1492R, which solved the problem of plant plastid ribosomal DNA unexpectedly amplified using regular 16S rDNA universal primers. In a similar project conducted by Sun *et al*. studying endophytic bacteria In rice roots, they

found that, different from the situation in maize roots, *Betaproteobacteria* and *Gammaproteobacteria* are the most dominant (82). Garbeva *et al*. studied the endophytic bacteria in potatoes using a combination method of fatty acid methyl ester (FAME) analysis, DGGE and sequencing of 16S rDNA, and found that *Alphaproteobacteria* and *Gammaproteobacteria* were dominant, with Firmicutes also detected (28). Sturz *et al*. also studied the endophytic bacteria that colonized in red clover, which is commonly used for grazing cattle and other animals (80). The most interesting discovery of their research was the recognition of some endophytes that always lead to the depression or promotion of clover growth.

3 Terminal Restriction Fragment Length Polymorphism (T-RFLP) and its application in bacterial communities' research

3.1 Terminal Restriction Fragment Length Polymophism

Many bacteria required special conditions, which are not discovered yet, for their growth; therefore most bacteria cannot be studied in the laboratory using cultivation-based methods (2), leading to the necessity of the introduction of cultivation-independent methods. Before T-RFLP, some cultivation-independent fingerprint techniques based on 16S rDNA were proposed to study the compositions and structure of the microbial communities, focusing on differences and dynamics, including Amplified Ribosomal DNA-Restriction Analysis (ARDRA) (89), Denaturing Gradient Gel Electrophoresis (DGGE) and Automated Ribosomal Intergenic Spacer Analysis (ARISA). Terminal Restriction Fragment Length Polymorphism (T-RFLP) was introduced by Larry Forney's group in 1997 as an extension technique of Restriction Fragment Length Polymorphism

(RFLP), as “a quantitative molecular technique” “for rapid analysis of microbial community diversity in various environments” (49). Doing T-RFLP to analyze microbial community, first a specific region of the bacterial 16S rDNA would be amplified from total communities DNA by PCR using a primer that was fluorescently labeled at its 5’ end. The PCR amplicons would be digested by selected restriction endonucleases, and the length of the terminal restriction fragments (T-RF) could be measured by an automatic DNA sequencer. One or more types of restriction enzymes could be used and the resulting T-RFLP patterns comprise of the length of all T-RFs from one or more restriction digestions. The T-RFLP patterns can be directly deposited into numerical analysis and further statistical analysis; and they can also be compared to existing bacterial 16S rDNA databases like RDP to find hits to confirm the bacterial source of specific T-RFs. Forney’s group did a computer simulation virtual T-RFLP and showed that T-RFLP has a high resolution ability to reveal the diversity of a model bacterial community, 233 unique T-RF lengths were obtained from 686 amplified sequences (49). The robustness and reproducibility of the T-RFLP technique has been evaluated and confirmed by Osborn *et al* (60).

3.2 Application of T-RFLP in microbial community research

T-RFLP is a powerful tool to assess the diversity of complex bacterial communities and especially to compare the bacterial communities from different environments. It has been applied mainly to, but not limited to, bacterial 16S rDNA. Fungal ribosomal genes(29, 41) and archaeal 16S rDNA (44, 48) have also been studied using T-RFLP. Some interesting and typical T-RFLP application in bacterial community research include: Hullar *et al*. who employed T-RFLP to study the dynamics of microbial communities in

stream habitats (38); Katsivela *et al*. who used T-RFLP to track the dynamics of microbial communities in bioremediation of petroleum waste (42); Noll *et al* who studied the bacterial community response to oxygen gradient using T-RFLP (59); Rasche *et al*. who applied T-RFLP to rhizosphere bacterial community research (68); and Thies *et al*. who analyzed the normal and disturbed vaginal microbial communities using T-RFLP (84). Statistical analysis methods have also been applied to support T-RFLP in the research of microbial communities. Clement *et al*. applied principle component analysis (PCA) to T-RFLP for comparisons of complex bacterial communities, trying to visualize the relationships among T-RFLP patterns (16). Blackwood *et al* applied clustering analysis for quantitative comparisons of bacterial communities to find bacterial groups of interest (9). Cao *et al*. introduced canonical correspondence analysis (CCA) to T-RFLP analysis of sediment microbial communities (11) and Blackwood and Paul used redundancy analysis (RDA) to assess the bacterial community structure in different agricultural systems (10), both trying to explore the environmental influences on the variation of microbial communities in different habitats. To date, T-RFLP has allowed great progress in microbial community research especially in analyzing the difference of communities in diverse habitats and the dynamics of microbiota.

4 Next generation sequencing techniques, 454 tag sequencing and their application in bacterial communities' research

4.1 DNA Sequencing and Next Generation Sequencing (NGS)

Since DNA was discovered as the carrier of the genetic information by Watson and Crick (91), the sequences of DNA have become the center of biological research. Sanger

introduced the “plus and minus” DNA sequencing method in 1975 (73) and improved it to “the dideoxy method” himself in 1977 (74). Since the Sanger sequencing method was developed, it has completely changed modern biological research and therefore dominated sequencing over three decades especially when automation of the Sanger sequencing principle was introduced in 1986 by Hood (78). However the limitation of Sanger sequencing methods including high cost and low efficiency showed that new techniques of sequencing were needed to sequence large DNAs. The needs for sequencing of complete genomes such as Human Genome Project (HGP) (1) and the involvement of metagenomics to understand environmental microorganisms greatly enhanced the development of new sequencing techniques, which are usually referred to as next-generation sequencing (76). These NGS methods differ from each other but commonly consist of template preparation, sequencing, imaging and post-sequencing alignment and assembly (55). The most popular NGS methods that are successfully realized into commercial products include 454 Sequencing (Roche Applied Science), Illumina (Solexa) Sequencing (Illumina) and SOLiD platform (Applied Bioscience). In Illumina Sequencing, to prepare the template, single-strand single-molecule primed DNA template was immobilized on a slide and then local clonal colonies were formed by bridge amplification. In each sequencing cycle, one nucleotide with one of four types of reversible terminator bases was added, and then the fluorescence labeling in the nucleotides was recorded by a camera. Finally the fluorescence dye with the terminator blocker was removed, leading to the next cycle (7). SOLiD is short for Sequencing by Oligonucleotide Ligation and Detection, and employs sequencing by ligation as core techniques (87). To prepare the template, DNA was amplified by emulsion PCR and the

beads were deposited on a glass slide instead of wells in 454 sequencing. A mixture of all possible oligonucleotides of a fixed length, n, were labeled according to the type of bases in the sequenced positions, then oligonucleotides were hybridized to the template and ligated by DNA ligase if the bases in oligonucleotides matched with the template sequence, leading to an informative signal. In the next round the original sequencing primer was shifted one nucleotide towards the 5' end. The complete sequencing could be finished in n rounds. The collections of all signals gave the consensus sequences of the DNA template after appropriate alignment.

4.2 454 pyrosequencing

454 pyrosequencing (53) was originally developed by 454 Life Sciences founded by Jonathan Rothberg, which was purchased by Roche Diagnostics in 2007. The first sequence of one individual was completed using 454 pyrosequencing to sequence the complete genome of James Watson (92). Roche 454 pyrosequencing uses a large-scale parallel pyrosequencing system. Although the size of 454 sequencing output is smaller than other NGS methods, compared to the long run time and short read length, 454 sequencing is much more time-efficient (10 hour per run) and generates long sequencing reads (400-500 bases) (55). 454 pyrosequencing has experienced its fast growth since Roche released the GS20 sequencing machine in 2005, which was the first NGS sequencer in market at that time. 454 pyrosequencing have more applications after Roche released the GS Junior System, a bench top version of Genome Sequencer FLX System in 2009.

The principle of 454 pyrosequencing is pyrophosphate-based sequencing (pyrosequencing) in picolitre-sized wells and the core technique is emulsion-PCR (emPCR) (53). Generally 454 pyrosequencing consists of three steps. The first step is the sample preparation. In this step, similar to other sequencing methods, the whole genome would be fragmented and then amplified with random PCR primers or a targeted DNA fragment would be amplified with specific primers, so enough material would be supplied for sequencing. The only difference of 454 pyrosequencing sample preparation relative to others is that we need to link two unique adapters to each end of DNA fragments, so the adapter-carrying DNA could be captured by the beads on which the complementary adapter sequences were immobilized. In the case of doing tagged pyrosequencing, the tags/barcodes would also be linked to the DNAs by PCR with adapters. The second step is loading DNA to beads and emulsion-PCR. In this step, the adapter carrying DNAs which were appropriately diluted would mix with capture beads, PCR reagents, emulsion oil in water to create a "water in oil" emulsion. Vigorous shaking of this combination would lead the water mixture to form droplets around the beads. Typically one droplet would contain only one DNA fragment, and in the droplet as a "micro-reactor" with PCR reagents, this DNA fragment would be amplified into millions of copies that are immobilized on the bead. The third step is pyrosequencing. Pyrosequencing is a sequencing by synthesis approach. After clean up, DNA-captured beads are deposited onto a 454 PicoTiterPlate with one bead into one picolitre-size well. Varied with different version of Genome Sequencer, each PicoTiterPlate comprises up to millions of wells, with each well leading to one sequencing read. The four types of 2'-deoxyribonucleoside triphosphate (dNTP) are added to the PicoTiterPlate sequentially in

a fixed order. A pyrophosphate molecule would be released after one dNTP is incorporated to the single DNA template immobilized on the bead complementarily; then in the presence of adenosine 5´ phosphosulfate the pyrophosphate molecules would be converted to ATP, which excites luciferin in a luciferase-mediated conversion to oxyluciferin and light. This reaction is quantitative, and the light signal would be recorded by a high-resolution charge-coupled device (CCD) and translated into sequence information.

4.3 NGS in microbial community research

NGS platforms produce large numbers of reads at low cost, leading to many applications including sequencing of regions of interest or whole genome of specific organisms or individuals, transcriptomes/cDNA library of cells, tissues and organisms, profiling of epigenetic marks, and species classification by metagenomics (55). NGS platforms are greatly useful in profiling of microbial communities, and both amplifying a fragment of 16S ribosomal DNA and sequencing the whole genome or genomic DNA from environments can be realized by NGS. Due to the relatively short sequencing reads, applications of pyrosequencing to identify microorganisms mainly focus on hypervariable regions within specific genes, especially 16S rDNA (62). 16S rDNA has nine hypervariable regions V1-V9, and therefore amplification of selected regions within 16S rDNA using conserved region primers can identify microbial species or genus with reference database such as Ribosomal Database Project (17), Greengenes (21) and ARB-SILVA (66) . 454 pyrosequencing was the first commercialized NGS platform and developed for metagenomics. 454 pyrosequencing generates longer sequencing reads than Illumina and SOLiD, and now 454 pyrosequencing is widely used in microbial

community research, including deep mine bacterial communities (25), soil bacterial communities (69), human body habitats (20), rhizosphere bacterial communities (86), Pathogenic bacteria in sewage treatment plants (93), endophytic bacteria in cultivated plants (52) and endophytic fungus (54). Some scientists have also proposed related algorithms to improve pyrosequencing analysis for research of microbial diversity and ecology, including PyroNoise (67) proposed to detect true sequences from sequencing noise, and UniFrac(51) proposed to solve the problem that pyrosequencing reads covered different variable regions of 16S rDNA.

Some comprehensive software packages have also been developed to analyze pyrosequencing data from microbial communities. Schloss's group introduced an open-source expandable software called Mothur to fill the bioinformatic needs of microbial community ecology (75). Mothur incorporated many developed analysis tools including defining Operational Taxonomic Units (OTU), UniFrac, Analysis of molecular variance (AMOVA) and Homogeneity of molecular variance (HOMOVA). Knight's group introduced Quantitative insights into microbial ecology (QIIME), an open-source software pipeline, to solve the problem of interpreting the raw sequencing data (12). Similar to Mothur, QIIME was designed for comparison and analysis of microbial communities, and incorporated the functions including OTU picking, taxonomic assignment, phylogenetic trees based on the representative sequence of OTUs and downstream statistical analysis. These two software packages have now been widely used in microbial community ecology study, as summarized by Kuczynski *et al* (45).

LITERATURE CITED

1. **Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, and a. et.** 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252:**1651-1656.
2. **Apajalahti, J., A. Kettunen, and H. Graham.** 2004. Characteristics of the gastrointestinal microbial communities, with special reference to the chicken. World's Poultry Science Journal **60:**223-232.
3. **Azevedo, J. L., W. Maccheroni Jr, J. O. Pereira, and W. L. de Araujo.** 2000. Endophytic microorganisms: a review on insect control and recent advances on tropical plants. Electronic Journal of Biotechnology **3:**15-16.
4. **Barac, T., S. Taghavi, B. Borremans, A. Provoost, L. Oeyen, J. V. Colpaert, J. Vangronsveld, and D. van der Lelie.** 2004. Engineered endophytic bacteria improve phytoremediation of water-soluble, volatile, organic pollutants. Nat Biotech **22:**583-588.
5. **Bell, C. R., G. A. Dickie, W. L. G. Harvey, and J. W. Y. F. Chan.** 1995. Endophytic bacteria in grapevine. Canadian Journal of Microbiology **41:**46-53.
6. **Bent, E, Chanway, and P. C.** 1998. The growth-promoting effects of a bacterial endophyte on lodgepole pine are partially inhibited by the presence of other rhizobacteria, vol. 44. National Research Council of Canada, Ottawa, ON, CANADA.

7. **Bentley, D. R.** 2006. Whole-genome re-sequencing. Current Opinion in Genetics & Development **16:**545-552.

8. **Berg, G., J. Hallmann, B. J. E. Schulz, C. J. C. Boyle, and T. N. Sieber.** 2006. Control of Plant Pathogenic Fungi with Bacterial Endophytes Microbial Root Endophytes, p. 53-69, vol. 9. Springer Berlin Heidelberg.

9. **Blackwood, C. B., T. Marsh, S.-H. Kim, and E. A. Paul.** 2003. Terminal Restriction Fragment Length Polymorphism Data Analysis for Quantitative Comparison of Microbial Communities. Applied and Environmental Microbiology **69:**926-932.

10. **Blackwood, C. B., and E. A. Paul.** 2003. Eubacterial community structure and population size within the soil light fraction, rhizosphere, and heavy fraction of several agricultural systems. Soil Biology and Biochemistry **35:**1245-1255.

11. **Cao, Y., G. Cherr, A. Córdova-Kreylos, T. Fan, P. Green, R. Higashi, M. LaMontagne, K. Scow, C. Vines, J. Yuan, and P. Holden.** 2006. Relationships between Sediment Microbial Communities and Pollutants in Two California Salt Marshes. Microbial Ecology **52:**619-633.

12. **Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight.** QIIME allows analysis of high-throughput community sequencing data. Nature Methods **7:**335-336.

13. **Catalán, A. I., F. Ferreira, P. R. Gill, and S. Batista.** 2007. Production of polyhydroxyalkanoates by Herbaspirillum seropedicae grown with different sole carbon sources and on lactose when engineered to express the lacZlacY genes. Enzyme and Microbial Technology **40:**1352-1357.

14. **Chanway, C. P.** 1997. Inoculation of Tree Roots with Plant Growth Promoting Soil Bacteria: An Emerging Technology for Reforestation. Forest Science **43:**99-112.

15. **Chelius, M. K., and E. W. Triplett.** 2001. The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. . Microbial Ecology **41:**252-263.

16. **Clement, B. G., L. E. Kehl, K. L. DeBord, and C. L. Kitts.** 1998. Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. Journal of Microbiological Methods **31:**135-142.

17. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research **37:**D141-D145.

18. **Conn, V. M., and C. M. M. Franco.** 2004. Analysis of the endophytic actinobacterial population in the roots of wheat (Triticum aestivum L.) by Terminal Restriction Fragment Length Polymorphism and sequencing of 16S rRNA Clones. Applied and Environmental Microbiology **70:**1787-1794.

19. **Cooley, M. B., W. G. Miller, and R. E. Mandrell.** 2003. Colonization of Arabidopsis thaliana with Salmonella enterica and Enterohemorrhagic Escherichia coli O157:H7 and Competition by Enterobacter asburiae. Applied and Environmental Microbiology **69:**4915-4926.

20. **Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight.** 2009. Bacterial community variation in human body habitats across space and time. Science **326:**1694-1697.

21. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Applied and Environmental Microbiology **72:**5069-5072.

22. **Dong, Z., M. J. Canny, M. E. McCully, M. R. Roboredo, C. F. Cabadilla, E. Ortega, and R. Rodes.** 1994. A nitrogen-fixing endophyte of sugarcane stems (A new role for the apoplast). Plant Physiology **105:**1139-1147.

23. **Duijff, B. J., V. Gianinazzi-Pearson, and P. Lemanceau.** 1997. Involvement of the outer membrane lipopolysaccharides in the endophytic colonization of tomato roots by biocontrol Pseudomonas fluorescens strain WCS417r. New Phytologist **135:**325-334.

24. **Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman.** 2005. Diversity of the human intestinal microbial flora. Science **308:**1635-1638.

25. **Edwards, R. A., B. Rodriguez-Brito, L. Wegley, M. Haynes, M. Breitbart, D. M. Peterson, M. O. Saar, S. Alexander, E. C. Alexander, and F. Rohwer.**

2006. Using pyrosequencing to shed light on deep mine microbial ecology. Bmc Genomics **7:**13.

26. **Elvira-Recuenco, M., and J. W. L. van Vuurde.** 2000. Natural incidence of endophytic bacteria in pea cultivars under field conditions. Canadian Journal of Microbiology **46:**1036-1041.

27. **Engelhard, M., T. Hurek, and B. Reinhold-Hurek.** 2000. Preferential occurrence of diazotrophic endophytes, Azoarcus spp., in wild rice species and land races of Oryza sativa in comparison with modern races. Environmental Microbiology **2:**131-141.

28. **Garbeva, P., L. van Overbeek, J. van Vuurde, and J. van Elsas.** 2001. Analysis of endophytic bacterial communities of potato by plating and denaturing gradient gel electrophoresis (DGGE) of 16S rDNA based PCR fragments. Microbial Ecology **41:**369-383.

29. **Genney, D. R., I. C. Anderson, and I. J. Alexander.** 2006. Fine-scale distribution of pine ectomycorrhizas and their extramatrical mycelium. New Phytologist **170:**381-390.

30. **Germaine, K. J., X. Liu, G. G. Cabellos, J. P. Hogan, D. Ryan, and D. N. Dowling.** 2006. Bacterial endophyte-enhanced phytoremediation of the organochlorine herbicide 2,4-dichlorophenoxyacetic acid. FEMS Microbiology Ecology **57:**302-310.

31. **Guo, X., M. W. van Iersel, J. Chen, R. E. Brackett, and L. R. Beuchat.** 2002. Evidence of association of salmonellae with tomato plants grown hydroponically

in inoculated nutrient solution. Applied and Environmental Microbiology **68:**3639-3643.

32. **Gyaneshwar, P., E. K. James, N. Mathan, P. M. Reddy, B. Reinhold-Hurek, and J. K. Ladha.** 2001. Endophytic Colonization of Rice by a Diazotrophic Strain of Serratia marcescens. Journal of Bacteriology **183:**2634-2645.

33. **Hallmann, J., A. QuadtHallmann, W. F. Mahaffee, and J. W. Kloepper.** 1997. Bacterial endophytes in agricultural crops. Canadian Journal of Microbiology **43:**895-914.

34. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Meth **5:**235-237.

35. **Head, I. M., J. R. Saunders, and R. W. Pickup.** 1998. Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. Microbial Ecology **35:**1-21.

36. **Heuer, H., M. Krsek, P. Baker, K. Smalla, and E. M. Wellington.** 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl. Environ. Microbiol. **63:**3233-3241.

37. **Hooper, L. V., M. H. Wong, A. Thelin, L. Hansson, P. G. Falk, and J. I. Gordon.** 2001. Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine. Science **291:**881-884.

38. **Hullar, M. A. J., L. A. Kaplan, and D. A. Stahl.** 2006. Recurring Seasonal Dynamics of Microbial Communities in Stream Habitats. Applied and Environmental Microbiology **72:**713-722.

39. **Ingham, S. C., M. A. Fanslau, R. A. Engel, J. R. Breuer, J. E. Breuer, T. H. Wright, J. K. Reith-Rozelle, and J. Zhu.** 2005. Evaluation of fertilization-to-planting and fertilization-to-harvest intervals for safe use of noncomposted bovine manure in Wisconsin vegetable production. Journal of Food Protection **68:**1134-1142.

40. **Islam, M., J. Morgan, M. P. Doyle, S. C. Phatak, P. Millner, and X. Jiang.** 2004. Fate of Salmonella enterica Serovar Typhimurium on carrots and radishes grown in fields treated with contaminated manure composts or irrigation water. Applied and Environmental Microbiology **70:**2497-2502.

41. **Johnson, D., P. J. Vandenkoornhuyse, J. R. Leake, L. Gilbert, R. E. Booth, J. P. Grime, J. P. W. Young, and D. J. Read.** 2004. Plant communities affect arbuscular mycorrhizal fungal diversity and community composition in grassland microcosms. New Phytologist **161:**503-515.

42. **Katsivela, E., E. Moore, D. Maroukli, C. Strömpl, D. Pieper, and N. Kalogerakis.** 2005. Bacterial community dynamics during <i>in-situ bioremediation of petroleum waste sludge in landfarming sites. Biodegradation **16:**169-180.

43. **Kerry, B. R.** 2000. Rhizosphere interactions and the exploitation of microbial agents for the biological control of plant-parasitic nematodes. Annual Review of Phytopathology **38:**423-441.

44. **Kotsyurbenko, O. R., K.-J. Chin, M. V. Glagolev, S. Stubner, M. V. Simankova, A. N. Nozhevnikova, and R. Conrad.** 2004. Acetoclastic and hydrogenotrophic methane production and methanogenic populations in an acidic West-Siberian peat bog. Environmental Microbiology **6:**1159-1173.

45. **Kuczynski, J., C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers, and R. Knight.** Experimental and analytical tools for studying the human microbiome. Nat Rev Genet **13:**47-58.

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

47. **Lee, S., M. Flores-Encarnacion, M. Contreras-Zentella, L. Garcia-Flores, J. E. Escamilla, and C. Kennedy.** 2004. Indole-3-Acetic Acid Biosynthesis Is Deficient in Gluconacetobacter diazotrophicus Strains with Mutations in Cytochrome c Biogenesis Genes. Journal of Bacteriology **186:**5384-5391.

48. **Leybo, A., A. Netrusov, and R. Conrad.** 2006. Effect of hydrogen concentration on the community structure of hydrogenotrophic methanogens studied by T-RELP analysis of 16S rRNA gene amplicons. Microbiology **75:**683-688.

49. **Liu, W.-T., T. L. Marsh, H. Cheng, and L. J. Forney.** 1997. Characterization of Microbial Diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16s rRNA. Applied and Environmental Microbiology **63:**4516-4522.

50. **Liu, Z. Z., C. Lozupone, M. Hamady, F. D. Bushman, and R. Knight.** 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Research **35:**10.

51. **Lodewyckx, C., J. Vangronsveld, F. Porteous, E. R. B. Moore, S. Taghavi, M. Mezgeay, and D. v. der Lelie.** 2002. Endophytic Bacteria and Their Potential Applications. Critical Reviews in Plant Sciences **21:**583-606.

52. **Manter, D. K., J. A. Delgado, D. G. Holm, and R. A. Stong.** Pyrosequencing Reveals a Highly Diverse and Cultivar-Specific Bacterial Endophyte Community in Potato Roots. Microbial Ecology **60:**157-166.

53. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.

54. **Maria, V. A., and A. Vannini.** Abundance and diversity of fungal endophytic community in an Italian beech forest: Pyrosequencing vs isolation method. Phytopathology **101:**S7-S7.

55. **Metzker, M. L.** 2009. Sequencing technologies [mdash] the next generation. Nature Review Genetics **11:**31-46.

56. **Moore, F. P., T. Barac, B. Borremans, L. Oeyen, J. Vangronsveld, D. van der Lelie, C. D. Campbell, and E. R. B. Moore.** 2006. Endophytic bacterial diversity in poplar trees growing on a BTEX-contaminated site: The characterisation of isolates with potential to enhance phytoremediation. Systematic and Applied Microbiology **29:**539-556.

57. **Muyzer, G., E. C. De Wall, and A. G. Uitterlinden.** 1993. Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain reaction-amplified Genes coding for 16s rRNA. Applied and Environmental Microbiology **59:**695-700.

58. **Newman, L. A., and C. M. Reynolds.** 2005. Bacteria and phytoremediation: new uses for endophytic bacteria in plants. Trends in Biotechnology **23:**6-8.

59. **Noll, M., D. Matthies, P. Frenzel, M. Derakshani, and W. Liesack.** 2005. Succession of bacterial community structure and diversity in a paddy soil oxygen gradient. Environmental Microbiology **7:**382-395.

60. **Osborn, A. M., E. R. B. Moore, and K. N. Timmis.** 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environmental Microbiology **2:**39-50.

61. **Perott, R.** 1926. On the limits of biological inquiry on soil science. Proceedings of International Society for Soil Science **2:**16.

62. **Petrosino, J. F., S. Highlander, R. A. Luna, R. A. Gibbs, and J. Versalovic.** 2009. Metagenomic Pyrosequencing and Microbial Identification. Clinical Chemistry **55:**856-866.

63. **Phillips, D. A., E. Martínez-Romero, G. P. Yang, and C. M. Joseph.** 2000. Release of nitrogen: A key trait in selecting bacterial endophytes for agronomically useful nitrogen fixation., p. 205-217. *In* J. K. Ladha and P. M. Reddy (ed.), The Quest for Nitrogen Fixation in Rice. International Rice Research Institute,, Manila.

64. **Pirttilä, A. M., P. Joensuu, H. Pospiech, J. Jalonen, and A. Hohtola.** 2004. Bud endophytes of Scots pine produce adenine derivatives and other compounds that affect morphology and mitigate browning of callus cultures. Physiologia Plantarum **121:**305-312.

65. **Pirttilä, A. M., H. Pospiech, H. Laukkanen, R. Myllylä, and A. Hohtola.** 2005. Seasonal variations in location and population structure of endophytes in buds of Scots pine. Tree Physiology **25:**289-297.

66. **Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Gl 枚 ckner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research **35:**7188-7196.

67. **Quince, C., A. Lanzen, T. P. Curtis, R. J. Davenport, N. Hall, I. M. Head, L. F. Read, and W. T. Sloan.** 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods **6:**639-U27.

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

69. **Roesch, L. F., R. R. Fulthorpe, A. Riva, G. Casella, A. K. M. Hadwin, A. D. Kent, S. H. Daroub, F. A. O. Camargo, W. G. Farmerie, and E. W. Triplett.** 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. Isme Journal **1:**283-290.

70. **Rosenblueth, M., and E. Martinez-Romero.** 2006.. Molecular Plant-Microbe Interactions **19:**827-837.

71. **Ryan, R. P., K. Germaine, A. Franks, D. J. Ryan, and D. N. Dowling.** 2008. Bacterial endophytes: recent developments and applications. FEMS Microbiology Letters **278:**1-9.

72. **Ryan, R. P., D. J. Ryan, Y.-C. Sun, F.-M. Li, Y. Wang, and D. N. Dowling.** 2007. An acquired efflux system is responsible for copper resistance in Xanthomonas strain IG-8 isolated from China. FEMS Microbiology Letters **268:**40-46.

73. **Sanger, F., and A. R. Coulson.** 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology **94:**441-448.

74. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America **74:**5.

75. **Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber.** 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology **75:**7537-7541.

76. **Shendure, J., and H. Ji.** 2008. Next-generation DNA sequencing. Nat Biotech **26:**1135-1145.

77. **Siciliano, S. D., N. Fortin, A. Mihoc, G. Wisse, S. Labelle, D. Beaumier, D. Ouellette, R. Roy, L. G. Whyte, M. K. Banks, P. Schwab, K. Lee, and C. W. Greer.** 2001. Selection of Specific Endophytic Bacterial Genotypes by Plants in Response to Soil Contamination. Applied and Environmental Microbiology **67:**2469-2475.

78. **Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood.** 1986. Fluorescence detection in automated DNA sequence analysis. Nature **321:**674-679.

79. **Strobel, G., B. Daisy, U. Castillo, and J. Harper.** 2004. Natural Products from Endophytic Microorganisms? Journal of Natural Products **67:**257-268.

80. **Sturz, A. V., B. R. Christie, B. G. Matheson, and J. Nowak.** 1997. Biodiversity of endophytic bacteria which colonize red clover nodules, roots, stems and foliage and their influence on host growth. Biol fertil soils **25:**13-19.

81. **Sturz, A. V., B. R. Christie, and J. Nowak.** 2000. Bacterial endophytes: Potential role in developing sustainable systems of crop production. Critical Reviews in Plant Sciences **19:**1-30.

82. **Sun, L., F. Qiu, X. Zhang, X. Dai, X. Dong, and W. Song.** 2007. Endophytic bacterial diversity in rice (Oryza sativa L.) roots estimated by 16S rDNA sequence analysis. Microbial Ecology **55:**415-424.

83. **Suzuki, M., M. S. Rappe, and S. J. Giovannoni.** 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Applied Environmental Microbiology **64:**4522-4529.

84. **Thies, F. L., W. König, and B. König.** 2007. Rapid characterization of the normal and disturbed vaginal microbiota by application of 16S rRNA gene terminal RFLP fingerprinting. Journal of Medical Microbiology **56:**755-761.

85. **Thrall, P. H., M. E. Hochberg, J. J. Burdon, and J. D. Bever.** 2007. Coevolution of symbiotic mutualists and parasites in a community context. Trends in Ecology & Evolution **22:**120-126.

86. **Uroz, S., M. Buee, C. Murat, P. Frey-Klett, and F. Martin.** Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environmental Microbiology Reports **2:**281-288.

87. **Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J. A. Malek, G. Costa, K. McKernan, A. Sidow, A. Fire, and S. M. Johnson.** 2008. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Research **18:**1051-1063.

88. **Van Aken, B., C. M. Peres, S. L. Doty, J. M. Yoon, and J. L. Schnoor.** 2004. Methylobacterium populi sp. nov., a novel aerobic, pink-pigmented, facultatively methylotrophic, methane-utilizing bacterium isolated from poplar trees (Populus deltoides×nigra DN34). International Journal of Systematic and Evolutionary Microbiology **54:**1191-1196.

89. **Vaneechoutte, M., R. Rossau, P. De Vos, M. Gillis, D. Janssens, N. Paepe, A. De Rouck, T. Fiers, G. Claeys, and K. Kersters.** 1992. Rapid identification of bacteria of the Comamonadaceae with amplified ribosomal DNA-restriction analysis (ARDRA). FEMS Microbiology Letters **93:**227-233.

90. **Verma, S. C., J. K. Ladha, and A. K. Tripathi.** 2001. Evaluation of plant growth promoting and colonization ability of endophytic diazotrophs from deep water rice. Journal of Biotechnology **91:**127-141.

91. **Watson, J. D., and F. H. C. Crick.** 1953. A Structure for Deoxyribose Nucleic Acid. Nature **171:**2.

92. **Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.-J. Chen, V. Makhijani, G. T. Roth, X. Gomes, K. Tartaro, F. Niazi, C. L. Turcotte, G. P. Irzyk, J. R. Lupski, C. Chinault, X.-z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D. M. Muzny, M. Margulies, G. M. Weinstock, R.**

A. Gibbs, and J. M. Rothberg. 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature **452:**872-876.

93. **Ye, L., and T. Zhang.** Pathogenic Bacteria in Sewage Treatment Plants as Revealed by 454 Pyrosequencing. Environmental Science & Technology **45:**7173-7179.

CHAPTER II

COMMUNITY TERMINAL RESTRICTION FRAGMENT LENGTH POLYMORPHISMS REVEAL INSIGHTS INTO THE DIVERSITY AND DYNAMICS OF LEAF ENDOPHYTIC BACTERIA

Abstract

Plant endophytic bacteria play an important role benefiting plant growth or being pathogenic to plants or organisms that consume those plants. Multiple species of bacteria have been found co-inhabiting plants, both cultivated and wild, with viruses and fungi. For these reasons, a general understanding of plant endophytic microbial communities and their diversity is necessary. A key issue is how the distributions of these bacteria vary with location, with plant species, with individual plants and with plant growing season. Five common plant species were collected monthly for four months in the summer of 2010, with replicates from four different sampling sites in the Tallgrass Prairie Preserve in Osage County, Oklahoma, USA. Metagenomic DNA was extracted from ground, washed plant leaf samples, and fragments of the bacterial 16S rDNA genes were amplified for analysis of terminal restriction fragment length polymorphism (T-RFLP).

We performed mono-digestion T-RFLP with restriction endonuclease *Dde*I, to reveal the structures of leaf endophytic bacterial communities, to identify the differences between plant-associated bacterial communities in different plant species or environments, and to explore factors affecting the bacterial distribution. We tested the impacts of three major factors on the leaf endophytic bacterial communities, including host plant species, sampling dates and sampling locations. **Results indicated that all of the three factors had significant impacts (α=0.05) on the distribution of leaf endophytic bacteria, with host species being the most important, followed by sampling dates and sampling locations.**

Introduction

Bacteria are associated with plants in many ways. Rhizosphere bacteria are associated with the roots of plants while those associated with leaves are designated phyllosphere bacteria. Within each of these spheres of plant influences, it is common to distinguish between those bacteria that are associated loosely with the outside of the roots or leaves, the epiphytes, from those that have colonized the internal parts of the organs, the endophytes. Rhizosphere bacteria have been extensively studied as have root endophytic bacteria (7, 31, 34). Numerous publications address the leaf epiphytic bacteria (4, 14, 22). Only a few studies have addressed leaf endophytic bacteria as part of phyllosphere bacteria (15). The diversity of leaf endophytic bacteria in different plants is largely unexplored, and is the main subject of this study. We want to understand what factors shape the communities of leaf endophytic bacteria.

A universally agreed definition of plant endophytic bacteria has not been established. In this study, we accept Hallmann's definition of endophytic bacteria (13): endophytic bacteria are those bacteria that "can be isolated from surface-disinfested plant tissue or extracted from within the plant and do not visibly harm the plant". Endophytic bacteria have been found in most plants, colonize the internal tissues and construct diverse relationships with their host plants. Endophytic bacteria can be beneficial to the host plant, including by growth promotion (28), biological control against plant pathogens (13), and bioremediation of the contaminated environment (28). They can also potentially become pathogens (7). Since endophytic bacteria are important to the entire ecosphere, it is imperative to gather a general understanding of endophytic microbial communities, their diversity, and their distribution among plant species, plant individuals and plant organs.

Traditionally, most studies of endophytic bacterial communities (5, 29, 30) are based on bacterial culture methods. However, most environmental bacteria are not culturable, as evidenced, for example, by the finding that culture-independent methods revealed a broader diversity of bacteria than did culture-dependent methods in a study of bacteria in the apple phyllosphere (40). In recent years, the study of endophytic bacteria often has employed culture-independent methods, most of which are based on the PCR amplification of bacterial 16S rDNA. Some notable studies of root endophytic bacteria (6, 31, 32) focused on single crop species, including maize and rice, because of their importance to food supply and safety. Several researchers have applied Terminal Restriction Fragment Length Polymorphism (T-RFLP) (21), a rapid fingerprint technique based on 16S rDNA PCR, to the evaluation of endophytic bacteria. T-RFLP can compare

multiple microbial communities fast and accurately, especially when high-throughput bacterial community characterization is needed.

In this project, we studied leaf endophytic bacteria in diverse environments from the Tallgrass Prairie Preserve (TGPP), Osage County, Oklahoma, USA (2), managed by The Nature Conservancy, which was the site of previous efforts by a Plant Virus Biodiversity and Ecology team to examine the diversity of viruses associated with plants growing in this setting (38). That study showed nucleotide sequence evidence of bacterial association with plants (23, 24, 27). We extracted total DNAs from plant samples obtained in the TGPP and amplified bacterial 16S rDNA sequences using bacterial rDNA specific primers. Rather than using multi-digestion T-RFLP with three or more restriction endonucleases, we performed mono-digestion T-RFLP with restriction endonuclease *Dde*I, to reveal the structures of leaf endophytic bacterial communities, to identify the differences between plant-associated bacterial communities in different plant species or environments, and to explore the factors affecting the bacterial distribution.

Materials and Methods

Plant Sampling.

Fresh healthy leaves were collected monthly from May to August, 2010, in the TGPP. Four sites were randomly chosen (Table 4). At each site, samples of 5 species of plants (*Asclepias viridis*, *Ambrosia psilostachya*, *Sorghastrum nutans*, *Panicum virgatum*, and *Ruellia humilis*) that are among the most frequent in the TGPP were collected. At each site, three individuals of *A. viridis* were identified and labeled with tags on May 14th

2010, and resampled on June 16th and July 14th (in August *A.viridis* samples were not found in the TGPP due to senescence).

Extraction of Total DNA from Plants.

Fresh leaf samples were washed with running tap water for at least 5 minutes to remove soil, dust and epiphytic organisms, followed by shaking in 75% ethanol twice each for 3 minutes, and then rinsed with running distilled water for 3 minutes. To validate the effect of the protocol, treated leaves were rinsed with 10 ml double distilled water for 3 minutes. The rinse water was collected and incubated on Lysogeny Broth (LB) plates at 37 °C overnight. No colonies were observed. Treated leaf samples were ground into a fine powder with liquid nitrogen. Then, 0.1 g of the homogenate was resuspended in a 1.5 ml microcentrifuge tube containing 1 ml CTAB extraction buffer [2%(w/v) cetyltrimethylammonium bromide, CTAB; 100 mM Tris-HCl (pH 8.0), 1.4 M NaCl, 20 mM EDTA, 1.5% polyvinyl-pyrolidone, PVP; 0.5% 2-mercaptoethanol] preheated to 65°C. Contents were mixed by inverting the tube several times, followed by incubating the tubes in a 60°C water bath for 60 min. The tube was centrifuged at 12,000 rpm for 5 min at 4°C and the supernatant was transferred to a new tube. DNA was then extracted twice with chloroform-isoamylalcohol (24:1 v/v) until the aqueous phase was clear. DNA was precipitated using 2 to 2.5 volumes of absolute ethanol, and 0.1 volume 3 M sodium acetate for 2 hours at -20°C, followed by centrifuging at 12,000 g for 10 min at 4°C, washed with 1 ml DNA wash solution (0.1 M trisodium citrate in 10% ethanol) twice (30 min incubation and 5 min centrifugation) and 1.5 ml 75% ethanol once (15 min incubation and 5 min centrifugation), then air dried. Finally DNA was resuspended in 50 µl DNase-free water.

PCR Amplification and T-RFLP.

Because the bacterial 16S rDNA sequences are highly similar to plant mitochondrial and chloroplast rDNA sequences, popular universal bacterial 16S rDNA primers are not appropriate for specific amplification of bacterial rDNA from plant DNA extracts (26). Primers 799F and 1492R (6) designed to exclude amplification of plastid 16S rDNA, were used in PCR. Each 50 µl PCR contained PCR buffer (Promega, MadisonWI), 2.5 mM $MgCl_2$, 200 µM each dNTP, 0.5 mg/ml BSA, 15 pmol of each primer, and 2.5 U Taq polymerase. Thermal cycling conditions were: an initial denaturation at 95°C for 3 min followed by 30 cycles of 94°C for 20 sec, 53°C for 40 sec, 72°C for 40 sec, and a final extension at 72°C for 7 min. The PCR yielded a 1.1 kbp mitochondrial product and a 0.74 kbp bacterial product. These were electrophoretically separated in an agarose gel. Bacterial rDNA fragments from multiple PCRs were pooled for restriction with *Dde*I (Promega)**.** Restriction digestion reactions were incubated at 37°C for 4 h, followed by 20 min at 65°C to denature the enzyme. Two microliters of the restricted PCR product were mixed with 0.75 µl of size standard LIZ1200 (ABI, Foster City, CA) and 7.25 µl of Hi-Di formamide (ABI). DNA fragments were scanned on an ABI 3730 automated DNA sequencer at Oklahoma State University's Recombinant DNA/Protein Core Facility. The T-RFLP data profiles were obtained and analyzed by using GeneMapper Software version 4.0 (ABI).

Data Processing and Statistical Analysis.

In 16S-rDNA-T-RFLP profiles, a baseline threshold of 50 relative fluorescence units was used to distinguish 'true peaks' from background noise. Considering T-RF drift (improperly sized T-RFs due to differences in fragment migration and purine content),

peaks were manually aligned using the method described by Culman *et al*. (9). After background removal, raw peak height was normalized to balance the uncontrolled differences in the amount of DNA between samples by dividing the peak height by the sum of all peak heights of each sample. Culman *et al*. (9) determined that relative peak heights are better than peak areas for comparisons in T-RFLP profile analysis, yielding greater signal to noise ratios.

All the T-RFLP data were arranged into a matrix with each row as a community sample and each column as the relative abundance of each T-RF. The matrix was analyzed by partial Canonical Correspondence Analyses (pCCA) using *Canoco for Windows* 4.5 (Plant Research International) (33). We performed three pCCAs: one using Sites as explanatory variables and Months and Species as covariables, a second using Months as explanatory variables and Species and Sites as covariables, and a third using Species as explanatory variables and Months and Sites as covariables. This allowed us to isolate the independent effects of each factor. For each analysis, we performed a permutation test of significance with 9,999 permutations, conditioned on the covariables.

Based on the complete T-RFLP data matrix, we calculated also the percentage of empty cells in the data matrix (8) as 100% x (total number of cells in the data matrix of T-RFs vs. samples - count of all cells with non-zero values)/ (total number of cells in data matrix). Multivariate Analysis of Variance (MANOVA) was conducted using SAS v9.2 (SAS Institute Inc.) and Hierarchical Clustering Analysis was carried out with R (R development core team, 2003).

Results

In this study, we used T-RFLP profiles to study the features of the distribution of leaf endophytic bacterial communities. Rather than using multiple restriction digestions and then comparing the combined T-RFLP profiles to entries in a pre-computed database, here we chose to use only one restriction endonuclease and the T-RFs with a certain length were treated as a special kind of OTU (Operational Taxonomic Unit) - Operational T-RFLP Unit, a unit that can be directly used to describe a community. In this manner we avoided the problems caused by T-RFs not referring to a known bacterial species in the database. This approach allows direct study of the complexity of and changes in distribution of leaf endophytic bacteria without requiring taxonomic identification. Engebretson *et al*. (12) suggested that four restriction endonucleases including *Bst*UI, *Dde*I, *Sau*96I, and *Msp*I had the highest frequency of resolving single populations from bacterial communities. To select the endonuclease with the highest power to resolve leaf endophytic bacterial communities, we cloned 16s rDNA PCR products and randomly selected and sequenced inserts from 50 colonies. Computer-simulated virtual digestions indicated that *Dde*I generated the most distinct T-RFs and thus had the highest resolution, so we chose *Dde*I to perform the mono-digestion T-RFLP to generate T-RFLP profiles from five species of plants (see **Materials and Methods**).

Osborn *et al*. (25) have demonstrated that T-RFLP is highly reproducible and robust in studying microbial communities and yields high-quality fingerprints consisting of fragments of precise sizes. In this research we also confirmed the reproducibility of T-RFLP to validate the application of T-RFLP to study endophytic bacterial communities.

We repeated the complete procedure from DNA extraction to final T-RFLP scanning, and the results indicated that the T-RFLP profiles from the same sample were indistinguishable (Figure 9).

General analysis of T-RFLP profiles of endophytic bacterial communities in *A. viridis*. In total, we obtained 36 *A. viridis* samples from four sites, sampled monthly from May to July with three samples for each site. T-RFLP profiles were generated for all and analyzed to identify T-RFs. The total number of T-RFs increased from May to July, suggesting that as the plant grows from May to July, endophytic bacteria become more diverse (Table 1). The richness of T-RFs (defined as the average number of T-RFs in a dataset) of samples from May, much lower than of those from June and July, indicated that from May to June, the complexity of the endophytic bacterial community increased three-fold. The percentage of empty cells is a measure of sharing of community components (8). Samples from May had the highest percentage, while samples from June had the lowest percentage, suggesting that in June different host plants share more common leaf endophytic bacterial species than they do in May, consistent with the leaf endophytic bacterial communities in June being more complex. Unlike the samples from different months, the samples from different sites did not show significant variation when the data were analyzed for presence/absences of individual T-RFs (Table 1) even though samples from site 4 appeared to have a lower diversity of leaf endophytic bacteria than others.

Between-site variation observed from T-RFLP patterns. Although the general level of diversity of leaf endophytic bacteria did not show variation among sites when presence/absence data were considered, the T-RFLP profiles of samples from different

sites suggested that the compositions and the relative abundances of individual T-RFs varied with the site/location of host plants, revealing a possible connection of leaf endophytic bacterial species with host locations. Figure 1 shows the T-RFLP patterns of two *A. viridis* plants both collected on July 14, 2010, but from different sites. In the sample from site 2, the T-RF 75bp was more prominent than the T-RF 85bp; while in the sample from site 3, the T-RF 85bp was more prominent. Other dominant T-RFs, including the T-RF 364bp and the T-RF 529bp, also show differences in relative abundance. The influence of host locations may contribute to these differences of endophytic bacterial communities. Alternatively, the differences could reflect sample to sample variation.

Seasonal variation observed from T-RFLP patterns. Temporal variations of leaf endophytic bacteria can also be observed in T-RFLP patterns, which reveal the development of different T-RFs during the growing season. We labeled three *A. viridis* plants at each site in order to track the dynamics of the leaf endophytic bacterial community of the same host plants. Figure 2 shows the comparison of T-RFLP patterns of one *A. viridis* individual from May to July. On May 14, the dominant T-RF in this bacterial community was the T-RF 85bp. On June 16, an increase of the relative abundance of the T-RF 529bp led this T-RF to share dominance of this bacterial community with the T-RF 85bp. On July 14th, the dominance of the T-RF 85bp had been replaced by the T-RF 75bp, which had a significant increase in relative abundance from May to July. The observations indicate that the leaf endophytic bacterial community changed as the host plant grew.

T-RFLP patterns reveal differences in leaf endophytic bacterial communities from different host plant species. Besides host plant location and sampling date, host plant species may also influence leaf endophytic bacterial communities because of their different physiological and biochemical features. Indeed, the T-RFLP patterns of *A. viridis*, *A. psilostachya*, and *P. virgatum* individuals were distinct (Figure 3). The dominant T-RFs (the group of the T-RFs which have an average proportion more than 3% of the total) for these three species (Table 5) reveal that each host species had its own characteristic group of dominant T-RFs. Especially the most dominant T-RFs differed among these three species. These observations indicate that the host species has properties determining the compositions of their leaf endophytic bacterial populations.

Partial Canonical Correspondence Analysis measures the influence of multiple factors on leaf endophytic bacterial communities. As described above, endophytic bacterial communities varied with the time of sampling, the locations of host plants and the species of host plants. To determine the relative importance of each factor, the relative abundances of each T-RF were used to conduct pCCA of T-RFLP profiles.

Figure 4 shows the pCCA of T-RFLP profiles of *A. viridis* treating sampling dates as the environmental factor with sampling locations and host species as covariables. Because the first pCCA axis is more important than the second axis, the differences between samples from May and the other two months are more significant than the differences between samples from June and July, a result which is consistent with the summary statistics of T-RFs (Table 1). This result implies rapid early changes in the development of endophytic bacterial communities, consistent with rapid plant growth of the host

species, *A. viridis*. Permutation tests revealed sampling date is a significant factor (p-value = 0.0001).

The pCCA result of T-RFLP profiles of *A. viridis* treating location of host plants as environmental factor with sampling dates and host species as covariables (Figure 5) indicates that the differences between samples from site 1 and other sites were stronger than the differences between sites 2 and 3. Permutation tests revealed location of host plants was a significant factor (p-value = 0.0005).

The pCCA result of treating host species as the environmental factor with sampling dates and locations as covariables in analyzing T-RFLP profiles using data from five host plant species shows that T-RF patterns are influenced by the host species identity (Figure 6). In the pCCA biplots, *S. nutans* and *P. virgatum* were close to each other, indicating that the leaf endophytic bacterial communities from these two species were similar to each other. Those of the other three host species were distinct from each other with *A. viridis* the most distinct, since the data point of *A. viridis* lay on the other end of the first axis. These results are consistent with the features of these host plant species: both *S. nutans* and *P. virgatum* are grass species; *A. viridis* is different from the other four species because it contains latex, giving it the common name "milkweed". Permutation tests revealed host species as a significant factor (p-value=0.0001).

We also studied the impacts of the sampling dates and host plant locations based on the 5-species dataset using pCCA. Results indicate that all of these three factors were significant with p-values < 0.01. The 5-species pCCA biplots confirm the inference we obtained from the *A. viridis* pCCA biplots, that samples from May were more distinct

from other samples considering sampling date as an environmental factor, and samples from Site 1 were more distinct from other samples considering sampling site as an environmental factor. After an analysis using all three factors as environmental factors, we were able also to partition the overall variation to reveal how much variation was contributed by each factor. Results calculated from pCCA eigenvalues indicated that host plant species contributed 49.8% of the overall variation, sampling date contributed 28.5%, and host plant locations contributed 14.2%. Thus although these three factors all significantly determined the structure of endophytic bacteria, host plant species was the most important factor, followed by sampling date and host locations.

The diversity of leaf endophytic bacteria was examined also by counting the number of T-RFs in each community. The average number of T-RFs (Table 2) over all samples of *R. humilis* was significantly smaller than those of *A. psilostachya*, *P. virgatum* and *A. viridis* by Tukey range test ($p = 0.0014$). This result indicates that *R. humilis* plants have a simpler endophytic bacterial community than the other species. This result further supports that the host plant species plays an important role in determining the diversity of endophytic bacteria. The average number of T-RFs (Table 2) appeared to have risen from May to July and then fallen from July to August. However, the Tukey test did not detect any significant differences among these four different months. The Tukey test also did not detect any significant differences among the average number of T-RFs in the four sites (Table 2). However we cannot rule out significant differences had a larger spatial scale been chosen. The tests agree with the pCCA results described above: the host plant species is the most important factor. Considering that average numbers of T-RFs are unweighted alpha diversity indices, the weighted alpha diversity indices (Shannon

indices) were also calculated based on the relative proportions of each T-RFs (Table 6). These indices also supported the conclusion that the host species was the most important factor.

The diversity of leaf endophytic bacteria can also be evaluated by hierarchical clustering of the frequencies of T-RFs in these five species (Figure 7). The frequency of a T-RF is defined as the fraction of samples of a host species that have the T-RF in question. A high frequency of a T-RF in one host species indicates that the bacterial species represented is a common component in that host species, and a low frequency means that the existence of the bacterial group represented is occasional. Complete linkage clustering of different host species based on the frequencies of T-RFs showed that *P. virgatum* and *S. nutans* were the closest to each other, and *A. viridis* and *R. humilis* were distinct from the other three species (Figure 7). These results are consistent with those obtained from the pCCA when treating host species as environmental factors. Complete linkage clustering of the T-RFs indicated different groups of the T-RFs, of which the major cluster containing the most frequent T-RFs is shown in Figure 8. This cluster contains some T-RFs that are highly frequent among multiple host species. For instance, the T-RF 355bp was highly frequent in *P. virgatum*, *S. nutans* and *A. psilostachya*, but rarely detected in *A. viridis* and *R. humilis*, indicating that T-RF 355bp represents bacterial groups which are sensitive to the different physical/ biochemical features of these two groups of host plant species. Some T-RFs have a high frequency in some host species but maintain a low frequency in other host species; this is interpreted to mean that the bacterial groups represented by these T-RFs are more likely to grow in the leaf endophytic bacterial communities of their preferred host species. (For complete

data of the frequencies of all T-RFs, see Table 7). An extreme example is the T-RF 493bp: this T-RF had a frequency of 61.5% in *A. psilostachya*, but was not detected in other host species. Some unique biochemical or physiological features of *A. psilostachya* may lead to a preferable inner-environment for the bacterial groups represented by the T-RF 493bp to grow, so that those bacteria are characteristic of the leaf endophytic bacterial communities in *A. psilostachya*.

We also calculated the average frequencies of the T-RFs over all the five host species based on the frequencies of the T-RFs in each species. The average frequency reflects the general distribution of endophytic bacteria among multiple species of host plants. In Table 7, the average frequencies of all recognized T-RFs were also compared: for example, the T-RF 529bp had an average frequency more than 80% in these five selected host species and was the most frequent T-RF.

Multivariate Analysis of Variance (MANOVA) of the T-RFLP profile also indicated that the three major factors are significant, consistent with the pCCA result. The T-RFLP profiles of all samples that include only those T-RFs present in highest proportions shown in Figure 8 were also used to test the three major factors by MANOVA. Generally, for the data including all samples, Wilk's Lambda Analysis and Hotelling-Lawley Trace Analysis both indicated that the three major factors (host species, dates and sampling sites) were significant factors at alpha = 0.05. For these nine T-RFs, at alpha = 0.05, the host species factor was significant for seven T-RFs; the sampling dates factor was significant for seven T-RFs; the sampling sites factor was significant for six T-RFs. In aggregate, these three major factors were all significant at alpha = 0.05 for four T-RFs: 75bp, 79bp, 236bp and 355bp. The three factor models for these four T-RFs gave R-

square coefficients greater than 0.9. Thus, the results of MANOVA were consistent with pCCA, again confirming the importance of the three major factors.

The average proportion per existence (APE) of all T-RFs found in five host species estimated the prevalence of T-RFs in diverse communities (Table 7**).** APE is defined as the average proportion of one T-RF over those host samples which contain this T-RF in their T-RFLP profiles, and was calculated by the sum of the relative proportions divided by the number of the samples containing this T-RF, as in the following formula:

$$\mathrm{APE} = \frac{\sum_{i=1}^{m} P_i}{n}$$

where *Pi* is the relative proportion of the T-RF in *i*th sample, *m* is the total number of samples, and *n* is the number of these which have the T-RF. APE reflects the diversity of leaf endophytic bacteria since T-RFs have different prevalences in different host species. For instance, T-RF 78bp had a proportion of 54.4% in *R. humilis* but only 7.2% in *S. nutans*; while T-RF 236bp made up 17.2% of the T-RFs in *S. nutans*, which was the highest among the five host species, but was not detected in *R. humilis* (Table 7). Since each T-RF represents a different group of bacteria, APE actually reflects that certain groups of bacteria are present in different proportions in different host species, consistent with the host species determining the compositions of the endophytic bacterial communities.

Some prominent T-RFs had relatively higher proportions than other T-RFs (Table 7), and these T-RFs represent the dominant bacterial groups in the endophytic bacterial communities. We compared the most abundant T-RFs, those which have average frequencies more than 0.3 over all five host species (Table 4). Some T-RFs were significantly different in APE among host species, making those T-RFs the characteristic T-RFs of the endophytic bacterial communities, for instance, the T-RF 75bp in *A. viridis* and the T-RF 78bp in *R. humilis*. That one bacterial group has a significantly higher or lower proportion in some host species indicates that this bacterial group is preferred or inhibited, respectively, by some physiological or biochemical properties of the host species.

Discussion

Hallman *et al*. defined endophytic bacteria as those bacteria that "can be isolated from surface-disinfested plant tissue or extracted from within the plant and do not visibly harm the plant" (13). Disinfestation by killing all the epiphytic bacteria may be effective when culture-dependent protocols are used, but is not appropriate in culture-independent protocols, such as the present one, since the DNA or RNA of dead epiphytes, if not removed, would still be amplified by bacteria-specific PCR. For those organs, like tubers, whose outer layers can be easily peeled off, endophytic bacteria can be isolated from inside of the plants unambiguously. However, peeling the epidermis off leaves, while possible, is not practical for a study like the present one. Therefore, to study leaf endophytic bacterial communities, it is critical to dislodge epiphytic bacteria from the

leaf surfaces as far as possible. We have dislodged epiphytes using methods similar to those reported by others (1, 18, 37, 40). Although we cannot be sure that we have removed all epiphytic bacteria, the observation that the complexities of the populations (Table 6) were substantially lower than those reported for leaf epiphytic bacteria (10, 16) suggests that most epiphytes have been removed.

Past studies have applied multiple enzyme digestion T-RFLP to environmental bacterial community research (3, 11, 35). Some studies have focused on the rhizosphere and phyllosphere bacterial communities using fingerprint techniques of 16S rRNA genes, especially the rhizosphere of single cultivated plant species including potato and rice (19, 36, 41) and the phyllosphere of soybean, rice and maize (4, 17). The present research is the first to apply single digestion T-RFLP to leaf endophytic bacteria in multiple host species. Multi-enzyme studies depend on a reliable T-RFLP database to deduce species information; however most T-RFLP databases are still developing, so that a large proportion of novel bacteria, which are highly abundant in the environment, may not be matched using current databases (22). Although closely related bacterial species will usually produce the same T-RF, one or more other distinct taxonomic groups may also produce the same T-RF. Therefore variation in abundance of a T-RF may be due to changes in taxonomic composition of the class and/or relative abundance of one or more members of the group. Multi-enzymes are used in an effort to make taxonomic assignments; however taxonomic assignments are not necessary for identification of the factorial influences on the leaf endophytic bacterial communities, as studied in this work. Single digestion T-RFLP peaks represent special OTUs (Operational T-RFLP Unit) that

provide information on the diversity of leaf endophytic bacteria in different environments.

In this research, we explored the diversity of leaf endophytic bacteria in selected plant species over time and the physical environment, in order to propose a model describing how multiple factors influence endophytic bacterial communities. Past studies have found the plant genotype and growth conditions have significant impacts on the rhizosphere bacterial communities (19, 36, 41) and on the phyllosphere bacterial communities (4, 20). Here we considered three major influencing factors: host plant species, time and sampling sites. The distributions of leaf endophytic bacteria must be influenced by many factors; however, we hypothesized that these three major factors include most variables affecting community composition. We analyzed leaf endophytic bacterial communities from samples differing in these factors by pCCA and MANOVA of T-RFs and comparisons of the average amounts of T-RFs present in samples.

The factor of host plant species is the plant genotype, which includes the effect of inner biochemical environment and the physiological features of the host plant. The coevolution and codivergence of host plants and leaf endophytic bacterial communities may also contribute to the similarities and differences in the leaf endophytic bacterial communities from different host species. In order to focus on the relative amounts of T-RF OTUs in different plants only in those plants in which they are found, the APE of a T-RF in one host species was defined as the average proportion of a T-RF in all the samples of one plant species which have this T-RF. Calculating APE rather than regular average proportion can avoid the problem of underestimation of the abundance of a T-RF in one host species due to non-infection of the bacterial species represented in some samples.

The APE of a T-RF can more accurately reflect the overall compositions of leaf endophytic bacterial communities in a plant species than can methods that include absence in the analysis. The expectation of a major influence of host plant species on the communities was supported by the APE analysis (Table 3), distinct T-RF patterns from each host species (Figure 1 and Table 5) and by the results of pCCA which assigned half of the total variation to plant species.

The time factor includes climate change, temperature and the dynamics of host plant growth. Jackson and Denney (18) studied the annual and seasonal variation of phyllosphere bacteria and found that compared to significant seasonal variation, the annual variation was not significant. Yadav *et al*. (39) also found that the mature leaves have higher populations of phyllosphere bacteria than young leaves. These studies motivated us to consider the seasonal variation of plant-associated bacteria. The pCCA examination of T-RFs treating sampling date as the environmental factor implicated it as a significant factor (Figure 2). The impacts of sampling date on the distribution of plant-associated bacteria were also seen in the average numbers of T-RFs at different sampling dates (Table 2). The temporal variations in relative abundance of different T-RFs suggest that during host plant growth, the structure of plant leaf-associated bacterial communities are also developing to respond to the changes of the inner biochemical environments of host plants and the variations of the weather and overall environment. The host species selected for study reach their best physiological condition in July after initial blooming in late April or May and then begin senescing and die in August or later. The ratios of standard deviation to the average value are smaller in June and July than those in May

and August, indicating that the plant-associated bacterial communities are more stable and complex when the host plants are growing rapidly.

The factor of physical environment includes the soil and geobiochemical conditions, the effect of surrounding plants and animals, and the burning and grazing history of the sampling field, records of the latter of which are available. Again, pCCA attributed a significant contribution of sampling site to the total variation (Figure 2b) consistent with pattern differences for the same plant species on the same date (Figure 1).

We recognize that the three targeted factors may not account for all the variation in the communities and that we did encounter a residual variation. Sources of this variation could include: occasional animal disturbance, insect-induced damages and other factors that cannot be measured accurately and parameterized in a mathematical model. Nevertheless, we suggest that the three-factor model is appropriate to describe the largest part of the distribution of plant-associated bacteria. The plant-associated bacterial communities are not static, but dynamic and evolve with host plants and environments.

Acknowledgements

Authors acknowledge the support of the Oklahoma Agricultural Experiment Station, whose Director has approved this publication, the R. J. Sirny Professorship at Oklahoma State University and the National Science Foundation through EPS-0447262. They thank Michael Anderson, Mostafa Elshahed for critical readings of the manuscript and Joshua Habiger for suggesting additional statistical analyses.

LITERATURE CITED

1. **Ali N, Sorkhoh N, Salamah S, Eliyas M, Radwan S**. The potential of epiphytic hydrocarbon-utilizing bacteria on legume leaves for attenuation of atmospheric hydrocarbon pollutants. J. Environ. Manage 93:113-120.
2. **Allen MS, Hamilton RG, Melcher U, Palmer MW**. 2009. Lessons from the prairie: Research at The Nature Conservancy's Tallgrass Prairie Preserve. Oklahoma Academy of Sciences, Stillwater, OK. 48 pages.
3. **Avaniss-Aghajani E, Jones K, Holtzman A, Aronson T, Glover N, Boian M, Froman S, Brunk C**. 1996. Molecular technique for rapid identification of mycobacteria. J. Clin. Microbiol. 34:98-102.
4. **Balint-Kurti P, Simmons SJ, Blum JE, Ballare CL, Stapleton AE**. 2010. Maize leaf epiphytic bacteria diversity patterns are genetically correlated with resistance to fungal pathogen infection. Mol. Plant Microbe Interact. 23:473-484.
5. **Bell CR, Dickie GA, Harvey WLG, Chan JWYF**. 1995. Endophytic bacteria in grapevine. Can. J. Microbiol. 41:46-53.
6. **Chelius MK, Triplett EW.** 2001. The diversity of archaea and bacteria in association with the roots of Zea mays L. Microb. Ecol. 41:252-263.
7. **Conn VM, Franco CMM**. 2004. Analysis of the endophytic actinobacterial population in the roots of wheat (Triticum aestivum L.) by terminal restriction fragment length polymorphism and sequencing of 16S rRNA clones. Appl. Environ. Microbiol. 70:1784-1794.

8. **Culman SW, Bukowski R, Gauch HG, Cadillo-Quiroz H, Buckley D**. 2009. T-REX: software for the processing and analysis of T-RFLP data. BMC Bioinformatics 10:171.

9. **Culman SW, Gauch HG, Blackwood CG, Thies JE**. 2008. Analysis of T-RFLP data using analysis of variance and ordination methods: A comparative study. J. Microbiol. Methods. 75:55-63.

10. **Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R, von Mering C, Vorholt JA.** 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc. Natl. Acad. Sci. USA. 106:16428-16433.

11. **Elvira-Recuenco M, van Vuurde JWL**. 2000. Natural incidence of endophytic bacteria in pea cultivars under field conditions. Can. J. Microbiol. 46:1036-1041.

12. **Engebretson JJ, Moyer CL.** 2003. Fidelity of select restriction endonucleases in determining microbial diversity by terminal-restriction fragment length polymorphism. Appl. Environ. Microbiol. 69:4823-4829.

13. **Hallmann J, Quadt-Hallmann A, Mahaffee WF, Kloepper JW**. 1997. Bacterial endophytes in agricultural crops. Can. J. Microbiol. 43:895-914.

14. **Hirano SS, Nordheim EV, Arny DC, Upper CD.** 1982. Lognormal distribution of epiphytic bacterial populations on leaf surfaces. Appl. Environ. Microbiol. 44:695-700.

15. **Hunter PJ, Hand P, Pink D, Whipps JM , Bending GD**. 2010. Both leaf properties and microbe-microbe interactions influence within-species variation in

bacterial population diversity and structure in the lettuce (Lactuca species) phyllosphere. Appl. Environ. Microbiol. 76:8117-8125.

16. **Ibekwe AM, Grieve CM.** 2004. Changes in developing plant microbial community structure as affected by contaminated water. FEMS Microbiol. Ecol. 48:239-248.

17. **Ikeda S, Okubo T, Anda M, Nakashita H, Yasuda M, Sato S, Kaneko T, Tabata S, Eda S, Momiyama A, Terasawa K, Mitsui H, Minamisawa H.** 2010. Community- and genome-based views of plant-associated bacteria: Plant-bacterial interactions in soybean and rice. Plant Cell Physiol. 51:1398-1410.

18. **Jackson C, Denney W.** 2011. Annual and seasonal variation in the phyllosphere bacterial community associated with leaves of the southern Magnolia (Magnolia grandiflora). Microb. Ecol. 61:113-122.

19. **Knauth S, Hurek T, Brar D, Reinhold-Hurek B.** 2005. Influence of different Oryza cultivars on expression of nifH gene pools in roots of rice. Environ. Microbiol. 7:1725-1733.

20. **Knief C, Ramette A, Frances L, Alonso-Blanco C, Vorholt JA**. 2010. Site and plant species are important determinants of the Methylobacterium community composition in the plant phyllosphere. ISME J. 4:719-728.

21. **Liu WT, Marsh TL, Cheng H, Forney LJ.** 1997. Characterization of Microbial Diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16s rRNA. Appl. Environ. Microbiol. 63:4516-4522.

22. **Lopez-Velasco G, Welbaum GE, Boyer RR, Mane SP, Ponder MA.** 2011. Changes in spinach phylloepiphytic bacteria communities following minimal

processing and refrigerated storage described using pyrosequencing of 16S rRNA amplicons. J. Appl. Microbiol. 110:1203-1214.

23. **Melcher UK, Muthukumar V, Wiley GB, Min BE, Palmer MW, Verchot-Lubicz J, Ali A, Nelson RS, Roe BA, Thapa V, Pierce ML.** 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from Ambrosia psilostachya. J. Virol. Methods. 152:49-55.

24. **Muthukumar V, Melcher UK, Pierce ML, Wiley GB, Roe BA, Palmer MW, Thapa V, Ali A, Ding T.** 2009. Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. Virus Res. 141:169-173.

25. **Osborn AM, Moore ERB, Timmis KN.** 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. 2:39-50.

26. **Rastogi G, Tech JJ, Coaker GL, Leveau JHJ.** 2010. A PCR-based toolbox for the culture-independent quantification of total bacterial abundances in plant environments. J. Microbiol. Methods. 83:127-132.

27. **Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, ChavarrÍA F, Shen G, Roe BA**. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Mol. Ecol. 19:81-88.

28. **Ryan RP, Germaine K, Franks A, Ryan DJ, Dowling DN**. 2008. Bacterial endophytes: recent developments and applications. FEMS Microbiol. Lett. 278:1-9.

29. **Stoltzfus JR , So R, Malarvithi PP, Ladha JK, de Brujin FJ.** 1998. Isolation of endophytic bacteria from rice and assessment of their potential for supplying rice with biologically fixed nitrogen. Plant Soil 194:25-36.

30. **Sturz AV, Christie BR, Matheson BG.** 1998. Associations of bacterial endophyte populations from red clover and potato crops with potential for beneficial allelopathy. Can. J. Microbiol. 44:162-167.

31. **Sturz AV, Christie BR, Matheson BG, Nowak J.** 1997. Biodiversity of endophytic bacteria which colonize red clover nodules, roots, stems and foliage and their influence on host growth. Biol. Fert. Soils. 25:13-19.

32. **Sun L, Qiu F, Zhang X, Dai X, Dong X, Song W.** 2007. Endophytic bacterial diversity in rice (Oryza sativa L.) roots estimated by 16S rDNA sequence analysis. Microb. Ecol. 55:415-424.

33. **ter Braak CJF, Smilauer P**. 1998. CANOCO reference manual and user's guide to Canoco for Windows : software for canonical community ordination (version 4). Wageningen : Centre for Biometry, Ithaca, NY.

34. **Ulrich A, Becker R.** 2006. Soil parent material is a key determinant of the bacterial community structure in arable soils. FEMS Microbiol. Ecol. 56:430-443.

35. **Ulrich K, Ulrich A, Ewald D**. 2008. Diversity of endophytic bacterial communities in poplar grown under field conditions. FEMS Microbiol. Ecol. 63:169-180.

36. **Weinert N, Meincke R, Gottwald C, Heuer H, Schloter M, Berg G, Smalla K**. 2010. Bacterial diversity on the surface of potato tubers in soil and the influence of the plant genotype. FEMS Microbiol. Ecol. 74:114-123.

37. **Wellner S, Lodders N, Kämpfer P.** Diversity and biogeography of selected phyllosphere bacteria with special emphasis on Methylobacterium spp. **Syst. Appl. Microbiol.** 34:621-630.

38. **Wren JD, Roossinck MJ, Nelson RS, Scheets K, Palmer MW, Melcher U.** 2006. Plant virus biodiversity and ecology. PLoS Biol. 4:e80.

39. **Yadav R, Karamanoli K, Vokou D.** 2011. Bacterial populations on the phyllosphere of Mediterranean plants: influence of leaf age and leaf surface. Front. Agric. China 5:60-63.

40. **Yashiro E, Spear RN, McManus PS.** 2011. Culture-dependent and culture-independent assessment of bacteria in the apple phyllosphere. J. Appl. Microbiol. 110:1284-1296.

41. **İnceoğlu Ö, Salles JF, van Overbeek L, van Elsas JD.** Effects of plant genotype and growth stage on the betaproteobacterial communities associated with different potato cultivars in two fields. Appl. Environ. Microbiol. 76:3675-3684.

Figure Captions

Figure 1 -3. Comparisons of T-RFLP profiles of endophytic bacterial communities. Relative fluorescence intensity (normalized to the most intense peak) is plotted against length of the T-RF. T-RFLP profiles represented the bacterial species compositions, indicating the influences from multiple factors.

Figure 1. T-RFLP profiles from two *A. viridis* individuals respectively in Site 2 and Site 3, both collected on July 14th, 2010.

Figure 2. T-RFLP profiles from one labeled *A. viridis* individual, collected respectively on May 14th, June 16th and July 14th, 2010.

Figure 3. T-RFLP profiles from 3 individuals respectively from *A. viridis*, *A. psilostachya* and *P. virgatum*. For the dominant T-RFs from these three plant species, see Table 5.

Figure 4 - 6. Partial Canonical Correspondence Analyses (pCCA) of T-RFLP profiles treating each of the three factors considered as the environmental factor. pCCA Axis1 and pCCA Axis 2 represented two most important canonical correlation which can explain the variation of those testing samples, of which pCCA Axis1 was the most important.

Figure 4. pCCA of T-RFLP profiles treating sampling date as the environmental factor.

Figure 5. pCCA of T-RFLP profiles treating sampling location as the environmental factor.

Figure 6. pCCA of T-RFLP profiles treating host plant species as the environmental factor.

Figure 7. The complete heatmap of the frequencies of T-RFs detected in five host species. showed the frequencies of all the T-RFs and the clustering results of the T-RFs and host species.

Figure 8. The first branch of the clustering of the T-RFs in Figure 7 containing most frequent T-RFs.

Figure 9. The comparisons of two T-RFLP patterns from the *Dde*I digestion products of the *Asclepias viridis* Sample 1 from Site 2 collected in June 16th , 2010, scanned on Aug 19th, 2010 (above) and Aug 30th 2010 (below). The T-RFLP patterns of the same sample scanned in different experiment were indistinguishable, indicating that the T-RFLP is highly reproducible.

Table 1. Summary statistics for T-RFs of *Asclepias viridis* samples from different months and sites.

Sample Variable[a]	Total T-RFs	Richness	Percent empty cells in matrix	Beta diversity[b]
May	27	6.8	77.2%	2.95
June	46	21.9	52.3%	1.10
July	59	20.0	68.7%	1.95
Site 1	45	15.3	65.9%	1.93
Site 2	44	15.4	64.9%	1.76
Site 3	44	15.0	65.9%	1.93
Site 4	33	13.8	58.2%	1.39

[a]For months, data summarized over all sites; for sites, data summarized over all months.

[b]Beta diversity was calculated as the total species number in all samples collected in May divided by the average species number in all samples collected in May.

Table 2. Average numbers of T-RFs from different host species, sampling dates or locations.

Samples	Average Number of T-RFs
Ambrosia psilostachya	17.38 +/- 4.98
Panicum virgatum	15.00 +/- 10.46
Asclepias viridis	14.89 +/- 7.04
Sorghastrum nutans	12.92 +/- 5.09
Ruellia humilis	5.50 +/- 2.72
Site 1 Samples	14.71 +/- 7.46
Site 2 Samples	13.86 +/- 6.94
Site 3 Samples	12.45 +/- 7.84
Site 4 Samples	14.60 +/- 8.24
May Samples	9.29 +/- 7.95
June Samples	14.72 +/- 6.16
July Samples	18.04 +/- 5.91
August Samples	12.73 +/- 7.47

Table 3. Average proportion per existence in five different host species of selected significant T-RFs (Average frequencies > 0.3).

T-RF (bp)	*A. psilostachya*	*P. virgatum*	*A. viridis*	*S. nutans*	*R. humilis*
75	0.05	0.04	0.18	0.05	0.11
77	0.00	0.02	0.05	0.05	0.07
78	0.04	0.30	0.12	0.07	0.54
79	0.11	0.14	0.15	0.08	0.30
85	0.18	0.13	0.14	0.12	0.09
94	0.08	0.00	0.01	0.04	0.00
236	0.03	0.07	0.02	0.17	0.00
350	0.05	0.09	0.07	0.12	0.09
352	0.09	0.04	0.04	0.06	0.00
355	0.09	0.20	0.00	0.15	0.03
529	0.14	0.08	0.22	0.09	0.15

Table 4. Locations of sampling sites in the TGPP.

Site No.	UTM location	Elevation (m)
Site 1	14 S 0736182 4070432	288
Site 2	14 S 0732625 4070095	300
Site 3	14 S 0730241 4080682	326
Site 4	14 S 0727969 4076948	299

Table 5. Dominant T-RFs from amplified 16S bactrial rDNA from three plant species.

Host species	Dominant T-RFs
Asclepias viridis	75bp, 77bp, 78bp, 79bp, 85bp, 89bp, 347bp, 350bp, 354bp, 529bp *
Ambrosia psilostachya	75bp, 79bp, 84bp, 85bp *, 94bp, 346bp, 348bp, 352bp, 355bp, 529bp
Panicum virgatum	78bp, 79bp, 85bp, 95bp, 236bp, 355bp *, 529bp

* indicates the most dominant T-RF in that species.

Table 6. Frequencies of all the T-RFs in 5 different host species and their average frequencies.

T-RF (bp)	*A. psilostachya*	*P. virgatum*	*A. viridis*	*S. nutans*	*R. humilis*	Average frequency
55	0.00	0.00	0.00	0.00	0.10	0.02
57	0.08	0.57	0.00	0.25	0.00	0.18
62	0.00	0.07	0.00	0.00	0.00	0.01
66	0.00	0.00	0.03	0.00	0.00	0.01
67	0.08	0.00	0.00	0.00	0.00	0.02
71	0.08	0.00	0.17	0.00	0.00	0.05
72	0.00	0.21	0.11	0.17	0.00	0.10
73	0.00	0.00	0.25	0.00	0.00	0.05
74	0.00	0.14	0.11	0.00	0.00	0.05
75	0.85	0.29	0.75	0.50	0.40	0.56
76	0.08	0.14	0.25	0.08	0.00	0.11
77	0.00	0.36	0.75	0.50	0.50	0.42
78	0.31	0.57	0.75	0.58	0.80	0.60
79	0.85	0.57	0.19	0.50	0.40	0.50
81	0.00	0.00	0.14	0.00	0.00	0.03
82	0.00	0.07	0.00	0.00	0.00	0.01
83	0.00	0.00	0.06	0.00	0.00	0.01
84	0.23	0.14	0.00	0.17	0.00	0.11
85	0.85	0.71	0.72	0.67	0.10	0.61

89	0.00	0.00	0.92	0.00	0.00	0.18
92	0.00	0.00	0.00	0.17	0.80	0.19
94	1.00	0.00	0.06	0.75	0.00	0.36
95	0.00	0.93	0.00	0.00	0.00	0.19
96	0.00	0.07	0.11	0.50	0.10	0.16
97	0.00	0.00	0.22	0.00	0.00	0.04
98	0.00	0.36	0.00	0.33	0.10	0.16
99	0.00	0.00	0.00	0.08	0.00	0.02
100	0.00	0.00	0.00	0.17	0.00	0.03
103	0.00	0.00	0.03	0.00	0.00	0.01
113	0.00	0.00	0.03	0.00	0.00	0.01
129	0.00	0.07	0.00	0.00	0.00	0.01
148	0.00	0.00	0.03	0.00	0.00	0.01
163	0.00	0.00	0.44	0.00	0.00	0.09
164	0.31	0.00	0.14	0.00	0.00	0.09
185	0.23	0.00	0.00	0.00	0.00	0.05
186	0.08	0.00	0.00	0.00	0.00	0.02
193	0.00	0.00	0.06	0.00	0.00	0.01
194	0.00	0.00	0.03	0.00	0.00	0.01
203	0.08	0.00	0.00	0.00	0.00	0.02
206	0.00	0.00	0.03	0.00	0.00	0.01
213	0.00	0.00	0.03	0.00	0.00	0.01
213.8	0.23	0.00	0.36	0.00	0.00	0.12
215	0.08	0.00	0.25	0.00	0.00	0.07

219	0.00	0.00	0.06	0.00	0.00	0.01
224	0.00	0.00	0.03	0.00	0.00	0.01
224.5	0.31	0.29	0.03	0.08	0.00	0.14
227	0.00	0.00	0.00	0.08	0.00	0.02
228	0.54	0.43	0.03	0.25	0.00	0.25
229	0.00	0.00	0.03	0.00	0.00	0.01
230	0.00	0.00	0.03	0.00	0.00	0.01
232	0.00	0.21	0.00	0.00	0.00	0.04
235	0.31	0.14	0.42	0.25	0.00	0.22
236	0.46	0.50	0.50	0.33	0.00	0.36
239	0.08	0.07	0.08	0.00	0.00	0.05
241	0.00	0.07	0.00	0.08	0.00	0.03
243	0.15	0.00	0.00	0.00	0.00	0.03
245.5	0.00	0.07	0.00	0.08	0.00	0.03
247.2	0.00	0.14	0.00	0.00	0.00	0.03
248	0.00	0.07	0.14	0.17	0.00	0.08
249	0.08	0.43	0.03	0.25	0.00	0.16
251	0.00	0.07	0.00	0.00	0.00	0.01
256	0.00	0.00	0.00	0.08	0.00	0.02
266	0.23	0.14	0.00	0.08	0.00	0.09
267	0.00	0.00	0.03	0.00	0.00	0.01
268	0.00	0.07	0.00	0.08	0.00	0.03
269	0.08	0.29	0.00	0.25	0.00	0.12
280	0.08	0.00	0.00	0.00	0.00	0.02

303	0.00	0.00	0.03	0.00	0.00	0.01
319.5	0.00	0.00	0.03	0.00	0.00	0.01
327.5	0.08	0.00	0.00	0.00	0.00	0.02
335	0.08	0.00	0.06	0.00	0.00	0.03
336	0.00	0.07	0.00	0.00	0.00	0.01
337	0.00	0.00	0.08	0.00	0.00	0.02
339	0.00	0.07	0.00	0.00	0.00	0.01
345	0.00	0.00	0.44	0.00	0.00	0.09
346	0.77	0.50	0.00	0.00	0.10	0.27
347	0.08	0.14	0.28	0.17	0.00	0.13
347.5	0.31	0.00	0.56	0.00	0.10	0.19
348	0.69	0.43	0.00	0.25	0.00	0.27
349	0.00	0.07	0.00	0.58	0.20	0.17
350	0.62	0.79	0.69	0.58	0.70	0.68
351	0.31	0.14	0.06	0.00	0.00	0.10
352	0.62	0.43	0.56	0.58	0.00	0.44
353	0.00	0.07	0.03	0.42	0.10	0.12
354	0.00	0.00	0.94	0.00	0.00	0.19
355	1.00	1.00	0.00	1.00	0.30	0.66
367	0.00	0.00	0.14	0.00	0.00	0.03
368	0.77	0.43	0.00	0.25	0.00	0.29
370	0.00	0.00	0.03	0.00	0.00	0.01
372	0.00	0.14	0.00	0.00	0.00	0.03
375	0.00	0.14	0.00	0.00	0.00	0.03

376	0.00	0.00	0.03	0.00	0.00	0.01
377	0.15	0.07	0.00	0.00	0.00	0.05
379	0.08	0.07	0.00	0.00	0.00	0.03
380	0.00	0.07	0.31	0.00	0.00	0.08
380.7	0.69	0.36	0.25	0.42	0.00	0.34
382	0.00	0.07	0.03	0.00	0.00	0.02
383	0.15	0.00	0.00	0.00	0.00	0.03
384	0.15	0.00	0.00	0.00	0.00	0.03
395	0.00	0.00	0.03	0.00	0.00	0.01
396	0.08	0.00	0.00	0.00	0.00	0.02
397	0.00	0.07	0.00	0.00	0.00	0.01
493	0.62	0.00	0.00	0.00	0.00	0.12
504	0.00	0.00	0.03	0.00	0.00	0.01
509	0.08	0.00	0.00	0.08	0.00	0.03
511	0.00	0.07	0.03	0.00	0.00	0.02
513	0.00	0.00	0.03	0.00	0.00	0.01
514	0.00	0.00	0.06	0.00	0.00	0.01
523	0.08	0.07	0.06	0.08	0.00	0.06
524	0.23	0.14	0.19	0.00	0.00	0.11
525	0.23	0.07	0.53	0.00	0.00	0.17
526	0.54	0.07	0.08	0.00	0.00	0.14
527	0.15	0.14	0.00	0.17	0.00	0.09
528	0.00	0.07	0.00	0.00	0.00	0.01
529	1.00	0.79	0.83	0.75	0.70	0.81

531	0.00	0.00	0.03	0.00	0.00	0.01
532	0.08	0.14	0.00	0.00	0.00	0.04
536	0.00	0.00	0.03	0.00	0.00	0.01
550	0.08	0.00	0.00	0.00	0.00	0.02
624	0.00	0.00	0.00	0.08	0.00	0.02
672	0.00	0.00	0.03	0.00	0.00	0.01
706	0.00	0.00	0.03	0.00	0.00	0.01

Table 7. Average Proportion per Existence (APE) of all the T-RFs in 5 different host species.

T-RF (bp)	*A. psilostachya*	*P. virgatum*	*A. viridis*	*S. nutans*	*R. humilis*
55	0.00	0.00	0.00	0.00	0.01
57	0.02	0.05	0.00	0.03	0.00
62	0.00	0.00	0.00	0.00	0.00
66	0.00	0.00	0.00	0.00	0.00
67	0.00	0.00	0.00	0.00	0.00
71	0.02	0.00	0.01	0.00	0.00
72	0.00	0.01	0.01	0.03	0.00
73	0.00	0.00	0.01	0.00	0.00
74	0.00	0.02	0.01	0.00	0.00
75	0.05	0.04	0.18	0.05	0.11
76	0.00	0.01	0.02	0.03	0.00
77	0.00	0.02	0.05	0.05	0.07
78	0.04	0.30	0.12	0.07	0.54
79	0.11	0.14	0.15	0.08	0.30
81	0.00	0.00	0.03	0.00	0.00
82	0.00	0.00	0.00	0.00	0.00
83	0.00	0.00	0.01	0.00	0.00
84	0.01	0.04	0.00	0.02	0.00
85	0.18	0.13	0.14	0.12	0.09
89	0.00	0.00	0.09	0.00	0.00

92	0.00	0.00	0.00	0.09	0.11
94	0.08	0.00	0.01	0.04	0.00
95	0.00	0.07	0.00	0.00	0.00
96	0.00	0.11	0.04	0.09	0.10
97	0.00	0.00	0.02	0.00	0.00
98	0.00	0.05	0.00	0.09	0.02
99	0.00	0.00	0.00	0.02	0.00
100	0.00	0.00	0.00	0.32	0.00
103	0.00	0.00	0.07	0.00	0.00
113	0.00	0.00	0.01	0.00	0.00
129	0.00	0.01	0.00	0.00	0.00
148	0.00	0.00	0.05	0.00	0.00
163	0.00	0.00	0.03	0.00	0.00
164	0.01	0.00	0.03	0.00	0.00
185	0.03	0.00	0.00	0.00	0.00
186	0.03	0.00	0.00	0.00	0.00
193	0.00	0.00	0.03	0.00	0.00
194	0.00	0.00	0.04	0.00	0.00
203	0.00	0.00	0.00	0.00	0.00
206	0.00	0.00	0.00	0.00	0.00
213	0.00	0.00	0.01	0.00	0.00
213.8	0.02	0.00	0.01	0.00	0.00
215	0.05	0.00	0.01	0.00	0.00
219	0.00	0.00	0.01	0.00	0.00

224	0.00	0.00	0.01	0.00	0.00
224.5	0.02	0.01	0.01	0.02	0.00
227	0.00	0.00	0.00	0.03	0.00
228	0.02	0.04	0.02	0.13	0.00
229	0.00	0.00	0.01	0.00	0.00
230	0.00	0.00	0.01	0.00	0.00
232	0.00	0.01	0.00	0.00	0.00
235	0.02	0.02	0.02	0.01	0.00
236	0.03	0.07	0.02	0.17	0.00
239	0.01	0.01	0.00	0.00	0.00
241	0.00	0.02	0.00	0.02	0.00
243	0.01	0.00	0.00	0.00	0.00
245.5	0.00	0.00	0.00	0.01	0.00
247.2	0.00	0.01	0.00	0.00	0.00
248	0.00	0.01	0.01	0.02	0.00
249	0.01	0.03	0.01	0.17	0.00
251	0.00	0.00	0.00	0.00	0.00
256	0.00	0.00	0.00	0.02	0.00
266	0.02	0.01	0.00	0.01	0.00
267	0.00	0.00	0.01	0.00	0.00
268	0.00	0.00	0.00	0.03	0.00
269	0.01	0.03	0.00	0.03	0.00
280	0.01	0.00	0.00	0.00	0.00
303	0.00	0.00	0.04	0.00	0.00

319.5	0.00	0.00	0.00	0.00	0.00
327.5	0.03	0.00	0.00	0.00	0.00
335	0.00	0.00	0.02	0.00	0.00
336	0.00	0.00	0.00	0.00	0.00
337	0.00	0.00	0.01	0.00	0.00
339	0.00	0.00	0.00	0.00	0.00
345	0.00	0.00	0.02	0.00	0.00
346	0.06	0.01	0.00	0.00	0.03
347	0.00	0.00	0.05	0.01	0.00
347.5	0.03	0.00	0.06	0.00	0.18
348	0.09	0.02	0.00	0.06	0.00
349	0.00	0.00	0.00	0.07	0.25
350	0.05	0.09	0.07	0.12	0.09
351	0.04	0.01	0.01	0.00	0.00
352	0.09	0.04	0.04	0.06	0.00
353	0.00	0.39	0.00	0.03	0.12
354	0.00	0.00	0.09	0.00	0.00
355	0.09	0.20	0.00	0.15	0.03
367	0.00	0.00	0.01	0.00	0.00
368	0.02	0.03	0.00	0.07	0.00
370	0.00	0.00	0.00	0.00	0.00
372	0.00	0.02	0.00	0.00	0.00
375	0.00	0.00	0.00	0.00	0.00
376	0.00	0.00	0.01	0.00	0.00

377	0.01	0.00	0.00	0.00	0.00
379	0.01	0.00	0.00	0.00	0.00
380	0.00	0.01	0.02	0.00	0.00
380.7	0.02	0.01	0.01	0.02	0.00
382	0.00	0.02	0.01	0.00	0.00
383	0.02	0.00	0.00	0.00	0.00
384	0.01	0.00	0.00	0.00	0.00
395	0.00	0.00	0.02	0.00	0.00
396	0.02	0.00	0.00	0.00	0.00
397	0.00	0.00	0.00	0.00	0.00
493	0.03	0.00	0.00	0.00	0.00
504	0.00	0.00	0.01	0.00	0.00
509	0.01	0.00	0.00	0.03	0.00
511	0.00	0.00	0.01	0.00	0.00
513	0.00	0.00	0.01	0.00	0.00
514	0.00	0.00	0.01	0.00	0.00
523	0.01	0.00	0.01	0.01	0.00
524	0.02	0.01	0.02	0.00	0.00
525	0.03	0.01	0.02	0.00	0.00
526	0.01	0.00	0.01	0.00	0.00
527	0.02	0.02	0.00	0.01	0.00
528	0.00	0.00	0.00	0.00	0.00
529	0.14	0.08	0.22	0.09	0.15
531	0.00	0.00	0.01	0.00	0.00

532	0.01	0.01	0.00	0.00	0.00
536	0.00	0.00	0.01	0.00	0.00
550	0.03	0.00	0.00	0.00	0.00
624	0.00	0.00	0.00	0.02	0.00
672	0.00	0.00	0.02	0.00	0.00
706	0.00	0.00	0.00	0.00	0.00

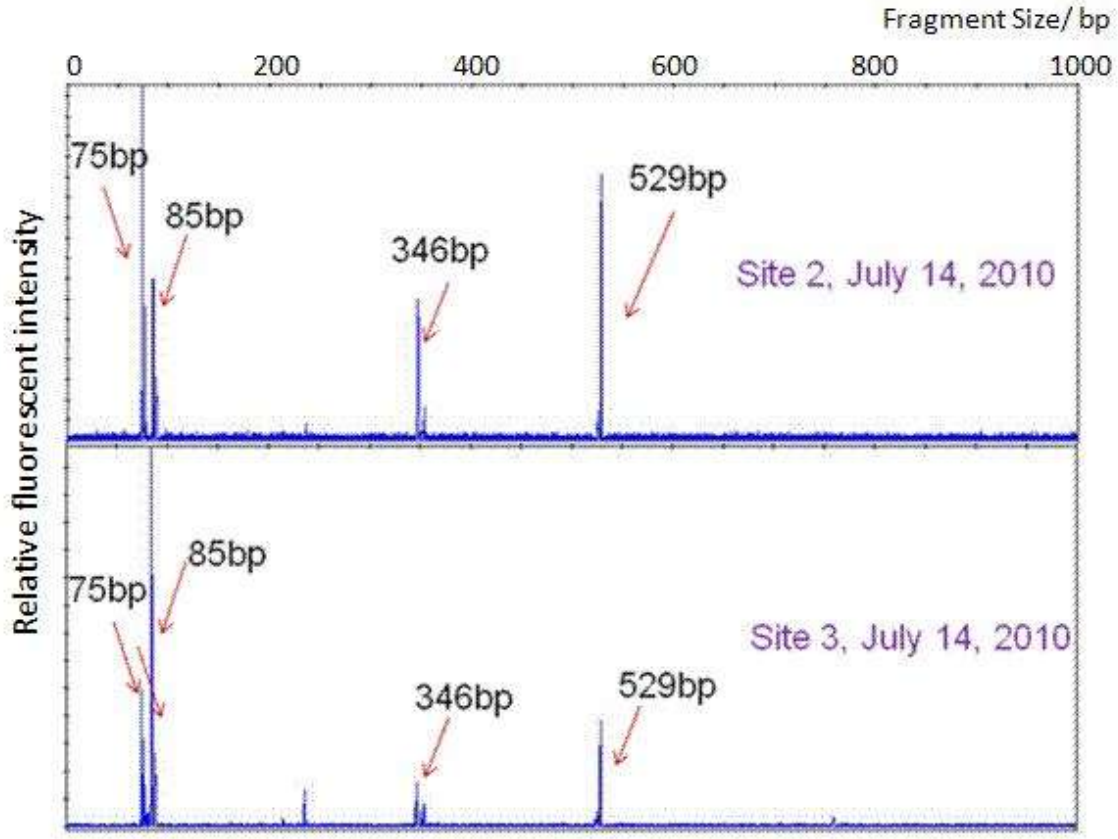

Fragment Size/ bp
0
200
400
600
800
1000
Relative fluorescent intensity
75bp
85bp
346bp
529bp
Site 2, July 14, 2010
75bp
85bp
346bp
529bp
Site 3, July 14, 2010

Figure 1.

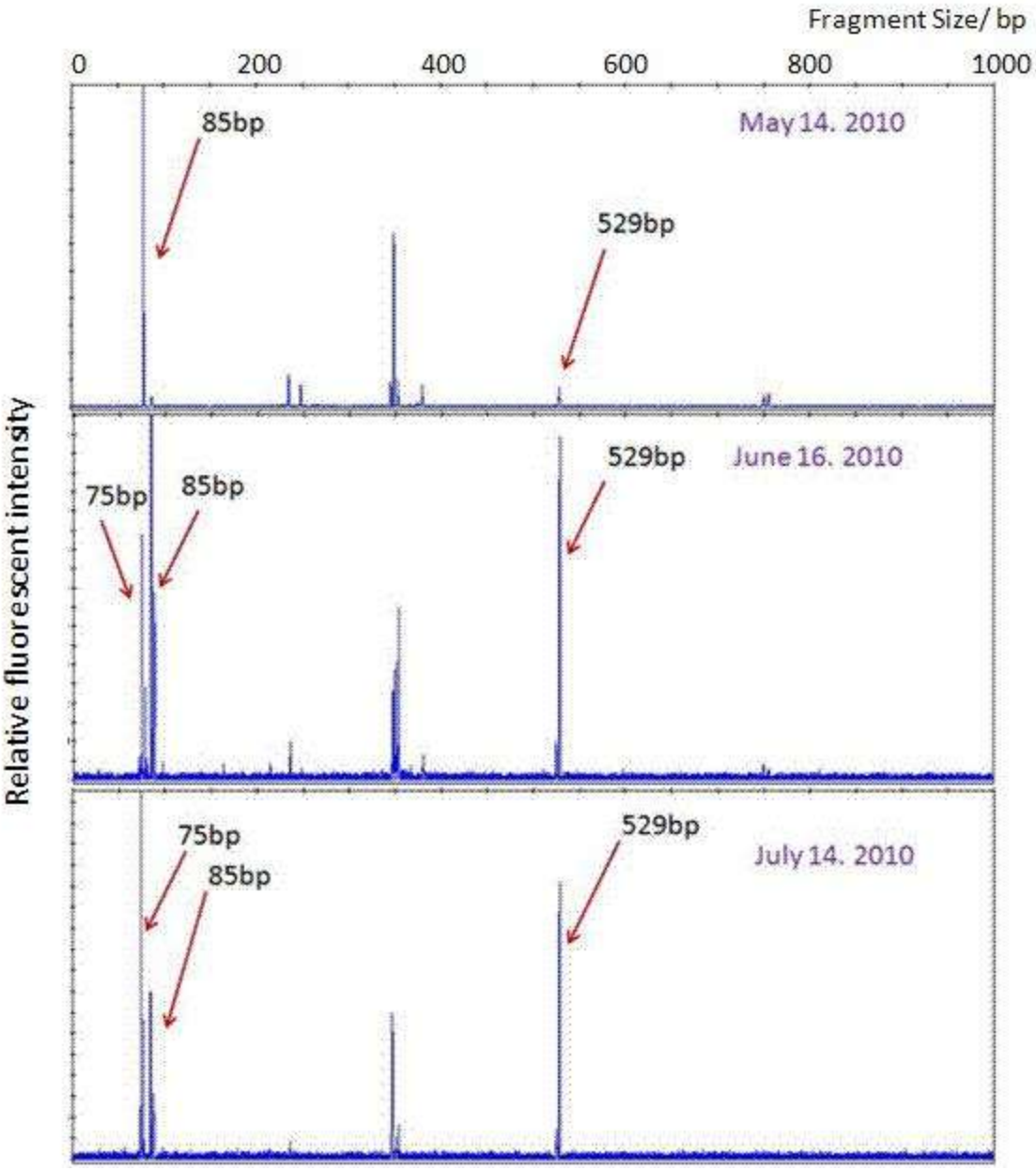

Fragment Size/ bp
0
200
400
600
800
1000
Relative fluorescent intensity
85bp
May 14. 2010
529bp
75bp
85bp
529bp
June 16. 2010
75bp
85bp
529bp
July 14. 2010

Figure 2.

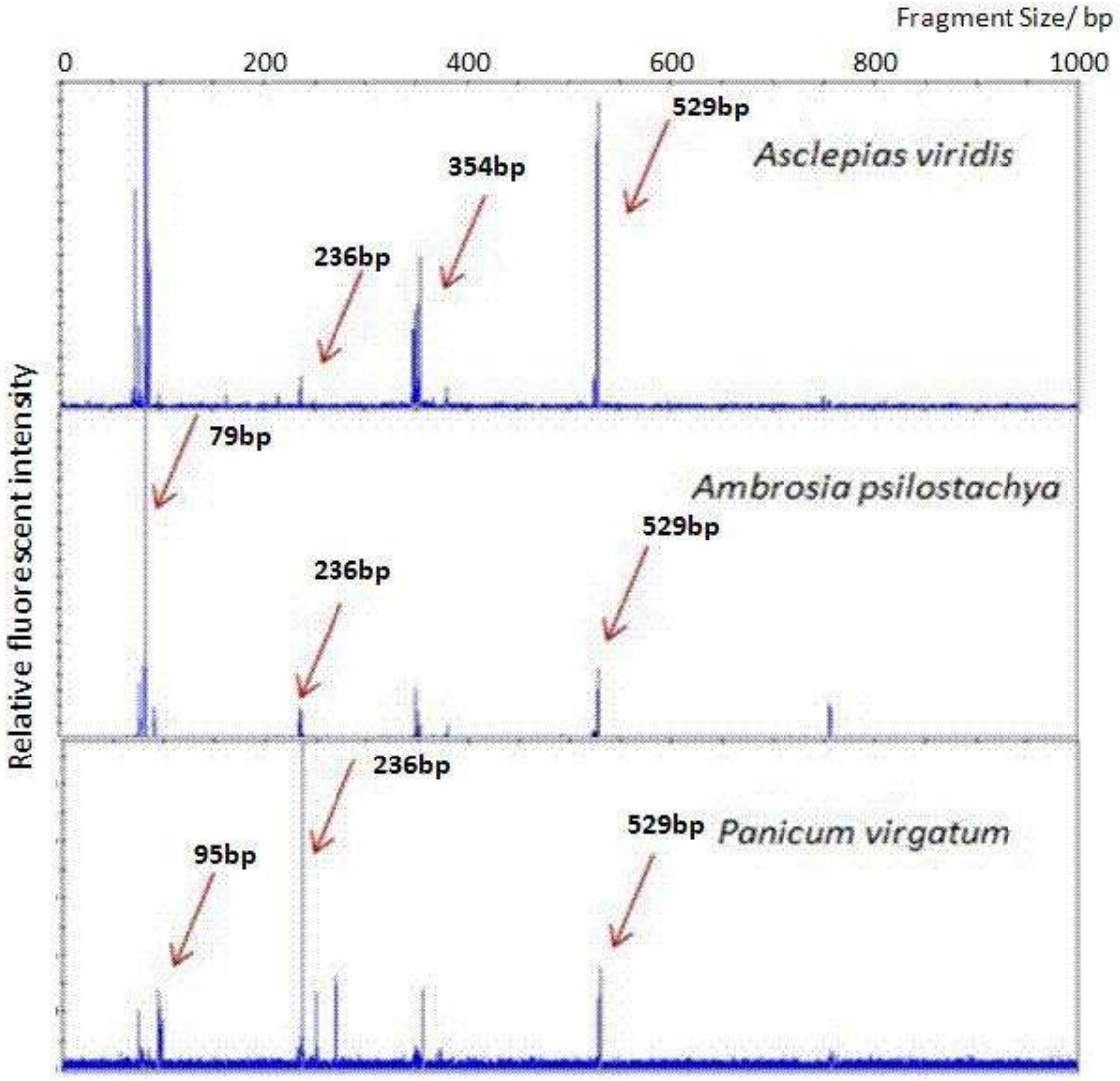

Fragment Size/ bp
0
200
400
600
800
1000
Relative fluorescent intensity
529bp
Asclepias viridis
354bp
236bp
79bp
Ambrosia psilostachya
529bp
236bp
236bp
529bp
Panicum virgatum
95bp

Figure 3.

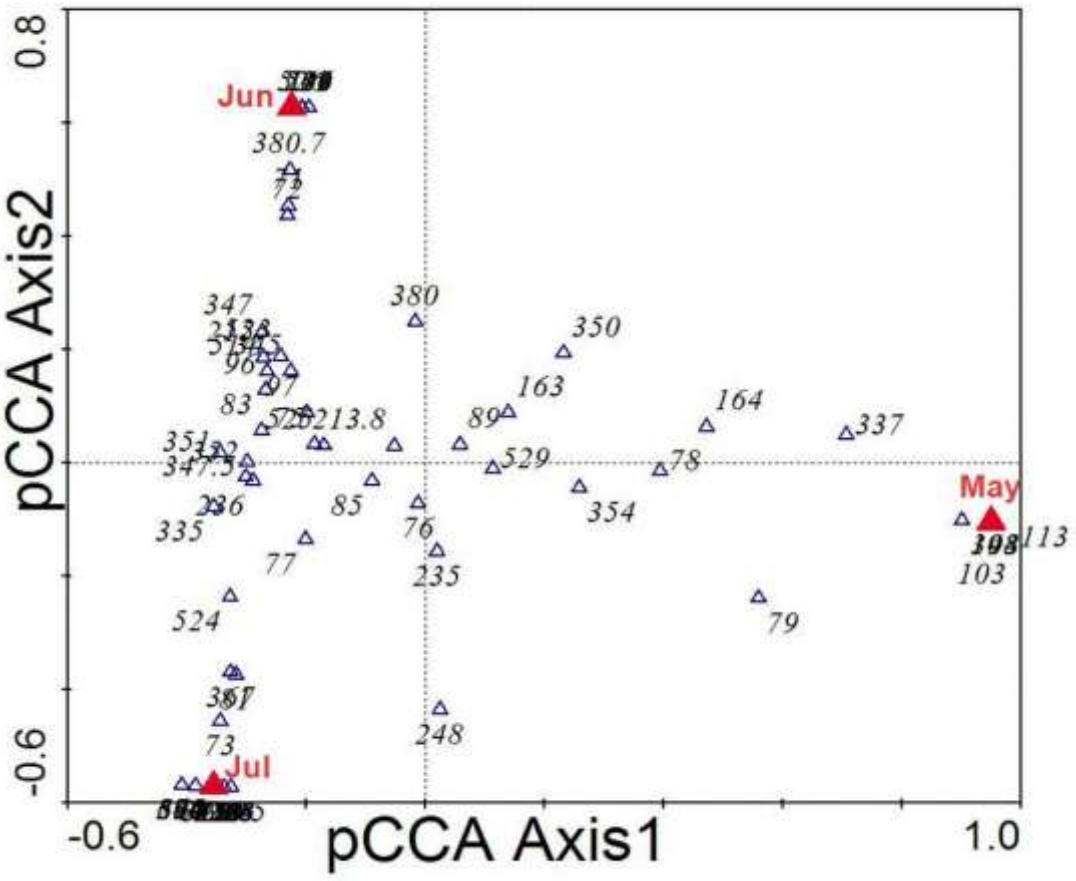

pCCA Axis2
pCCA Axis1
0.8
-0.6
-0.6
1.0
Jun
380.7
May
Jul
380
350
163
164
337
89
213.8
529
78
354
85
76
235
77
335
524
79
103
113
248
73
347

Figure 4.

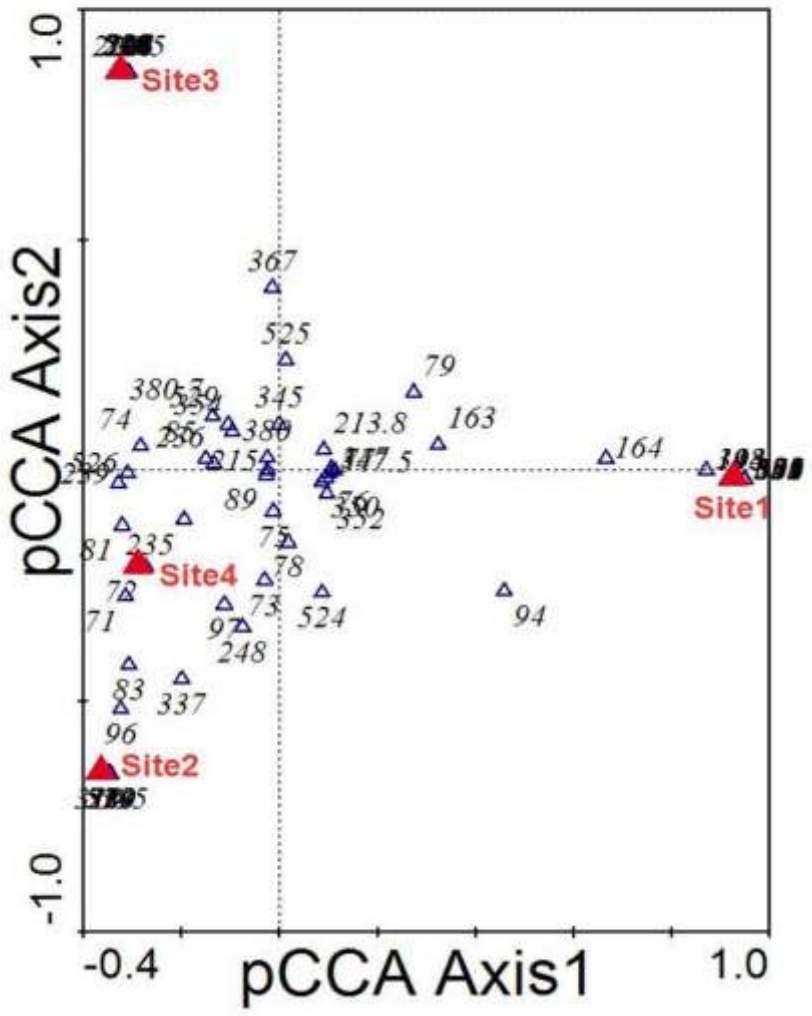

1.0
-1.0
pCCA Axis2
-0.4
pCCA Axis1
1.0
Site3
Site1
Site4
Site2
367
525
79
345
213.8
163
164
89
235
524
94
248
337
96

Figure 5.

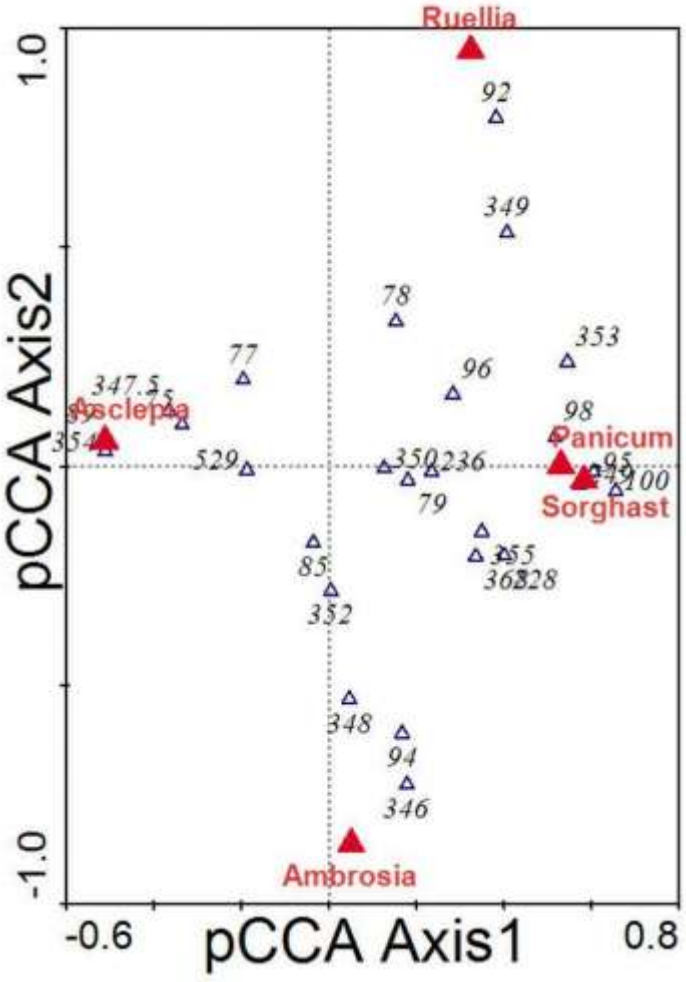

Ruellia
92
349
78
353
77
96
347.5
Asclepias
98
Panicum
354
529
350
236
95
100
Sorghast
79
85
352
348
94
346
Ambrosia
1.0
-1.0
pCCA Axis2
-0.6
pCCA Axis1
0.8

Figure 6.

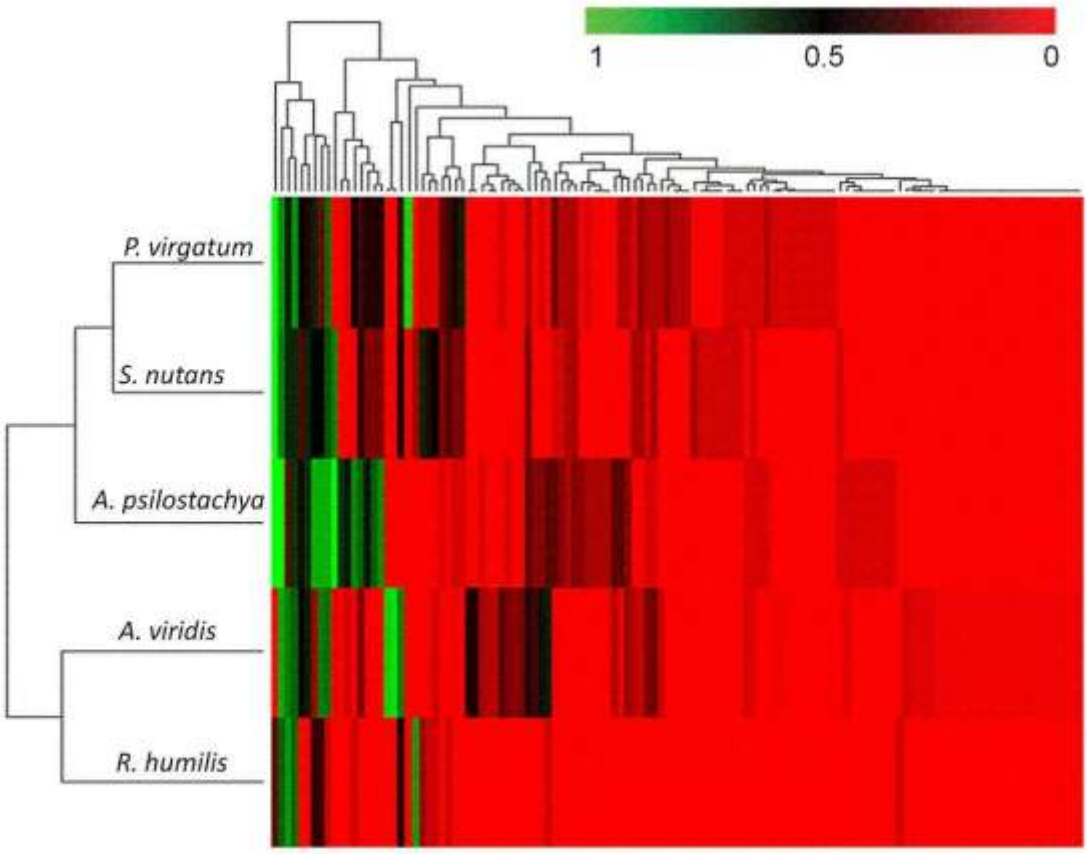


Figure 7.

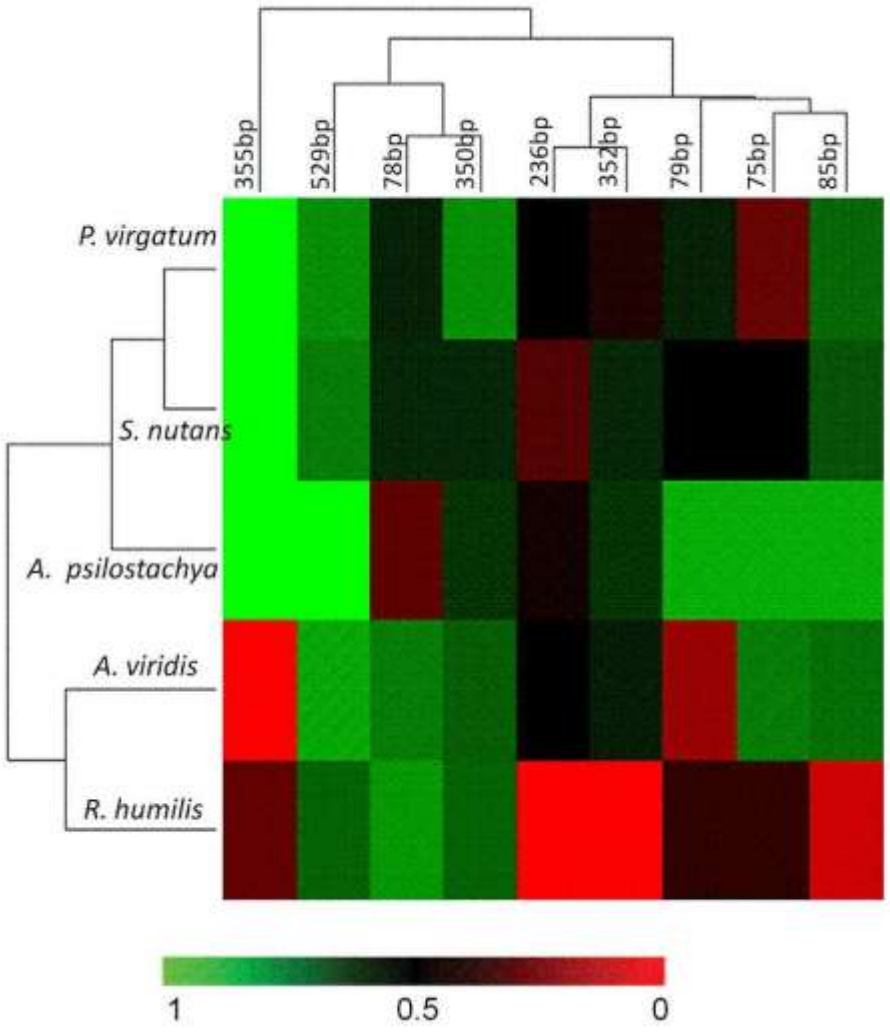
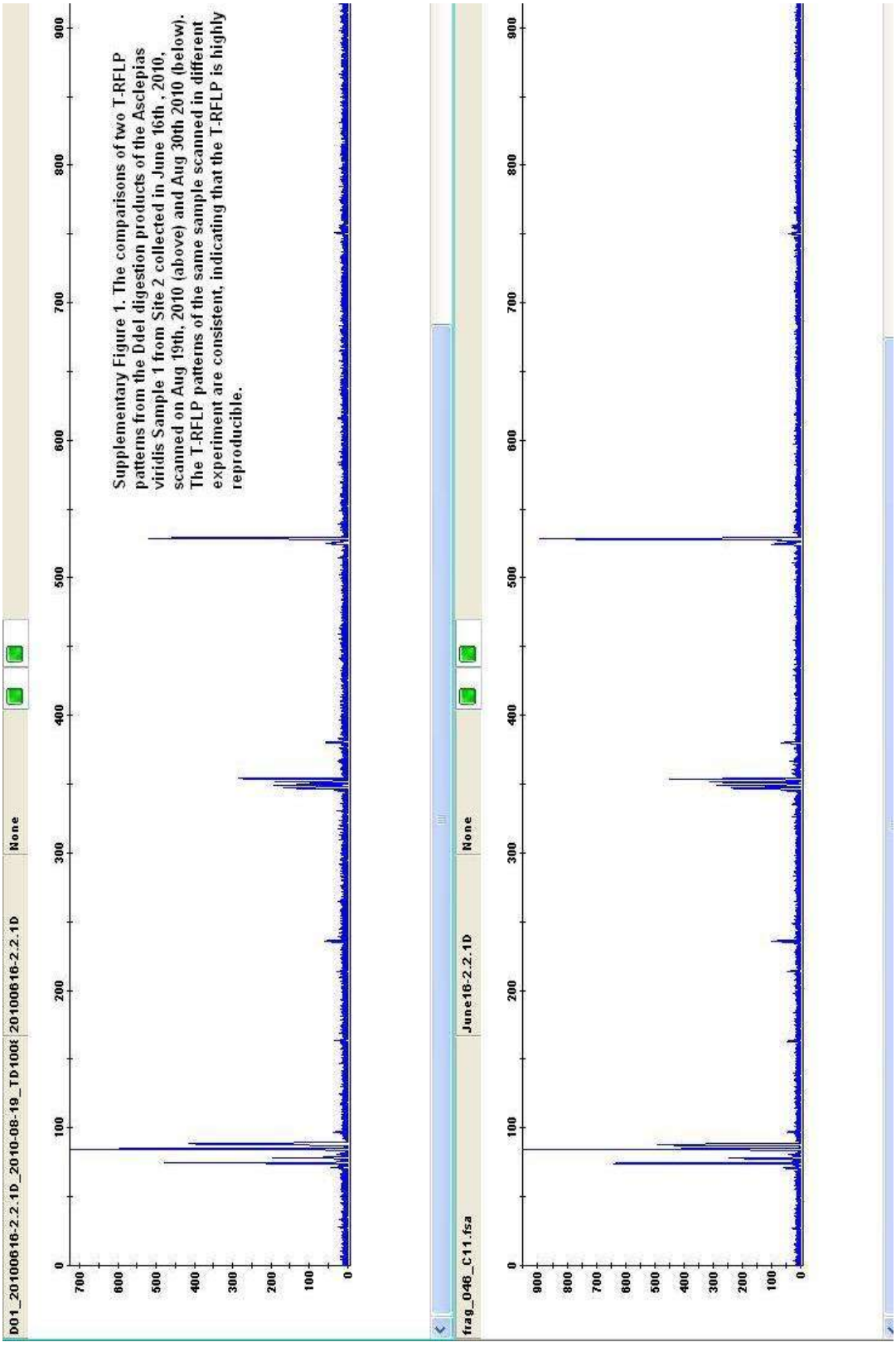

355bp
529bp
78bp
350bp
236bp
352bp
79bp
75bp
85bp
P. virgatum
S. nutans
A. psilostachya
A. viridis
R. humilis
1
0.5
0

Figure 8.



Supplementary Figure 1. The comparisons of two T-RFLP patterns from the DdeI digestion products of the Asclepias viridis Sample 1 from Site 2 collected in June 16th , 2010, scanned on Aug 19th, 2010 (above) and Aug 30th 2010 (below). The T-RFLP patterns of the same sample scanned in different experiment are consistent, indicating that the T-RFLP is highly reproducible.

Figure 9.

CHAPTER III

DISCOVERY OF THE DIVERSITY OF LEAF ENDOPHYTIC BACTERIAL COMMUNITIES USING HIGH-THROUGHPUT PYROSEQUENCING

Abstract

To reveal the compositions of leaf endophytic bacterial communities in non-cultivated plants and to evaluate environmental influence on the bacterial communities, plant samples were collected from the Tallgrass Prairie Preserve, Oklahoma, and total DNA was extracted from surface sterilized plants and later fragments of bacterial 16S rDNA were amplified by PCR. Tagged 454 pyrosequencing of amplicons gave 64,952 sequencing reads, which defined 3,229 OTUs at 97% sequence similarities and represented bacterial groups from 16 phyla. *Proteobacteria* was the dominant phylum in the communities, followed by the phyla *Bacteriodetes* and *Actinobacteria*. Bacteria from four classes of *Proteobacteria* were detected with *Alphaproteobacteria* as the dominant class. Host plant species and collecting date had significant influences on the compositions of the leaf endophytic bacterial communities: The proportion of *Alphateobacteria* was much higher in the communities from *Asclepias viridis* than from other plant species and expanded greatly in July. The most dominant bacterial groups,

represented by OTUs, showed host-specific patterns, indicating mutual selection between host plants and endophytic bacteria. Leaf endophytic bacterial compositions were dynamic, varying with the host plant growing season via three main types of trend. In summary, pyrosequencing has proved a powerful tool to illustrate the compositions, diversities and variations of the leaf endophytic bacterial communities.

Introduction

Although current microbiological research activities mainly focus on pathogenic plant-associated bacteria that are highly relevant to food safety, agricultural production and bioterrorism prevention, another type of plant-associated bacteria – endophytic bacteria has become a hot spot of microbiological research (18, 25, 26). Endophytic bacteria are harbored inside healthy plant tissues but do not lead to pathogenic reactions (13), and play important roles in phytoremediation (1, 11, 21, 22), biological control against insects or pathogenic microorganisms (10, 26), and plant growth promotion (2, 5, 17). Endophytic bacteria may also be pathogenic to other plants, animals, especially cattle, and human beings (7, 12, 15). Although endophytic bacteria are important in the entire ecosphere, their diversity and the mechanism of their interaction with host plants are not well known. Root endophytic bacteria together with rhizosphere bacteria contribute greatly to plant nutrient intake and make up a major part of soil bacterial diversity (13, 29). As a result many researches have addressed root endophytic bacteria especially in cultivated plants. By contrast, more remains unknown to reveal about leaf endophytic bacteria. We applied Terminal Restriction Fragment Length Polymorphism (T-RFLP) to study leaf endophytic bacterial communities in the previous chapter. We collected plant

samples from multiple host plant species at different sites in four consecutive months during the whole growing seasons, and statistical analysis of the T-RFLP data allowed us to study the influences of environmental factors quantitatively. We found that three environmental factors including host plant species, collecting date and sampling sites all have significant impacts on the leaf endophytic bacterial communities, and the T-RFLP patterns also identified some dominant T-RFs which represent specific bacterial groups. To understand which bacterial groups those T-RFs represent, in this research, high-throughput tagged 454 pyrosequencing (20) was applied to reveal the diversity of leaf endophytic bacterial communities using 16S rDNA sequence information. Specifically we sought answers to the following questions: what the pyrosequencing of the 16S rDNA fragments would tell us about the diversity of the leaf endophytic bacterial communities; what the 16S rDNA sequence information reveals about the environmental impacts on the leaf endophytic bacterial communities; and how the information helps to interpret the T-RFLP data.

Methods and Materials

Plant sampling & DNA extraction.

Leaf samples of five species *Ambrosia psilostachaya*, *Asclepias viridis*, *Panicum virgatum*, *Sorghastrum nutans* and *Ruellia humilis* in the Tallgrass Prairie Preserve, Osage County, Oklahoma were collected at four different sites in each month from May to August, 2010 as described in the previous chapter. To eliminate a contribution of epiphytic bacteria to the analysis of leaf endophytic bacteria, the surface of plant samples were rinsed well first with tap water and then with 70% ethanol. The efficacy of this

treatment was validated by plating. Total DNA (including plant and bacterial DNA) was extracted from ground leaf material as previously described in chapter 2.

PCR for sequencing.

For DNA from each plant sample, a PCR was conducted using a pair of primers which incorporated a unique barcode (14) so that the sequences of the PCR amplicons could be assigned to the correct samples based on the barcodes. To avoid PCR amplification of plant plastid DNA, which is highly similar to bacterial 16S rDNA, a pair of primers, 799F and 1492R (6), was used to amplify the bacterial 16S rDNA. The resulting amplicon fragments covered several variable regions of 16S rDNA including V5 to V8 and part of V4 (23). The forward primers (5'-CGTATCGCCTCCCTCGCGCCATCAGXXXXXXXXCAAACMGGATTAGATACCCKG-3') contained the 454 sequencing primer A, the barcode (14) represented by XXXXXXXXX, a 2-base linker sequence CA, and the primer 799F. The reverse primer (5'- CTATGCGCCTTGCCAGCCCGCTCAGTCGGCTACCTTGTTACGACTT-3') contained the 454 sequencing primer B, a 2-base linker sequence TC, and the primer 1492R. PCR products were separated by electrophoresis in 2 % agarose to remove amplicons of plant mitochondrial DNA. The bacterial amplicon DNA was purified from the gel using Qiaquick Gel Purification Kit (Qiagen, Hilden, Germany).

Sequencing technique 454: library preparation

Purified DNA was subjected to pyrosequencing (20). The length of amplicons was estimated by electrophoresis and the concentration was estimated by Quant-iT Picogreen dsDNA Reagent and Kits (Invitrogen, Carlsbad CA) using a spectrofluorometer. An

equimolar mix of amplicon libraries from all plant samples was made and prepared for pyrosequencing by emulsion bead PCR using the recommended Lib-A kit and protocol (454 Life Sciences, Branford CT). Pyrosequencing on the Genome Sequencer Junior Titanium Series (Roche, Indianapolis IN) was conducted at the Oklahoma State University Recombinant DNA/Protein Core Facility.

Sequence analysis - QIIME and Mothur and Virtual Restriction .

Sequencing data were processed using a combination of Mothur (28) and Quantitatve Insights Into Microbial Ecology (QIIME) toolkit software (4). Bacterial sequences were quality trimmed and those sequences >80 bp in length with an average quality score >25 and no ambiguous bases were included in the analyses. OTUs were defined from the bacterial sequences using a 97% sequence similarity threshold, and a representative sequence was picked for each OTU by selecting the most abundant sequence from that OTU. Taxonomy assignment was done by using BLAST for each OTU (representative sequence) against the Silva database (24). Alpha diversity, beta diversity and unweighted UniFrac (19) were calculated for the comparisons of the compositions of all communities. Virtual restriction of sequencing reads was done in Perl.

Results

Pyrosequencing summary: OTU and Taxonomy

From 81 plant samples, we obtained 64952 sequences for further analyses after the quality filter described above with an average of 797 sequences for each of the 81 plant samples (Table 1). A species-level bacterial phylotype is often defined as an OTU whose

members share 97% or greater identity in their 16S rDNA sequence (27). Among the 64,952 sequences representing the overall leaf endophytic bacterial communities, we distinguished 3,229 distinct OTUs at 97% sequence similarity cutoff ($OTU_{0.03}$). The average number of $OTU_{0.03}$ per sample was 122. $OTU_{0.03}$ were assigned to 16 phyla (Figure 1) with only a minor fraction (298 $OTU_{0.03}$, 7906 sequences) that could not be classified. Proteobacteria was the most abundant phylum and 61.38% (1982/3229) of the $OTU_{0.03}$ belonged to it, followed by Bacteriodetes with 393 $OTU_{0.03}$ and Actinobacteria with 331 $OTU_{0.03}$. Representatives of four classes were found within Proteobacteria: *Alphaproteobacteria*, 1123 $OTU_{0.03}$; *Gammaproteobacteria*, 339 $OTU_{0.03}$; *Betaproteobacteria*, 267 $OTU_{0.03}$ and *Deltaproteobacteria*, 59 $OTU_{0.03}$.

Among the 14 most common $OTU_{0.03}$, each of which represented more than 1000 sequences, all but one was from Proteobacteria with the only exception being from Bacteriodetes (Table 2). *Pseudomonas*, *Sphingomonas* and *Methylobacterium* were the most abundant genera.

By defining the OTUs from the complete sequencing data at different levels of sequence identity, we constructed rarefaction curves to monitor the bacterial diversity at different taxonomy levels (Figure 3). Similar to an OTU defined at 97% 16S rDNA sequence identity representing species level differences, it is commonly accepted that OTUs defined at 94%, 92% and 90% sequence identity represented genus, family and order level, respectively (27). In the comparison of the four OTU levels of the entire sequence dataset (Figure 3a), our sequencing effort revealed probably most of the orders/genus/families present since their rarefaction curves were approaching an asymptote.

Comparing the number of sequences represented by each $OTU_{0.03}$ (Figure 7), we found that only a few OTUs represented a large number of sequences while most OTUs represented only a few sequences. Totally 14 $OTU_{0.03}$ have more than 1000 sequences, representing those dominant bacteria; 2863 $OTU_{0.03,}$ which is 88.70% of all defined $OTU_{0.03}$, have fewer than 10 sequences. This result indicated that the basic structure for leaf endophytic bacterial communities is that of a few dominant bacteria with a large number of low occurrence bacteria.

All the representative sequences of each $OTU_{0.03}$ were compared to the reference Silva database to calculate the similarity indices of those OTUs to their best hits (Figure 8). Of all 3228 $OTU_{0.03}$, the similarity indices of 1323 OTUs to their best hits are lower than 97%, which means that by the definition of $OTU_{0.03}$ these representative sequences are novel sequences and these $OTU_{0.03}$ represent novel bacterial species. Even when the threshold was lowered to 90%, still 655 $OTU_{0.03}$ were regarded as novel sequences which may represent novel bacterial species that have not been described before.

Environmental influences: communities in different host plants

We collected leaf endophytic bacterial communities from diverse environments: host plant of different species, four consecutive collecting months and four different sites. Studying the sequencing data as a function of those environmental factors can help us to understand the environmental influences on leaf endophytic bacterial communities.

In the five plant species we examined, samples from *P. virgatum* had the most $OTU_{0.03}$, followed by *A. viridis*, and then *S. nutans*, *R. humilis* and *A. psilostachya* (Table 3). The broad range of the $OTU_{0.03}$ numbers in different plant species indicated that the features

of the host plant have great impacts on the diversity of the endophytic bacteria that it harbors. In the comparison, mentioned above, of the fourteen most abundant $OTU_{0.03}$ in the five plant species (Figure 2a), some $OTU_{0.03}$ including two *Sphingomonas* $OTU_{0.03}$ (1327 & 2021), all three *Methylobacterium* $OTU_{0.03}$ (2857, 3184 & 1252) and *Hymenobacter* (2934) are significantly more abundant in *A. viridis* than in other plant species; especially the OTU2021- *Sphingomonas* seems to be *A. viridis*-specific. OTU2245-*Pseudomonas* is significantly more abundant in *R. humilis*. OTU1486-*Pantoea* is significantly more abundant in *P. virgatum* and *S. nutans*. This result suggested that some bacterial groups play important roles in specific host plants.

The diversities of different bacterial communities can also be compared using rarefaction curves (Figure 3). We compared the rarefaction curves of the $OTU_{0.03}$ of the samples collected from different plant species, months or sites. The leaf endophytic bacterial communities harbored in *P. virgatum* showed a higher diversity than those in other plant species because at any given number of sequences retrieved, more $OTU_{0.03}$ were defined in *P. virgatum* than other plant species; bacterial communities harbored in *A. psilostachya* showed the lowest diversity (Figure 3b). This result is consistent with the comparison of the $OTU_{0.03}$ counts (Table 3).

Besides the dominant $OTU_{0.03}$ comparisons and the rarefaction curves, the number of the sequences represented by diverse taxonomic units also gives us an insight into the environmental influences (Figure 4). Comparing the samples from five host plant species, the abundance of *Alphaproteobacteria* is higher in *A. viridis* than in the other four plant species; by contrast, the proportion of *Betaproteobacteria* and *Gammaproteobacteria* is

lower in *A. viridis* than in other plant species. We also noticed that the abundance of *Actinobacteria* is lower in *A. psilostachya* and *A. viridis* than in other plant species.

Environmental influences: the dynamics during the host plant growing season

Similar to the comparison above, sequence information from samples collected in four consecutive months during the host plant growing season allowed us to study the dynamics of leaf endophytic bacteria. In the four consecutive months that we collected samples, the number of $OTU_{0.03}$ decreased from May to August, indicating that the diversity of leaf endophytic bacterial communities decreased during the whole growing season (Table 3). In the comparison of the most abundant $OTU_{0.03}$ mentioned above in the four consecutive months (Figure 2b), three main trend types were observed: Type I: the abundance of some bacterial species-level phylotypes including $OTU_{0.03}$ 980, 1327, 2476, 2857, 3184, 1252, 2573 and 1724, increased from May to July then fell to August; Type II: the abundance of some bacterial phylotypes including $OTU_{0.03}$ 2245 and 2934 decreased from May to August during the whole blooming season; Type III: the abundance of some bacterial phylotypes including $OTU_{0.03}$ 2528, 1486 and 1333 increased from May to June, then fell to a relatively low number, and finally rose to the highest number in the whole growing season. This result indicated that the dynamics of leaf endophytic bacteria are related to the plant growing activity in multiple ways. One exception is the OTU2021-Sphingomonas, which almost only appeared in the samples collected in May, and was seldom observed in samples collected later.

Comparing the samples collected in different months by rarefaction curves (Figure 3c), the samples collected in May showed the highest diversity while the July samples showed the lowest counts. This differs from what we concluded by comparing the $OTU_{0.03}$

numbers above (Table 3), because we obtained more sequences from the samples collected in July, which lead to more defined $OTU_{0.03}$. In similar comparisons regarding different sites, no obvious difference was observed (Figure 3d), indicating that, unlike plant species or collecting date, the collecting site had no significant impact on the diversity of the leaf endophytic bacterial communities.

The comparison of the number of sequences represented by each $OTU_{0.03}$ in four consecutive months showed that the abundance of *Alphaproteobacteria* expanded greatly in July to consolidate its dominant position but then fell in August (Figure 5).

Interpretations of T-RFLP based on Sequencing

The previous study of leaf endophytic bacterial communities using single digestion T-RFLP with *Dde*I (9), focused on quantification of environmental influences and discovery of the prominent T-RFs. The sequences obtained here allow virtual restriction digestion of the reads, providing a link between the T-RFLP results and the sequencing. In the 64,951 total sequencing reads, *Dde*I restriction sites were found in 56,438 reads or 86.89% of the overall sequences. This indicates that single restriction T-RFLP using *Dde*I can characterize all but a few members of leaf endophytic bacterial communities. Thus, it is appropriate to identify the probable bacterial groups corresponding to the prominent T-RFs in the previous study. Virtual *Dde*I digestion of all the sequencing reads from which the barcode-linked primer sequences had been trimmed was performed. The resulting virtual terminal restriction fragments (VT-RFs) were compared with the observed T-RFs (Figure 6). In T-RFLP we used the fluorescently labeled forward primer 799F-TA-6FAM (6FAM -5'-AAT AAA TCA TAA AAC MGG ATT AGA TAC CCK G) and all primer sequences had been trimmed from the sequences prior to virtual restriction; so the

difference in lengths between VT-RFs and T-RFs can be predicted as the length of the primer 799F-TA-6FAM in nucleotide unit equivalents. The molecular weight of 6FAM at the 5' end of DNA is 537.5 Dalton or between the equivalents of 1 to 2 nt. Thus, the length in nucleotide units of 799F-TA-6FAM is approximately 32 to 33 nt and corresponding pairs of T-RFs and VT-RFs should differ by that amount. Slight deviations from this expectation are possible due to dependence of the actual mass on base pair composition and slight variations, independent of length, of the migration distance of individual T-RFs in capillary electrophoresis. The latter is exemplified by the need to do manual alignment adjustment of tRF peaks as explained in Chapter 2. In the comparison of the T-RFs with the VT-RFs for sample Am05 (Figure 6), three groups of RFs were observed in both T-RFLP and VT-RFs. Group 1 included the peaks of 80bp, 85bp and 94bp in T-RFLP, corresponding to the peaks of 49nt, 50nt and 55nt in virtual restriction. Group 2 included the peaks of 345bp and 355 bp, corresponding to the peaks of 319nt, 320nt, 322nt and 325nt in virtual restriction. Group 3 consists of the peak of 529bp in T-RFLP and its corresponding peaks around the center 497nt in virtual restriction. The difference of length between T-RF and corresponding VT-RF is a little bit shorter in the smaller fragments than larger ones with a narrow range from 29 to 32 nt. The virtual restriction of the complete OTU representative sequences gave us a full list of the length of VT-RFs of each OTU. Since multiple OTUs can generate VT-RFs of the same length, one T-RF may refer to diverse sources of bacteria; we can find the most probable source of bacteria through the survey of all OTUs which generate the T-RF (Table 5). The results showed that the most frequent T-RFs found in chapter 2 were probably from

Sphingomonas and *Methylobacterium*, which is consistent with the dominant OTU taxonomy result above.

Discussion

A few of next generation sequencing methods have been successfully commercialized including 454 pyrosequencing (20), Illumina sequencing (3) and SOLiD platform (30). Although the 454 pyrosequencing can only generate 0.45 Giga Bases per run, which is much smaller than 18 Giga Bases of Illumina and 30 Giga Bases of SOLiD, it can generate longer reads (500 bases) than Illumina (75-100 bases) and SOLiD (50 bases). This feature is important in 16S rDNA research: a fragment of 500 bases covers several important variable regions in 16S rDNA containing large phylogenetic information, and can be directly used for sequence analysis. So we can avoid the assembly challenge by using 454 pyrosequencing. Short reads assembly would be needed if using Illumina sequencing or SOLiD platform and the highly conserved sequences of 16S rDNA would lead to many artificial DNA recombinants, resulting in overestimation of the leaf endophytic bacterial communities.

The pyrosequencing of the whole leaf endophytic bacterial communities gave us 64561 sequences, which defined 3229 OTUs at 97% sequence identity. These 3229 $OTU_{0.03}$ represent 16 phyla, 31 classes, 49 orders, 109 families and 222 genera. The rarefaction curves representing OTUs defined as 94% or above sequence identity approached the asymptote (Figure 3a), indicating that the major diversity of leaf endophytic bacterial communities has been explored. By survey of the most abundant OTUs we were able to

recognize the most dominant bacterial species; although specific bacterial species have dramatic variations in different host species or in different months (Figure 3), no sharp change was observed at higher levels of taxonomy and the whole leaf endophytic bacterial communities remained roughly stable.

In the leaf endophytic bacterial communities, evidence of 16 phyla was discovered with *Proteobacteria*, *Bacteriodetes* and *Actinobacteria* as the most dominant. These phyla especially *Bacteriodetes* are the most dominant in soil, showing the close relation between leaf endophytic bacteria and rhizosphere bacteria. As the major decomposer of chitin, polymeric carbon and other organic detritus, *Bacteriodetes* and *Actinobacteria* probably originated from cattle faeces or soil, which invade the host plants from the phyllosphere. In *Proteobacteria*, four classes were found and *Alphaproteobacteria* was the most dominant. *Alphaproteobacteria* comprises most phototrophic genera such as *Rhodobacterales* and *Rhodospirillales*, so sunshine might be the reason that this class dominates leaf endophytic bacterial community. Under *Alphaproteobacteria* some genera of special interest were found, including *Rhizobium*, which is an important genus of soil bacteria that fixes nitrogen, and *Rickettsia*, which spends most of its life in insects and causes a large range of human and plant diseases. *Rickettsia* is one of the closet living relatives to the bacteria that originated the mitochondria organelle in eukaryotes. Janssen summarized that *Proteobacteria*, *Actinobacteria* and *Acidobacteria* are the most dominant phyla in soil (16), and Costello *et al* (8) has revealed that *Actinobacteria* and *Firmicutes* are the most dominant phyla in human body habitats; both showed the bacterial communities adapted to their harboring environment. Similarly the leaf

endophytic bacterial communities adapted highly to the leaf environment and showed close relationships with the soil bacterial communities.

In the previous research we discussed the environmental influences on endophytic bacterial communities by applying partial Canonical Correlation Analysis (pCCA) on T-RFLP data, and found that all three environmental factors including host plant species, collecting date and sampling sites have significant impacts ($p<0.05$). The sequencing data reveals more about how the environment shaped the leaf endophytic bacterial communities. The rarefaction curves allow evaluation of diversity level by how many $OTU_{0.03}$ are defined within a certain number of sequencing reads. The comparison of rarefaction curves of bacterial communities harbored in different host plants (Figure 3b) shows that the samples from *P. virgatum* harbored a higher diversity of leaf endophytic bacteria than other plants. The compositions of the leaf endophytic bacterial communities harbored in different host plant species differed at both higher and lower taxonomic levels. Although the phylum of *Proteobacteria* is dominant in all five host plant species, the abundance of the class of *Alphaproteobacteria* in *A. viridis* is significantly higher than in other plant species (Figure 4a). At a lower level of taxonomy, in the comparisons of the 14 most dominant species ($OTU_{0.03}$), some species from the genus *Sphingomonas* and *Methylobacterium* were more abundant in *A. viridis*, especially the OTU2021, which is almost specifically harbored in *A. viridis.* Since some *Sphingomonads* have special biodegradative and biosynthetic features and *Methylobacterium* can use methanol as well as C2, C3 and C4 compounds to grow, the special milky liquid from *A. viridis* (common name as milkweed) might be the reason for the bacteria to prefer *A. viridis* as habitats. Similarly, collecting months also showed great

impacts on the leaf endophytic bacterial communities and since this kind of variation is over time it can be regarded as the dynamics of leaf endophytic bacteria. From the rarefaction curve (Figure 3c) and the number of $OTU_{0.03}$ in different months (Table 3), the samples of May harbored the highest diversity of leaf endophytic bacterial communities. Over the whole growing season of host plants, the diversity of endophytic bacterial communities had a decreasing trend. This phenomenon may be related to the growth condition of host plants: as the host plants keep growing during the season their metabolic and physiological features develop gradually, so that the specific leaf endophytic bacterial groups which the host plants preferred, as we discussed above, would be selected to remain and propagate, leading to the diversity reduction of the whole bacterial communities. The dynamics of the bacterial communities can also be evaluated by the change of the composition at different taxonomic levels. The class of *Alphaproteobacteria* had a large expansion in July (Figure 4b), which might be related to the strong sunshine and high temperature since this class comprises most phototrophic bacteria. At the species level, we also noticed that the abundance of those dominant $OTU_{0.03}$ varied in different months (Figure 2b). As discussed above, three types of main trends (I, II and III) were observed. Considering most host plants are growing to their best condition in June or July when the temperature or humidity are the best, bacteria of Type I are positively selected by the host plants and bacteria of Type II are negatively selected. As for bacteria of Type III, probably extreme weather condition especially extreme hot temperature was the major limiting factor for their growth.

Admittedly, the analysis of pyrosequencing data required much higher computing capacity than T-RFLP did, and the results are more difficult to visualize and convey.

However, compared to T-RFLP in which we detected 122 T-RFs (which we may treat as Operational Taxonomic Units), pyrosequencing has a higher resolution to elucidate leaf endophytic bacterial communities, so the diversity revealed would be closer to its reality in nature. T-RFLP showed us the abundance of those dominant bacterial groups. However from each sequencing read we can recognize those minor bacterial groups even when they were represented by just one sequencing read. Leading to easily readable data visualization and high mass of information, a combination of these two tools would be greatly useful.

Acknowledgements

Authors acknowledge the support of the Oklahoma Agricultural Experiment Station, whose Director has approved this publication, the R. J. Sirny Professorship at Oklahoma State University and the National Science Foundation through EPS-0447262. They thank Mostafa Elshahed for critical readings of the manuscript and Jimmy Davis for suggesting additional sequencing analyses.

LITERATURE CITED

1. **Barac, T., S. Taghavi, B. Borremans, A. Provoost, L. Oeyen, J. V. Colpaert, J. Vangronsveld, and D. van der Lelie.** 2004. Engineered endophytic bacteria improve phytoremediation of water-soluble, volatile, organic pollutants. Nature Biotechnology **22:**583-588.
2. **Bent, E, Chanway, and P. C.** 1998. The growth-promoting effects of a bacterial endophyte on lodgepole pine are partially inhibited by the presence of other rhizobacteria, vol. 44. National Research Council of Canada, Ottawa, ON, CANADA.
3. **Bentley, D. R.** 2006. Whole-genome re-sequencing. Current Opinion in Genetics & Development **16:**545-552.
4. **Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth **7:**335-336.
5. **Chanway, C. P.** 1997. Inoculation of tree roots with plant growth promoting soil bacteria: An emerging technology for reforestation. Forest Science **43:**99-112.
6. **Chelius, M. K., and E. W. Triplett.** 2001. The Diversity of archaea and bacteria in association with the roots of Zea mays L. . Microbial Ecology **41:**252-263.

7. **Cooley, M. B., W. G. Miller, and R. E. Mandrell.** 2003. Colonization of Arabidopsis thaliana with Salmonella enterica and Enterohemorrhagic Escherichia coli O157:H7 and Competition by Enterobacter asburiae. Applied and Environmental Microbiology **69:**4915-4926.

8. **Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight.** 2009. Bacterial community variation in human body habitats across space and time. Science **326:**1694-1697.

9. **Ding, T., U. Melcher, and M. Palmer.** 2012. Community Terminal Restriction Fragment Length Polymorphisms Reveal insights into the Diversity and Dynamics of Leaf Endophytic Bacteria, p. 28, Community Terminal Restriction Fragment Length Polymorphisms Reveal insights into the Diversity and Dynamics of Leaf Endophytic Bacteria, Submitted.

10. **Duijff, B. J., V. Gianinazzi-Pearson, and P. Lemanceau.** 1997. Involvement of the outer membrane lipopolysaccharides in the endophytic colonization of tomato roots by biocontrol Pseudomonas fluorescens strain WCS417r. New Phytologist **135:**325-334.

11. **Germaine, K. J., X. Liu, G. G. Cabellos, J. P. Hogan, D. Ryan, and D. N. Dowling.** 2006. Bacterial endophyte-enhanced phytoremediation of the organochlorine herbicide 2,4-dichlorophenoxyacetic acid. FEMS Microbiology Ecology **57:**302-310.

12. **Guo, X., M. W. van Iersel, J. Chen, R. E. Brackett, and L. R. Beuchat.** 2002. Evidence of association of salmonellae with tomato plants grown hydroponically

in inoculated nutrient solution. Applied and Environmental Microbiology **68:**3639-3643.

13. **Hallmann, J., A. QuadtHallmann, W. F. Mahaffee, and J. W. Kloepper.** 1997. Bacterial endophytes in agricultural crops. Canadian Journal of Microbiology **43:**895-914.

14. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Meth **5:**235-237.

15. **Ingham, S. C., M. A. Fanslau, R. A. Engel, J. R. Breuer, J. E. Breuer, T. H. Wright, J. K. Reith-Rozelle, and J. Zhu.** 2005. Evaluation of fertilization-to-planting and fertilization-to-harvest intervals for safe use of noncomposted bovine manure in wisconsin vegetable production. Journal of Food Protection **68:**1134-1142.

16. **Janssen, P. H.** 2006. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. Applied and Environmental Microbiology **72:**1719-1728.

17. **Lee, S., M. Flores-Encarnacion, M. Contreras-Zentella, L. Garcia-Flores, J. E. Escamilla, and C. Kennedy.** 2004. Indole-3-Acetic Acid Biosynthesis Is Deficient in Gluconacetobacter diazotrophicus Strains with Mutations in Cytochrome c Biogenesis Genes. Journal of Bacteriology **186:**5384-5391.

18. **Lodewyckx, C., J. Vangronsveld, F. Porteous, E. R. B. Moore, S. Taghavi, M. Mezgeay, and D. v. der Lelie.** 2002. Endophytic Bacteria and Their Potential Applications. Critical Reviews in Plant Sciences **21:**583-606.

19. **Lozupone, C., and R. Knight.** 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Applied and Environmental Microbiology **71:**8228-8235.

20. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.

21. **Moore, F. P., T. Barac, B. Borremans, L. Oeyen, J. Vangronsveld, D. van der Lelie, C. D. Campbell, and E. R. B. Moore.** 2006. Endophytic bacterial diversity in poplar trees growing on a BTEX-contaminated site: The characterisation of isolates with potential to enhance phytoremediation. Systematic and Applied Microbiology **29:**539-556.

22. **Newman, L. A., and C. M. Reynolds.** 2005. Bacteria and phytoremediation: new uses for endophytic bacteria in plants. Trends in Biotechnology **23:**6-8.

23. **Petrosino, J. F., S. Highlander, R. A. Luna, R. A. Gibbs, and J. Versalovic.** 2009. Metagenomic Pyrosequencing and Microbial Identification. Clinical Chemistry **55:**856-866.

24. **Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research **35:**7188-7196.

25. **Rosenblueth, M., and E. Martinez-Romero.** 2006. Bacterial Endophytes and Their Interactions with Hosts. Molecular Plant-Microbe Interactions **19:**827-837.

26. **Ryan, R. P., K. Germaine, A. Franks, D. J. Ryan, and D. N. Dowling.** 2008. Bacterial endophytes: recent developments and applications. FEMS Microbiology Letters **278:**1-9.

27. **Schloss, P. D., and J. Handelsman.** 2004. Status of the Microbial Census. Microbiology and Molecular Biology Reviews **68:**686-691.

28. **Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber.** 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology **75:**7537-7541.

29. **Siciliano, S. D., N. Fortin, A. Mihoc, G. Wisse, S. Labelle, D. Beaumier, D. Ouellette, R. Roy, L. G. Whyte, M. K. Banks, P. Schwab, K. Lee, and C. W. Greer.** 2001. Selection of Specific Endophytic Bacterial Genotypes by Plants in

Response to Soil Contamination. Applied and Environmental Microbiology **67:**2469-2475.

30. **Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J. A. Malek, G. Costa, K. McKernan, A. Sidow, A. Fire, and S. M. Johnson.** 2008. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Research **18:**1051-1063.

Figure Captions

Figure 1. Bacterial family distribution of $OTU_{0.03}$ recovered from endophytic bacteria of the TGPP. 3229 $OTU_{0.03}$ were assigned into 16 phyla with 298 $OTU_{0.03}$ cannot be classified. *Proteobacteria* is the most dominant phylum, and *Alphaproteobacteria* is the most dominant class.

Figure 2. Frequencies of detection of the 14 most abundant $OTU_{0.03}$ from TPP endophytic bacteria as a function of a) plant source and b) month of sampling

Figure 3. Rarefaction curves plotting the number of recovered $OTU_{0.03}$ as a function of the number of sequences obtained a) for different taxonomic levels, b) host plant species, c) month of sampling and d) site of sampling. Rarefaction curves were calculated by using Mothur.

Figure 4. Distribution of $OTU_{0.03}$ (percentage of total sequences) of TPP endophytic bacteria by family as a function of host plant. Not all taxa listed are visible due to their underrepresentation in the dataset.

Figure 5. Distribution of $OTU_{0.03}$ (percentage of total sequences) of TPP endophytic bacteria by family as a function month of sampling. Not all taxa listed are visible due to their underrepresentation in the dataset.

Figure 6. The virtual restriction digestions of the sequences from Sample Am05 compared to the T-RFLP pattern from the same sample. Arrows indicate correspondence between of T-RF and VT-RF pairs.

Figure 7. The number of sequences represented by each defined $OTU_{0.03.}$

Figure 8. The similarity indices of all defined $OTU_{0.03}$ to their best hits in the reference silva database.

Table 1. Summary description of sequencing effort: the number of sequences collected, the sequencing quality, and the levels of bacterial diversity discovered.

No. of plant samples	81
Total no. of classifiable sequences	64561
Average length of sequence reads, bp (range)	536 (200-644)
Total no. of $OTU_{0.03}$ across all samples	3229
Average no. of sequences per sample	797
Average no. of $OTU_{0.03}$ per sample	122

Table 2. List of the most abundant $OTU_{0.03}$. The most abundant $OTU_{0.03}$ here are defined as those $OTU_{0.03}$ represent more than 1000 sequences.

OTU ID	Class	Family	Genus	No. of Sequences	No. of Samples
980	Gammaproteobacteria	Pseudomonadaceae	Pseudomonas	4456	46 (56.79%)
1327	Alphaproteobacteria	Sphingomonadaceae	Sphingomonas	3293	68 (83.95%)
2476	Alphaproteobacteria	Sphingomonadaceae	Sphingomonas	3154	64 (79.01%)
2245	Gammaproteobacteria	Pseudomonadaceae	Pseudomonas	3125	66 (81.48%)
2857	Alphaproteobacteria	Methylobacteriaceae	Methylobacterium	2736	65 (80.25%)
3184	Alphaproteobacteria	Methylobacteriaceae	Methylobacterium	2569	67 (82.72%)
2528	Betaproteobacteria	Burkholderiales_incertae_sedis	Aquabacterium	2360	64 (79.01%)
1486	Gammaproteobacteria	Enterobacteriaceae	Pantoea	2086	45 (55.56%)
2934	Sphingobacteria	Cytophagaceae	Hymenobacter	1462	63 (77.78%)
1252	Alphaproteobacteria	Methylobacteriaceae	Methylobacterium	1406	60 (74.07%)
1333	Alphaproteobacteria	Caulobacteraceae	Caulobacter	1311	56 (69.14%)
2573	Alphaproteobacteria	Rhizobiaceae	Rhizobium	1285	44 (54.32%)
1724	Gammaproteobacteria	Xanthomonadaceae		1211	43 (53.09%)
2021	Alphaproteobacteria	Sphingomonadaceae	Sphingomonas	1061	12 (14.81%)

Table 3. The number of $OTU_{0.03}$ defined in the samples from five plant species and from four consecutive months.

In different plant host species					In different collecting months			
Ambrosia	Asclepias	Panicum	Sorghastrum	Ruellia	May	June	July	August
607	1140	1662	950	785	1299	1299	1291	1130

Table 4. The most frequent T-RFs and the most probable source of bacterial group responding to those T-RFs.

T-RFs (bp)	Virtual Digestion Fragments (bp)	The Most Probable Source
529	497	*Proteobacteria*/*Alphaproteobacteria*/*Sphingomonadales*/*Sphingomonadaceae*/*Sphingomonas*
350	320	*Proteobacteria*/*Alphaproteobacteria*/*Rhizobiales*/*Methylobacteriaceae*/*Methylobacterium*
355	325	*Proteobacteria*/*Alphaproteobacteria*/*Sphingomonadales*/*Sphingomonadaceae*
85	55	*Bacteroidetes*/*Sphingobacteria*/*Sphingobacteriales*/*Chitinophagaceae* or *Cytophagaceae* or *Sphingobacteriaceae*

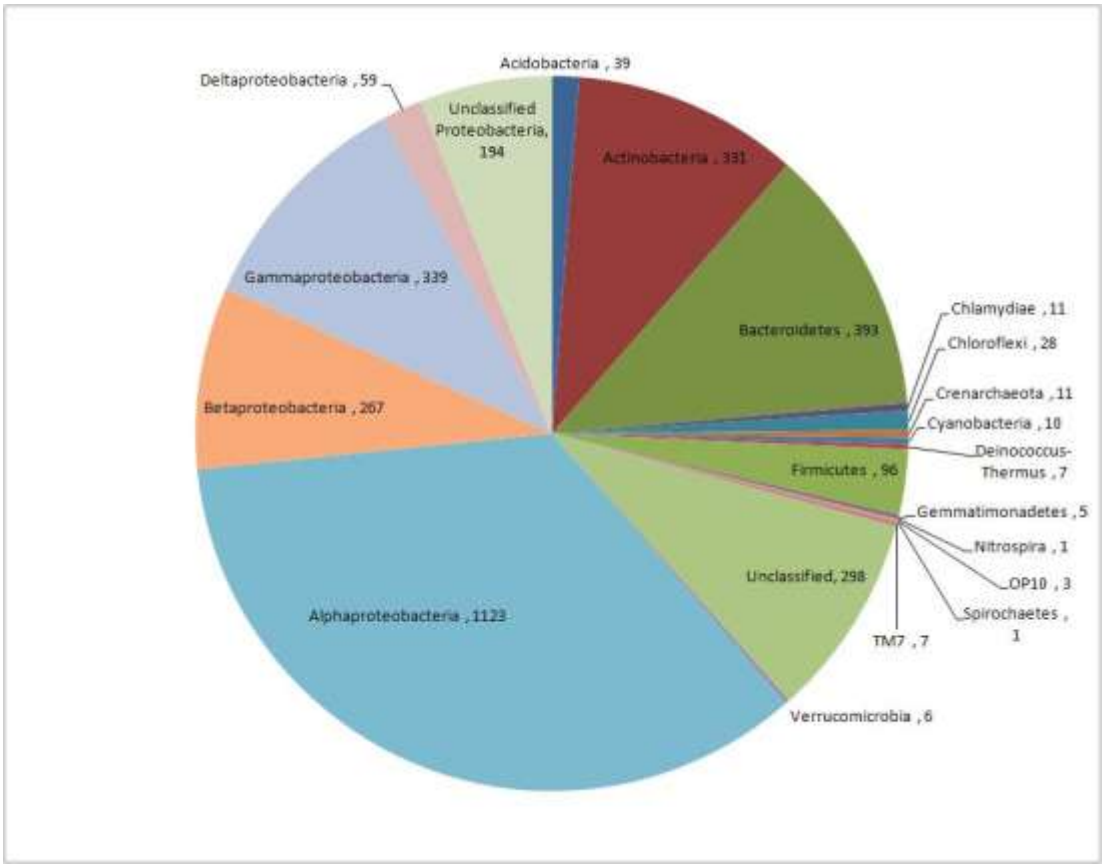

Acidobacteria , 39
Deltaproteobacteria , 59
Unclassified Proteobacteria, 194
Actinobacteria , 331
Gammaproteobacteria , 339
Bacteroidetes , 393
Chlamydiae , 11
Chloroflexi , 28
Crenarchaeota , 11
Betaproteobacteria , 267
Cyanobacteria , 10
Deinococcus-Thermus , 7
Firmicutes , 96
Gemmatimonadetes , 5
Nitrospira , 1
Unclassified, 298
OP10 , 3
Alphaproteobacteria , 1123
Spirochaetes , 1
TM7 , 7
Verrucomicrobia , 6

Figure 1.

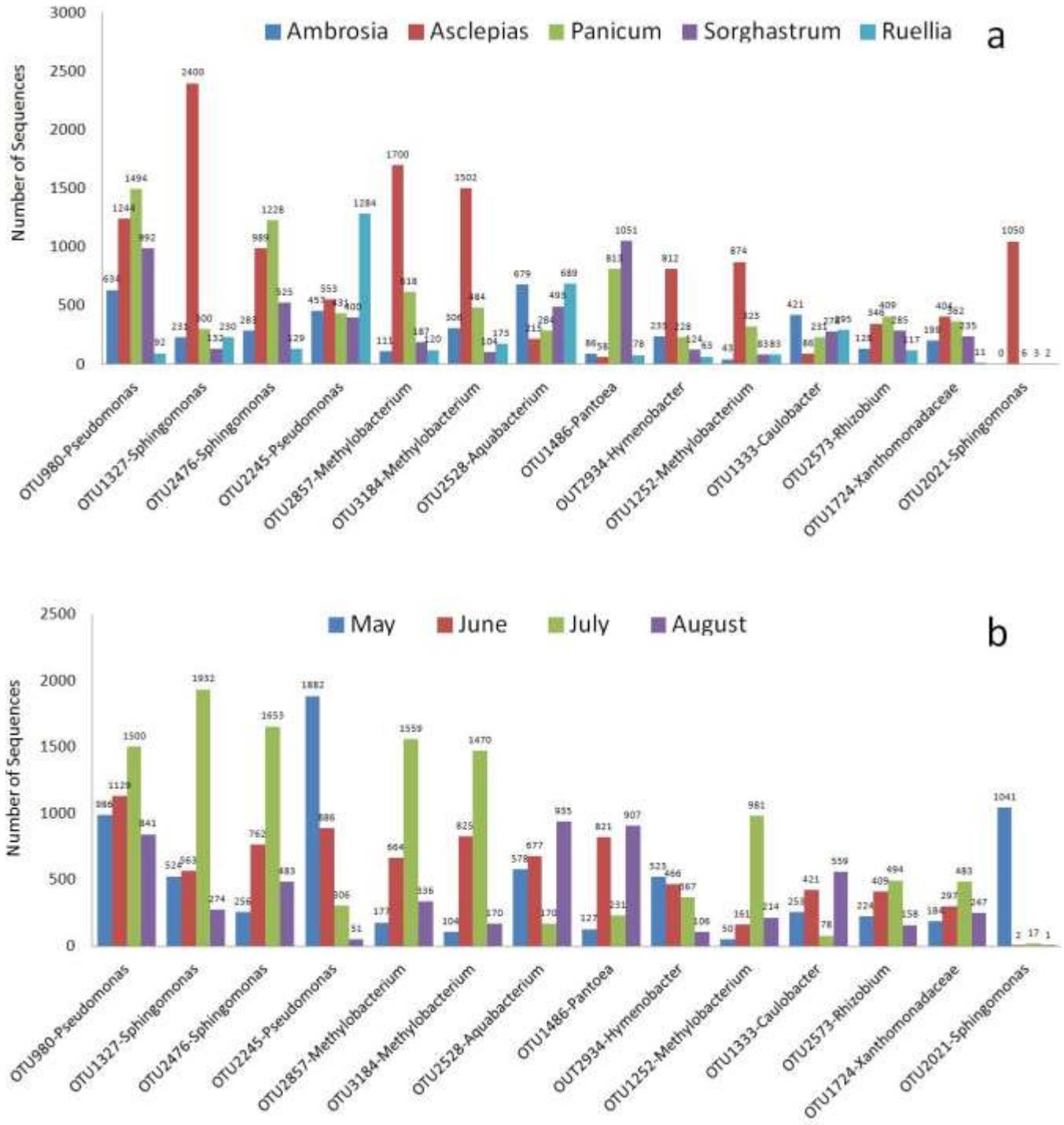
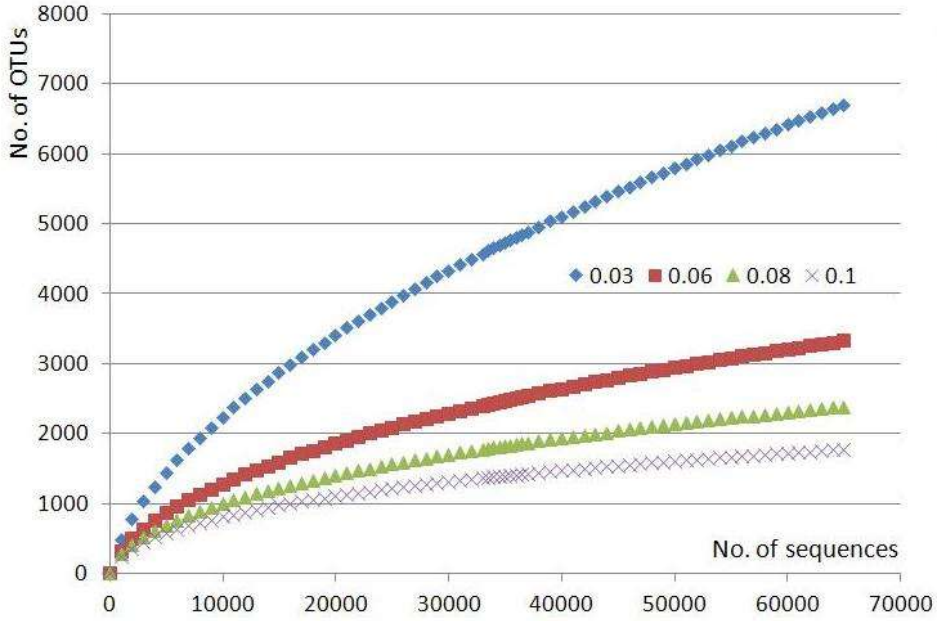

Ambrosia
Asclepias
Panicum
Sorghastrum
Ruellia
a
Number of Sequences
3000
2500
2000
1500
1000
500
0
OTU980-Pseudomonas
OTU1327-Sphingomonas
OTU2476-Sphingomonas
OTU2245-Pseudomonas
OTU2857-Methylobacterium
OTU3184-Methylobacterium
OTU2528-Aquabacterium
OTU1486-Pantoea
OUT2934-Hymenobacter
OTU1252-Methylobacterium
OTU1333-Caulobacter
OTU2573-Rhizobium
OTU1724-Xanthomonadaceae
OTU2021-Sphingomonas
May
June
July
August
b
Number of Sequences
2500
2000
1500
1000
500
0
OTU980-Pseudomonas
OTU1327-Sphingomonas
OTU2476-Sphingomonas
OTU2245-Pseudomonas
OTU2857-Methylobacterium
OTU3184-Methylobacterium
OTU2528-Aquabacterium
OTU1486-Pantoea
OUT2934-Hymenobacter
OTU1252-Methylobacterium
OTU1333-Caulobacter
OTU2573-Rhizobium
OTU1724-Xanthomonadaceae
OTU2021-Sphingomonas

Figure 2.

A.


No. of OTUs
8000
7000
6000
5000
4000
3000
2000
1000
0
0.03
0.06
0.08
0.1
No. of sequences
0
10000
20000
30000
40000
50000
60000
70000

B.

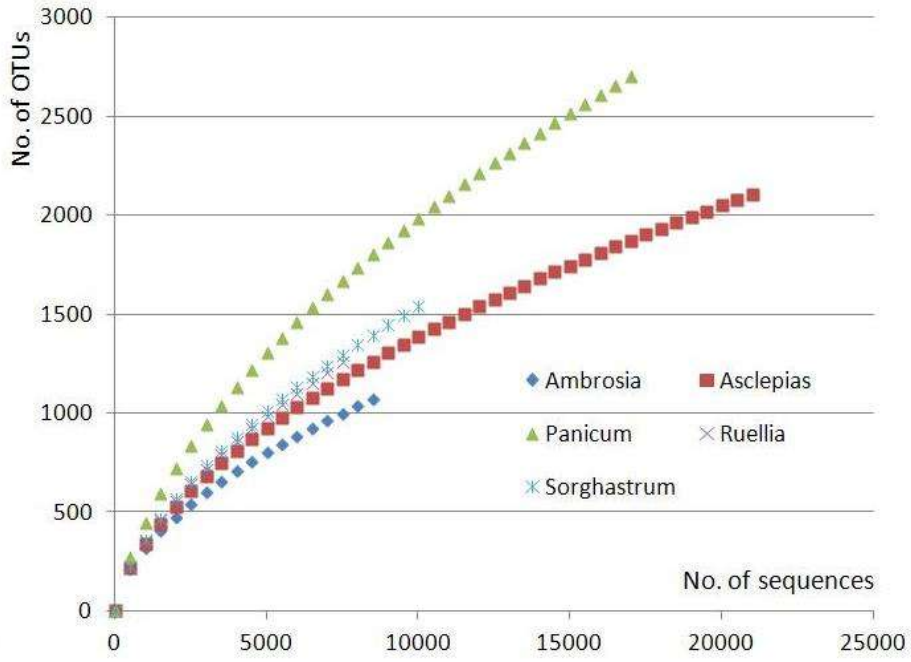
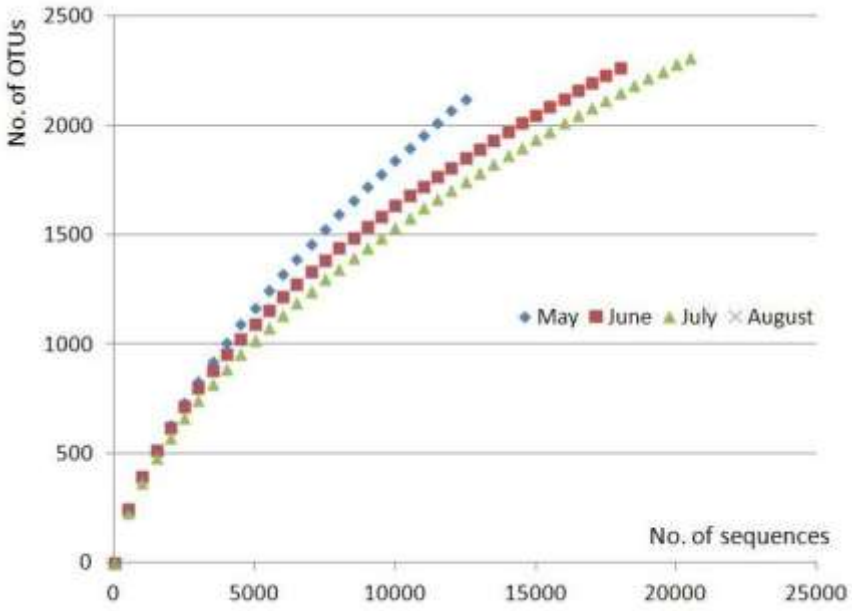

No. of OTUs
3000
2500
2000
1500
1000
500
0
Ambrosia
Asclepias
Panicum
Ruellia
Sorghastrum
No. of sequences
0
5000
10000
15000
20000
25000

Figure 3.

C.



D.

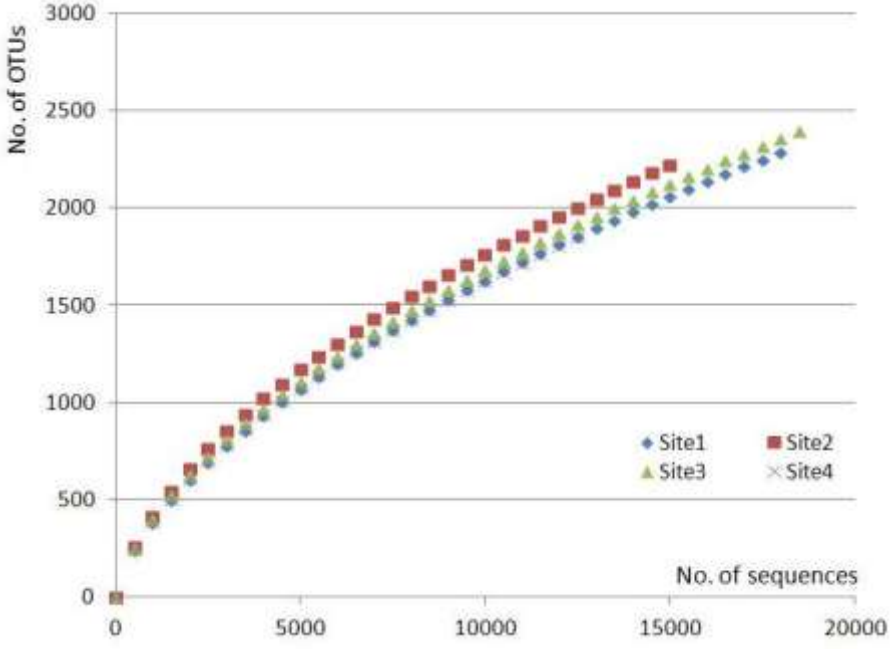


Figure 3. (Continue)

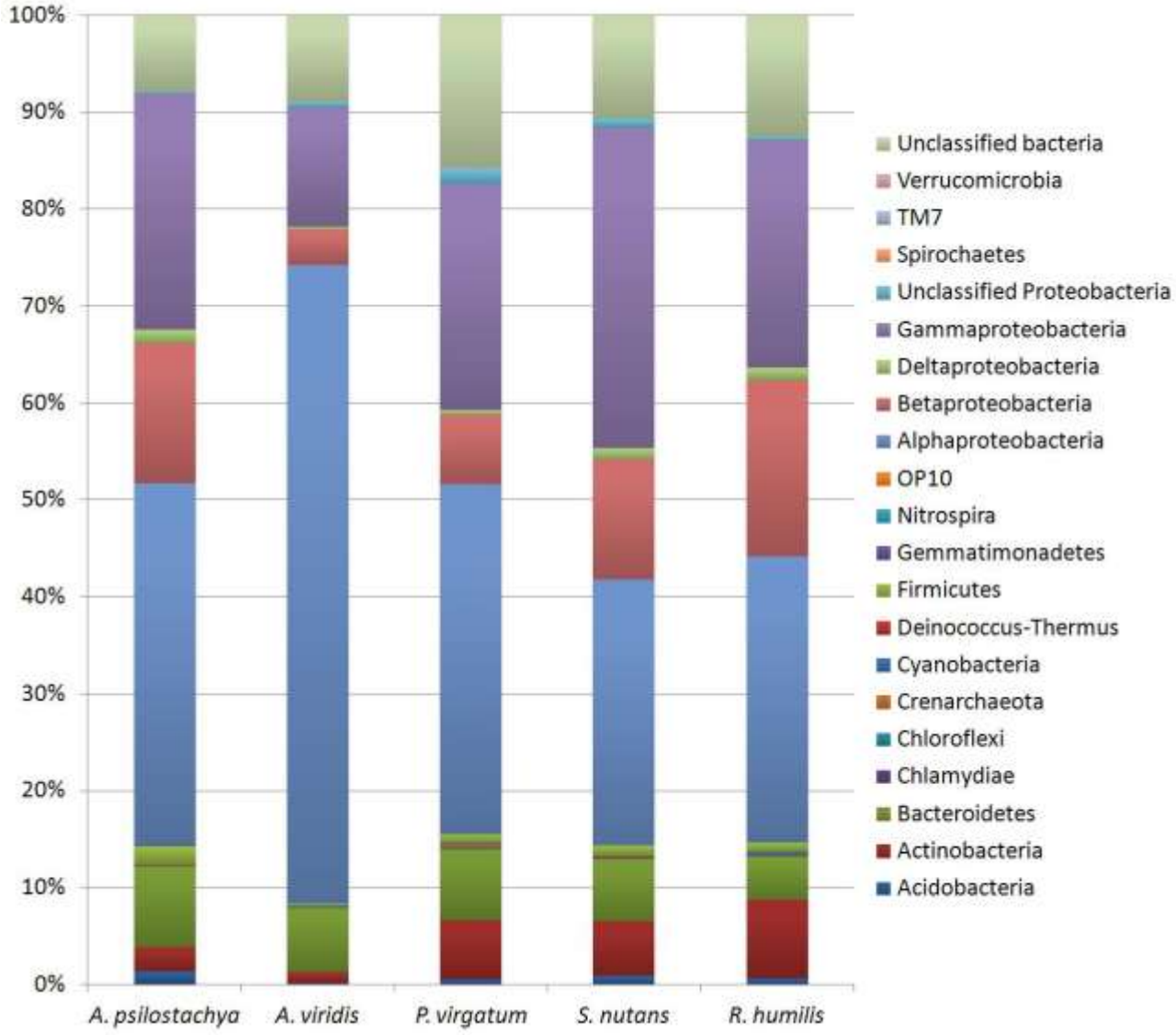

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A. psilostachya
A. viridis
P. virgatum
S. nutans
R. humilis
Unclassified bacteria
Verrucomicrobia
TM7
Spirochaetes
Unclassified Proteobacteria
Gammaproteobacteria
Deltaproteobacteria
Betaproteobacteria
Alphaproteobacteria
OP10
Nitrospira
Gemmatimonadetes
Firmicutes
Deinococcus-Thermus
Cyanobacteria
Crenarchaeota
Chloroflexi
Chlamydiae
Bacteroidetes
Actinobacteria
Acidobacteria

Figure 4

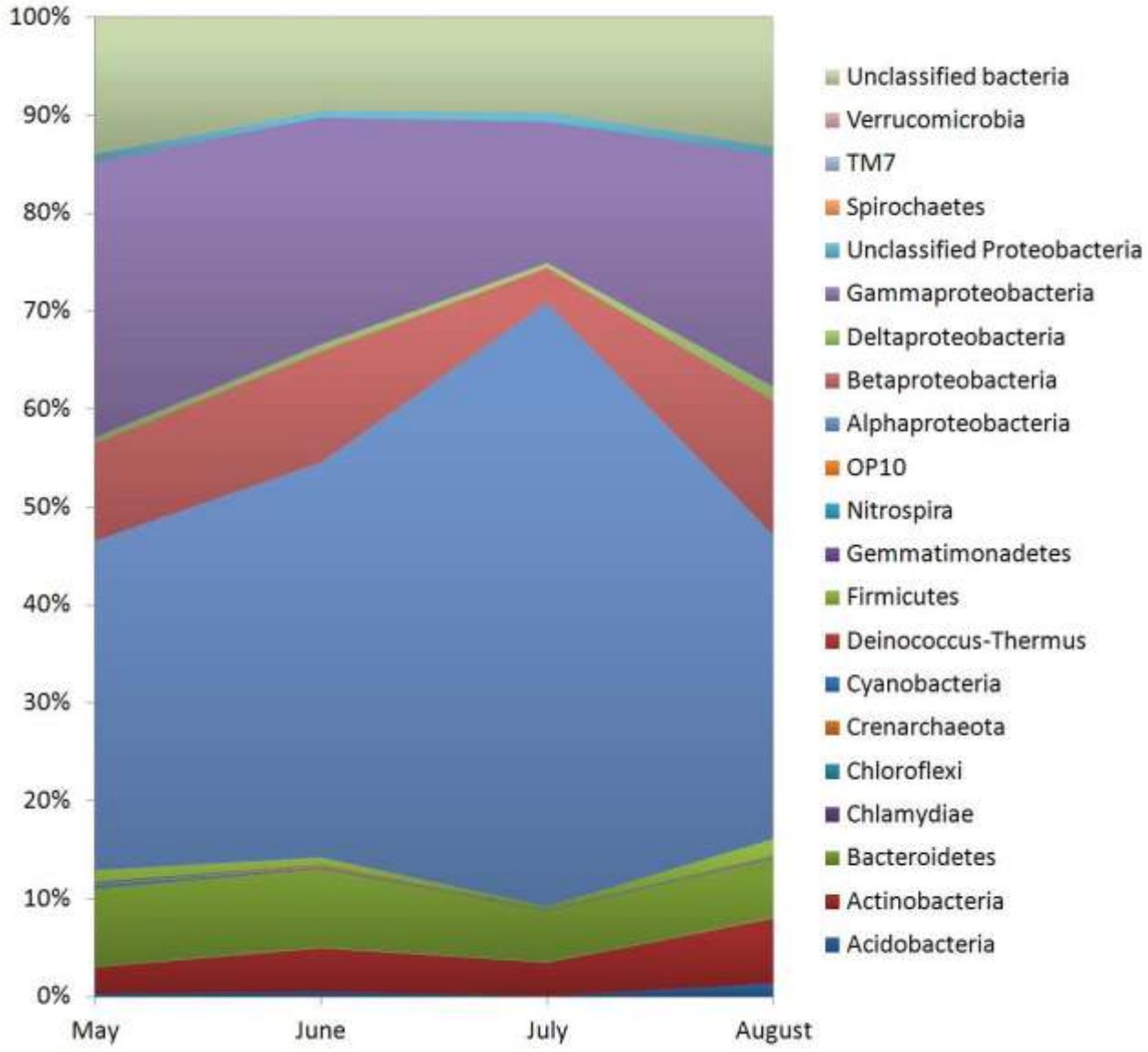

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
May
June
July
August
Unclassified bacteria
Verrucomicrobia
TM7
Spirochaetes
Unclassified Proteobacteria
Gammaproteobacteria
Deltaproteobacteria
Betaproteobacteria
Alphaproteobacteria
OP10
Nitrospira
Gemmatimonadetes
Firmicutes
Deinococcus-Thermus
Cyanobacteria
Crenarchaeota
Chloroflexi
Chlamydiae
Bacteroidetes
Actinobacteria
Acidobacteria

Figure 5.

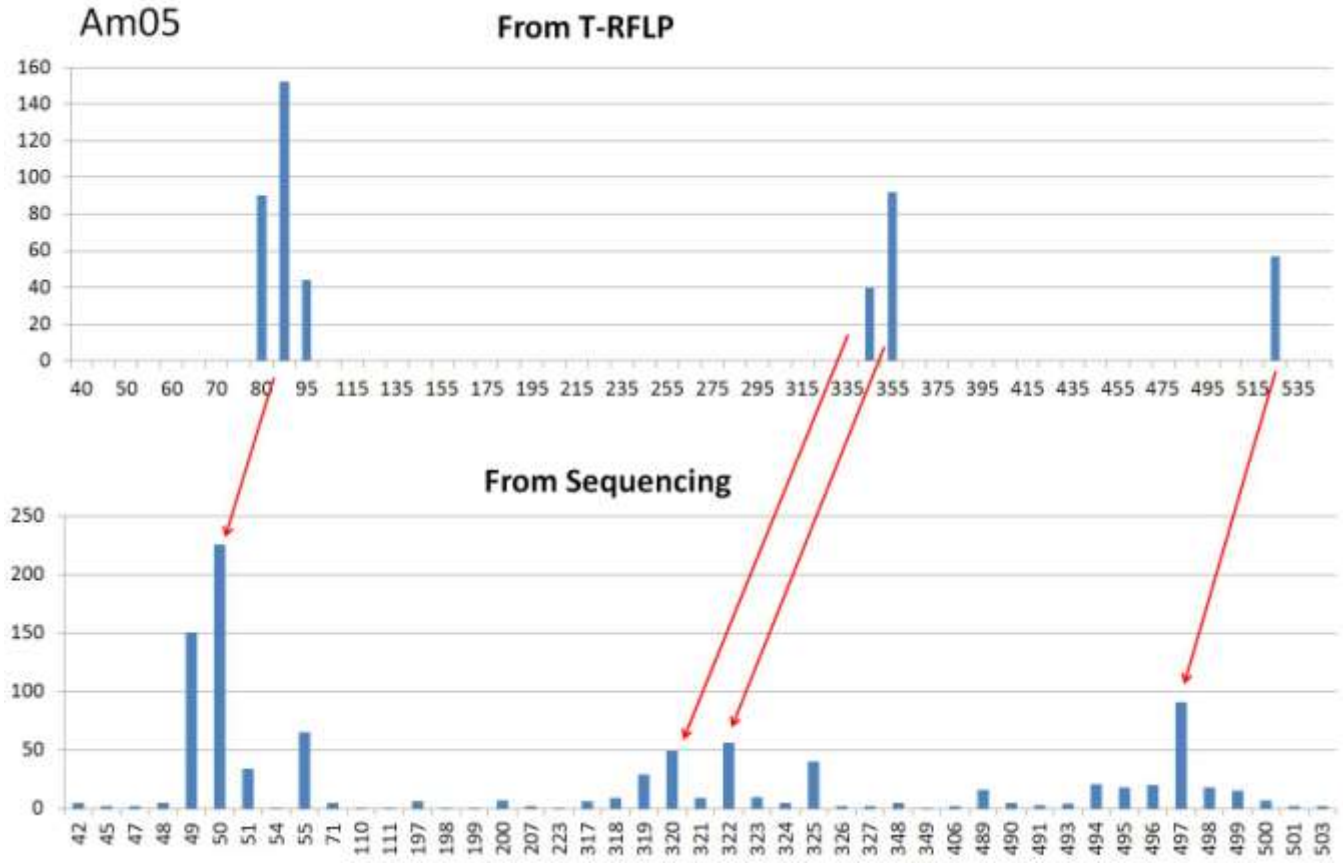

Am05
From T-RFLP
160
140
120
100
80
60
40
20
0
40 50 60 70 80 95 115 135 155 175 195 215 235 255 275 295 315 335 355 375 395 415 435 455 475 495 515 535
From Sequencing
250
200
150
100
50
0
42 45 47 48 49 50 51 54 55 71 110 111 197 198 199 200 207 223 317 318 319 320 321 322 323 324 325 326 327 348 349 406 489 490 491 493 494 495 496 497 498 499 500 501 503

Figure 6.

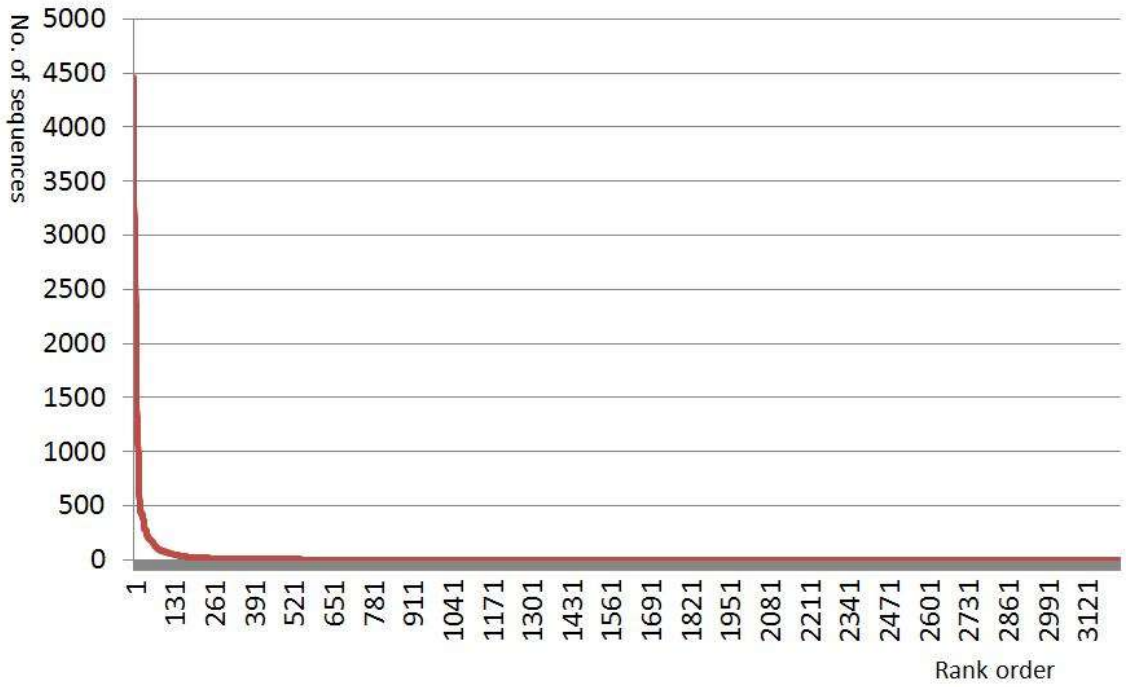


Figure 7.

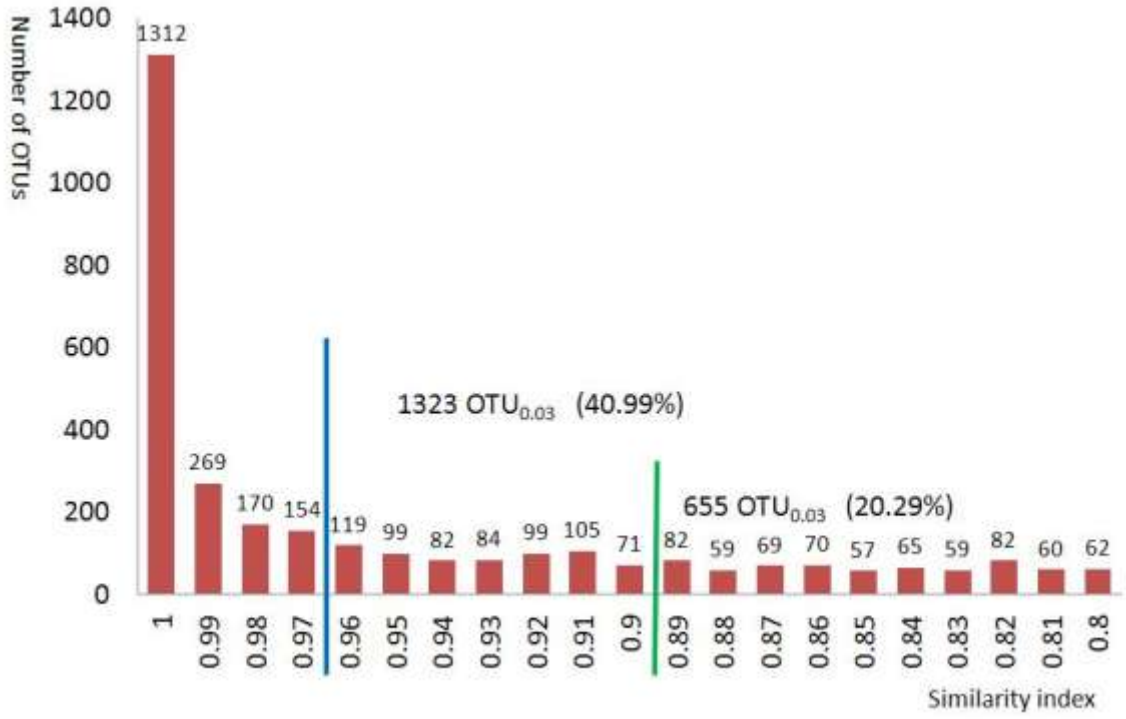


Figure 8.

APPPENDICES

Pictures and common names for the five plant species used in this Ph.D. thesis. (Courtesy of Natural Resources Conservation Services, USDA)

Asclepias viridis (Antelope-horn milkweed)

Ambrosia psilostachya (Cuman ragweed)

Ruellia humilis (Fringeleaf wild petunia)

Sorghastrum nutans (Indiangrass)

Panicum virgatum (Switchgrass)

VITA

Tao Ding

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE ANALYSIS OF THE DIVERSITY AND DISTRIBUTION OF LEAF ENDOPHYTIC BACTERIAL COMMUNITIES

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

2012, Completed the requirements for the Doctor of Philosophy in Biochemistry and Molecular Biology at Oklahoma State University.
2006, Bachelor of Science, University of Science and Technology of China, Hefei, China,

Professional Memberships:

American Society for Microbiology
Biochemistry and Molecular Biology Graduate Student Association at Oklahoma State University.

Publications:

Tao Ding, Micheal W. Palmer, Ulrich Melcher, *Community terminal restriction fragment length polymorphisms reveal insights into the diversity and dynamics of leaf endophytic bacteria*, Submitted

Melcher, Michael W. Palmer, Graham B. Wiley, **Tao Ding**, Bruce A. Roe, *Detection of members of the Tombusviridae in the Tallgrass Prairie Preserve, Osage County, Oklahoma, USA*, Virus Research, Volume 160, Issues 1–2, September 2011, Pages 256-263

Vijay Muthukumar, Ulrich Melcher, Marlee Pierce, Graham B. Wiley, Bruce A. Roe, Michael W. Palmer, Vaskar Thapa, Akhtar Ali, **Tao Ding**, *Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions*, Virus Research, Volume 141, Issue 2, May 2009, Pages 169-173

Presentations:

Poster, "Community Terminal Restriction Fragment Length Polymorphisms Reveal insights into the Diversity and Dynamics of Endophytic Bacteria", American Society for Microbiology 111th Annual Meeting, New Orleans, LA (2011)

Name: Tao Ding Date of Degree: December, 2012
Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THE ANALYSIS OF THE DIVERSITY AND DISTRIBUTION OF LEAF ENDOPHYTIC BACTERIAL COMMUNITIES

Pages in Study: 126 Candidate for the Degree of Doctor of Philosophy
Major Field: Biochemistry and Molecular Biology

Scope and Method of Study:

This research project aimed to reveal the basic composition of leaf endophytic bacterial communities, to detect the dominant and significant bacterial groups and to study the environmental influences on the structure of leaf endophytic bacterial communities; specifically, to see the relationship between host plants and endophytic bacteria and to track the dynamics of the endophytic bacteria during the host plant growing season. This research employed cultivation-independent methods to analyze bacterial communities, including total DNA extraction, polymerase chain reaction (PCR), terminal restriction fragment length polymorphism (T-RFLP) and tagged 454 pyrosequencing. Research data was analyzed using statistical software including R, SAS and CANOCO. Sequencing data was analyzed using Mothur and quantitative insight into microbial ecology (QIIME).

Findings and Conclusions:

T-RFLP has helped to study the environmental influences on the leaf endophytic bacterial communities quantitatively. Three major environmental factors, including host plant species, sampling date and collecting locations, were all tested significant using the profiles of the proportion of terminal restriction fragments (T-RF) by partial Canonical Correlation Analysis (pCCA). Dominant T-RFs were detected and host-specific T-RFs were also defined.
Tagged 454 pyrosequencing allowed revealing the leaf endophytic bacterial communities at a deeper level. Sequences (64,591) of the 16S rDNA fragments were obtained, and after alignment and distance calculation were categorized into 3,291 Operational Taxonomic Units (OTUs) at 97% similarity level. Bacteria species from 16 phyla were detected with the dominant group, from *Proteobacteria*, represented by 1982 OTUs, followed by *Bacteriodetes* and *Actinobacteria*. Environmental influences were also evaluated. Host-specific OTUs were recognized. Three main types of trends of the OTU dynamics during the host plant growing season were observed. *Alphaproteobacteria* was significantly more abundant in *Asclepias viridis* among the five host plant species, and also expanded greatly in July in the whole leaf endophytic bacterial community. Pyrosequencing data were also used to identify the dominant T-RFs, showing that the result of T-RFLP is consistent with pyrosequencing.

ADVISER'S APPROVAL DR. ULRICH MELCHER