

SOIL MICROBIAL DYNAMICS AND ASSOCIATIVE NITROGEN FIXATION
IN KANSAN TALLGRASS PRAIRIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Steven William Culman

August 2008

© 2008 Steven William Culman

SOIL MICROBIAL DYNAMICS AND ASSOCIATIVE NITROGEN FIXATION
IN KANSAS TALLGRASS PRAIRIES

Steven William Culman, Ph. D.

Cornell University 2008

Soil ecosystem properties and processes which simultaneously maintain native fertility and sustain plant yields are of principal interest in sustainable agriculture. Native prairies in Kansas are relevant in this context, as they have been annually hayed with no fertilization or detectable decline in yield or soil fertility. In contrast, intensive wheat production has resulted in significant reductions in soil fertility and now requires intensive inputs to maintain yield. This study aimed to shed light on the soil microbiological differences between these two contrasting agricultural systems in an attempt to gain insight into possible mechanisms driving nutrient and energy efficiencies in these hayed prairie ecosystems. The objectives of this study were: i) to identify major differences in soil bacterial and nitrogen fixing communities between prairies and adjacent annual wheat fields, ii) to determine if dramatic losses of soil organic carbon (SOC) are a result of obsolete farming practices, or from plant community composition, and iii) to document the relative contribution of associative N-fixation to total plant N in three C₄ prairie grasses. Soil analyses, microbial biomass, and terminal restriction fragment length polymorphism analyses (T-RFLP) revealed that bacterial and nitrogen fixing communities that were correlated with soil chemical, physical, and biological properties indicative of higher soil quality in prairie sites. In addition, SOC loss was documented in annual agriculture fields, even in the absence of tillage, demonstrating the large role that prairie plant communities play in maintaining soil fertility. Finally, evidence of associative N fixation was found in

prairie grasses which may help alleviate N limitations and sustain long-term exports of N. Two additional studies were conducted to advance T-RFLP methodology. The first study was an evaluation of statistical multivariate analyses for T-RFLP data and yielded insight into which analyses were most appropriate given research objectives and dataset complexity. The second study yielded *T-REX*, a free, online software for rapid and less-biased analyses of T-RFLP data. Collectively, the results of this work suggest a greater synchrony of plant nutrient demand in prairies, which may help to explain the greater nutrient use efficiencies seen in these systems relative to wheat.

BIOGRAPHICAL SKETCH

Steven William Culman, was born to Linda and Gene Culman one fine October morning in Cincinnati, OH. Steve grew up with a fondness for sunshine, the outdoors and (like any hard-core soil scientist) the taste of dirt (commonly called “soil” in the halls of Bradfield). As a kid, Steve spent most of his time eating copious amounts of peanut butter and jelly sandwiches, jockeying for attention in an eight-member family, and avoiding daily thrashings from his older brother. Steve survived high school and attended Thomas More College, in Kentucky, earning a B.A. in Biology in 1999. He spent his last semester on a study abroad program in Australia, and it was here that he was first exposed to the ideas and writings of Wes Jackson, Bill Mollison and many other sustainable agriculture pioneers. These writings and ideas left such a deep impression on Steve that he stayed in Australia after his studies to earn his permaculture designer’s certificate. He returned home for a few years of working various jobs such as a cook, an outdoor educator, and a laboratory technician. Steve came to Cornell in 2001. He spent fourteen months as a Fulbright Scholar in Nepal, assessing the effects of soil solarization on the microbial community in Nepal’s rice-wheat cropping systems and earned his master’s degree in soil science in 2005. Political instability in Nepal and shifting research interests toward perennial systems led Steve to the Land Institute—to work with the same organization that inspired him to pursue graduate school in the first place. Steve hopes he can continue collaborating with them in the future. For the time being, he lives in Vallejo, CA with his beautiful wife Nicole and trusty side-kick, Rufus.

To visionaries everywhere who cast aside convention and choose not
the safe and well traveled path, but the one in their hearts.
Your courage gives us hope and makes this world extraordinary.

ACKNOWLEDGMENTS

As anyone who completes a graduate degree can attest, the number of people who make the process possible far exceed those that you can acknowledge. The essential list of ‘VIPFS’s (very important players for Steve) follows.

I would first like to thank Gary Fick and Hugh Gauch. There is absolutely no doubt without these two gentlemen, I would not have completed this degree. Gary, I really appreciate your confidence and support over these past few years. I have learned a lot about teaching and mentoring from you, and I cannot thank you enough for taking me on as a student and giving me this opportunity. Hugh, thank you so much all the time you took to teach me statistics and about the scientific process in general. Your friendship and sense of service to others and to science is an inspiration that would transform academia for the better, if it were modeled by all.

I owe a great deal to Dan Buckley and his entire lab for taking me in as a stray and generously offering me their lab resources. Dan, I really appreciate the time you spent with me, helping me think through problems and finding the best approaches. Tyrrell Nelson and Florence Hsu were tremendously helpful in the lab and were kind enough to laugh at all my bad jokes over the years. I also am indebted to Laurie Drinkwater and her amazing lab group, for good discussions and technical consulting—especially Meagan Schipanski and Julie Grossman.

The Land Institute has been an amazing source of academic, intellectual and spiritual inspiration. I feel tremendously enriched to have been able to interact with so many committed and creative folks. I’d especially like to thank Jerry Glover—a rare trail-

blazer that I've found in science today—whose courage and commitment make the Land Institute's message just plain infectious. Tim Crews from Prescott College offered invaluable guidance and technical help to me throughout my program. I'd like to thank Howard Ferris and Tianna Dupont for all of their help and hospitality with my greenhouse experiment at the University of California Davis. They made my transition to Davis smooth and I could never have completed the study without them.

I'd also like to thank so many good folks in Crop and Soil Science and the University as a whole for providing an engaging environment and a wealth of resources. People and groups who have been especially impactful were, Chris Jones, Barbara Sledziona, Janice Thies and the past and present members of her lab, Kathy Howard, Chris Peters, Francoise Vermeulen, Cornell NWAEG, the Biogeochemistry Program and Mann Library. And let me individually acknowledge Thanwalee 'Jijy' Sooksa-nguan. The omission of her name from the acknowledgements section of my masters thesis came with consequences I wish not on my worst of enemies! I can only hope this may right the wrong I've been hearing about over the past 3 years. Thank you all.

I'd like to gratefully acknowledge the Land Institute's Graduate Fellows program, the Department of Crop and Soil Science, and the Cornell Biogeochemistry and Biocomplexity Small Grant Program for funding to make this research possible.

Finally, I would like to thank my family, especially my parents, Gene and Linda, for their unconditional love and support, for helping me keep a healthy perspective and for reminding me why I do what I do. At last, I'd like to thank my best friend and wife, Nicole, who has been with me through it all. Your love is the most beautiful thing I know. I could not have done this without you. Thank you.

TABLE OF CONTENTS

Biographical Sketch.....	iii
Dedication.....	iv
Acknowledgements.....	v
List of Figures.....	xi
List of Tables.....	xiv
List of Abbreviations.....	xv
Chapter I: Introduction.....	1
Chapter II. Differences in Soil Properties and Bacterial and Free-living Diazotrophic Community Structure Following 75 Years of High-input Wheat versus No-input Harvested Perennial Grasslands in Kansas	
• Introduction.....	6
• Methods.....	8
• Results and Discussion.....	15
○ Soil Properties.....	15
○ Microbial Biomass.....	17
○ 16S T-RFLP.....	21
○ <i>nif</i> T-RFLP.....	36
• Conclusions	45
• References.....	46

Chapter III. Soil Bacterial and Free-living Diazotrophic Community Dynamics

Following the Conversion of Native Tallgrass Prairie to Annual Wheat in Kansas

- Introduction.....53
- Materials and Methods.....57
- Results and Discussion.....63
 - Soil Properties.....63
 - Microbial Biomass.....65
 - 16S T-RFLP.....66
 - *nif* T-RFLP.....76
- References.....81

Chapter IV. Associative Nitrogen Fixation in Three Perennial C4 Prairie Grasses: A

Greenhouse Study

- Introduction.....86
 - History of Associative N fixation Research.....87
 - Two Case Studies.....90
 - Non-N-related Plant Growth Promotion.....92
 - Justification and Significance of Study.....93
- Materials and Methods.....95
- Results.....100
 - Experiment 1.....101
 - Experiment 2.....106
- Discussion.....106

- References.....112

Chapter V. Analysis of T-RFLP Data Using Analysis of Variance and Ordination

Methods: A Comparative Study

- Abstract.....117
- Introduction.....118
- Materials and Methods.....119
- Results.....129
 - Analysis of Variance.....129
 - Empirical Ordination Results.....131
 - Theoretical Ordination Results.....137
- Discussion.....140
- References.....146

Chapter VI. T-REX: Software for the Labeling, Processing, and Analysis of T-RFLP

Data

- Background.....149
- Implementation.....153
 - User Profile and Project Management.....154
 - Uploading Data and Labeling Procedure.....154
 - Viewing and Editing Individual Samples.....156
 - Export Labeled Data to Use Elsewhere.....157
 - Procedure to Filter out Noise from True Peaks158

- Automated Alignment of Peaks158
- Grouping Samples into Environments.....159
- Data Matrix Construction and AMMI analysis159
- Summary of Results and Output162
- Example Dataset.....162
- Results and Discussion.....163
- Conclusions.....165
- References.....166

LIST OF FIGURES

2.1	Relationship of microbial biomass carbon to depth	18
2.2	Relationship of $\delta^{13}\text{C}$ microbial biomass to depth	20
2.3	Legend of symbols used to depict T-RFLP samples	22
2.4	16S T-RFLP: complete dataset, all depths (0 – 100cm).....	23
2.5	16S T-RFLP: complete dataset, surface 3 depths (0 – 40cm).....	25
2.6	16S T-RFLP: complete dataset, surface depth (0 – 10cm).....	26
2.7	16S T-RFLP: June 2007, all depths (0 – 100cm).....	27
2.8	16S T-RFLP: June 2007, surface depths (0 – 10cm) and bottom depths (40 – 100cm).....	29
2.9	16S T-RFLP: June 2007, surface depth (0 – 10cm).....	30
2.10	16S T-RFLP: complete dataset averaged by date, bottom depths (40 – 100cm).....	30
2.11	16S T-RFLP: complete dataset averaged by date and site, all depths (0 – 100cm) and surface four depths (0 – 60cm).....	31
2.12	<i>nif</i> T-RFLP: June 2007, all depths (0 – 100cm).....	40
2.13	<i>nif</i> T-RFLP: June 2007, surface depths (0 – 40cm).....	41
2.14	<i>nif</i> T-RFLP: June 2007 averaged by site, all depths (0 – 100cm) and surface four depths (0 – 60cm).....	43
3.1	Map of Niles conversion study site.....	58
3.2	Relationship of microbial biomass carbon to depth	65
3.3	Legend of symbols used to depict T-RFLP samples	67

3.4	16S T-RFLP: complete dataset, all depths (0 – 100cm).....	68
3.5	16S T-RFLP: 2005 dataset averaged by block, all depths (0 – 100cm).....	70
3.6	16S T-RFLP: June 2007 dataset, all depths (0 – 100cm).....	72
3.7	16S T-RFLP: June 2007 dataset, surface depths (0 – 40cm).....	74
3.8	16S T-RFLP: June 2007 dataset, bottom depths (40 – 100cm).....	75
3.9	<i>nif</i> T-RFLP: June 2007, all depths (0 – 100cm).....	77
3.10	<i>nif</i> T-RFLP: June 2007, surface depths (0 – 40cm).....	78
3.11	<i>nif</i> T-RFLP: June 2007, bottom depths (40 – 100cm).....	79
4.1	Graphic depiction of methods.....	99
4.2	Experimental pot and nutrient solution watering set up.....	100
4.3	Effects of N level and inoculation in big bluestem.....	103
4.4	Effects of N level and inoculation in indianguass.....	104
4.5	Effects of N level and inoculation in switchgrass.....	105
4.6	Effects of clipping and inoculation in big bluestem.....	107
4.7	Changes with time and inoculation in clipped big bluestem.....	108
5.1	Ordinations of fungal community T-RFLP data with environment-centered PCA and AMMI analysis.....	133
5.2	T-RFLP bacterial community data analyzed with AMMI and NMS with Sørensen distance measure.....	136
6.1	Screenshot of T-REX Homepage.....	154
6.2	Screenshot of T-REX Samples Summary page.....	157
6.3	Screenshot of T-REX Create Data Matrix/ Run AMMI page.....	160
6.4	Screenshot of tables 1 and 2 from the AMMI analysis.....	161

6.5 Screenshot of tables 3 and 4 from the AMMI analysis.....162

LIST OF TABLES

2.1	Soil properties of field sites in regional study.....	16
2.2	MRPP analysis of complete bacterial dataset.....	33
2.3	MRPP analysis of June 2007 bacterial dataset.....	34
2.4	Average <i>nif</i> H T-RF richness with different fluors.....	37
2.5	Distribution of variance in <i>nif</i> dataset with different fluors.....	38
2.6	Average <i>nif</i> H T-RF richness with different fluors and management histories.....	38
2.7	MRPP analysis of June 2007 <i>nif</i> dataset.....	42
3.1	Soil properties 3 years after no-till conversion.....	64
3.2	MRPP analysis of complete bacterial dataset.....	69
3.3	MRPP analysis of June 2007 bacterial dataset.....	73
3.4	MRPP analysis of June 2007 <i>nif</i> dataset.....	78
4.1	Nitrogen budget for annually hayed bottomland tallgrass prairie.....	93
4.2	Summary of significant differences found in experiment 1.....	101
5.1	T-RFLP datasets used in statistical comparison study.....	120
5.2	Characteristics of T-RFLP datasets used in statistical comparison study.....	122
5.3	Percent of variation in T-RFLP datasets from Analysis of Variance.....	130
5.4	Results of empirical comparisons between ordination methods.....	132

LIST OF ABBREVIATIONS

T-RFLP = Terminal restriction fragment length polymorphism

T-RF = Terminal restriction fragment

E = Environment

T×E = Terminal restriction fragment × Environment interaction

AMMI = Additive Main Effects and Multiplicative Interaction Model

PCA = Principal component analysis

CA = Correspondence Analysis

DCA = Detrended Correspondence Analysis

NMS = Nonmetric Multidimensional Scaling

IPC = Interaction principal component

MBC = microbial biomass carbon

SOC = soil organic carbon

BNF = biological nitrogen fixation

CHAPTER 1

INTRODUCTION

In 2005, the United Nations released the Millennium Ecosystem Assessment (MEA 2005), the most comprehensive report to date with the goal of assessing the state of the planet and to establish a basis for actions needed to improve the conservation and use of ecosystems. Over two thousand authors, scientists and experts in a vast range of fields, reported their sobering findings to anyone who was willing to listen. To quote the first of the four main findings: “Over the past 50 years, humans have changed ecosystems more rapidly and extensively than in any comparable period of time in human history, largely to meet rapidly growing demands for food, fresh water, timber, fiber, and fuel. This has resulted in a substantial and largely irreversible loss in the diversity of life on Earth.”

It is difficult to comprehend the weight and scale of such a statement with words meticulously chosen over the course of this five-year synthesis. It is also difficult to deny that we find ourselves living in a period of history that is somewhat paradoxical—we are armed with enough technical knowledge to simultaneously cause such destruction, measure the results of these actions, and predict the implications of continuing on our current trajectory.

Although agriculture is only one human activity out of many that is leading to the degradation of ecosystems outlined in the MEA, it is a substantial and ubiquitous activity that spans all arable regions of the globe. Given how intimately agriculture interfaces with the soil, water and atmosphere, improving the ways we grow food, fiber and fuel has the potential to bring about great human and ecosystem benefit.

The Land Institute is one of many organizations working to improve modern methods of food production by focusing on the development of perennial grain crops. The overarching mission of the Land Institute is to design agricultural ecosystems that mimic the native tallgrass prairies of Kansas—perennial herbaceous mixtures of seed-producing grasses, legumes and forbs. Such an agriculture could bring many environmental advantages (e.g., reduced soil erosion and herbicide use from perennial cover, reduced nutrient run-off due to extensive rooting systems), as well as many agronomic advantages (e.g., reduced input and on-farm fuel costs, reduced irrigation due to greater use of soil water reserves). To date, progress has been made, but much more work needs to be done. Perhaps one of the largest hurdles will be to move perennial grain breeding and its concepts from a ‘high-risk and fringe’ endeavor, to one which makes sound scientific sense, appropriately reflected in research priorities, academic dialogue and funding allocations.

The majority of the work in this dissertation is part of a larger research project catalyzed by the Land Institute. Seven graduate students from six universities have collaborated for the past 2 years examining hayed, native prairies in north central Kansas. Research by this team has assessed soil properties, bacterial, nematode, and mycorrhizal communities, nutrient budgets, plant communities, above ground insect communities and historic watershed N runoff. This dissertation specifically explored the soil biological components in this study, namely bacterial community dynamics. It is organized by the following chapters.

Chapter 2 in this dissertation outlines the experimental design, rationale, and overall motivation for a comparative study between hayed, native prairies and annual wheat

fields. It presents data on the effects of the two different management histories on key soil properties, as well as on bacterial and nitrogen fixing communities. Results of nematode analyses are mentioned briefly to discuss the bacterial findings in a larger context.

Chapter 3 presents the results of the conversion study—a study in which native prairies were converted to annual no-till agriculture without tillage. Like chapter 2, this chapter reports the results of soil data and bacterial and nitrogen fixing community data. However, instead of long-term management histories, these results document the first three years after the conversion of native prairie to annual crops.

Chapter 4 reports the results of a greenhouse study conducted to document associative nitrogen fixation in the three most dominant C4 grasses in the prairie sites. From long-term nitrogen budgets, it seems likely that there is considerable non-symbiotic nitrogen fixation contributing to the N removed in hayed biomass and this study was an initial attempt to document and quantify this phenomenon.

Chapter 5 is a statistical comparison of several multivariate ordination methods for the analysis of terminal restriction fragment length polymorphism (T-RFLP) microbial community datasets. This study empirically tested and theoretically compared seven statistical methods with ten diverse soil datasets. Results of the empirical findings and theoretical considerations are discussed.

Chapter 6 outlines T-REX, an online tool for the analysis of T-RFLP data. T-REX was developed in conjunction with Cornell Bioinformatics Service Unit to facilitate a streamlined, more robust analysis of these data. The many features and the benefits

offered by this software are presented. An example dataset is presented to demonstrate the software's functionality.

Adapting and creating an agriculture that fosters biological synergy and eliminates the degradation of our natural resources will be a major challenge for human civilization over the next century and beyond. It is hoped that this dissertation not only advances the understanding of these perennial grassland systems, but also raises awareness and interest in the field of perennial agriculture, as a potential solution to address some of the most pressing issues facing us today.

REFERENCES

MEA. 2005. Millennium Ecosystem Assessment: Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC.

CHAPTER 2
DIFFERENCES IN SOIL PROPERTIES AND BACTERIAL AND FREE-LIVING
DIAZOTROPHIC COMMUNITY STRUCTURE FOLLOWING 75 YEARS OF
HIGH-INPUT WHEAT VERSUS NO-INPUT HARVESTED PERENNIAL
GRASSLANDS IN KANSAS¹

Introduction

On a fundamental level, the human endeavor of agriculture is nothing more than the harvesting of nutrients. We manage plant communities to take elements from the atmosphere and soil and convert them to human usable forms (food, fiber, fuel) via photosynthesis. In this context, evaluating the efficiency (input vs. output) of an agricultural system's use of nutrients and energy can provide a useful framework in assessing its performance and be an indicator of the overall sustainability of that system.

Much of industrial agriculture, as it is practiced today, is characterized by both low nutrient use and energy efficiencies. For example, worldwide nitrogen use efficiency for cereals is approximately 33% (Raun and Johnson 1999). Despite increased energy efficiencies in the last several decades, the United States agriculture still consumes 1.7 quadrillion Btu (Schnepf 2004) while producing just 1.66 quadrillion Btu of food energy (ERS 2004). In addition, pollution of water sources, atmospheric degradation, and practically irreversible degradation of soil are common consequences of food production today. These issues led the Millennium Ecosystem Assessment to conclude

¹ Data from this chapter will be submitted to the Proceedings of the National Academy of Sciences.

that agriculture may be one of the largest threats to biodiversity and ecosystem function of any single human activity (Assessment 2005).

Looking to soil ecosystem properties and processes which simultaneously maintain fertility and sustain plant yields are of principal interest in sustainable agriculture. As agricultural systems, perennial grasslands have long been recognized for the environmental benefits they provide, such as reduced soil erosion (Lindstrom et al. 1994), reduced N and C loss (Gebhart et al. 1994, Knapp et al. 1998, Tilman et al. 2002) and greater C sequestration (Tilman et al. 2002). However perennial grasslands do not provide a direct human food source as the cereals (annual grasses) do, and as a result, have generally not been considered as a viable solution to meet direct human food demands.

Recent interest in biofuels has demonstrated the utility of perennial grasslands in producing fuel. Tillman et al. (Tilman et al. 2006) reported that high-diversity, low-input perennial grasslands yielded more usable biofuel energy, greater carbon sequestration and less pollution than corn grain ethanol or soy-biodeisel. Their study drew from ten years of aboveground biomass harvest data which showed that yield increased with both increased diversity and time. Other reports have argued that perennial grasslands offer a viable and often advantageous option over annual crops for biofuel production, especially when considering soil C sequestration potential of perennial systems (Schmer et al. 2008).

In a study comparing long-term nutrient removals from landscapes, Glover et al. (unpublished) showed perennial grasslands provided comparable levels of harvested N in biomass as adjacent high-input wheat fields provided in harvested grain. Over the

approximately 75-year management history of the two systems, roughly 26% more N has been harvested from the perennial grasslands than from wheat despite the absence of fertilizer inputs. In addition, soil fertility indicators (total N, organic matter, water stable aggregates) are significantly higher in the prairie sites than in the annual agriculture sites. These findings indicate much tighter nutrient cycling in the perennial grasslands, and suggest fundamental differences in soil biology. The authors suggested these attributes make perennial grasslands a model for a more sustainable agricultural system.

Here we build on the Glover et al. (unpublished) study by comparing key soil biological, chemical and physical properties between the native tallgrass prairies and the adjacent, annual agricultural fields. Our overall objective is to elucidate soil biological properties and processes, related to communities, nutrient cycling and overall food web structure that are correlated with the sustained long-term nutrient export in prairies. This chapter looks at the role of bacterial communities in these systems. Its specific objectives are to i) determine differences in key soil variables with depth, ii) determine the relative influence of experimental factors shaping bacterial community abundance and structure and, iii) examine nitrogen-fixing bacterial structure in these systems.

Materials and Methods

Site Descriptions and Soil Sampling. The field sites in this study were located in five counties of North Central Kansas as described by Glover et al. (2008). Specific field site names and respective locations were: **Buckeye**, Dickinson Co. N^o 39.2.344, W^o 97.7.798; **Niles**, Ottawa Co. N^o 38.58.145, W^o 97.28.616; **Goessel**, McPherson Co. N^o

38.15.333, W' 97.22.307; **New Cambria**, Saline Co. N' 38.53.54, W' 97.32.615; **Five Creek**, Clay Co. N' 38'22.665, W' 97.18.788. Five native, bottom-land prairie meadows (perennial grasslands) with adjacent annually cropped wheat fields were identified and sampled. Relatively consistent management of the two systems had been practiced for 75 years or more.

Soils were sampled three total times: i) June 18 – 22, 2006, ii) October 5 – 9, 2006, and iii) June 17 – 20, 2007. Four centimeter diameter cores were taken to a depth of one meter in a 25 m transect across wheat and prairie sites. Five cores were taken from each field and separated into sections by depth: 0 – 10 cm, 10 – 20 cm, 20 – 40 cm, 40 – 60 cm, 60 – 80 cm, 80 – 100 cm. The five samples from each depth were bulked and mixed until homogeneous. Soils were then frozen at – 20°C for molecular analyses or stored at 4°C for all other analyses.

The field sites consisted of five native prairie meadows (also called 'perennial grasslands') paired with annual agricultural fields. Prairie sites have never been tilled or fertilized and management consisted exclusively of annual mowing in June or July for hay removal. Annual wheat field management followed typical practices for the region (KSUAES 1996, 1997). This study's experimental design consisted of four factors: sampling date, depth, management history and site. Specifically, there were 3 sampling dates, 6 depths, 2 management histories, and 5 sites sampled, totaling 180 samples. Our research objectives led us to focus mainly on the differences between management histories (perennial grasslands vs. annual wheat), but also how these differences change with depth. Differences detected between sampling date and field site were of limited interest here. Multiple sampling dates were taken to verify trends seen at any one individual sampling period and multiple sites were sampled in an

attempt to capture consistent differences between these two systems, hence making our observations more robust.

Soil Properties. Soil properties were analyzed both at The Land Institute (TLI) and at the Soil Testing Laboratory at Kansas State University (KSU). Analyses at TLI included: pH (Robertson et al. 1999), bulk density by oven drying at 105 °C, percent clay by the hydrometer method (Elliott et al. 1999), water stable aggregates (WSA) by wet-sieving (Seybold and Herrick 2001), and readily oxidizable carbon (ROC) (Weil et al. 2003). Analyses at KSU included SOM by the Walkley-Black procedure, soil organic carbon (SOC) and total N by dry combustion on a LECO CN 2000 combustion analyzer, total P by a modified Kjeldahl digestion and total K by flame atomic absorption. Further details on analyses performed at KSU can be found at MEAS (MEAS 1998).

Microbial Biomass. Microbial biomass carbon (MBC) was measured with the simultaneous chloroform fumigation extraction (sCFE) method (Fierer and Schimel 2003). Briefly, 10g of soil from each sample were weighed into two, 70ml glass vials, one labeled 'chloroform', the other 'non-chloroform'. 40ml of 0.05M K₂SO₄ were added to both vials and the chloroform vial received 0.5ml of EtOH-free CH₃Cl. Blanks (both chloroform and non-chloroform vials without soil) were also prepared. Vials were sealed and shaken at 150rpm for 4 hours. Extracts were centrifuged for 15 minutes at 1500 rpm and the supernatant was vacuum filtered through 0.45 µm Watman filter paper. Microbial biomass extracts were bubbled for 30 minutes with air to remove any residual CH₃Cl and stored at – 20 °C until analysis. Dissolved organic carbon (DOC) and the natural abundance ¹³C values of DOC were determined using an O.I. Analytical Model 1010 TOC Analyzer (OI Analytical, College Station, TX)

interfaced to a PDZ Europa 20-20 isotope ratio mass spectrometer (Sercon Ltd., Cheshire, UK) at the University of California Davis Stable Isotope Facility.

MBC was calculated as the difference between chloroform and non-chloroform (control) samples divided by a K_{EC} -factor of 0.35 (Sparling et al. 1990). Analyzed blank samples contained no to extremely small amounts of C, so no correction was made for this insignificant analytical artifact. The ^{13}C isotope composition was expressed in parts per thousand (‰) relative to the International PeeDee Belemnite (PDB), where $\delta^{13}\text{C} = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 1000$ and R is the molar ratio of $^{13}\text{C}/^{12}\text{C}$. The $\delta^{13}\text{C}$ (‰) of MBC was calculated as follows: $\delta^{13}\text{C}_{\text{MB}} = [(\delta^{13}\text{C}_c \times C_c) - (\delta^{13}\text{C}_{\text{nc}} \times C_{\text{nc}})] / (C_c - C_{\text{nc}})$, where C_c and C_{nc} is MBC (DOC kg ha⁻¹) extracted from the chloroform and non-chloroform samples, and $\delta^{13}\text{C}_c$ and $\delta^{13}\text{C}_{\text{nc}}$ is the ^{13}C natural abundance of the chloroform and non-chloroform extracts (‰), respectively (Ryan and Aravena 1994).

Molecular Analyses (PCR and T-RFLP)

16S rRNA gene. Soil DNA was extracted from 0.25 g soil per sample using the MoBio PowerSoil™ DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA). DNA extracts were quantified and diluted with nuclease-free water to 2 ng μl^{-1} . DNA was then amplified by polymerase chain reaction (PCR) using the fluorescently-labeled forward primer 27f (5'-[6FAM] AGA GTT TGA TCM TGG CTC AG-3') and the unlabeled reverse primer 1492r (5'-TAC GGY TAC CTT GTT ACG ACT T-3') (Invitrogen, Carlsbad, CA). These primers target the bacterial 16S rDNA genes in the extracted soil DNA and the amplification results in products of approximately 1500 bp. Three, 50 μl reactions of each sample were amplified using a PTC 200 thermal

cycler (MJ Research, Waltham, MA) as follows: initial denaturation at 95°C for 5 min; 27 cycles of denaturation at 95°C for 45 s, annealing at 56°C for 45 s, and extension at 72°C for 1 min; and a final extension step at 72°C for 10 min. Reaction concentrations were: 0.05 U μl^{-1} *AmpliTaq Gold*® DNA polymerase (Applied Biosystems, Foster City, CA), 1x PCR buffer, and 2.0 mM MgCl_2 , 0.2 mM deoxy-nucleotide triphosphates (dNTPs), 0.1 $\mu\text{g } \mu\text{l}^{-1}$ bovine serum albumin (BSA), both primers at 0.1 μM , nuclease free water, and 5 μl of DNA template (10 ng reaction⁻¹). Amplified DNA products were verified by electrophoresis on a 1.0% agarose gel.

Following PCR, amplified DNA (three 50 μl reactions per sample) was pooled and quantified. DNA concentrations were adjusted to 30 ng μl^{-1} . Two, 30 μl restriction enzyme digests were prepared per sample using *HhaI* and *Sau96 I* restriction enzymes (New England Biolabs, Ipswich, MA). Reaction concentrations were: 5 U enzyme (either *HhaI* or *Sau96 I*), 1x of the respective buffer, 0.1 $\mu\text{g } \mu\text{l}^{-1}$ BSA, nuclease-free water, and 15 μl of amplified DNA (450 ng reaction⁻¹). Restriction digestion was carried out in a MJ Research PTC 200 thermal cycler at 37°C for 4.5 h with a final step of 70°C for 15 min to stop the reaction. Complete digestion of the DNA was verified by electrophoresis on a 1.5% agarose gel.

Digested DNA was purified using a PERFORMA® DTR Edge Plate (Edge BioSystems, Gaithersburg, MD) and lyophilized. DNA was resuspended in a 10 μl mix containing 9.85 μl of formamide and 0.15 μl of LIZ 500 size standard (Applied Biosystems). Terminal fragment-size analysis was performed using a 3730 ABI electrophoretic capillary sequencer in conjunction with the Genemapper Software (Applied Biosystems) at Cornell University's Biotechnology Resource Center, Ithaca, NY.

nifH. T-RFLP analyses were also performed to characterize free-living diazotrophic populations in soils. T-RFLP targeting the *nifH* gene was performed on all 2006 samples; however, the electrophoretic traces were very poor quality and rendered unusable. After further protocol optimizations, *nifH* T-RFLP data for June 2007 yielded satisfactory results. Re-analysis of 2006 samples was not possible, as too little extracted soil DNA remained after laboratory optimizations. Hence, only the results of June 2007 will be discussed and presented here. The methods used were as described above, with the following changes. Soil DNA extracts were amplified by PCR using the fluorescently-labeled forward primer *nifH*-b1-112F (5'-[PET] GGC TGC GAT CCC AAG GCT GA-3') (Applied Biosystems) and the fluorescently-labeled reverse primer CDHP Nif723R (5'-[6FAM] GAT GTT CGC GCG GCA CGA ADT RNA TSA-3') (Invitrogen). These primers target *nifH*, the structural gene for nitrogenase reductase, in the extracted soil DNA and the amplification results in products of approximately 700 bp. Three, 50 μ l reactions of each sample were amplified as follows: initial denaturation at 95°C for 10 min; 35 cycles of denaturation at 95°C for 30 s, annealing at 61°C for 30 s, and extension at 72°C for 45 s; and a final extension step at 72°C for 10 min. Reaction concentrations were: 0.05 U μ l⁻¹ *AmpliTaq Gold*® DNA polymerase, 1x PCR buffer, and 2.5 mM MgCl₂, 0.8 mM dNTPs, 0.5 μ g μ l⁻¹ BSA, both primers at 0.25 μ M, nuclease free water and 50 ng DNA template reaction⁻¹. Amplified DNA was digested with the restriction enzyme *MspI* (New England Biolabs) in the following reaction concentrations: 5 U enzyme, 1x of supplied buffer, 0.1 μ g μ l⁻¹ BSA, nuclease-free water and 450 ng DNA reaction⁻¹.

Statistical Analyses

Analysis of Variance was performed on the soil nutrient data and microbial biomass data using PROC MIXED procedure in SAS v.9 (Cary, NC). Depth and management history were treated as fixed effects and site a random effect. All soil variables were measured at all sampling dates, except ROC, microbial biomass and WSA, which were only measured at the June 2007 sampling. Soil variables exhibited consistent results across the three sampling periods with no sampling date interactions. Therefore, only June 2007 soil data are presented here, with significance differences determined at $\alpha = 0.05$ level of probability. All soil data were converted into mass per hectare, in order to account for differences in soil volume and bulk density. Least squared means are reported for all soils data.

The T-RFLP data analyzed in this study were uncommonly complex in regard to number of underlying environmental gradients (sampling date, depth, management history, and site), sample heterogeneity and percent variation from interaction signal. As a result of this complexity, nonmetric multidimensional scaling (NMS) analyses with the Sørensen distance measure were used to analyze all T-RFLP data. (See Chapter 5 for further discussion selecting an appropriate ordination analysis for T-RFLP.) NMS analyses were performed in PC-ORD v.5 (MjM Software Design, Gleneden Beach, OR) with 2 axes selected, 50 runs with real data, <0.0011 stability criterion, 50 iterations to evaluate stability.

Multiple-response Permutations Procedures (MRPP) were also employed to test significance among the experimental factors within the datasets (Mielke 1984, McCune and Grace 2002). This procedure creates *p*-values to determine statistical

significance between groups in a factor, as well as an A statistic. The A statistic, also called “the chance-corrected within-group agreement,” describes within-group homogeneity and is independent of sampling size. When all items within each group are identical, $A = 1$; if the items within each group equals what is expected by chance, $A = 0$. Negative A values result from cases with less agreement between groups than is expected by chance.

Results and Discussion

Soil Properties. Key soil properties for all 5 sites are found in Table 2.1. Here, we assume (based on soil horizonation and textural properties) that prior to agricultural conversion, the soil properties between paired agricultural and prairie sites were not different. Under this assumption, the long-term effects of annual agriculture on the soil are dramatic. The annual wheat fields have significantly lower amounts of soil organic matter (SOM), soil organic carbon (SOC), readily oxidizable carbon (ROC), total soil N, and water stable aggregates (WSA) compared to perennial grassland sites at every depth measured through 60cm of soil (Table 2.1). Annual agriculture has decreased the amount of SOC by 28% and total N by 27% in the upper 60cm of soil, despite the fact that agricultural plots have received approximately $70 \text{ kg N ha}^{-1} \text{ yr}^{-1}$.

In contrast, the annual haying of prairie sites, which have never received fertilization, has not resulted in the same rates of degradation. Historic hay yield data from the same five counties in KS as the field sites show prairie productivity has been maintained and even increased over the same time span (Figure 5 in supplementary,

Table 2.1. Soil properties of field sites in this study*

Depth (cm)	Management History	pH	clay	SOM (Mg ha ⁻¹)	SOC (Mg ha ⁻¹)	ROC (kg ha ⁻¹)	Total N (Mg ha ⁻¹)	Total P (kg ha ⁻¹)	Total K (kg ha ⁻¹)	WSA	BD (Mg m ⁻³)
0 – 10	PR	5.9	21	62.0	36.8	1040	3.1	427.6	431.4	0.93	1.22
	W	5.4	24	43.6	25.3	766	2.2	545.8	451.2	0.67	1.48
		N.S.	N.S.	0.005	0.004	0.008	0.005	0.019	N.S.	<0.001	0.026
10 – 20	PR	5.6	25	59.8	35.3	964	3.0	457.0	388.2	0.90	1.44
	W	5.5	27	42.5	24.8	716	2.1	505.1	442.7	0.67	1.66
		N.S.	N.S.	0.002	0.002	0.006	0.003	N.S.	N.S.	<0.001	0.030
20 – 40	PR	5.7	33	55.5	32.2	812	2.7	920.3	843.3	0.85	1.56
	W	6.0	32	40.4	23.7	615	2.0	850.6	883.2	0.68	1.64
		N.S.	N.S.	<0.001	0.001	0.004	0.001	N.S.	N.S.	<0.001	0.099
40 – 60	PR	6.0	32	51.1	29.1	660	2.5	851.3	867.3	0.80	1.64
	W	6.4	30	38.3	22.7	515	1.9	795.3	830.6	0.70	1.66
		N.S.	N.S.	0.002	0.012	0.025	0.010	N.S.	<0.001	<0.001	N.S.
60 – 80	PR	6.2	32	46.8	26.0	508	2.2	842.7	884.2	0.75	1.74
	W	6.6	32	36.1	21.7	414	1.8	827.2	911.0	0.71	1.68
		N.S.	N.S.	0.043	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
80 – 100	PR	6.5	31	42.4	22.9	357	1.9	739.3	850.3	0.69	1.66
	W	6.8	33	34.0	20.6	314	1.7	826.2	944.6	0.72	1.64
		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.

* The first two rows at each depth display the least squares means of the fives sites; the third row contains the p-value of those means. N.S. = not statistically significant ($\alpha = 0.10$); SOM = soil organic matter; SOC = soil organic carbon; ROC = readily-oxidizable carbon; Total P = total phosphorus; Total K = total potassium; WSA = water stable aggregates; BD = bulk density; PR = prairie; W = wheat.

Glover et al., unpublished). If the annual mowing of the prairie resulted in significant degradation in soil fertility, yields would be expected to decline over time. These data are consistent with other reports of long-term harvests of unfertilized grasslands.

Shortridge (Shortrid 1973) found no decline in hay yields were reported from unfertilized Kansan prairies after 55 years of annual harvest. Similarly, unfertilized grasslands at the Rothamsted Research Park Grass experiment have been hayed twice-annually for 150 years without experiencing yield declines (Jenkinson et al. 1994, Silvertown et al. 1994) or reductions in total soil N over the last 120 years (Jenkinson et al. 2004). Fifty years or more of annual, unfertilized grass harvests did not reduce SOC or Total N in the upper 2 meters of soil compared to non-harvested grasslands in a Russian Chernozem (Mikhailova et al. 2000, Mikhailova and Post 2006).

Soil physical properties were also influenced by management history. Bulk density and water stable aggregates (WSA), important physical indicators of soil health and plant root activity, were significantly reduced in annual wheat fields compared to perennial grasslands. These stark differences in soil fertility—cumulative artifacts of decades of ecological processes—suggest fundamental differences in plant community functioning, nutrient cycling and associated soil biology between these two systems.

Microbial Biomass. Microbial biomass carbon (MBC) was measured at the June 2007 sampling date. Perennial grasslands exhibited significantly greater amounts of MBC than annual wheat fields at all depths, except 80 – 100cm (Figure 2.1). Values of MBC in the surface soils of grasslands are typically at least twice as great as in the surface of cultivated fields (Jenkinson and Powelson 1976, Lynch and Panting 1980b, a, Schimel et al. 1985, Acosta-Martinez et al. 2007), but few studies have reported MBC through multiple depths. In a restored prairie chronosequence, Allison et al. (Allison et al.

2007a) found microbial biomass (measured by phospholipid fatty acid (PLFA) analysis) to decrease with depth and time since prairie restoration began. At 25cm and below, no differences were detected.

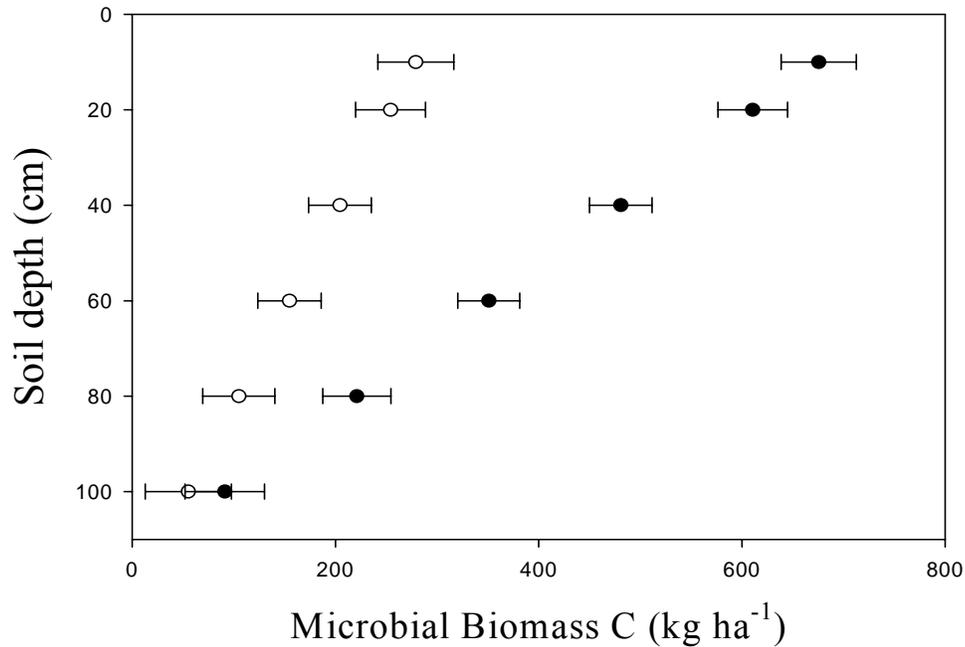


Figure 2.1. Relationship of microbial biomass carbon (MBC) to depth in prairie sites (closed circles) and annual wheat sites (open circles). Significant differences were detected at all depths except 80 – 100cm.

Microbial biomass is a chief component of the active SOM pool (Smith and Paul 1990) and has been shown to be strongly correlated with root biomass and SOM lability across many ecosystems (Wardle 1992, Paterson 2003). It has been used as a measure of belowground resource availability (Waldrop et al. 2006), an indirect measurement of belowground inputs from plants via root exudation and rhizodeposition. The greater MBC in the prairie sites coincides with the greater root

biomass (Figure 2 from Glover et al, unpublished), suggesting greater availability of resources due to root activity (exudation and rhizodeposition). Buyanovsky et al. (Buyanovsky et al. 1987) compared decomposition rates and soil organic matter accumulation in annual wheat vs. native prairie. They reported CO₂ losses from litter decay in wheat fields twice as great as prairie sites and estimated belowground biomass decay constants almost three times as great in wheat fields. They attributed these differences to several factors including greater oxidative potential from tillage and higher soil temperatures in wheat fields, and the separation in time of C accumulation (spring) and C mineralization (summer) in wheat. They suggested the greater synchrony of plant and microbial activity in the prairie likely leads to increased competition for nutrients in the summer, especially N and water (see water use data, Glover et al., unpublished, Figure 2.3, in supplementary) and ultimately lowers decomposition potential in these systems. However, greater MBC in prairie sites could also be attributed to the increased diversity, as increased plant diversity has been shown to increase microbial biomass (Zak et al. 2003, Waldrop et al. 2006). In addition, MBC is known to fluctuate seasonally (Wardle 1992, Steenwerth et al. 2006). More measurements over the growing season would be needed to make this assessment more robust.

Isotope ratio mass spectrometry was used to determine $\delta^{13}\text{C}$ values of the microbial biomass. This approach takes advantage of the difference in photosynthetic pathways (and resulting $\delta^{13}\text{C}$ values) between C₄-dominated prairies and C₃ annual wheat. Since prairies dominated this region for the last 10,000 years or more, SOC should be primarily derived from C₄ photosynthesis. However, since the conversion of prairies into agricultural fields, C₃ annual wheat has historically been the primary crop grown in this region. Hence, new additions of C in annual fields should have C₃ $\delta^{13}\text{C}$

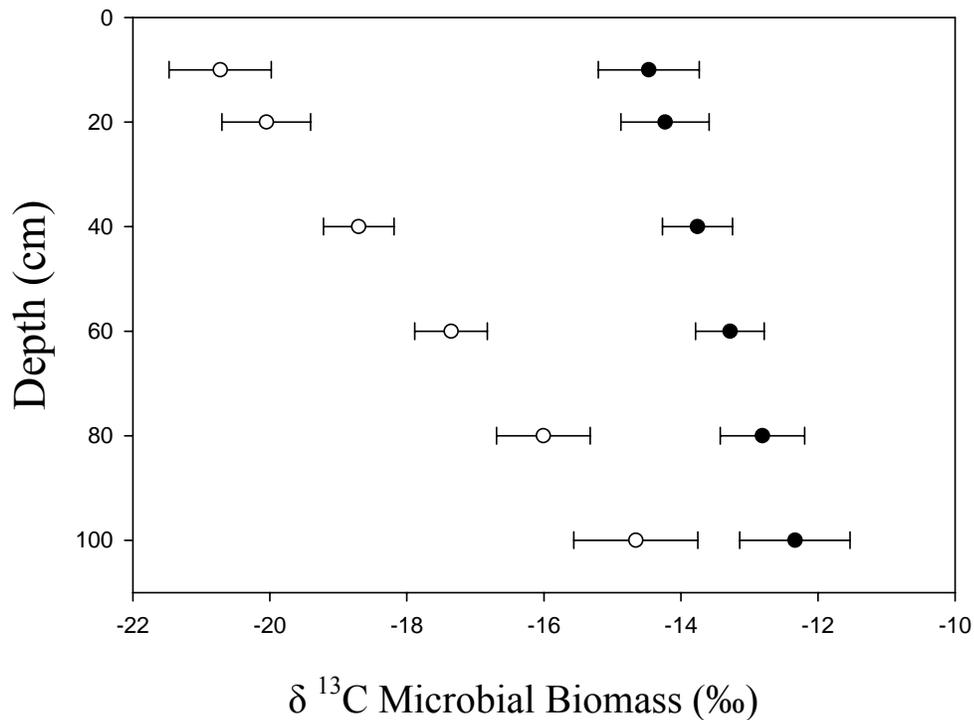


Figure 2.2. Relationship of $\delta^{13}\text{C}$ microbial biomass to depth in prairie sites (closed circles) and annual wheat sites (open circles). Significant differences were detected at all depths.

signatures (roughly -26‰), relative to the background of C4 $\delta^{13}\text{C}$ values (roughly -13‰).

Figure 2.2 illustrates the $\delta^{13}\text{C}$ values for MBC. In the surface depths, clear differences in $\delta^{13}\text{C}$ values exist between prairie and annual wheat, indicating the MBC was derived from the different photosynthetic pathways in these two systems. The C4-dominated prairies yielded $\delta^{13}\text{C}$ values much higher than the C3 annual wheat fields. However, $\delta^{13}\text{C}$ values from deeper profiles in the wheat fields increasingly become more enriched in ^{13}C , indicating the MBC at these depths was derived increasingly more from C4 photosynthesis. This trend suggests two phenomena. First, very little

C3-derived carbon from annual wheat is reaching these lower soil profiles. The wheat root biomass data reported by Glover et al. (unpublished) supports this, by showing a lack of rooting activity at these lower depths. Since isotopic signatures of microbial biomass have been shown to shift within weeks after the incorporation of litter with a different signature (Gregorich et al. 2000, Potthoff et al. 2003, John et al. 2004), $\delta^{13}\text{C}$ of microbial biomass is generally viewed to reflect the signature of the most recent input (Dijkstra et al. 2006). Second, the more enriched $\delta^{13}\text{C}$ values in wheat field MBC at these lower depths suggests this carbon source was derived from SOC sequestered by C4 prairie plants prior to agricultural conversion. This phenomenon would help explain the lower SOC values measured in annual wheat fields relative to prairie sites—heterotrophic bacteria are mineralizing remnant SOC that is not being replaced through plant deposition, resulting in a net SOC loss in the annual wheat fields.

16S rRNA gene T-RFLP. Soil bacterial community composition was characterized with terminal restriction fragment length polymorphism (T-RFLP) (Liu et al. 1997), and summarized by nonmetric multidimensional scaling (NMS). NMS analysis of the complete dataset revealed sampling date and depth to be the two largest drivers in general bacterial structure. Differences between years (June 2007 and June, October 2006) were the most consistent differences observed (Figure 2.4a). Depth also appears to be a large driver of structure, although there was great variability between depth profiles (Figure 2.4b). Differences between management history and site were not observed when the entire dataset was analyzed (Figure 2.4c and 2.4d).

<p style="text-align: center;"><u>Sampling Date</u></p> <p>◇ June 2006 ○ October 2006 + June 2007</p>	<p style="text-align: center;"><u>Depth</u></p> <p>● 0–10cm ◆ 20–40cm ■ 10–20cm ► 60–80cm ▼ 40–60cm ◆ 80–100cm</p>
<p style="text-align: center;"><u>Management History</u></p> <p>△ Annual Wheat ▲ Perennial Grassland</p>	<p style="text-align: center;"><u>Site</u></p> <p>⊠ Niles ⊙ Buckeye ⊗ Five Creek ⊕ New Cambria × Goessel</p>

Figure 2.3. Legend of symbols used to depict T-RFLP samples in nonmetric multidimensional scaling (NMS) analyses.

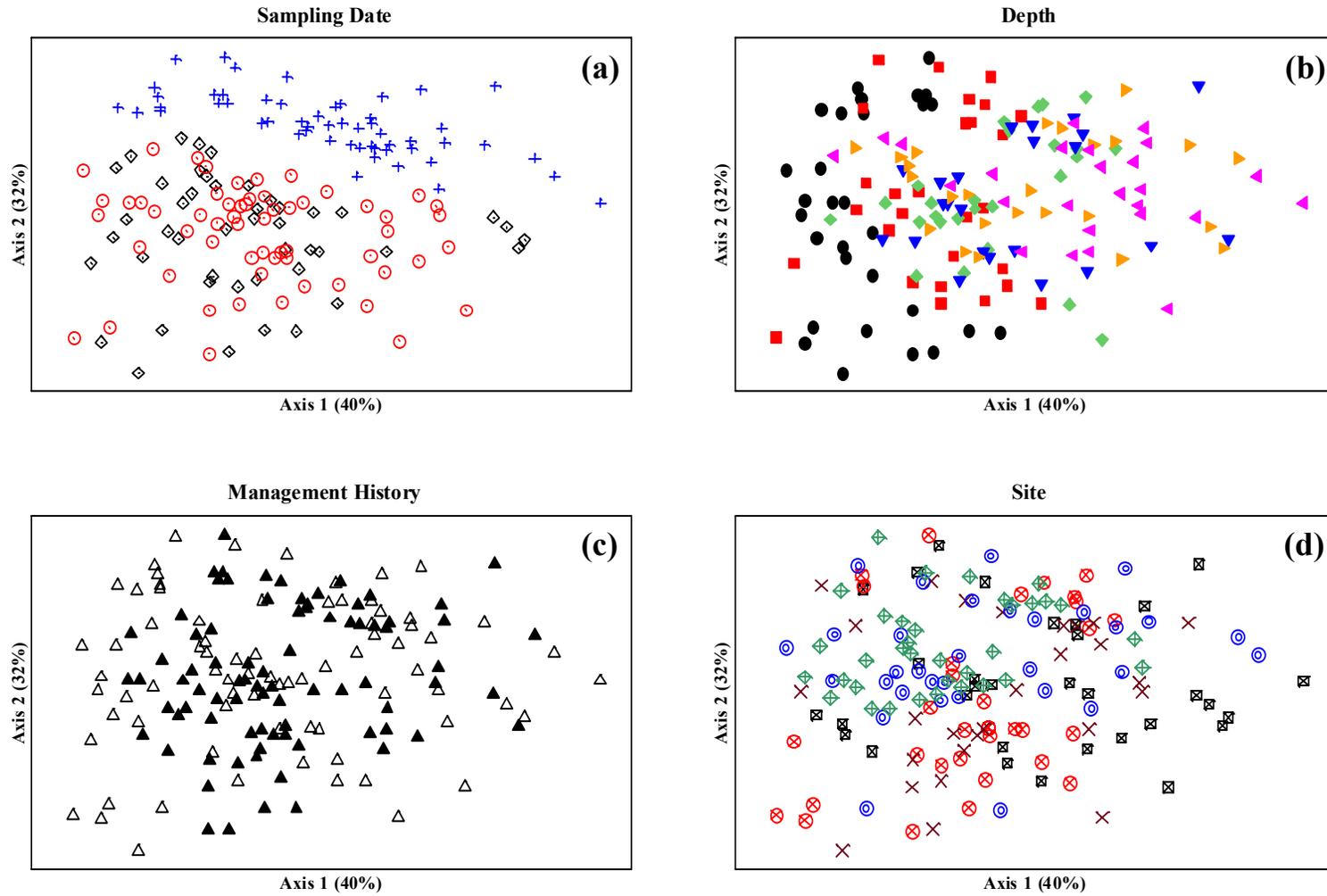


Figure 2.4. NMS analysis of complete bacterial T-RFLP dataset (June 2006, October 2006, and June 2007). The four panels display the same data represented by the four experimental factors: sampling date (a), depth (b), management history (c), and site (d). Figure 2.3 contains the symbol legend.

Since sampling date and depth were the largest drivers of community structure in our experiment, we attempted to control for, or eliminate some of this variation in order to reveal more subtle patterns pertaining to management history. Depth was controlled for first, by decomposing the dataset into two separate matrices: one with the top 3 depths (0 – 40cm) and the other with the bottom 3 depths (40 – 100cm). This resulted in similar ordinations produced with all depths (data not shown), and proved no more discriminatory in respect to management history. However, when NMS analysis parameters were set to produce a 3-dimensional solution instead of a 2-dimensional solution, differences were observed in regards to both management history and site. A 3-dimensional solution of the top 3 depths showed differences between sampling date (Figure 2.5a), depth (Figure 2.5b) and management history (Figure 2.5c). A 3-dimensional solution of the bottom 3 depths showed sampling date and site to be major drivers with neither depth nor management history grouping consistently (data not shown). Even though a 3-dimensional solution proved to be more discriminatory than 2-dimensional solution with these data, it is much more conventional to report T-RFLP data in 2-dimensions. This convention likely results from the difficulty in accurately representing 3 axes on a 2-dimensional surface, among other things. In light of this, 2-dimensional solutions were sought whenever possible.

Completing eliminating the factor of depth from the analysis, by analyzing only individual depths revealed that management history drives community structure at 0 – 10cm (Figure 2.6) and 10 – 20cm (data not shown). Differences in management history at lower depths were not consistently observed. Detecting differences between

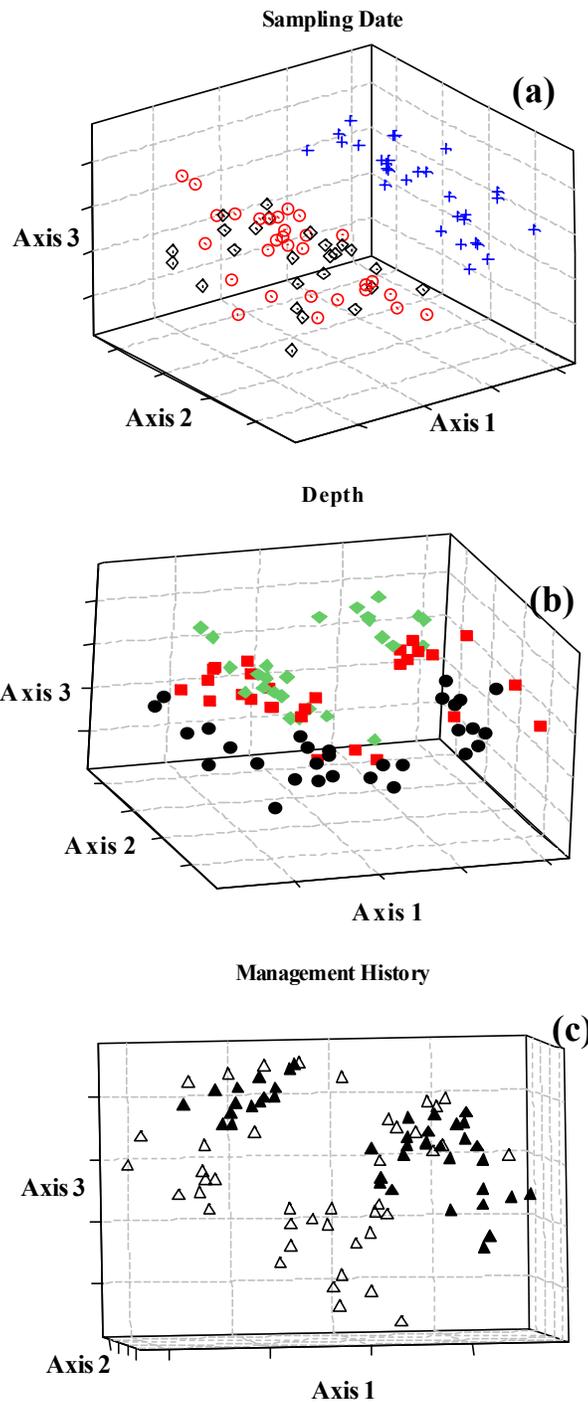


Figure 2.5. NMS analysis of the surface 3 depths (0 – 40cm) in the complete bacterial T-RFLP dataset (June 2006, October 2006, and June 2007). The three panels display the same data represented by three experimental factors: sampling date (a), depth (b), and management history (c). Figure 2.3 contains the symbol legend.

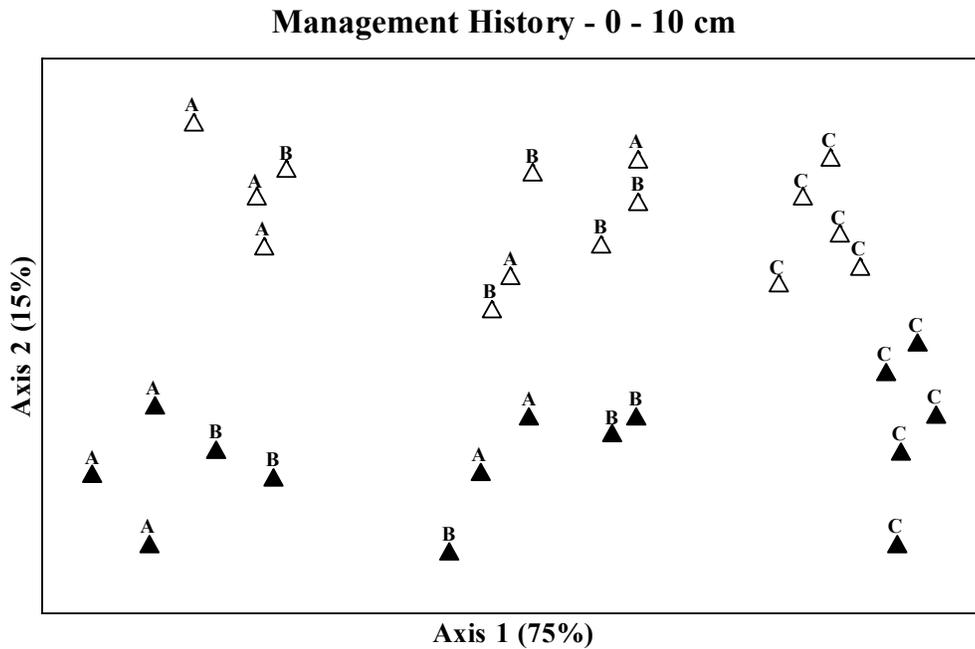


Figure 2.6. NMS analysis of the 0 – 10cm depth in the complete bacterial T-RFLP dataset (June 2006, October 2006, and June 2007) showing differences between annual wheat fields (open triangles) and perennial grasslands (closed triangles) with respect to sampling date (A = June 06; B = Oct 06; C = June 07).

management histories were made possible only when completely or partially eliminating variability from depth. This suggests that sampling time and depth are larger drivers of community structure than management history and site. Sampling time appeared to drive structure at all depths, while differences in depth were strongest in the surface (0 – 40 cm).

Bacterial community structure differences between sampling dates were controlled by analyzing each sampling date individually. These 3 analyses (June 2006, Oct 2006 and June 2007) revealed the same general patterns between all three dates. Only June 2007 data are presented here. NMS analyses of June 2007 bacterial communities show that removing the effect of sampling date reveals more consistent groupings with

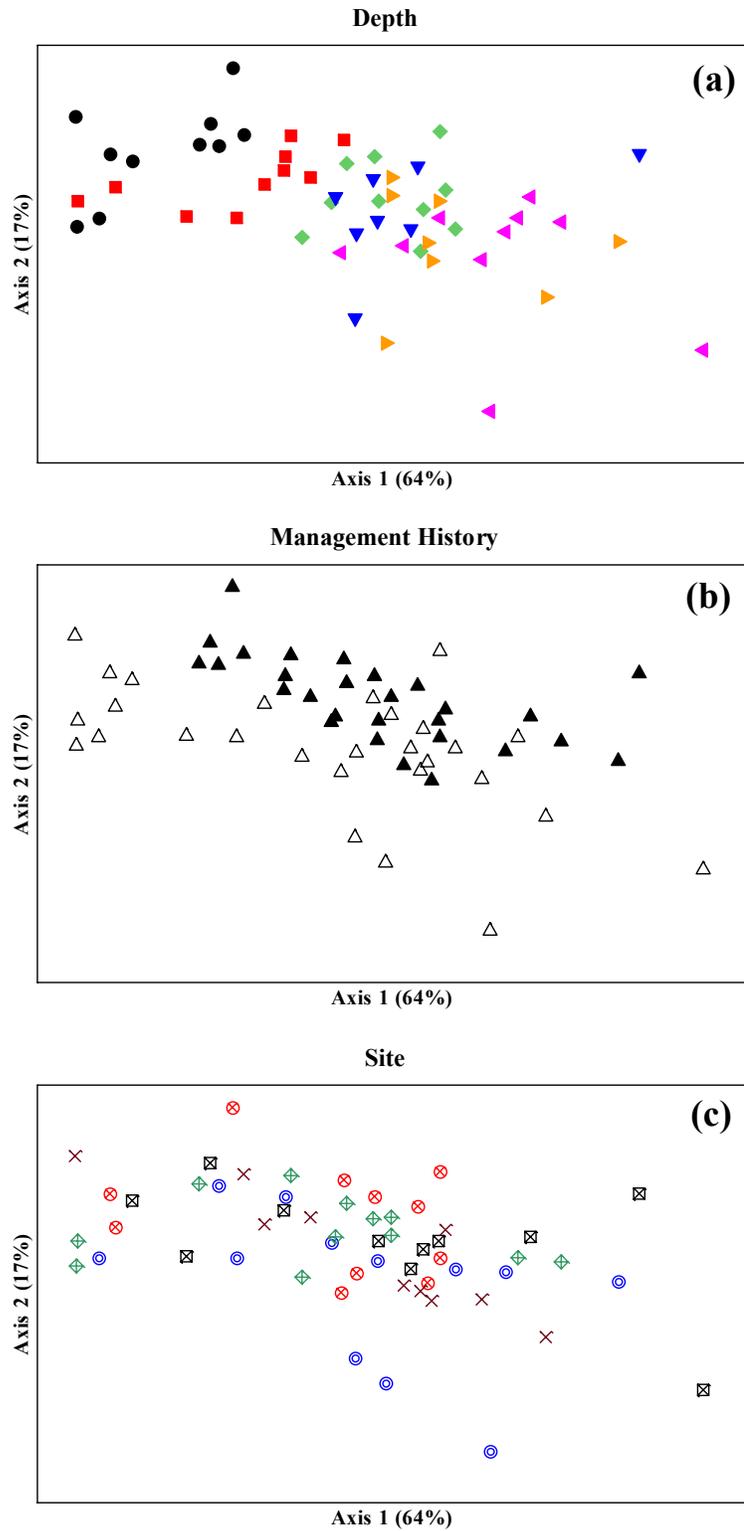


Figure 2.7. NMS analysis of the June 2007 bacterial T-RFLP dataset. The three panels display the same data represented by three experimental factors: depth (a), management history (b), and site (c). Figure 2.3 contains the symbol legend.

depth than when all dates are analyzed together (Figure 2.7a), but also that management history has a profound effect on community structure (Figure 2.7b). Consistent differences in site were not detected here (Figure 2.7c), indicating that depth and management history are the primary drivers in community structure when sampling date variability is eliminated. When the June 2007 dataset was decomposed into upper and lower depths, NMS analyses yield results consistent with the above findings—that depth and management history drive structure in the upper depths (Figure 2.8a) and that site differences drive structure in the lower depths (Figure 2.8b). NMS analyses of individual depths of the June 2007 dataset (i.e., analysis after the factors of sampling time and depth have been removed), show detectable differences in both site and management history in the upper 3 depths (only 0 – 10cm shown in Figure 2.9), while only differences in site were detected in the lower 3 depths (data not shown).

A second approach to reduce the effect of sampling date on the ordinations was to simply average sampling dates for each sample. Ordinations from datasets with averaged sampling dates yielded results that were similar to non-averaged complete datasets, where trends in depth and management histories were observed but variable (data not shown). However, averaging sampling dates resulted in ordinations with groupings based on site differences more consistent than with previous approaches. Site differences were most pronounced in lower depths (Figure 2.10).

The final approach taken to analyze general bacterial community structure was to average both site and sampling date. This effectively reduced dataset variability due to these two factors and allowed for an analysis to focus on the two factors of primary

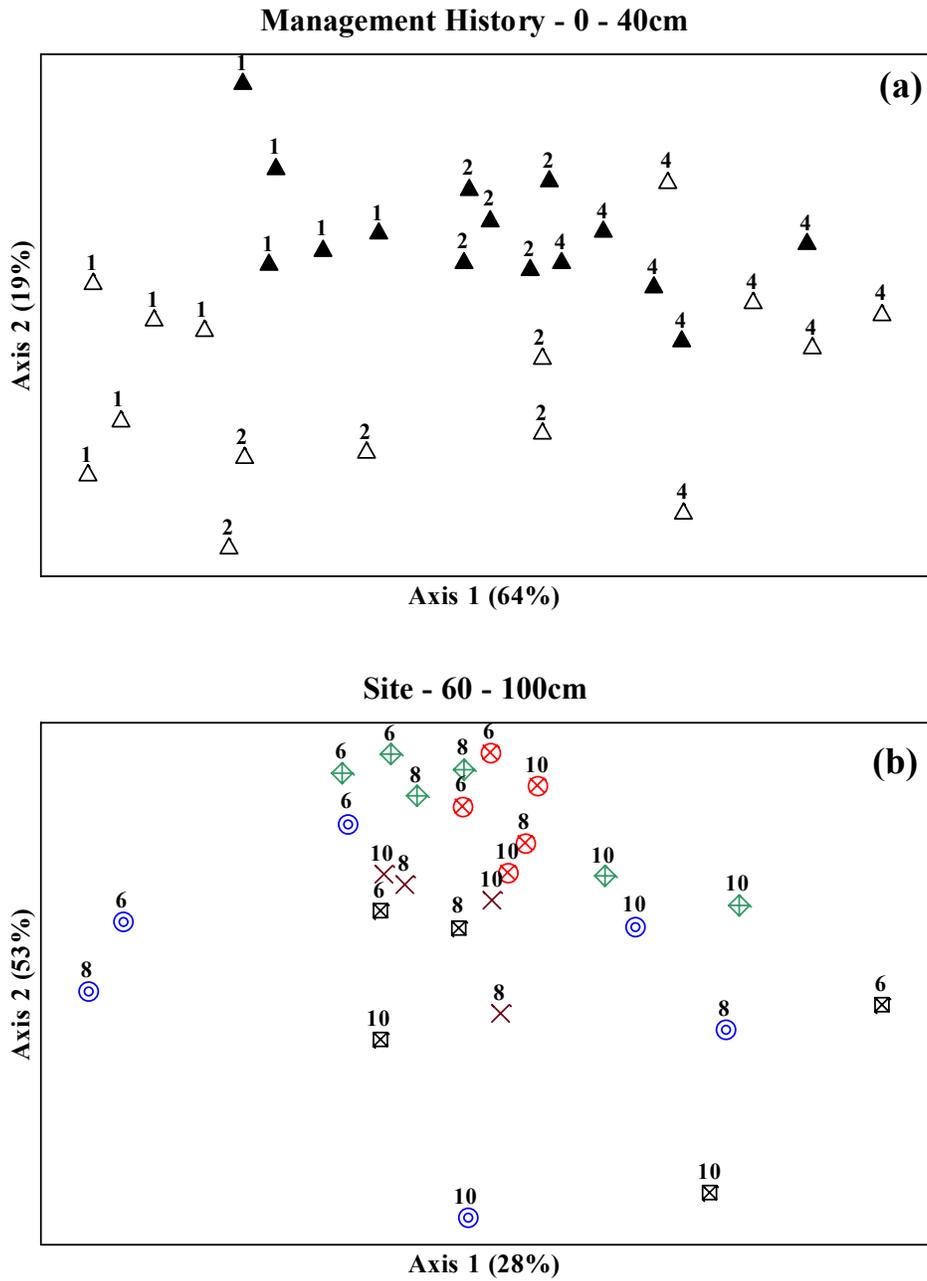


Figure 2.8. NMS analyses of the surface 3 depths (a) showing differences in management history and of the bottom 3 depths (b) showing differences in site in the June 2007 bacterial T-RFLP dataset. 1 = 0 – 10cm; 2 = 10 – 20cm; 4 = 20 – 40cm; 6 = 40 – 60cm; 8 = 60 – 80cm; 10 = 80 – 100cm. Figure 2.3 contains the symbol legend.

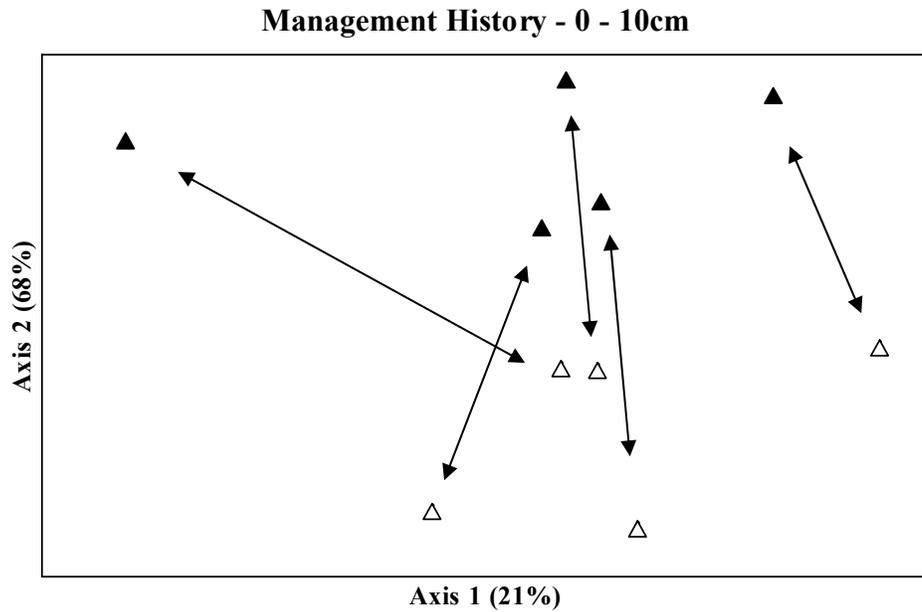


Figure 2.9. NMS analysis of the 0 – 10cm depth in the June 2007 bacterial T-RFLP dataset showing differences between annual wheat fields (open triangles) and perennial grasslands (closed triangles) with respect to site. Arrows point to samples with different management histories, but from corresponding sites.

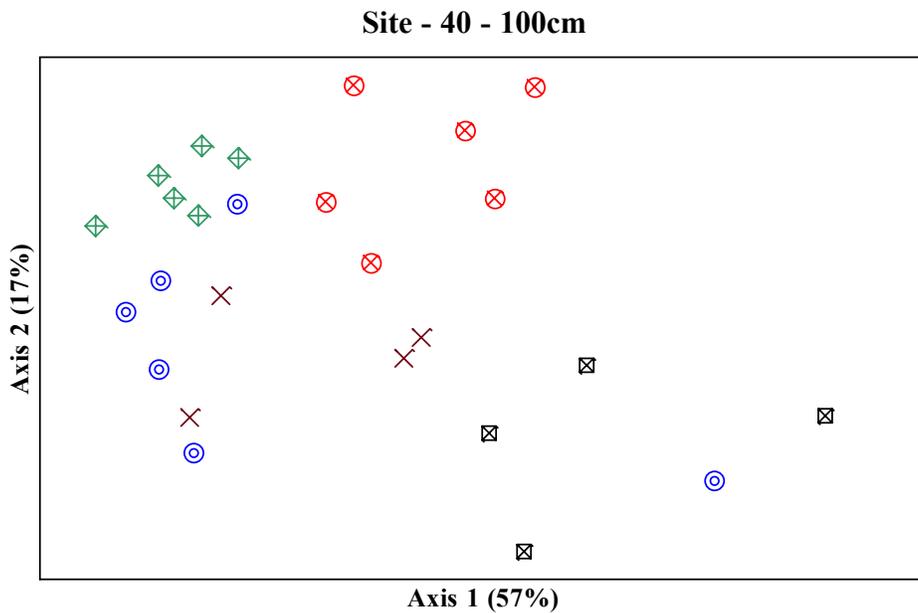


Figure 2.10. NMS analysis of the 40 – 100cm depths in the complete bacterial T-RFLP dataset averaged by date showing differences between sites. Figure 2.3 contains the symbol legend.

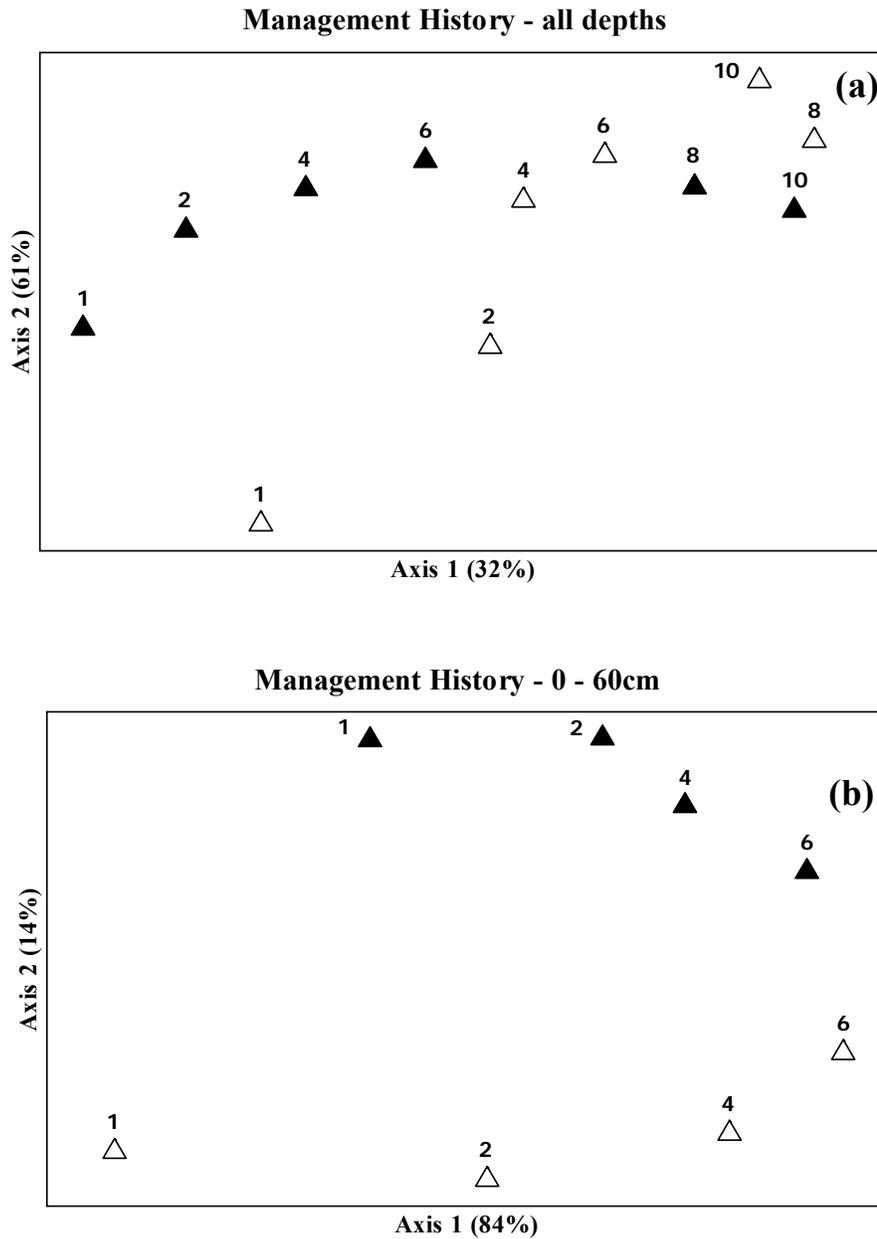


Figure 2.11. NMS analyses of the complete bacterial T-RFLP dataset averaged by both sampling date and site showing differences between annual wheat fields (open triangles) and perennial grasslands (closed triangles). Top panel (a) shows analysis with all depths (0 – 100cm); bottom panel (b) shows analysis with top 4 depths (0 – 60cm). 1 = 0 – 10cm; 2 = 10 – 20cm; 4 = 20 – 40cm; 6 = 40 – 60cm; 8 = 60 – 80cm; 10 = 80 – 100cm.

interest—depth and management history. Ordinations produced from these datasets show the effect of both depth and management history in shaping community structure. Trends with management history are consistent to 40 – 60cm depth, after which management history appears to have little effect (Figure 2.11a). Removing the bottom 2 depths and analyzing the top 4 soil depths indicates the effect of management history on communities to a depth of 60cm (Figure 2.11b). With variability from site and sampling date minimized by averaging, differences in management history down to a depth of 60cm are consistent with the differences found in C and N data in Table 2.1. This confirms an expected relationship, as bacterial populations have been found to be strongly shaped by the quantity and quality of soil carbon and nitrogen pools (Paul and Clark 1996).

MRPP was used to test significance in the datasets with all sampling dates (Table 2.2) and with the June 2007 dataset (Table 2.3). These datasets were also analyzed when decomposed into upper and lower 3 depths and at each individual depth. MRPP results largely confirm findings of the NMS analyses—that differences in depth and management history strongly influence bacterial community structure at the surface depths, and that site differences drive structure at lower depths. However NMS analyses and their interpretations were much more conservative than the results yielded from MRPP. For example, MRPP results in Table 2.2 shows statistically significant differences between all four factors when all 6 depths are analyzed simultaneously. (These results are based on the same data used to produce the ordinations in Figure 2.4.) However, these NMS ordinations show strong trends only with depth and sampling date, not with management history or site. This example exhibited the most discordance among these two classes of analyses, but with large datasets, MRPP often yielded significant differences, particularly in regard to site, that

Table 2.2. MRPP analysis of bacterial community differences- June '06, Oct '06, June '07

Dataset Section Analyzed	Enzyme	Depth		History		Site		Date	
		A	p-value	A	p-value	A	p-value	A	p-value
All Depths	<i>HhaI</i>	0.086	<0.001	0.010	<0.001	0.040	<0.001	0.087	<0.001
	<i>Sau96 I</i>	0.077	<0.001	0.015	<0.001	0.053	<0.001	0.082	<0.001
10-40cm	<i>HhaI</i>	0.070	<0.001	0.037	<0.001	0.027	0.001	0.110	<0.001
	<i>Sau96 I</i>	0.059	<0.001	0.038	<0.001	0.049	<0.001	0.113	<0.001
40-100cm	<i>HhaI</i>	0.012	0.021	0.001	0.317	0.093	<0.001	0.091	<0.001
	<i>Sau96 I</i>	0.016	0.011	0.004	0.137	0.112	<0.001	0.094	<0.001
10cm	<i>HhaI</i>	-	-	0.106	<0.001	-0.015	0.750	0.125	<0.001
	<i>Sau96 I</i>	-	-	0.072	<0.001	0.015	0.254	0.157	<0.001
20cm	<i>HhaI</i>	-	-	0.037	0.004	0.004	0.384	0.135	<0.001
	<i>Sau96 I</i>	-	-	0.058	0.001	0.055	0.024	0.125	<0.001
40cm	<i>HhaI</i>	-	-	0.016	0.087	0.070	0.008	0.178	<0.001
	<i>Sau96 I</i>	-	-	0.019	0.074	0.072	0.009	0.170	<0.001
60cm	<i>HhaI</i>	-	-	0.001	0.380	0.091	0.002	0.121	<0.001
	<i>Sau96 I</i>	-	-	-0.007	0.639	0.188	<0.001	0.103	<0.001
80cm	<i>HhaI</i>	-	-	-0.008	0.777	0.087	0.001	0.068	0.001
	<i>Sau96 I</i>	-	-	0.004	0.285	0.116	0.001	0.091	0.001
100cm	<i>HhaI</i>	-	-	-0.011	0.882	0.081	0.001	0.106	<0.001
	<i>Sau96 I</i>	-	-	-0.005	0.640	0.062	0.007	0.093	<0.001

Table 2.3. MRPP analysis of bacterial community differences- June '07

Dataset Section Analyzed	Enzyme	Depth		History		Site*	
		A	p-value	A	p-value	A	p-value
All Depths	<i>HhaI</i>	0.149	<0.001	0.061	<0.001	0.061	<0.001
	<i>Sau96 I</i>	0.170	<0.001	0.022	0.004	0.034	0.006
10-40cm	<i>HhaI</i>	0.128	<0.001	0.060	<0.001	0.038	0.047
	<i>Sau96 I</i>	0.155	<0.001	0.059	0.059	0.016	0.226
40-100cm	<i>HhaI</i>	-0.007	0.616	0.004	0.304	0.200	<0.001
	<i>Sau96 I</i>	0.014	0.177	0.008	0.212	0.134	<0.001
10cm	<i>HhaI</i>	-	-	0.200	0.002	-	-
	<i>Sau96 I</i>	-	-	0.175	0.002	-	-
20cm	<i>HhaI</i>	-	-	0.099	0.004	-	-
	<i>Sau96 I</i>	-	-	0.136	0.006	-	-
40cm	<i>HhaI</i>	-	-	0.069	0.048	-	-
	<i>Sau96 I</i>	-	-	0.082	0.013	-	-
60cm	<i>HhaI</i>	-	-	-0.022	0.737	-	-
	<i>Sau96 I</i>	-	-	-0.038	0.868	-	-
80cm	<i>HhaI</i>	-	-	-0.056	0.916	-	-
	<i>Sau96 I</i>	-	-	-0.010	0.609	-	-
100cm	<i>HhaI</i>	-	-	-0.004	0.470	-	-
	<i>Sau96 I</i>	-	-	0.010	0.342	-	-

* Values for site at individual depths were not determined, as MRPP analysis requires groups to have more than 2 items (wheat vs. grassland).

were not supported by NMS ordinations. In these instances, the more conservative NMS analyses were given greater weight in overall data interpretation.

In this study, four factors were examined in regard to bacterial community composition—sampling date, depth, management history and site—all four of which have been shown to alter communities in published literature. Bacterial community structure has been shown to change due to sampling date (Feng et al. 2003, Culman et al. 2006, Rumberger et al. 2007), management practices and management history (Drijber et al. 2000, Hedlund 2002, Feng et al. 2003, Gomez et al. 2004, Acosta-

Martinez et al. 2007), and differences in site (Hackl et al. 2004, McCulley and Burke 2004, Ulrich and Becker 2006). Other studies have found no detectable differences in that above factors (Mummey and Stahl 2003, Kennedy et al. 2005), suggesting that factors driving bacterial community dynamics are dependant on site characteristics and the relative strength of the measured environmental gradient/s.

Although rarely explored in most studies, depth has also been shown to dramatically affect microbial community structure (Kuske et al. 2002, Feng et al. 2003, Fierer et al. 2003, Allison et al. 2007a). Deeper soil profiles have been characterized by lower abundances of fungi compared to bacteria and higher abundances of actinomycetes and Gram-positive bacteria compared to Gram-negative bacteria (Zelles and Bai 1994, Feng et al. 2003, Fierer et al. 2003, Allison et al. 2007b). Only a few published studies have simultaneously looked at differences in microbial community structure with depth and site differences (Fierer et al. 2003) or management histories (Allison et al. 2007b). These studies have found depth to be a stronger determinate of community structure than site or management differences.

To our knowledge, this is the first study that has examined microbial community data with four factors simultaneously, making it particularly insightful in regards to factors shaping bacterial communities in these systems. All four factors were shown to exhibit influence on community structure, although the relative strengths of these factors varied. In general, bacterial dynamics in the top 40cm of soil were driven by sampling date > depth > management history > site; the lower depths (40 – 100cm) were driven by sampling date > site > depth > management history.

***nifH* T-RFLP**

T-RFLP analysis was also used to target *nifH* genes in soil, the functional enzyme found in bacteria that fix atmospheric nitrogen into ammonia. A slight modification was made for *nifH* T-RFLP analysis, where both the forward and reverse primers were labeled with different fluorophores (fluor). (This differs from the more traditional approach of just using one labeled-primer set). This was done in an attempt to increase resolution and overall information pertaining to the nitrogen-fixing community. T-RFLP analyses with two different fluorescently-labeled primers have been reported in the past with studies that claim this approach increased information and community resolution (Nilsson and Strom 2002, Gruter et al. 2006). However these reports failed to provide specifics on how their reported results from analyses with dual-labeled primer sets differed from results from analyses with just one labeled primer set.

To address such a comparison, *nifH* T-RFLP datasets were analyzed three separate ways: on datasets with the forward FAM-labeled primer only, on datasets with the reverse PET-labeled primer only, and on datasets containing both FAM and PET-labeled primers. These analyses showed that the total number of *nifH* terminal restriction fragments (T-RFs) in a sample (richness) was affected by which primer (forward or reverse) contained the fluor (Table 2.4). *nifH* T-RF richness with PET was significantly greater than with FAM, indicating that the region of the *nifH* gene closest to the reverse primer contained overall greater sequence variability (in regard to MspI restriction sites) than the region closer to the forward (FAM) primer.

Table 2.4. Average *nifH* T-RF richness with different fluors.

Depth	FAM	PET
0 – 10cm	19	33
10 – 20cm	20	30
20 – 40cm	16	23
40 – 60cm	18	26
60 – 80cm	23	30
80 – 100cm	17	23

Despite differences in *nifH* T-RF richness, NMS analyses of datasets with the three possible fluor combinations produced largely similar ordinations, indicating that *nifH* community structure was not greatly affected by fluor-primer selection. Although overall interpretations did not change between the fluor combinations, NMS analyses on FAM datasets were generally the most inconsistent in regard to known experimental factors. PET datasets were generally the most consistent and datasets with both fluors together were intermediate (data not shown). These empirical findings are supported by the distribution of variance within these datasets. Table 2.5 shows results from analysis of variance with PET having the highest relative contribution of variance from interaction signal and the highest interaction signal: noise ratio. FAM had the lowest interaction signal and signal: noise ratio and FAM + PET was intermediate. In this study, using dual fluorescently-labeled primers with *nifH* T-RFLP proved to be no more discriminatory than using a single fluor, although the selection of the labeled primer (forward/reverse) made a small, but insignificant difference. As a result of these findings, ordinations of *nifH* T-RFLP data reported here are with the PET fluor only.

Table 2.5. Distribution of variance in *nifH* dataset with different fluors.

Source of Variance	FAM	PET	FAM + PET
	-----%-----		
Main Effects			
T-RFs	73.5	68.5	71.0
Environments	0.9	1.2	0.9
Interaction Effects			
Interaction Signal	3.8	5.3	4.7
Interaction Noise	21.8	25.0	23.4

Differences in *nifH* T-RF richness between annual wheat and perennial grasslands were not found in any of the fluor combinations (Table 2.6). The only significant factor was that *nifH* T-RF richness was significantly affected by depth ($p = 0.046$) in the PET dataset. However, despite a lack of difference in T-RF richness between management histories, community structure between management histories was significantly different.

Table 2.6. Average *nifH* T-RF richness with different fluors and management histories.

	Annual Wheat		Perennial Grassland	
Depth	FAM	PET	FAM	PET
0 – 10cm	18	30	20	35
10 – 20cm	19	27	21	31
20 – 40cm	19	24	14	23
40 – 60cm	18	24	19	27
60 – 80cm	24	31	21	29
80 – 100cm	15	19	18	26

Figure 2.12 shows NMS analysis of *nifH* T-RFLP at all depths, indicating trends in all three factors—depth, management history and site. Of these factors, management

history appears to be the most influential in driving *nifH* community structure (Figure 2.12b). Unlike general bacterial community structure (Figure 2.7a), differences in depth with *nifH* (Figure 2.12a) were not consistent, suggesting that depth plays a lesser role in shaping these communities. Site differences with *nifH* communities (Figure 2.12c) were also more consistent than in general bacterial communities (Figure 2.7c), indicating the relative importance of site (e.g., pH, soil texture, nutrient pools, biotic interactions) in driving *nifH* community structure.

Decomposing the *nifH* dataset into upper and lower depths reveals that differences between perennial grasslands and annual wheat are the dominant driver in nitrogen-fixing bacterial community structure in the upper depths (Figure 2.13a). Site differences also influence *nifH* community structure, but become more significant in lower depths (NMS data not shown, MRPP data in Table 2.7). Depth has no significant effect on community structure in the top 40cm (Figure 2.13, Table 2.7) or in the lower depths (Table 2.7) when the dataset is decomposed. Instead, differences between management history and site are the primary drivers in this ecologically important functional gene.

Finally, site variation was minimized by averaging over site, which enabled an analysis focused on depth and management history. The results confirmed earlier findings—that management history shapes nitrogen fixing community structure in the surface depths only and that depth plays a lesser role in *nifH* community structure than with bacterial community structure (Figure 2.14).

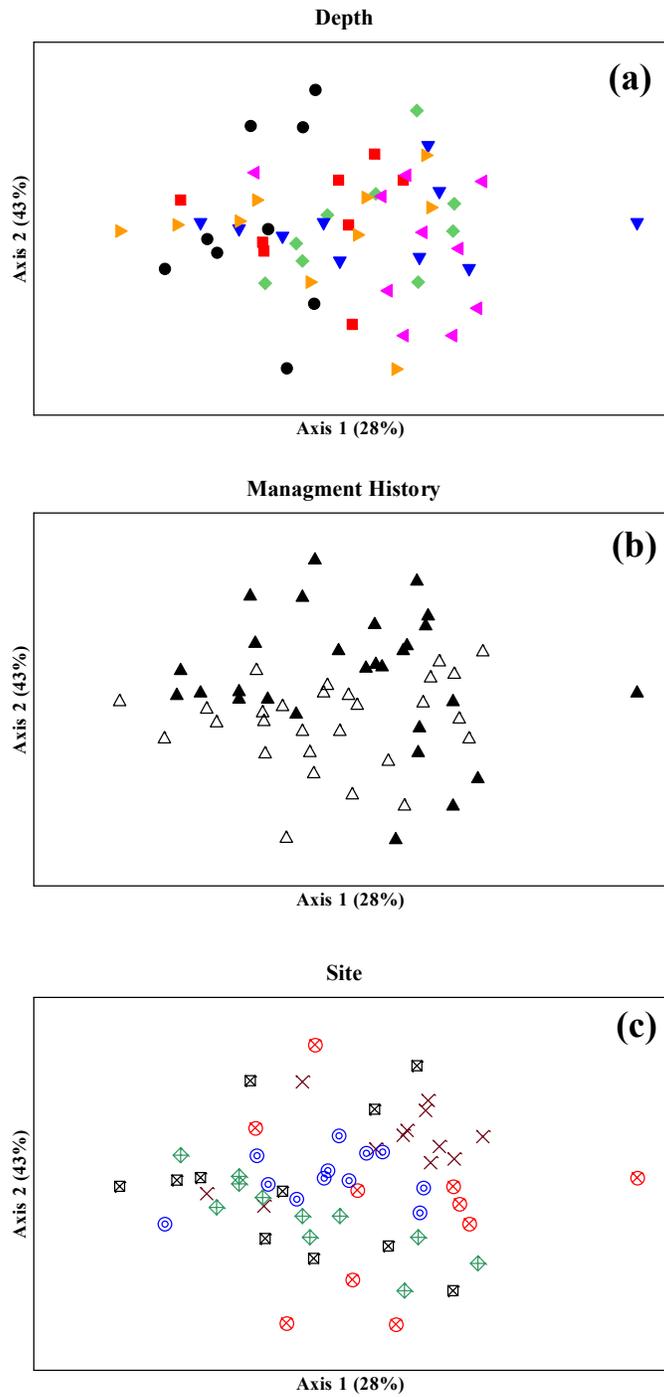


Figure 2.12. NMS analysis of the June 2007 *nifH* T-RFLP dataset. The three panels display the same data represented by three experimental factors: depth (a), management history (b), and site (c). Figure 2.3 contains the symbol legend.

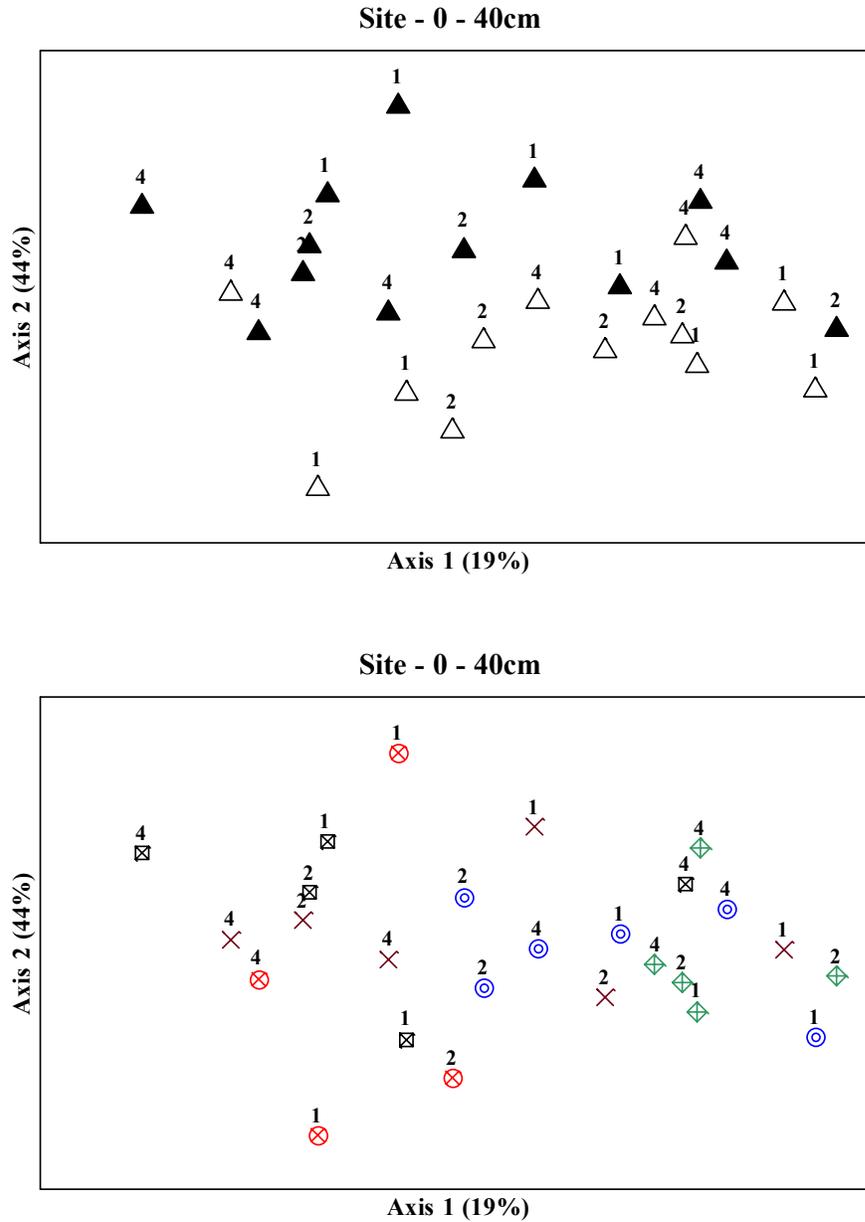


Figure 2.13. NMS analysis of the surface 3 depths of June 2007 *nifH* T-RFLP dataset. The top panel (a) shows differences in management history and the bottom panel (b) shows differences in site. 1 = 0 – 10cm; 2 = 10 – 20cm; 4 = 20 – 40cm. Figure 2.3 contains the symbol legend.

Table 2.7 MRPP analysis of nitrogen-fixing bacterial differences- June '07 (G)

Dataset Section Analyzed	Depth		Treatment		Site	
	A	p-value	A	p-value	A	p-value
All Depths	0.022	0.019	0.014	0.006	0.062	<0.001
10-40cm	0.014	0.120	0.038	<0.001	0.062	0.001
40-100cm	0.008	0.252	0.008	0.184	0.121	<0.001
10cm	-	-	0.058	0.015	-	-
20cm	-	-	0.066	0.031	-	-
40cm	-	-	-0.021	0.749	-	-
60cm	-	-	-0.029	0.746	-	-
80cm	-	-	-0.051	1.000	-	-
100cm	-	-	0.006	0.353	-	-

Unfortunately, the main factors that drive *nifH* gene diversity have yet to be identified (Zehr et al. 2003) and reports on free-living N₂ fixing communities in grasslands are few. Patra et al. (Patra et al. 2006) found that intensity of grazing, not plant species, affected N₂ fixing communities in unfertilized grasslands. In another study, intensive grazing increased N-related microbial enzyme activity and altered the composition of these groups (N₂ fixing, nitrate-reducing and ammonium oxidizing bacteria), demonstrating that differences in management can affect functional gene diversity (Patra et al. 2005).

The implications of altered *nifH* gene structure on rates of N fixation are unclear, but free-living and associative N fixation could alleviate N-limitation on non-fertilized agroecosystems (Patra et al. 2007). Various studies have reported associative N fixation with the same temperate C4 grass species found in our study (Tjepkema and Burris 1976, Morris et al. 1985, Brejda et al. 1994); however, robust N fixation rates

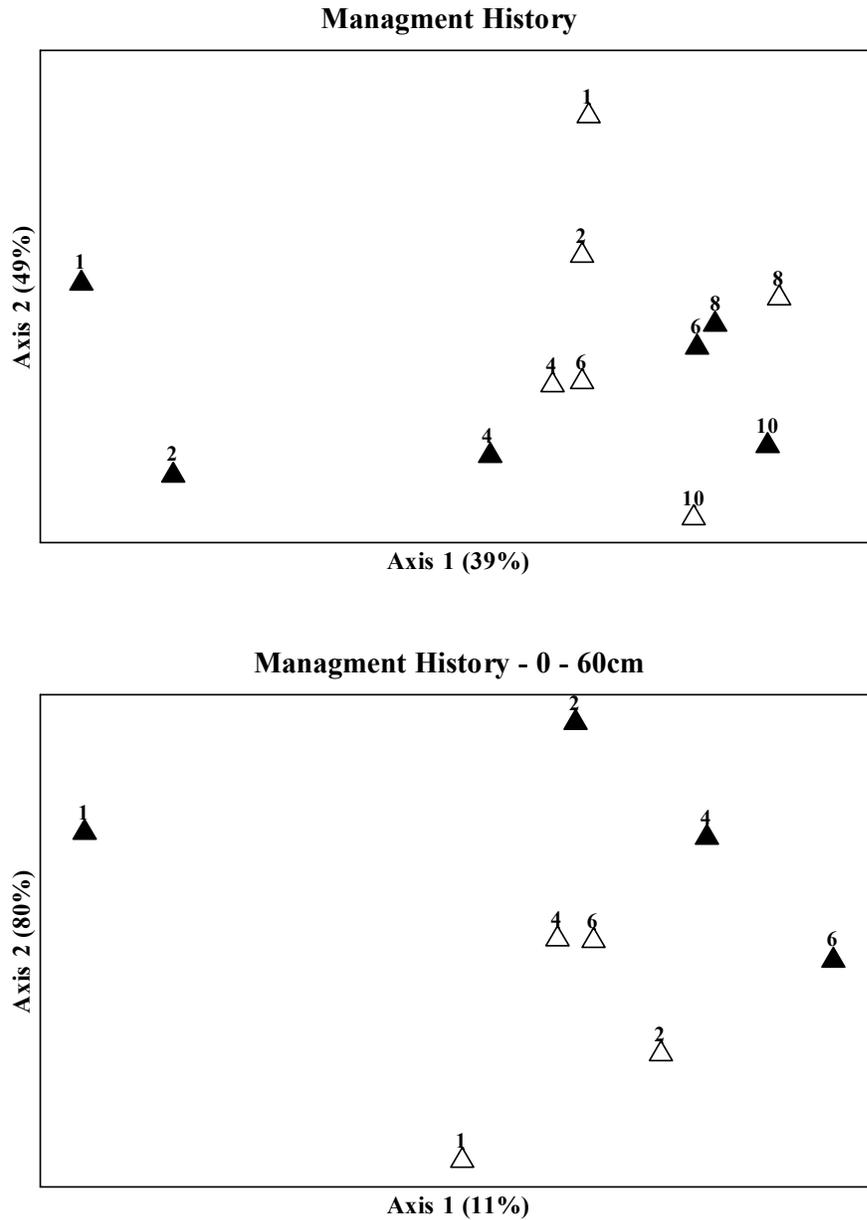


Figure 2.14. NMS analyses of the *nifH* T-RFLP dataset averaged by site showing differences between annual wheat fields (open triangles) and perennial grasslands (closed triangles). Top panel (a) shows analysis with all depths (0 – 100cm); bottom panel (b) shows analysis with top 4 depths (0 – 60cm). 1 = 0 – 10cm; 2 = 10 – 20cm; 4 = 20 – 40cm; 6 = 40 – 60cm; 8 = 60 – 80cm; 10 = 80 – 100cm.

of these grasses has yet to be quantified. Future work on linking *nifH* gene diversity with N fixation potential is warranted in these systems.

The influence of three factors—depth, management history and site—on free-living nitrogen fixing bacteria was simultaneously examined. The relative influences on these communities mostly mirrored the findings with general bacterial community structure. However the influence of depth was the major exception, it played only a minor role, relative to management history and site differences. Again, these data to our knowledge are unique in regard to insight into the effects of multiple factors in shaping nitrogen fixing bacterial communities. In general, nitrogen fixing bacterial dynamics in the top 40cm of soil were driven by management history > site > depth; the lower depths (40 – 100cm) were driven by site > management history > depth.

There were multiple differences between each treatment at every factor in this study's experimental design. For example, salient differences in management histories between annual wheat and perennial grasslands include differences in plant species, plant diversity, rooting architecture, tillage, and nitrogen fertilization. Examples of the treatment differences in the other factors include: sampling date (temperature, precipitation, crop phenology), depth (pH, soil texture, soil moisture, nutrient status), and site (microclimate patterns, soil properties). As a result, making a direct causal link with changes in microbial communities between these two systems is not possible here. Instead, this study represents the first step in elucidating differences in soil biology between these two different agricultural systems.

Collaborative work on nematode communities by Tianna DuPont and Howard Ferris at UC Davis have shown equal or greater differences in nematodes communities

relative to bacterial communities with respect to our experimental factors. In addition, their data suggest that perennial grasslands exhibit greater fungal decomposition pathways, fewer plant parasitic nematodes, and greater food web complexity and stability than found in annual wheat fields (S.T. DuPont, unpub. M.S. thesis, UC Davis, 2008).

Conclusions

A model agroecosystem would provide sufficient nutrient cycling to meet plant demands, while exhibiting the greater stability commonly found in natural ecosystems. This stability can be defined through essential ecosystem functions, such as conservation of soil fertility, tighter mineral cycling and effective pest management. Long-term yield data and current soil properties suggest these prairies are more capable of meeting plant nutrient demand *and* exhibiting greater stability than cultivated annual wheat. The soil biological data presented here show large differences in bacterial (and nematode) community structure and abundance, which suggest fundamental differences in belowground processes between prairies and wheat sites. These differences have likely contributed to the increased nutrient removal and energy efficiencies, and enhanced soil fertility found in the prairie ecosystems. Future work will be needed to tease apart these factors in order to determine the causal relationships and characteristics which are most important in driving these efficiencies.

REFERENCES

- Acosta-Martinez, V., M. M. Mikha, and M. F. Vigil. 2007. Microbial communities and enzyme activities in soils under alternative crop rotations compared to wheat-fallow for the Central Great Plains. *Applied Soil Ecology* **37**:41-52.
- Allison, V. J., Z. Yermakov, M. E. Miller, J. D. Jastrow, and R. Matamala. 2007a. Assessing Soil Microbial Community Composition Across Landscapes: Do Surface Soils Reveal Patterns? *Soil Science Society of America Journal* **71**:730-734.
- Allison, V. J., Z. Yermakov, R. M. Miller, J. D. Jastrow, and R. Matamala. 2007b. Using landscape and depth gradients to decouple the impact of correlated environmental variables on soil microbial community composition. *Soil Biology & Biochemistry* **39**:505-516.
- Assessment, M. E. 2005. *Ecosystems and Human Well-being: Synthesis*. Island Press, Washington, DC.
- Brejda, J. J., R. J. Kremer, and J. R. Brown. 1994. Indications of Associative Nitrogen-Fixation in Eastern Gamagrass. *Journal of Range Management* **47**:192-195.
- Buyanovsky, G. A., C. L. Kucera, and G. H. Wagner. 1987. Comparative Analyses of Carbon Dynamics in Native and Cultivated Ecosystems. *Ecology* **68**:2023-2031.
- Culman, S. W., J. M. Duxbury, J. G. Lauren, and J. E. Thies. 2006. Microbial community response to soil solarization in Nepal's rice-wheat cropping system. *Soil Biology & Biochemistry* **38**:3359-3371.
- Dijkstra, P., A. Ishizu, R. Doucett, S. C. Hart, E. Schwartz, O. V. Menyailo, and B. A. Hungate. 2006. C-13 and N-15 natural abundance of the soil microbial biomass. *Soil Biology & Biochemistry* **38**:3257-3266.
- Drijber, R. A., J. W. Doran, A. M. Parkhurst, and D. J. Lyon. 2000. Changes in soil microbial community structure with tillage under long-term wheat-fallow management. *Soil Biology & Biochemistry* **32**:1419-1430.
- Elliott, E. T., J. W. Heil, E. F. Kelly, and H. C. Monger. 1999. Soil structural and other physical properties. Pages 74-85 in G. P. Robertson, D. C. Coleman, C. S. Bledsoe, and P. Sollins, editors. *Standard Soil Methods for Long-Term Ecological Research*. Oxford University Press, New York, New York.
- ERS. 2004. Economic Research Service, <http://www.ers.usda.gov/>. USDA.

- Feng, Y., A. C. Motta, D. W. Reeves, C. H. Burmester, E. van Santen, and J. A. Osborne. 2003. Soil microbial communities under conventional-till and no-till continuous cotton systems. *Soil Biology & Biochemistry* **35**:1693-1703.
- Fierer, N., and J. P. Schimel. 2003. A Proposed Mechanism for the Pulse in Carbon Dioxide Production Commonly Observed Following the Rapid Rewetting of a Dry Soil. *Soil Science Society of America Journal* **67**:798–805.
- Fierer, N., J. P. Schimel, and P. A. Holden. 2003. Variations in microbial community composition through two soil depth profiles. *Soil Biology & Biochemistry* **35**:167-176.
- Gebhart, D. L., H. B. Johnson, H. S. H. S. Mayeux, and H. W. Polley. 1994. The CRP increases soil organic- carbon. *Journal of Soil and Water Conservation* **49**:488-492.
- Gomez, E. D., J. L. Garland, and M. S. Roberts. 2004. Microbial structural diversity estimated by dilution-extinction of phenotypic traits and T-RFLP analysis along a land-use intensification gradient. *Fems Microbiology Ecology* **49**:253-259.
- Gregorich, E. G., B. C. Liang, C. F. Drury, A. F. Mackenzie, and W. B. McGill. 2000. Elucidation of the source and turnover of water soluble and microbial biomass carbon in agricultural soils. *Soil Biology & Biochemistry* **32**:581-587.
- Gruter, D., B. Schmid, and H. Brandl. 2006. Influence of plant diversity and elevated atmospheric carbon dioxide levels on belowground bacterial diversity. *Bmc Microbiology* **6**.
- Hackl, E., S. Zechmeister-Boltenstern, L. Bodrossy, and A. Sessitsch. 2004. Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. *Applied and Environmental Microbiology* **70**:5057-5065.
- Hedlund, K. 2002. Soil microbial community structure in relation to vegetation management on former agricultural land. *Soil Biology & Biochemistry* **34**:1299-1307.
- Jenkinson, D. S., J. M. Potts, J. N. Perry, V. Barnett, K. Coleman, and A. E. Johnston. 1994. Trends in Herbage Yields over the Last Century on the Rothamsted Long-Term Continuous Hay Experiment. *Journal of Agricultural Science* **122**:365-374.
- Jenkinson, D. S., P. R. Poulton, A. E. Johnston, and D. S. Powlson. 2004. Turnover of nitrogen-15-labeled fertilizer in old grassland. *Soil Science Society of America Journal* **68**:865-875.

- Jenkinson, D. S., and D. S. Powlson. 1976. Effects of Biocidal Treatments on Metabolism in Soil .5. Method for Measuring Soil Biomass. *Soil Biology & Biochemistry* **8**:209-213.
- John, B., B. Ludwig, M. Potthoff, and H. Flessa. 2004. Carbon and nitrogen mineralization after maize harvest between and within maize rows: a microcosm study using C-13 natural abundance. *Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde* **167**:270-276.
- Kansas State University Agricultural Experiment Station and Cooperative Extension Service (KSUAES). 1997. *Wheat Production Handbook C-529*.
- Kansas State University Agricultural Experiment Station and Cooperative Extension Service (KSUAES). 1996. *Kansas Crop Planting Guide L-818*.
- Kennedy, N. M., D. E. Gleeson, J. Connolly, and N. J. W. Clipson. 2005. Seasonal and management influences on bacterial community structure in an upland grassland soil. *Fems Microbiology Ecology* **53**:329-337.
- Knapp, A. K., S. L. Conard, and J. M. Blair. 1998. Determinants of soil CO₂ flux from a sub-humid grassland: Effect of fire and fire history. *Ecological Applications* **8**:760-770.
- Kuske, C. R., L. O. Ticknor, M. E. Miller, J. M. Dunbar, J. A. Davis, S. M. Barns, and J. Belnap. 2002. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Applied and Environmental Microbiology* **68**:1854-1863.
- Lindstrom, M. J., T. E. Schumacher, and M. L. Blecha. 1994. Management considerations for returning CRP lands to crop production. *Journal of Soil and Water Conservation* **49**:420-425.
- Liu, W., 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.
- Lynch, J. M., and L. M. Panting. 1980a. Cultivation and the Soil Biomass. *Soil Biology & Biochemistry* **12**:29-33.
- Lynch, J. M., and L. M. Panting. 1980b. Variations in the Size of the Soil Biomass. *Soil Biology & Biochemistry* **12**:547-550.
- McCulley, R. L., and I. C. Burke. 2004. Microbial community composition across the Great Plains: Landscape versus regional variability. *Soil Science Society of America Journal* **68**:106-115.

- McCune, B., and J. B. Grace. 2002. *Analysis of Ecological Communities*. MjM Software Design, Glenden Beach, OR.
- MEAS. 1998. Missouri Agricultural Experiment Station: Recommended Chemical Soil Test Procedures for the North Central Region.
- Mielke, P. W., Jr. 1984. Meteorological applications of permutation techniques based on distance functions. *in* P. R. Krishnaiah and P. K. Sen, editors. *Handbook of Statistics*. Elsevier Science Publishers.
- Mikhailova, E. A., R. B. Bryant, Vassenev, II, S. J. Schwager, and C. J. Post. 2000. Cultivation effects on soil carbon and nitrogen contents at depth in the Russian Chernozem. *Soil Science Society of America Journal* **64**:738-745.
- Mikhailova, E. A., and C. J. Post. 2006. Organic carbon stocks in the Russian Chernozem. *European Journal of Soil Science* **57**:330-336.
- Morris, D. R., D. A. Zuberer, and R. W. Weaver. 1985. Nitrogen-Fixation by Intact Grass Soil Cores Using (N-2)-N-15 and Acetylene-Reduction. *Soil Biology & Biochemistry* **17**:87-91.
- Mummey, D. L., and P. D. Stahl. 2003. Spatial and temporal variability of bacterial 16S rDNA-based T-RFLP patterns derived from soil of two Wyoming grassland ecosystems. *Fems Microbiology Ecology* **46**:113-120.
- Nilsson, W. B., and M. S. Strom. 2002. Detection and identification of bacterial pathogens of fish in kidney tissue using terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes. *Diseases of Aquatic Organisms* **48**:175-185.
- Paterson, E. 2003. Importance of rhizodeposition in the coupling of plant and microbial productivity. *European Journal of Soil Science* **54**:741-750.
- Patra, A. K., L. Abbadie, A. Clays-Josserand, V. Degrange, S. J. Grayston, N. Guillaumaud, P. Loiseau, F. Louault, S. Mahmood, S. Nazaret, L. Philippot, F. Poly, J. I. Prosser, and X. Le Roux. 2006. Effects of management regime and plant species on the enzyme activity and genetic structure of N-fixing, denitrifying and nitrifying bacterial communities in grassland soils. *Environmental Microbiology* **8**:1005-1016.
- Patra, A. K., L. Abbadie, A. Clays-Josserand, V. Degrange, S. J. Grayston, P. Loiseau, F. Louault, S. Mahmood, S. Nazaret, L. Philippot, E. Poly, J. I. Prosser, A. Richaume, and X. Le Roux. 2005. Effects of grazing on microbial functional groups involved in soil N dynamics. *Ecological Monographs* **75**:65-80.
- Patra, A. K., X. Le Roux, L. Abbadie, A. Clays-Josserand, F. Poly, P. Loiseau, and F. Louault. 2007. Effect of microbial activity and nitrogen mineralization on free-

- living nitrogen fixation in permanent grassland soils. *Journal of Agronomy and Crop Science* **193**:153-156.
- Paul, E. A., and F. E. Clark. 1996. *Soil Microbiology and Biochemistry*. Academic Press, NY.
- Potthoff, M., N. Loftfield, F. Buegger, B. Wick, B. John, R. G. Joergensen, and H. Flessa. 2003. The determination of delta C-13 in soil microbial biomass using fumigation-extraction. *Soil Biology & Biochemistry* **35**:947-954.
- Raun, W. R., and G. V. Johnson. 1999. Improving nitrogen use efficiency for cereal production. *Agronomy Journal* **91**:357-363.
- Robertson, G. P., P. Sollins, B. G. Ellis, and K. Lajtha. 1999. Exchangeable ions, pH, and cation exchange capacity. Pages 106-114 in G. P. Robertson, D. C. Coleman, C. S. Bledsoe, and P. Sollins, editors. *Standard Soil Methods for Long-Term Ecological Research*. Oxford University Press, New York, New York.
- Rumberger, A., I. A. Merwin, and J. E. Thies. 2007. Microbial community development in the rhizosphere of apple trees at a replant disease site. *Soil Biology & Biochemistry* **39**:1645-1654.
- Ryan, M. C., and R. Aravena. 1994. Combining 13C natural abundance and fumigation-extractoin methods to investigate soil microbial biomass turnover. *Soil Biology & Biochemistry* **26**:1583-1585.
- Schimel, D. S., D. C. Coleman, and K. A. Horton. 1985. Soil Organic-Matter Dynamics in Paired Rangeland and Cropland Toposequences in North-Dakota. *Geoderma* **36**:201-214.
- Schmer, M. R., K. P. Vogel, R. B. Mitchell, and R. K. Perrin. 2008. Net energy of cellulosic ethanol from switchgrass. *Proceedings of the National Academy of Sciences of the United States of America* **105**:464-469.
- Schnepf, R. 2004. *Energy Use in Agriculture: Background and Issues*. Congressional Research Service, The Library of Congress.
- Seybold, C. A., and J. E. Herrick. 2001. Aggregate stability kit for soil quality assessments. *Catena* **44**:37-45.
- Shortrid, J. R. 1973. Prairie Hay in Woodson County, Kansas - Crop Anomaly. *Geographical Review* **63**:533-552.
- Silvertown, J., M. E. Dodd, K. McConway, J. Potts, and M. Crawley. 1994. Rainfall, Biomass Variation, and Community Composition in the Park Grass Experiment. *Ecology* **75**:2430-2437.

- Smith, J. S., and E. A. Paul. 1990. The significance of soil microbial biomass estimations. *in* J. M. Bollag and G. Stotzky, editors. *Soil Biochemistry*. Marcel Dekker, Inc., New York.
- Sparling, G. P., C. W. Feltham, J. Reynolds, A. W. West, and P. Singleton. 1990. Estimation of soil microbial C by a fumigaion-exatraction method: use on soils of high organic matter content and a reassessment of the KEC-factor. *Soil Biology & Biochemistry* **22**:301-307.
- Steenwerth, K. L., L. E. Jackson, E. A. Carlisle, and K. M. Scow. 2006. Microbial communities of a native perennial bunchgrass do not respond consistently across a gradient of land-use intensification. *Soil Biology & Biochemistry* **38**:1797-1811.
- Tilman, D., K. G. Cassman, P. A. Matson, R. Naylor, and S. Polasky. 2002. Agricultural sustainability and intensive production practices. *Nature* **418**:671-677.
- Tilman, D., J. Hill, and C. Lehman. 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. *Science* **314**:1598-1600.
- Tjepkema, J. D., and R. H. Burris. 1976. Nitrogenase Activity Associated with Some Wisconsin Prairie Grasses. *Plant and Soil* **45**:81-94.
- Ulrich, A., and R. Becker. 2006. Soil parent material is a key determinant of the bacterial community structure in arable soils. *Fems Microbiology Ecology* **56**:430-443.
- Waldrop, M. P., D. R. Zak, C. B. Blackwood, C. D. Curtis, and D. Tilman. 2006. Resource availability controls fungal diversity across a plant diversity gradient. *Ecology Letters* **9**:1127-1135.
- Wardle, D. A. 1992. A Comparative-Assessment of Factors Which Influence Microbial Biomass Carbon and Nitrogen Levels in Soil. *Biological Reviews of the Cambridge Philosophical Society* **67**:321-358.
- Weil, R. R., K. R. Islam, M. A. Stine, J. B. Gruver, and S. E. Sampson-Liebig. 2003. Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. *American Journal of Alternative Agriculture* **18**:1-15.
- Zak, D. R., W. E. Holmes, D. C. White, A. D. Peacock, and D. Tilman. 2003. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology* **84**:2042-2050.
- Zehr, J. P., B. D. Jenkins, S. M. Short, and G. F. Steward. 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology* **5**:539-554.

Zelles, L., and Q. Y. Bai. 1994. Fatty-Acid Patterns of Phospholipids and Lipopolysaccharides in Environmental-Samples. *Chemosphere* **28**:391-411.

CHAPTER 3
SOIL BACTERIAL AND FREE-LIVING DIAZOTROPHIC COMMUNITY
DYNAMICS FOLLOWING THE CONVERSION OF NATIVE TALLGRASS
PRAIRIE TO ANNUAL WHEAT IN KANSAS²

Introduction

Approximately 50% of the world's surface area has been converted to grazed land or cultivated crops resulting in large losses in soil fertility and dramatic shifts in flora and fauna and the ecosystem services they provide (MEA 2006). Of current world cropland, annual-tilled, monocultures produce more than 70% of our food and fiber needs. However, intensive production of annual-tilled crops severely impacts soil fertility, water quality, biodiversity and ecosystem function (McLaughlin and Mineau 1995, Tilman 1999, Turner and Rabalais 2003, Roberts et al. 2007). In particular severe losses of soil organic matter (SOM) up to 2 Mg ha⁻¹ yr⁻¹ in US croplands threaten water retention, nutrient availability and yield potential (Johnston 1976, Lucas et al. 1977, Pimentel et al. 1995).

Organic matter is a dynamic property of soils. The balance between plant production and microbial oxidation largely determines soil organic matter (SOM) and soil organic carbon (SOC) status and accumulation in soils. Plant- and animal-derived C feeds a diverse complex of soil biota collectively known as the soil food web. When organic material is added, soil food webs are more active and perform important functions and services such as the production of polysaccharides and non-humic compounds that form aggregates binding SOM and increase soil porosity, and mineralization of

² Data from this chapter will be submitted to Ecological Applications.

nutrients to plant-available forms (Paul and Clark 1996). SOC may accumulate more readily when higher trophic level soil organisms are abundant due to greater C conserved in biomass, gradual transformation to more recalcitrant forms of C, and protection in smaller aggregate size fractions (Fu et al. 2000).

Aside from the artificial drainage of wetlands, the management practices that most critically affect SOC accumulation and cycling are tillage and cropping (Davidson and Ackerman 2006). SOC declines rapidly after virgin soil is cultivated; soils lose 20–40% of soil carbon over the first few decades of agricultural use (Mann 1986, Davidson and Ackerman 2006). Tillage breaks up organic residue and aerates the soil making SOM more accessible to microbial oxidation and destroys macro-aggregates that physically protect labile C (Reicosky et al. 1997).

Conversion of cropping systems from conventional-till to no-till cropping is generally thought to build SOC. In a global analysis of 67 long-term agricultural experiments, West and Post found an average increase of $57 \text{ g C m}^{-2} \text{ yr}^{-1}$ after conversion to no-till (2002). However, a recent review has challenged some of the results of this study as biased by shallow sampling methods typical of most studies today (Baker et al. 2007). Of the 140 comparisons made in the West and Post analysis, none sampled below 30cm. In another large analysis of no-till practices on SOC in Canada, changes in SOC results were largely contingent on the sampling depth (VandenBygaart et al. 2003). In studies that sampled to 30 cm or less, 37 of 45 no-till treatments found more SOC in the no-till treatments than in conventional till, with a mean annual SOC gain of $0.38 \pm 0.72 \text{ t ha}^{-1} \text{ year}^{-1}$. However, in studies that sampled below 30cm, 35 out of 51 trials reported less SOC in no-till than in conventional till, with a mean annual SOC loss of $-0.23 \pm 0.97 \text{ t ha}^{-1} \text{ year}^{-1}$ (VandenBygaart et al. 2003).

Baker et al. proposed that the differences observed regarding depth could result from a changed soil environment and subsequent change in plant growth. No-till soil has been shown to have lower soil temperatures (Johnson and Lowery 1985, Drury et al. 2005, Fabrizzi et al. 2005) and increase bulk densities and penetration resistance (Larney and Kladivko 1989, Vyn and Raimbault 1993, Fabrizzi et al. 2005) relative to conventional till. These differences likely impact plant growth, as both wheat (Qin et al. 2004) and maize (Qin et al. 2005) show differences in root length density under no-till systems. Under no-till, greater root densities were found in the upper 5cm, while lower root densities were found in the lower soil profiles compared with systems that were conventionally tilled (Qin et al. 2004, 2005).

Since soils have lost a significant amount of C since tillage began, it is logical to think that ceasing or reducing tillage will lead to SOC accrual. However, Baker et al. (2007) suggest that a re-examination of these findings may be appropriate as no-till agriculture may only change the distribution of SOC in the soil profile. Baker et al. (2007) suggest two other major factors that have likely lead to major losses in SOC: i) the agricultural conversion of native perennial plant communities to primarily annual crops and ii) the increased rate of SOC oxidation, due to more soil drainage, increased aeration and application of mineral N fertilizer.

A recent study by Fornara and Tilman (2008) may shed light regarding changes in plant community composition from perennials to annuals. In a 12-year-long experiment on agriculturally degraded soil, the authors found that 500% more soil C and 600% more soil N accrued under high diversity mixtures of perennial grasslands than under monoculture plots of those same species. The higher levels of soil C and N

resulted from increases in soil C and N inputs due to greater root biomass and greater root biomass accumulation. Their results also suggested that the joint presence of a C4 grass and a legume species was a key driver of soil C and N accrual in both high and low diversity plots (Fornara and Tilman 2008). These results suggest that plant community composition can have major impacts of SOC storage and accrual/ loss.

This study was conducted to examine the relationship of SOC changes from perennial grasslands to annual wheat. Grasslands generally maintain high levels of SOM and associated soil food web functioning relative to agricultural fields (Gebhart et al. 1994, Potter et al. 1999). Recent work by Glover et al. (unpublished) demonstrates that perennial grasslands not only maintain soil quality and ecosystem function but can also maintain yield efficiencies comparable to those of annual wheat, and thus seem to balance a supposed tradeoff between productivity and environmental benefits.

Comparison of long-term (75 years or more) annual wheat and prairie sites showed that perennial fields yielded comparable amounts of protein and nitrogen as high input wheat fields, while maintaining soil quality, and supporting diverse above and belowground communities associated with nutrient cycling, pollination and pest suppression. Chapter 2 demonstrated that these increased efficiencies are related to differences in soil biological communities and hypothesized that differences were driven by belowground inputs.

The overall objective of this study was to determine which aspects of current agricultural systems lead to degradation of soil fertility and lower yield efficiencies compared to perennial grasslands. We questioned whether differences in SOM and associated soil food webs between perennial grasslands and annual crops found by Glover et al. (unpublished) were due to obsolete farming practices, effects of

differences in plant community composition on SOC, or both? This chapter specifically deals with the effects of conversion on the soil bacterial communities. The specific objectives are i) to determine if conversion from perennial grassland to annual cropping significantly affects bacterial community abundance and structure and ii) to examine if this conversion affects free-living nitrogen-fixing bacterial structure in these systems.

Materials and Methods

Site Description and Soil Sampling. The field site was located in Niles, in Ottawa County, Kansas (N^o 38.58.145, W^o 97.28.616). In 2004, three research blocks 20m x 20 m were established in a section of native tallgrass prairie. Each research block was divided in half and two treatments (prairie meadow or no-till annual cropping) were randomly assigned to each 10 m x 20 m long plot (Figure 3.1). Prior to this experiment, the prairie had undergone relatively consistent management for 75 years or more—hayed once annually in June/July with the hay removed offsite and fed to livestock. The prairie has never received fertilization, but has been burned periodically. In the fall of 2004, no-till annual cropping plots received 2 applications of glyphosate and in 2005 were planted into soybean. The annual plots were planted into sorghum in 2006 and in wheat in 2007. Prairie management has remained consistent to pre-experiment practices. Annual wheat field management followed typical practices for the region (KSUAES 1996, 1997).

This study's experimental design consisted of four factors: sampling date, depth, treatment and block. Specifically, there were 7 sampling dates, 6 depths, 2 treatments and 3 blocks sampled, totaling 252 samples. Our research objectives led us to focus

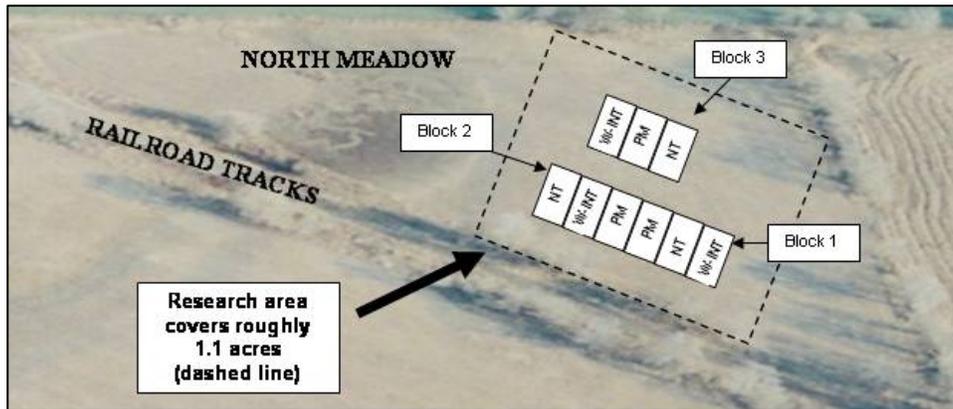


Figure 3.1. Field map of Niles conversion study site.

mainly on the differences between treatment (prairie vs. annual wheat), but also how these differences change with depth. Differences detected between sampling date and block were of limited interest here. In the regional study (Chapter 2) we assume that 75 years or more of consistent management have allowed these systems to reach near equilibrium. However, since conversion of prairie to annual wheat has occurred within the last three years, we assume that many belowground properties will be rapidly changing. Our sampling scheme was intended to be frequent enough to monitor these changes when they become detectable. Unlike the regional study, the component of time in this study does more than conceptually represent a confirmation of previous findings; each additional sampling period represents a greater divergence between these two systems. We expected that each sampling had a greater chance of observing detectable differences between the two systems than the previous samplings.

Soils were sampled seven total times: i) May 5, 2005, ii) June 2, 2005, iii) October 8, 2005, iv) April 27, 2006, v) June 20, 2006, vi) October 6, 2006, and iii) June 20, 2007. Four centimeter diameter cores were taken to a depth of one meter from each plot and

separated into sections by depth: 0 – 10 cm, 10 – 20 cm, 20 – 40 cm, 40 – 60 cm, 60 – 80 cm, 80 – 100 cm. Samples from each depth were bulked and mixed until homogeneous. Soils were then frozen at – 20°C for molecular analyses or stored at 4°C for all other analyses.

Soil Properties. Soil properties were analyzed both at The Land Institute (TLI) and at the Soil Testing Laboratory at Kansas State University (KSU). Analyses at TLI included: pH (Robertson et al. 1999), bulk density by oven drying at 105°C, percent clay by the hydrometer method (Elliott et al. 1999), water stable aggregates (WSA) by wet-sieving (Seybold and Herrick 2001), and readily oxidizable carbon (ROC) (Weil et al. 2003). Analyses at KSU included SOM by the Walkley-Black procedure, soil organic carbon (SOC) and total N by dry combustion on a LECO CN 2000 combustion analyzer, total P by a modified Kjeldahl digestion and total K by flame atomic absorption. Further details on analyses performed at KSU can be found at MEAS (1998).

Microbial Biomass. Microbial biomass carbon (MBC) was measured with the simultaneous chloroform fumigation extraction (sCFE) method (Fierer and Schimel 2003). Briefly, 10g of soil from each sample were weighed into two, 70ml glass vials, one labeled ‘fumigated’, the other ‘non-fumigated’. Forty ml of 0.05M K₂SO₄ were added to both vials and the fumigated vial received 0.5 ml of EtOH-free CH₃Cl. Blanks (both fumigated and non-fumigated vials without soil) were also prepared. Vials were sealed and shaken at 150 rpm for 4 h. Extracts were centrifuged for 15 minutes at 1500 rpm and the supernatant was vacuum filtered through 0.45 µm Watman filter paper. Microbial biomass extracts were bubbled for 30 minutes with air to remove any residual CH₃Cl and stored at – 20 °C until analysis. Dissolved organic

carbon (DOC) and the natural abundance ^{13}C values of DOC were determined using an O.I. Analytical Model 1010 TOC Analyzer (OI Analytical, College Station, TX) interfaced to a PDZ Europa 20-20 isotope ratio mass spectrometer (Sercon Ltd., Cheshire, UK) at the University of California Davis Stable Isotope Facility.

MBC was calculated as the difference between fumigated and non-fumigated (control) samples divided by a K_{EC} -factor of 0.35 (Sparling et al. 1990). Analyzed blank samples contained no to extremely small amounts of C, so no correction was made for this insignificant analytical artifact. Isotope ratio mass spectrometry was also used to determine $\delta^{13}\text{C}$ values of the microbial biomass. However, repeated problems in instrumentation at the UC Davis Stable Isotope Facility rendered these data non-meaningful. Since the extractant was completely used in the first run, re-analysis of the sample was not possible. Therefore, $\delta^{13}\text{C}$ data of microbial biomass are not included here.

Molecular Analyses (PCR and T-RFLP)

16S rRNA gene. Soil DNA was extracted from 0.25 g soil per sample using the MoBio PowerSoil™ DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA). DNA extracts were quantified and diluted with nuclease-free water to $2 \text{ ng } \mu\text{l}^{-1}$. DNA was then amplified by polymerase chain reaction (PCR) using the fluorescently-labeled forward primer 27f (5'-[6FAM] AGA GTT TGA TCM TGG CTC AG-3') and the unlabeled reverse primer 1492r (5'-TAC GGY TAC CTT GTT ACG ACT T-3') (Invitrogen, Carlsbad, CA). These primers target the eubacterial 16S rDNA genes in the extracted soil DNA and the amplification results in products of approximately 1500 bp. Three, 50 μl reactions of each sample were amplified using a PTC 200 thermal cycler (MJ Research, Waltham, MA) as follows: initial denaturation at 95°C

for 5 min; 27 cycles of denaturation at 95°C for 45 s, annealing at 56°C for 45 s, and extension at 72°C for 1 min; and a final extension step at 72°C for 10 min. Reaction concentrations were: 0.05 U μl^{-1} *AmpliTaq Gold*® DNA polymerase (Applied Biosystems, Foster City, CA), 1x PCR buffer, and 2.0 mM MgCl_2 , 0.2 mM deoxy-nucleotide triphosphates (dNTPs), 0.1 $\mu\text{g } \mu\text{l}^{-1}$ bovine serum albumin (BSA), both primers at 0.1 μM , nuclease free water, and 5 μl of DNA template (10 ng reaction⁻¹). Amplified DNA products were verified by electrophoresis on a 1.0% agarose gel.

Following PCR, amplified DNA (three 50 μl reactions per sample) was pooled and quantified. DNA concentrations were adjusted to 30 ng μl^{-1} . Two, 30 μl restriction enzyme digests were prepared per sample using HhaI and Sau96 I restriction enzymes (New England Biolabs, Ipswich, MA). Reaction concentrations were: 5 U enzyme (either HhaI or Sau96 I), 1x of the respective buffer, 0.1 $\mu\text{g } \mu\text{l}^{-1}$ BSA, nuclease-free water, and 15 μl of amplified DNA (450 ng reaction⁻¹). Restriction digestion was carried out in a MJ Research PTC 200 thermal cycler at 37°C for 4.5 h with a final step of 70°C for 15 min to stop the reaction. Complete digestion of the DNA was verified by electrophoresis on a 1.5% agarose gel.

Digested DNA was purified using a PERFORMA® DTR Edge Plate (Edge BioSystems, Gaithersburg, MD) and lyophilized. DNA was resuspended in a 10 μl mix containing 9.85 μl of formamide and 0.15 μl of LIZ 500 size standard (Applied Biosystems). Terminal fragment-size analysis was performed using a 3730 ABI electrophoretic capillary sequencer in conjunction with the Genemapper Software (Applied Biosystems) at Cornell University's Biotechnology Resource Center, Ithaca, NY.

nifH. T-RFLP analyses were also performed to characterize free-living diazotrophic populations in soils. The methods used were as described above, with the following changes. Soil DNA extracts were amplified by PCR using the fluorescently-labeled forward primer *nifH*-b1-112F (5'-[PET] GGC TGC GAT CCC AAG GCT GA-3') (Applied Biosystems) and the fluorescently-labeled reverse primer CDHP Nif723R (5'-[6FAM] GAT GTT CGC GCG GCA CGA ADT RNA TSA-3') (Invitrogen). These primers target *nifH*, the structural gene for nitrogenase reductase, in the extracted soil DNA and the amplification results in products of approximately 700 bp. Three, 50 μ l reactions of each sample were amplified as follows: initial denaturation at 95°C for 10 min; 35 cycles of denaturation at 95°C for 30 s, annealing at 61°C for 30 s, and extension at 72°C for 45 s; and a final extension step at 72°C for 10 min. Reaction concentrations were: 0.05 U μ l⁻¹ *AmpliTaq Gold*® DNA polymerase, 1x PCR buffer, and 2.5 mM MgCl₂, 0.8 mM dNTPs, 0.5 μ g μ l⁻¹ BSA, both primers at 0.25 μ M, nuclease free water and 50 ng DNA template reaction⁻¹. Amplified DNA was digested with the restriction enzyme *MspI* (New England Biolabs) in the following reaction concentrations: 5 U enzyme, 1x of supplied buffer, 0.1 μ g μ l⁻¹ BSA, nuclease-free water and 450 ng DNA reaction⁻¹.

Statistical Analyses. Analysis of Variance was performed on the soil nutrient data and microbial biomass data using PROC MIXED procedure in SAS v.9 (Cary, NC). Depth and treatment were treated as fixed effects and block as a random effect. All soil variables were measured at the June 2007 sampling with significance differences determined at $\alpha = 0.05$ level of probability. All soil data were converted into mass per hectare, in order to account for differences in soil volume and bulk density. Least squared means are reported for all soils data.

The T-RFLP data analyzed in this study were uncommonly complex in regard to number of underlying environmental gradients (sampling date, depth, treatment, and block), sample heterogeneity and percent variation from interaction signal. As a result of this complexity, nonmetric multidimensional scaling (NMS) analyses with Sørensen distance measure were used to analyze all T-RFLP data. (See Chapter 5 for further discussion selecting an appropriate ordination analysis for T-RFLP.) NMS analyses were performed in PC-ORD v.5 (MjM Software Design, Gleneden Beach, OR) with 2 axes selected, 50 runs with real data, 0.0001 stability criterion, 50 iterations to evaluate stability.

Multiple-response Permutations Procedures (MRPP) was also employed to test significance among the experimental factors within the datasets (Mielke 1984, McCune and Grace 2002). The test relies on calculating the average distance within each group and then calculating the probability that this weighted mean within group distance is greater than expected by chance. This creates a p -value and an agreement statistic (A). The p -value determines if the specified experimental factor has significantly affected community structure, while the A describes within-group homogeneity and is independent of sampling size. When all items within each group are identical, $A = 1$; if the items within each group equals what is expected by chance, $A = 0$. Negative A values result from cases with less agreement between groups than is expected by chance (Mielke 1984, McCune and Grace 2002).

Results and Discussion

Soil Properties. Key soil properties for the top 40 cm of soil from the field site are found in Table 3.1. Three years after conversion from prairie, most of soil properties

show no detectable differences between the background prairie and annual, no-till plots. However, ROC, a very labile pool of soil carbon, declined under no-till annual agriculture. Soil sampling prior to agricultural conversion showed no differences in any measured soil properties between outlined plots (data not shown), therefore reductions in ROC were due to the imposed treatments—likely resulting from reduced inputs from the crop roots relative to their perennial counterparts, greater SOC oxidation, or both. Globally, temperate grasslands average greater than 9 times the root biomass of croplands with 37 times greater root/shoot ratio (Jackson et al. 1996). This increased biomass may result in greater C inputs via root turnover and rhizodeposition. However, in a recent review of long-term yield trials, Khan et al. (2007) argued that a long-term fertilization with mineral N results in net SOC loss. Newly converted annual plots were fertilized with mineral N, also likely increasing SOC loss via increased microbial SOC oxidation.

Table 3.1. Soil properties to 40 cm depth, 3 years after no-till conversion of prairie to annual wheat

	No Till		Prairie Meadow	
	Mean	Std Dev	Mean	Std Dev
SOM (%)	12.10	0.51	12.37	0.27
Total N (%)	0.62	0.00	0.62	0.01
SOC (%)	7.04	0.12	7.27	0.17
ROC (%)*	2.02	0.03	2.12	0.04
BD (Mg m⁻³)	3.7	0.15	3.72	0.20
WSA	2.63	0.03	2.61	0.05

* Only significant differences detected were with ROC at $\alpha = 0.05$; SOM = soil organic matter; SOC = soil organic carbon; ROC = readily-oxidizable carbon; WSA = water stable aggregates; BD = bulk density.

Microbial Biomass. Prairie plots exhibited significantly greater amounts of microbial biomass carbon (MBC) than annual wheat fields at the first three depths (Figure 3.2). However, below 40 cm, there were no detectable differences in MBC. Surface MBC in grassland soils are typically at least twice as great as in the surface of cultivated fields (Jenkinson and Powlson 1976, Lynch and Panting 1980b, a, Schimel et al. 1985, Acosta-Martinez et al. 2007). But all of these measurements involved studies where agricultural fields had been converted via tillage. To our knowledge, this is the first study to convert native prairie into an annual agricultural crop, without the use of tillage. Controlling for the often compounding effects of tillage has enabled this study to examine the effects of plant community on soil carbon status and microbial dynamics.

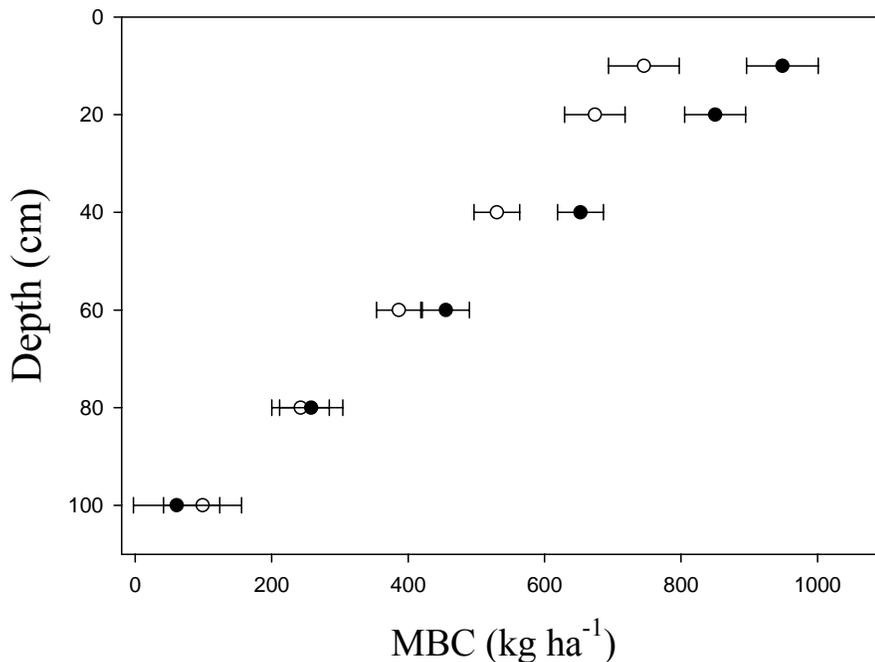


Figure 3.2. Relationship of microbial biomass carbon (MBC) to depth in prairie plots (closed circles) and annual wheat plots (open circles). Significant differences were detected in the top three depths.

Microbial biomass has been shown to be strongly correlated with root biomass, and has been used as a measure of belowground resource availability (Wardle 1992, Paterson 2003, Waldrop et al. 2006), an indirect measurement of belowground inputs from plants via root exudation and rhizodeposition. The reduction of MBC in the surface depths likely resulted, in part, from the change in rooting structure and activity in the plant community. Reduction of MBC in the annual wheat plots could also be attributed to the decrease in diversity, as decreased plant diversity has been shown to decrease microbial biomass (Zak et al. 2003, Waldrop et al. 2006). In addition, MBC is known to fluctuate seasonally (Wardle 1992, Steenwerth et al. 2006). More measurements over the growing season would be needed to make this assessment more robust. Additional measurements of microbial biomass with $\delta^{13}\text{C}$ values would potentially be insightful into SOC dynamics of these systems.

16S rRNA gene T-RFLP. Terminal restriction fragment length polymorphism (T-RFLP) analysis of the complete dataset (all sampling dates) revealed sampling date and depth to be the two largest drivers in general bacterial structure (Figure 3.4, Table 3.2). The largest difference observed was between the June 2007 and the samplings from the 2005 and 2006 years (Figure 3.4a). Depth also appears to be a large driver of structure, as fairly consistent separation based on depth profiles can be observed along the second axis (Figure 3.4b). Differences between treatments were not observed when the entire dataset was analyzed (Figure 3.4c, Table 3.2). Like with the regional study (Chapter 2), differences in block were not visually apparent with NMS analysis (Figure 3.4d), but were statistically significant with MRPP (Table 3.2). The influence of block as a driver of community structure was stronger at lower depths, but was insignificant in the surface depths (Table 3.2). These trends are consistent with findings in the regional study.

<p style="text-align: center;"><u>Sampling Date</u></p> <p>  May '05  April '06  June '05  June '06  Oct '05  Oct '06  June '07 </p>	<p style="text-align: center;"><u>Depth</u></p> <p>  0 – 10cm  20 – 40cm  10 – 20cm  60 – 80cm  40 – 60cm  80 – 100cm </p>
<p style="text-align: center;"><u>Treatment</u></p> <p>  Annual Wheat  Perennial Grassland </p>	<p style="text-align: center;"><u>Block</u></p> <p>  Block 1  Block 2  Block 3 </p>

Figure 3.3. Legend of symbols used to depict T-RFLP samples in NMS analyses.

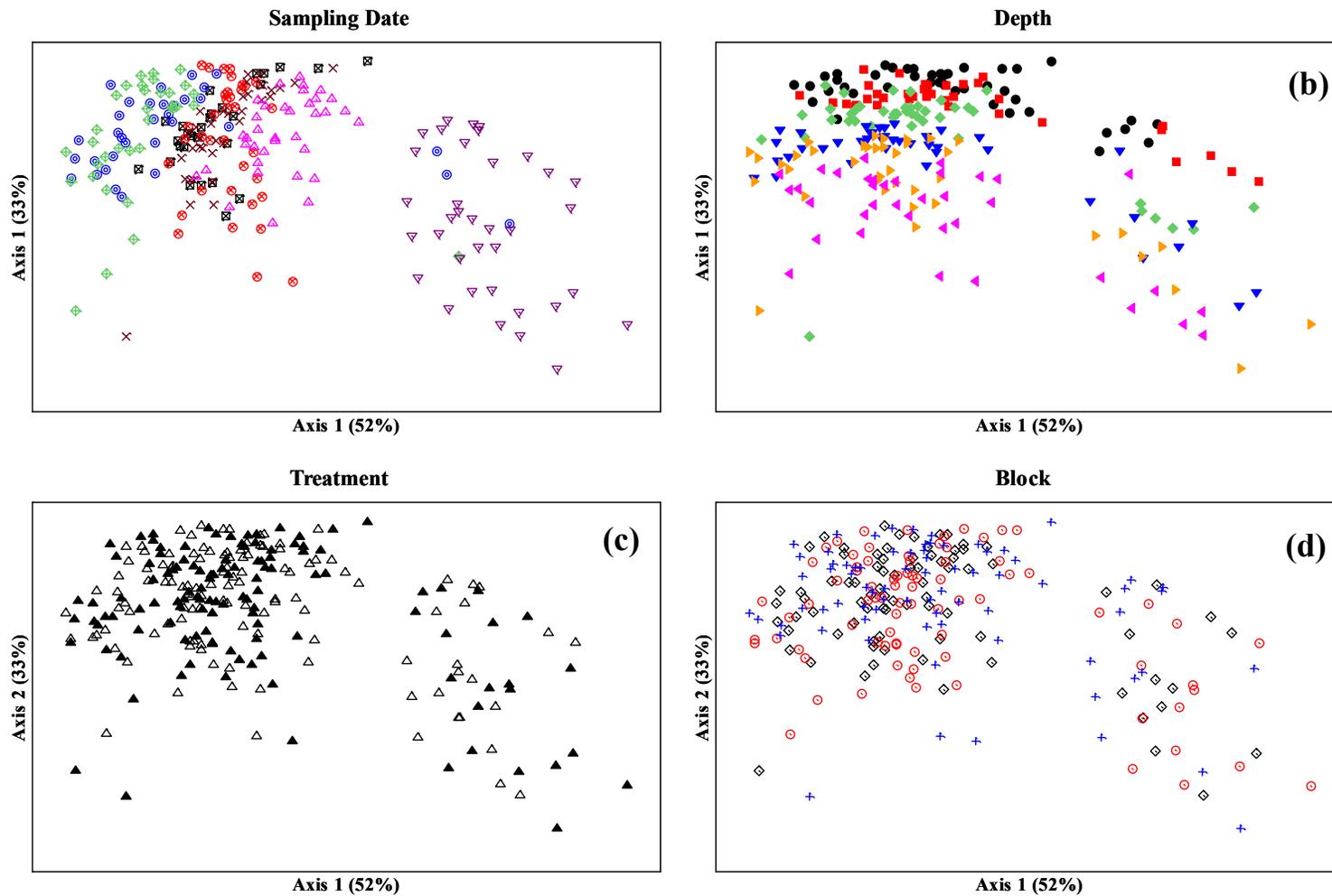


Figure 3.4. NMS analysis of complete bacterial T-RFLP dataset (May 2005 – June 2007). The four panels display the same data represented by the four experimental factors: sampling date (a), depth (b), treatment (c), and block (d). Figure 3.3 contains the symbol legend.

Table 3.2. MRPP analysis of bacterial community differences- all sampling dates (May 2005 – June 2007)

Dataset Section Analyzed	Enzyme	Depth		Treatment		Block		Date	
		A	p-value	A	p-value	A	p-value	A	p-value
All Depths	<i>HhaI</i>	0.151	<0.001	-0.001	0.932	0.010	<0.001	0.134	<0.001
	<i>Sau96 I</i>	0.135	<0.001	-0.002	1.000	0.009	0.001	0.172	<0.001
10-40cm	<i>HhaI</i>	0.087	<0.001	-0.004	1.000	0.002	0.233	0.232	<0.001
	<i>Sau96 I</i>	0.090	<0.001	-0.005	1.000	0.002	0.240	0.238	<0.001
40-100cm	<i>HhaI</i>	0.052	<0.001	0.000	0.480	0.032	<0.001	0.188	<0.001
	<i>Sau96 I</i>	0.042	<0.001	-0.003	0.845	0.025	<0.001	0.238	<0.001
10cm	<i>HhaI</i>	-	-	-0.010	0.977	-0.002	0.510	0.361	<0.001
	<i>Sau96 I</i>	-	-	-0.010	0.960	-0.002	0.491	0.359	<0.001
20cm	<i>HhaI</i>	-	-	-0.017	1.000	-0.004	0.534	0.359	<0.001
	<i>Sau96 I</i>	-	-	-0.016	1.000	0.006	0.294	0.357	<0.001
40cm	<i>HhaI</i>	-	-	-0.012	1.000	0.018	0.085	0.325	<0.001
	<i>Sau96 I</i>	-	-	-0.013	1.000	0.011	0.185	0.337	<0.001
60cm	<i>HhaI</i>	-	-	-0.002	0.544	0.049	0.001	0.230	<0.001
	<i>Sau96 I</i>	-	-	-0.008	0.871	0.036	0.013	0.289	<0.001
80cm	<i>HhaI</i>	-	-	-0.004	0.651	0.050	<0.001	0.229	<0.001
	<i>Sau96 I</i>	-	-	-0.010	0.944	0.025	0.052	0.284	<0.001
100cm	<i>HhaI</i>	-	-	-0.005	0.678	0.025	0.041	0.213	<0.001
	<i>Sau96 I</i>	-	-	-0.012	0.978	0.025	0.059	0.301	<0.001

Since differences between treatments were of primary interest here, controlling for, or eliminating variation from sampling date, depth and/or block was performed in an attempt to reveal more subtle patterns pertaining to treatment. Analyzing the 2005 T-RFLP dataset revealed consistent groupings of bacterial communities only by depth (Figure 3.5), but showed no signs of treatment effects (data not shown). Depth was controlled by decomposing the dataset into the top 3 depths (0 – 40cm) and the bottom 3 depths (40 – 100cm), as well as analyzing each depth individually. These approaches were no more discriminatory, as detectable shifts in bacterial community structure

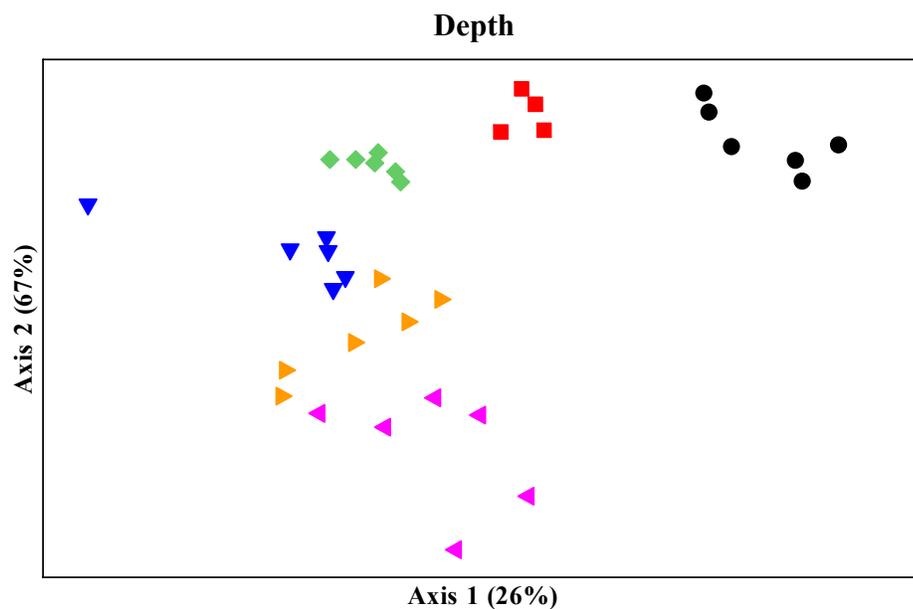


Figure 3.5. NMS analyses of the 2005 sampling year. Each data point represents the average of the 3 blocks. Figure 3.3 contains the symbol legend.

based on treatment were not found in the first year (data not shown). Similar analyses were performed on the 2006 datasets and yielded similar results—depth was a strong driver in bacterial community structure, but treatment differences were not detectable (data not shown).

The lack of detectable treatment differences in the first two years of sampling directed the focus to the June 2007 sampling date, as it represented the largest divergence between the annual agricultural system and the native grassland in regard to time. Analysis of the entire June 2007 sampling time revealed similar patterns to analyses of earlier datasets—bacterial communities were significantly affected by depth, but not by treatment or block (Figure 3.6, Table 3.3). Dataset variation due to depth was minimized by decomposing the June 2007 data matrix into the surface three depths (Figure 3.7), three bottom depths (Figure 3.8), and into individual depths (NMS analyses not shown). These approaches failed to produce ordinations that resulted in consistent treatment differences between prairie grasslands and no-till wheat. Despite the lack of treatment differences, the effects of block were observed at the lower depths (with *HhaI* at 40 – 100cm in Table 3.3, and suggestive in Figures 3.6 and 3.8). In particular, block 3 (represented by blue crosses) in Figures 3.6 and 3.8 demonstrated unique grouping patterns from blocks 1 and 2. These groupings reflected physical distances in the site, as block 3 was furthest in the field from both blocks 1 and 2. This finding demonstrates that although soil textural properties did not vary between the three blocks, other factors leading to field site heterogeneity were large enough to alter these communities.

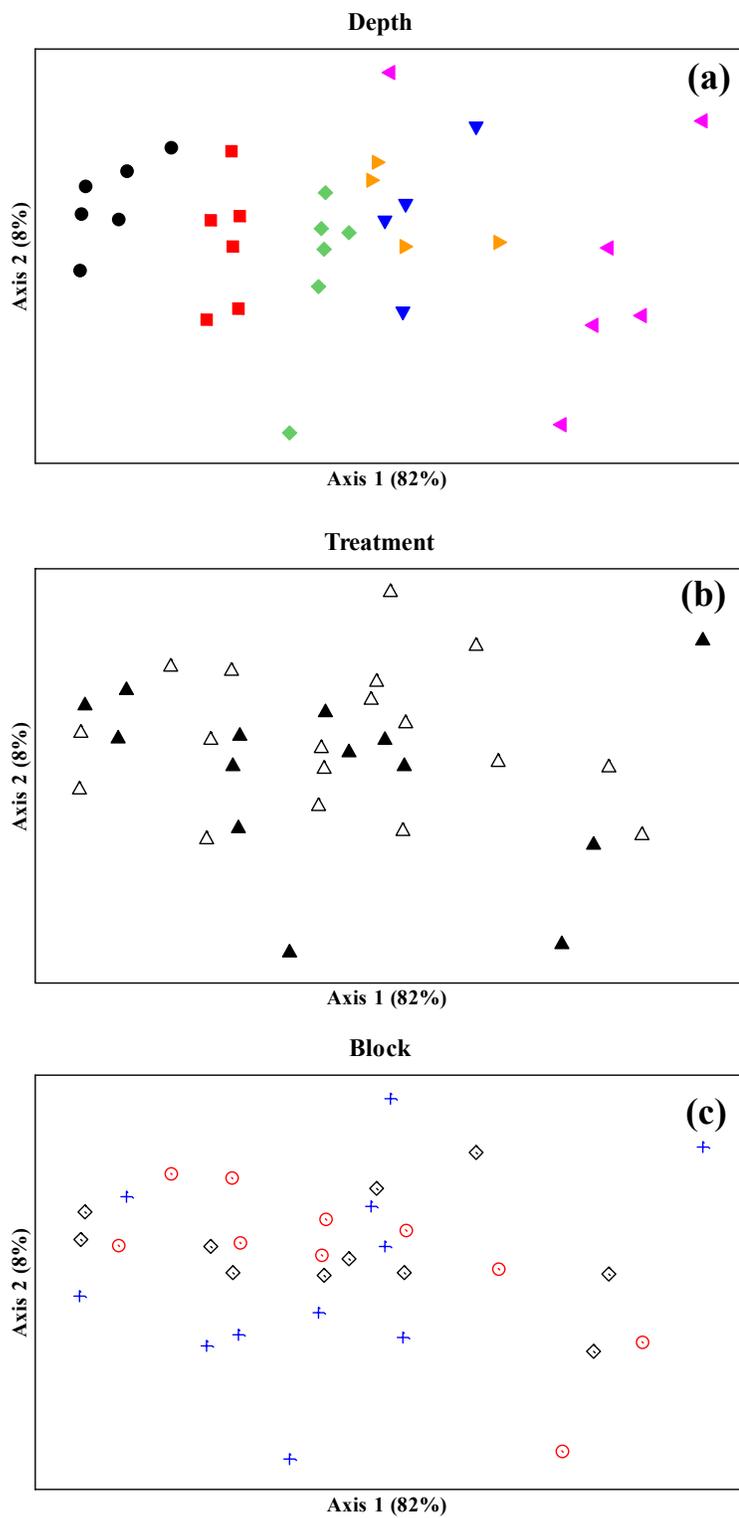


Figure 3.6. NMS analysis of the June 2007 bacterial T-RFLP dataset. The three panels display the same data represented by three experimental factors: depth (a), treatment (b), and block (c). Figure 3.3 contains the symbol legend.

Table 3.3. MRPP analysis of bacterial community differences- June '07.

Dataset Section		Depth		Treatment		Block*	
Analyzed	Enzyme	A	p-value	A	p-value	A	p-value
All Depths	<i>HhaI</i>	0.328	<0.001	-0.009	0.793	-0.006	0.576
	<i>Sau96 I</i>	0.353	<0.001	-0.005	0.542	-0.016	0.832
10-40cm	<i>HhaI</i>	0.290	<0.001	-0.031	1.000	0.003	0.391
	<i>Sau96 I</i>	0.364	<0.001	-0.025	0.925	-0.029	0.755
40-100cm	<i>HhaI</i>	0.097	0.017	0.015	0.210	0.055	0.081
	<i>Sau96 I</i>	0.070	0.029	0.045	0.040	-0.012	0.593
10cm	<i>HhaI</i>	-	-	0.001	0.432	-	-
	<i>Sau96 I</i>	-	-	-0.056	0.818	-	-
20cm	<i>HhaI</i>	-	-	-0.077	0.844	-	-
	<i>Sau96 I</i>	-	-	0.007	0.460	-	-
40cm	<i>HhaI</i>	-	-	0.005	0.425	-	-
	<i>Sau96 I</i>	-	-	0.085	0.061	-	-
60cm	<i>HhaI</i>	-	-	0.048	0.255	-	-
	<i>Sau96 I</i>	-	-	-0.002	0.467	-	-
80cm	<i>HhaI</i>	-	-	ND	ND	-	-
	<i>Sau96 I</i>	-	-	ND	ND	-	-
100cm	<i>HhaI</i>	-	-	-0.072	0.954	-	-
	<i>Sau96 I</i>	-	-	-0.021	0.658	-	-

* Values for block at individual depths were not determined, as MRPP analysis requires groups to have more than 2 items (wheat vs. prairie). ND = no data.

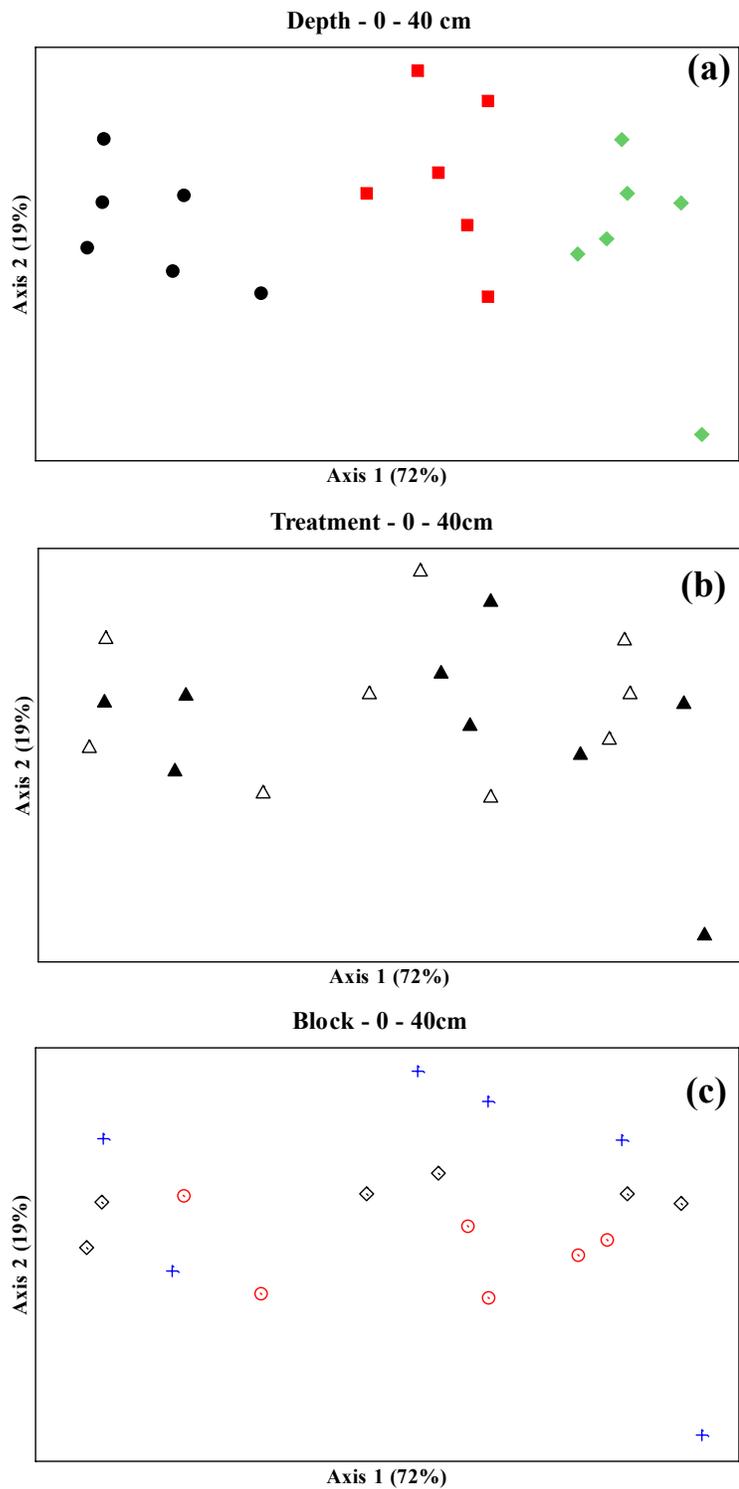


Figure 3.7. NMS analysis of the surface three depths of the June 2007 bacterial T-RFLP dataset. The three panels display the same data represented by three experimental factors: depth (a), treatment (b), and block (c). Figure 3.3 contains the symbol legend.

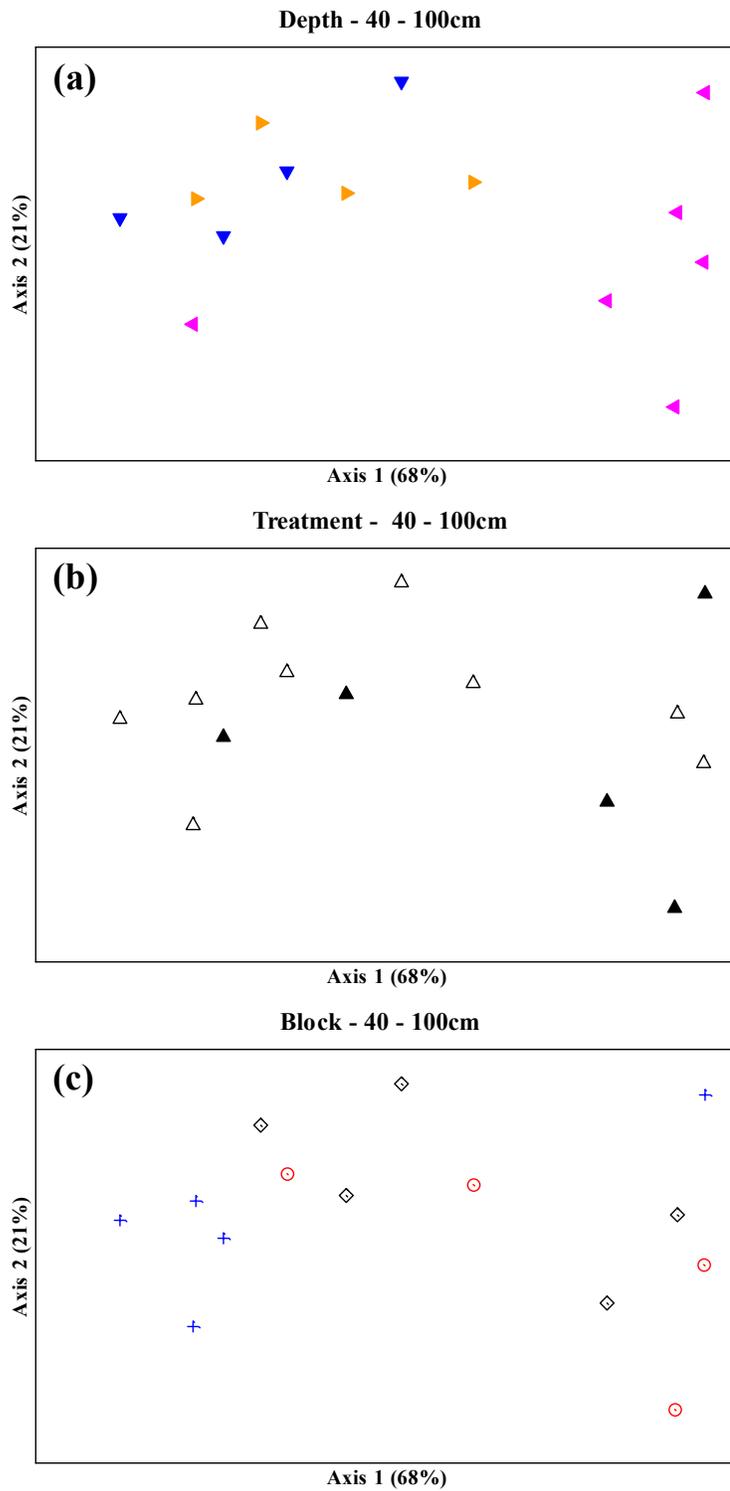


Figure 3.8. NMS analysis of the bottom three depths of the June 2007 bacterial T-RFLP dataset. The three panels display the same data represented by three experimental factors: depth (a), treatment (b), and block (c). Figure 3.3 contains the symbol legend.

nifH T-RFLP

T-RFLP analysis of the June 2007 sampling revealed that both depth and block were significant in shaping nitrogen fixing communities (Figure 3.9, Table 3.4). Although the effect is significant, depth has a much less consistent effect on the nitrogen fixing communities (Figure 3.9a), compared to the general bacterial communities observed in Figure 3.6a. The distinction between block 3 and blocks 1 and 2 appears to be greater in the nitrogen fixing communities (Figure 3.9c) than was observed in the general bacterial communities (Figure 3.6c). Differences in treatment in the nitrogen fixing community were not found with this dataset.

Analysis of the surface three depths revealed significant differences in depth, and suggestive but non-significant differences in treatment (Table 3.4, Figure 3.10). Block differences at these depths were also suggestive (NMS analysis not shown) but not significant (Table 3.4). When the bottom three depths were analyzed together, depth was significant ($\alpha = 0.10$) with MRPP analysis, but the trends were weak in the NMS ordination (Figure 3.11). However, differences in block, not only between block 3 and blocks 1 and 2, but between all three blocks were apparent at these lower depths (Figure 3.11), to a greater extent than in general bacterial structure (Figure 3.8c).

This study differed conceptually from the regional study (Chapter 2) in two major ways. First, the study's design was limited to one field site, so it did not contain the large source of site variability. Second, this study did not have a long-term imposed treatment history between the two systems, as the conversion occurred just three years prior. Data from the regional study demonstrated that management history and site variability were both large drivers in soil microbial dynamics in these systems.

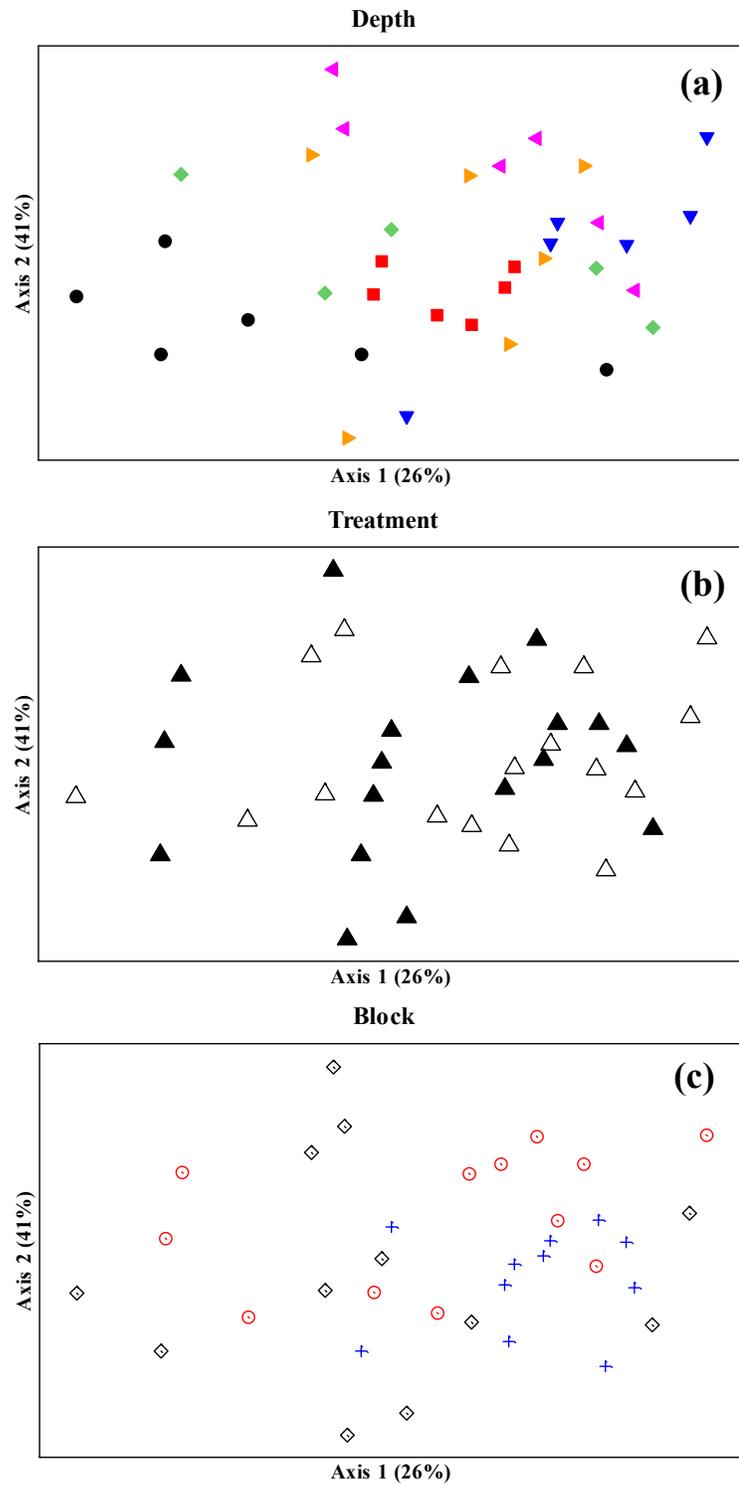


Figure 3.9. NMS analysis of the June 2007 *nif* T-RFLP dataset. The three panels display the same data represented by three experimental factors: depth (a), treatment (b), and block (c). Figure 3.3 contains the symbol legend.

Table 3.4. MRPP analysis of nitrogen-fixing bacterial differences- June '07

Dataset Section Analyzed	Depth		Treatment		Block	
	A	p-value	A	p-value	A	p-value
All Depths	0.118	<0.001	-0.008	0.901	0.031	0.004
10-40cm	0.074	0.002	-0.013	0.801	0.019	0.178
40-100cm	0.040	0.051	-0.021	0.956	0.110	<0.001
10cm	-	-	-0.020	0.728	-	-
20cm	-	-	-0.023	0.762	-	-
40cm	-	-	ND	ND	-	-
60cm	-	-	-0.028	0.791	-	-
80cm	-	-	-0.087	0.926	-	-
100cm	-	-	-0.072	0.898	-	-

ND = no data.

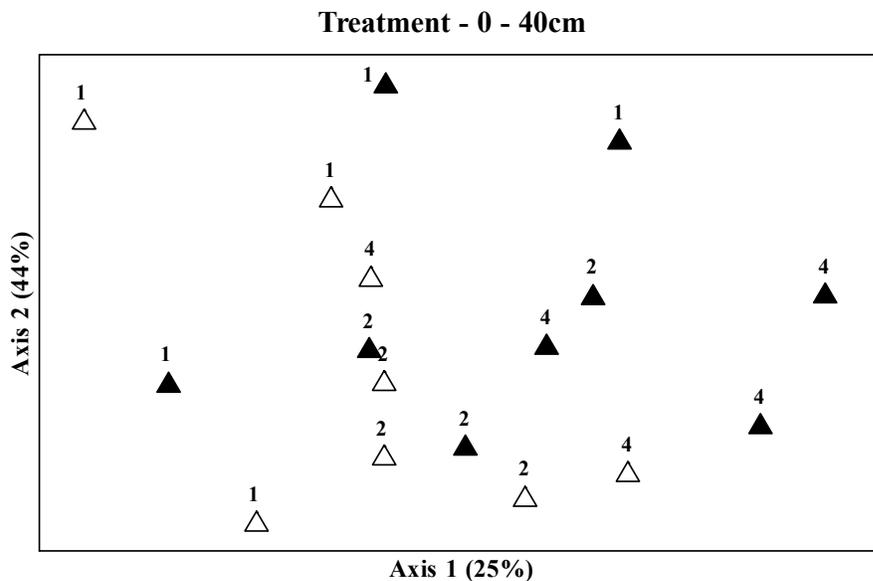


Figure 3.10. NMS analysis of the surface 3 depths of June 2007 *nifH* T-RFLP dataset showing relationship between treatments. 1 = 0 – 10cm; 2 = 10 – 20cm; 4 = 20 – 40cm. Figure 3.3 contains the symbol legend.

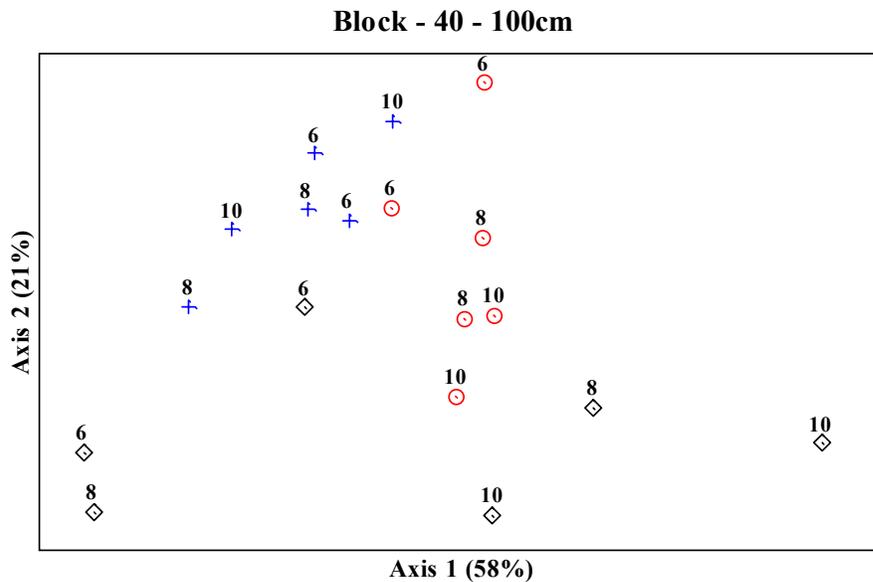


Figure 3.11. NMS analysis of the bottom 3 depths of June 2007 *nif* T-RFLP dataset, showing the relationship between blocks. 6 = 40 – 60cm; 8 = 60 – 80cm; 10 = 80 – 100cm. Figure 3.3 contains the symbol legend.

These two large sources of variability were not present in this study. As a result, differences in microbial communities were predominately shaped by depth. This has been reported in previous studies (Kuske et al. 2002, Feng et al. 2003, Fierer et al. 2003, Allison et al. 2007). Depth more strongly shaped the general bacterial community than diazotroph communities, as ordination plots of bacterial communities resulted in more consistent and definitive groupings by depth than for diazotrophs. However differences between blocks were more prevalent in diazotroph communities than in general bacterial communities, and were most prevalent at lower depths in both communities.

The main factors shaping *nifH* gene diversity and structure are not well understood (Zehr et al. 2003) and reports on free-living N₂ fixing communities in grasslands are few. To our knowledge, this work is the first report on free-living soil nitrogen fixing communities through depth. These results, in conjunction with the regional study, show that even in the absence of large drivers such as site and long-term management history, free-living nitrogen fixing communities are less influenced by depth and more influenced by field-level soil heterogeneity than are general bacterial communities.

In this study, consistent treatment differences between newly converted agricultural fields and native prairie were not found in either the general bacterial or the free-living nitrogen fixing communities. Three years after the conversion of these prairie plots to no-till annual cropping has resulted in declines in microbial biomass abundance, but not in detectable shifts in bacterial or nitrogen fixing community structure.

Collaborative work on nematode communities by Tianna DuPont and Howard Ferris at UC Davis have shown that significant differences in nematodes communities between the two treatments. Many of the same trends in nematode trophic groups and food web indices found in the regional study are also beginning to emerge in this conversion study (S.T. DuPont, unpublished data). Although it is unclear, our knowledge of soil microbial dynamics suggest that given enough time, differences in bacterial and diazotroph community structure will be observed between these two systems. However, tillage which is known to dramatically alter microbial communities (Paul and Clark 1996, Drijber et al. 2000) has been controlled for here. The relative influence of these plant communities in the absence of tillage should shed insight into the role they play in shaping soil organic carbon status and soil biota. Documenting these interactions will only be possible with continued monitoring of these changes over time.

REFERENCES

- Acosta-Martinez, V., M. M. Mikha, and M. F. Vigil. 2007. Microbial communities and enzyme activities in soils under alternative crop rotations compared to wheat-fallow for the Central Great Plains. *Applied Soil Ecology* **37**:41-52.
- Allison, V. J., Z. Yermakov, M. E. Miller, J. D. Jastrow, and R. Matamala. 2007. Assessing Soil Microbial Community Composition Across Landscapes: Do Surface Soils Reveal Patterns? *Soil Science Society of America Journal* **71**:730–734.
- Baker, J. M., T. E. Ochsner, R. T. Venterea, and T. J. Griffis. 2007. Tillage and soil carbon sequestration - What do we really know? *Agriculture Ecosystems & Environment* **118**:1-5.
- Davidson, E. A., and I. L. Ackerman. 2006. Changes in soil carbon inventories following cultivation of previously untilled soils. *Biogeochemistry* **20**.
- Drijber, R. A., J. W. Doran, A. M. Parkhurst, and D. J. Lyon. 2000. Changes in soil microbial community structure with tillage under long-term wheat-fallow management. *Soil Biology & Biochemistry* **32**:1419-1430.
- Drury, C. F., C. S. Tan, T. W. Welacky, T. O. Oloya, A. S. Hamill, and S. E. Weaver. 2005. Red clover and tillage influence on soil temperature, water content, and corn emergence. *Agronomy Journal* **91**:101–108.
- Elliott, E. T., J. W. Heil, E. F. Kelly, and H. C. Monger. 1999. Soil structural and other physical properties. Pages 74-85 in G. P. Robertson, D. C. Coleman, C. S. Bledsoe, and P. Sollins, editors. *Standard Soil Methods for Long-Term Ecological Research*. Oxford University Press, New York, New York.
- Fabrizzi, K. P., F. O. Garcia, J. L. Costa, and L. I. Picone. 2005. Soil water dynamics, physical properties and corn and wheat response to minimum and no-tillage systems in the southern Pampas of Argentina. *Soil and Tillage Research* **81**:57–69.
- Feng, Y., A. C. Motta, D. W. Reeves, C. H. Burmester, E. van Santen, and J. A. Osborne. 2003. Soil microbial communities under conventional-till and no-till continuous cotton systems. *Soil Biology & Biochemistry* **35**:1693-1703.
- Fierer, N., and J. P. Schimel. 2003. A Proposed Mechanism for the Pulse in Carbon Dioxide Production Commonly Observed Following the Rapid Rewetting of a Dry Soil. *Soil Science Society of America Journal* **67**:798–805.

- Fierer, N., J. P. Schimel, and P. A. Holden. 2003. Variations in microbial community composition through two soil depth profiles. *Soil Biology & Biochemistry* **35**:167-176.
- Fornara, D. A., and D. Tilman. 2008. Plant functional composition influences rates of soil carbon and nitrogen accumulation. *Journal of Ecology* **96**:314–322.
- Fu, S. L., D. C. Coleman, P. F. Hendrix, and D. A. Crossley. 2000. Responses of trophic groups of soil nematodes to residue application under conventional tillage and no-till regimes. *Soil Biology & Biochemistry* **32**:1731-1741.
- Gebhart, D. L., H. B. Johnson, H. S. H. S. Mayeux, and H. W. Polley. 1994. The CRP increases soil organic- carbon. *Journal of Soil and Water Conservation* **49**:488-492.
- Jackson, R. B., J. Canadell, J. R. Ehleringer, H. A. Mooney, O. E. Sala, and E. D. Schulze. 1996. A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**:389-411.
- Jenkinson, D. S., and D. S. Powlson. 1976. Effects of Biocidal Treatments on Metabolism in Soil .5. Method for Measuring Soil Biomass. *Soil Biology & Biochemistry* **8**:209-213.
- Johnson, M. D., and B. Lowery. 1985. Effect of three conservation tillage practices on soil temperature and thermal properties. *Soil Science Society of America Journal* **49**:1547–1552.
- Johnston, A. 1976. Experiments on continuous growth of arable crops at Rothamsted and Woburn experimental stations: Effects on crop yields and soil analysis and recent modifications in purpose and design. *Annales agronomiques* **27**:927-956.
- Kansas State University Agricultural Experiment Station and Cooperative Extension Service (KSUAES). 1997. *Wheat Production Handbook C-529*.
- Kansas State University Agricultural Experiment Station and Cooperative Extension Service (KSUAES). 1996. *Kansas Crop Planting Guide L-818*.
- Khan, S. A., R. L. Mulvaney, T. R. Ellsworth, and C. W. Boast. 2007. The Myth of Nitrogen Fertilization for Soil Carbon Sequestration. *Journal of Environmental Quality* **36**:1821-1832.
- Kuske, C. R., L. O. Ticknor, M. E. Miller, J. M. Dunbar, J. A. Davis, S. M. Barns, and J. Belnap. 2002. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Applied and Environmental Microbiology* **68**:1854-1863.

- Larney, F. J., and E. J. Kladvko. 1989. Soil strength properties under four tillage systems at three long-term study sites in Indiana. *Soil Science Society of America Journal* **53**:1539–1545.
- Lucas, R. E., J. B. Holtman, and L. J. Connor. 1977. Pages 333-351 *in* W. Lockeretz, editor. *Agriculture and Energy*. Academic Press, New York.
- Lynch, J. M., and L. M. Panting. 1980a. Cultivation and the Soil Biomass. *Soil Biology & Biochemistry* **12**:29-33.
- Lynch, J. M., and L. M. Panting. 1980b. Variations in the Size of the Soil Biomass. *Soil Biology & Biochemistry* **12**:547-550.
- Mann, L. K. 1986. Changes in Soil Carbon Storage after Cultivation. *Soil Science* **142**:279-288.
- McCune, B., and J. B. Grace. 2002. *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, OR.
- McLaughlin, A., and P. Mineau. 1995. The impact of agricultural practices on biodiversity. *Agriculture Ecosystems & Environment* **55**:201-212.
- MEA. 2006. *Millennium Ecosystem Assessment Synthesis Report Strengthening Capacity to Manage Ecosystems Sustainably for Human well Being*. Island Press, Washington, D.C.
- MEAS. 1998. *Missouri Agricultural Experiment Station: Recommended Chemical Soil Test Procedures for the North Central Region*.
- Mielke, P. W., Jr. 1984. Meteorological applications of permutation techniques based on distance functions. *in* P. R. Krishnaiah and P. K. Sen, editors. *Handbook of Statistics*. Elsevier Science Publishers.
- Paterson, E. 2003. Importance of rhizodeposition in the coupling of plant and microbial productivity. *European Journal of Soil Science* **54**:741-750.
- Paul, E. A., and F. E. Clark. 1996. *Soil Microbiology and Biochemistry*. Academic Press, NY.
- Pimentel, D., C. Harvey, P. Resosudarmo, K. Sinclair, D. Kurz, M. McNair, S. Crist, L. Shpritz, L. Fitton, R. Saffouri, and R. Blair. 1995. Environmental and Economic Costs of Soil Erosion and Conservation Benefits. *Science* **267**:1117-1123.
- Potter, K. N., H. A. Torbert, H. B. Johnson, and C. R. Tischler. 1999. Carbon storage after long-term grass establishment on degraded soils. *Soil Science* **164**:718-725.

- Qin, R., P. Stamp, and W. Richner. 2004. Impact of Tillage on Root Systems of Winter Wheat. *Agron J* **96**:1523-1530.
- Qin, R., P. Stamp, and W. Richner. 2005. Impact of Tillage and Banded Starter Fertilizer on Maize Root Growth in the Top 25 Centimeters of the Soil. *Agron J* **97**:674-683.
- Reicosky, D. C., W. A. Dugas, and H. A. Torbert. 1997. Tillage-induced soil carbon dioxide loss from different cropping systems. *Soil and Tillage Research* **41**:105-118.
- Roberts, E. M., P. B. English, J. K. Grether, G. C. Windharn, L. Somberg, and C. Wolff. 2007. Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. *Environmental Health Perspectives* **115**:1482-1489.
- Robertson, G. P., P. Sollins, B. G. Ellis, and K. Lajtha. 1999. Exchangeable ions, pH, and cation exchange capacity. Pages 106-114 in G. P. Robertson, D. C. Coleman, C. S. Bledsoe, and P. Sollins, editors. *Standard Soil Methods for Long-Term Ecological Research*. Oxford University Press, New York, New York.
- Schimel, D. S., D. C. Coleman, and K. A. Horton. 1985. Soil Organic-Matter Dynamics in Paired Rangeland and Cropland Toposequences in North-Dakota. *Geoderma* **36**:201-214.
- Seybold, C. A., and J. E. Herrick. 2001. Aggregate stability kit for soil quality assessments. *Catena* **44**:37-45.
- Sparling, G. P., C. W. Feltham, J. Reynolds, A. W. West, and P. Singleton. 1990. Estimation of soil microbial C by a fumigaion-exatraction method: use on soils of high organic matter content and a reassessment of the KEC-factor. *Soil Biology & Biochemistry* **22**:301-307.
- Steenwerth, K. L., L. E. Jackson, E. A. Carlisle, and K. M. Scow. 2006. Microbial communities of a native perennial bunchgrass do not respond consistently across a gradient of land-use intensification. *Soil Biology & Biochemistry* **38**:1797-1811.
- Tilman, D. 1999. Global environmental impacts of agricultural expansion: The need for sustainable and efficient practices. *Proceedings of the National Academy of Sciences of the United States of America* **96**:5995-6000.
- Turner, R. E., and N. N. Rabalais. 2003. Linking landscape and water quality in the Mississippi river basin for 200 years. *Bioscience* **53**:563-572.

- VandenBygaart, A. J., E. G. Gregorich, and D. A. Angers. 2003. Influence of agricultural management on soil organic carbon: a compendium and assessment of Canadian studies. *Canadian Journal of Soil Science* **83**:363–380.
- Vyn, T. J., and B. A. Raimbault. 1993. Long-term effect of five tillage systems on corn response and soil structure. *Agronomy Journal* **85**:1074–1079.
- Waldrop, M. P., D. R. Zak, C. B. Blackwood, C. D. Curtis, and D. Tilman. 2006. Resource availability controls fungal diversity across a plant diversity gradient. *Ecology Letters* **9**:1127-1135.
- Wardle, D. A. 1992. A Comparative-Assessment of Factors Which Influence Microbial Biomass Carbon and Nitrogen Levels in Soil. *Biological Reviews of the Cambridge Philosophical Society* **67**:321-358.
- Weil, R. R., K. R. Islam, M. A. Stine, J. B. Gruver, and S. E. Sampson-Liebig. 2003. Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. *American Journal of Alternative Agriculture* **18**:1-15.
- West, T. O., and W. M. Post. 2002. Soil organic carbon sequestration rates by tillage and crop rotation: A global data analysis. *Soil Science Society of America Journal* **66**:1930-1946.
- Zak, D. R., W. E. Holmes, D. C. White, A. D. Peacock, and D. Tilman. 2003. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology* **84**:2042-2050.
- Zehr, J. P., B. D. Jenkins, S. M. Short, and G. F. Steward. 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology* **5**:539-554.

CHAPTER 4
ASSOCIATIVE NITROGEN FIXATION IN THREE
PERENNIAL C4 PRAIRIE GRASSES: A GREENHOUSE STUDY³

I. Introduction

Biological nitrogen fixation (BNF) is a process that takes gaseous nitrogen (N₂) and reduces it to ammonia (NH₃) or another reactive form of nitrogen that is critical for plant growth. BNF is an energetically expensive reaction that is mediated by diazotrophs—bacteria and archaeobacteria containing nitrogenase enzymes. This phenomenon has been of keen interest to agricultural researchers for many decades, as it has potential to increase crop yield, decrease cost of production and lead to greater agricultural sustainability overall. A great deal of progress has been made in understanding and managing symbiotic N fixation in legume-*Rhizobium* associations in agriculture, and it is hoped that research on non-symbiotic/associative N fixation with graminaceous, agronomic plants can yield similar results.

There are several factors that make it difficult to study these diazotrophs and the phenomenon of N fixation. Unlike legume-*Rhizobium* associations, these non-symbiotic/associative diazotrophic bacteria do not form specialized structures and could be living in close association with the plant in the rhizosphere (epiphytic) or actually within the plant roots (endophytic) (James, 2000). Also, within the last decade molecular tools have revealed the diversity of diazotrophs to be much greater than previously believed (Hamelin et al., 2002, Poly et al., 2001). There is mounting evidence which suggests that many of our model, culturable diazotrophic organisms

³ The author is unsure if the work reported in this chapter is sufficient for a stand-alone publication. If so, this chapter could be submitted to *Journal of Rangeland Ecology and Management*, or similar journal.

(e.g. *Azotobacter*, *Azospirillum*, *Azoarcus*) comprise a small non-representative minority of total diazotrophs found in the soil (Hamelin et al., 2002, Poly et al., 2001, Tan et al., 2003, Zehr et al., 2003). Our inability to culture these organisms severely limits our ability to study them or attempt to manipulate these systems. Future research will be needed to gain a more comprehensive knowledge of the majority of the organisms involved in N-fixation.

The over-arching goal of associative N fixation research is to promote and improve the contributions of N to agronomic grasses from nitrogen (N) fixing bacteria. Since N fertilizers are often the greatest input cost for production and because N is often the most limiting nutrient in agricultural systems, development of agronomic crops that could subsidize some of their N need through BNF would likely have significant impacts on agricultural and natural ecosystems on a global scale. However, to date, research efforts have not been able to significantly improve the amount of N supplied to agronomic grasses through associative N fixation (Rao et al., 1998).

History of Associative N Fixation Research

The development of the acetylene reduction (AR) method in the late 1960's led to the widespread use of this method to assess and quantify N fixation. The method has great sensitivity and is relatively inexpensive to perform, giving a researcher a 'snapshot' in time (minutes to days) of N fixing activity (Myrold et al., 1999). Studies using the AR method have shown two consistent patterns among perennial C4 grasses: associative N fixation is commonly occurring and is highly variable depending on genotype and environment.

The discovery of the *Azotobacter paspali*—*Paspalum notatum* (Bahia grass) association by Dobereiner et al. (1972) generated great interest in the scientific community about N fixation in grasses. A subsequent study from the same research

group (Day et al., 1975) looked at nitrogenase activity in several tropical perennial forage species including, *Urochloa maxima* [*Panicum maximum*] (guinea grass), *Pennisetum purpureum* (elephant grass), *U. mutica* [*Brachiaria mutica*] (para grass), *Digitaria eriantha* (common finger grass), *Cynodon dactylon* (bermuda grass), and *Melinis minutiflora* (molasses grass). They showed several of these grasses to have potential to fix economically significant levels of N, but concluded that forage genotypes can vary widely in the degree of N fixation they support.

AR showed associative N fixation occurred in almost all of the tropical perennial forage grasses that were screened in a study in Zimbabwe (*Andropogon eucomus*, *U. mutica*, *Digitaria gazensis*, *Hyparrhenia filipendula*, *H. altissima*, *Paspalum urvillei*, *Melinis repens* [*Rhynchelytrum repens*], *Setaria sphacelata* and *Sporobolus pyramidalis*) although the results were variable and the estimated contributions of total plant N from these data were low (Maasdorp, 1987).

In the US, Tjepkema and Burris (1976) used AR to examine N fixation in seven prairie grasses, *Panicum virgatum* (switchgrass), *Sporobolus heterolepis* (prairie dropseed), *Spartina pectinata* (prairie cordgrass), *Hesperostipa spartea* [*Stipa spartea*] (porcupine grass), *Andropogon gerardi* (big bluestem), *Schizachyrium scoparium* [*A. scoparius*] (little bluestem), and *Poa pratensis* (Kentucky bluegrass), across 16 time frames in a growing season. They found most N fixation rates to be fairly low and to occur mostly in the first two species. These rates of N fixation were correlated with soil moisture.

These early studies and others with maize (Von Bulow and Dobereiner, 1975), rice (Yoshida and Ancajas, 1971, Yoshida and Ancajas, 1973) and other agronomic crops created great interest and optimism in the literature about the associative N fixation in both perennial and annual agronomic grasses. However, AR is limited in its

ability to measure N fixation over an entire growing season, and it became increasingly clear that this technique would not allow for reliable extrapolation of the total amount of N accumulated in the plant from BNF. Isotopic ^{15}N studies hold a great advantage over AR, in this regard, as they can provide an integrative measure of total N derived from the atmosphere. ^{15}N studies can and have been used in a variety of ways to measure associative N fixation, but usually involve labeling soil ^{15}N or atmospheric ^{15}N (isotopic dilution), or by taking advantage of the natural enrichment of ^{15}N in soil (natural abundance). It has become more popular in recent years as this technology has become more affordable and the sensitivity of mass spectrometers has increased (Myrold et al., 1999).

In the last 20 years isotopic ^{15}N studies (often complemented with AR and/or other methods) have settled some disputes by giving more accurate estimates of total N accumulated in plants through BNF. Morris et al. (1985) used AR and ^{15}N to examine N fixation in 9 rangeland plants across 25 field sites in Texas. They found N fixation to vary greatly based on genotype, soil moisture and technique used, and extrapolated fixation rates to range from 0 – 20 kg N ha⁻¹ yr⁻¹ for the entire study. Another study using ^{15}N with *P. purpureum*, *Urochloa brizantha* [*Brachiaria brizantha*] and *U. ruziziensis* [*B. ruziziensis*] (Reis et al., 2001) echoed the results of Morris et al. with regard to genotype variation in N fixation. However, estimated BNF inputs were higher than in the Morris et al. study, with the highest *Pennisetum* and *Brachiaria* estimates of the accumulated N from BNF at 41% and 20%, respectively.

A recent paper by Dalton et al. (2004) utilized several techniques to confirm N fixation in two widespread species of dune grass on the Oregon coast, *Ammophila arenaria* (European beachgrass) and *Leymus mollis* [*Elymus mollis*] (American dunegrass). Their results indicate that N-fixation likely contributes to the great success of these grasses on nutrient poor sand dunes.

There are several other reports like those listed above, essentially documenting the ability of specific grasses to support associative N fixation. There are however, relatively few model systems where investigation has been quite extensive. Two of these model systems will be discussed below in greater depth, as they are the only cases that have reported direct evidence of fixed N transferred from a specific diazotroph to the host plant. This was possible through inoculation experiments with a non-nitrogen fixing (Nif^-) mutant (Hurek et al., 2002, Sevilla et al., 2001), and they represent possibly the most promising cases that associative and/or endophytic diazotrophs could be exploited for better crop performance.

Two Case Studies

One of the most well studied endophytic N fixation associations in grasses is with sugarcane (*Saccharum* sp.) and *Acetobacter diazotrophicus* (synon. *Gluconacetobacter diazotrophicus*). Four million hectares of sugar cane (*Saccharum* spp.) are grown in Brazil each year. In many areas the sugar cane is cultivated without N input and yet high yields of this crop are still obtained (Triplett, 1996). ^{15}N and N balance techniques have shown that sugarcane can attain up to 80% of its needs through N fixation, with 150 and even 200 kg N ha⁻¹ yr⁻¹ attained through associative BNF (Boddey and Dobereiner, 1995, Dobereiner et al., 1993, Lima et al., 1987, Urquiaga et al., 1992). The large levels of N fixation observed in sugarcane are largely thought to be from N fixation from *A. diazotrophicus* and the *Herbaspirillum* genus (Dobereiner et al., 1993, Lima et al., 1987).

A. diazotrophicus is a small, Gram-negative, aerobic rod. It has been shown to have high tolerance for oxygen, can grow on a 10% sucrose solution as its sole carbon source, prefers a very low pH for optimum growth, and shows little to no inhibition of nitrogenase activity in the presence of nitrates and/or ammonia (Boddey and

Dobereiner, 1995, Boddey et al., 1991, James, 2000). More importantly, researchers have demonstrated its capacity to directly transfer half of the N it fixes to amylolytic yeast (*Lypomyces kononenkoae*) in a mixed culture which suggests that it would be capable of a similar transfer in the sugar cane plant (Cojho et al., 1993).

Additionally, the relationship between *Acetobacter* and sugar cane is now viewed by many as an effective experimental model. Studies in India using *A. diazotrophicus* as an inoculum in sugar cane effectively increased yields in four varieties when applied in association with vesicular arbuscular mycorrhizae. The researchers in this case believe that this practice can be substituted for recommended application of urea-nitrogen in this system (Muthukumarasamy et al., 1999). There is an extensive body of literature involving sugarcane and BNF, especially from Brazil. For a through review, see Baldani and Baldani (2005).

Another intensively studied diazotroph is *Azoarcus* sp. Strain BH72, an endophytic Gram-negative N₂-fixing bacterium. It was originally isolated from kallar grass (*Leptochloa fusa* Kunth) found growing in the saline-sodic soils typical of Pakistan (Reinhold-Hurek et al., 1993). Several other diazotrophs have been isolated from the rhizosphere of kallar grass but the Gram-negative rods of *Azoarcus* predominate inside the roots (Reinhold et al., 1986). Similar to the relationship between sugarcane and *Acetobacter*, *Azoarcus* is found only in the xylem of the plant and never in living cells (Hurek et al., 1994b).

Bacteria of the genus *Azoarcus* have a strictly aerobic type of metabolism and have been found to fix N microaerobically (Hurek and Reinhold-Hurek, 1995). Some studies indicate that *Azoarcus* may fix N₂ *in planta* and transfer the fixed N to the kallar host plant (Hurek et al., 2002). In fact, in various green house and field experiments conducted by Malik et al. (1997), kallar grass has been shown to fix up to 26% of its N content. This potential plant-associated “symbiosis” merits further study

as a potential endophytic N-fixing microsymbiont for cereals. It has already been shown that other species of *Azoarcus* have been found to colonize other grasses such as rice in both the field and laboratory and that the presence of *Azoarcus* cells in rice seedlings significantly promotes the growth of the plants (Hurek et al., 1994b).

Non-N-related Plant Growth Promotion

In addition to possible transfer of fixed N to plants, a body of literature has emerged over the last few decades examining the possible role diazotrophs play in plant growth promotion. This has been demonstrated in numerous experiments, as Nif mutants (unable to fix N₂) of *Azospirillum*, *Azoarcus*, *Acetobacter* and *Pseudomonas* often promote plant growth (Bashan et al., 1989, Bastian et al., 1998, Fuentes-Ramirez et al., 1993, Hurek et al., 1994a, Lifshitz et al., 1986). Research in this area has focused on the production and export of phytohormones by diazotrophs to the host plant (Dobbelaere et al., 2003). Phytohormones are plant growth regulators, vital to the regulation of plant growth and development. Auxins, cytokinin-like and gibberellin-like substances have all been proposed as possible ways diazotrophs promote plant growth, since inoculations with diazotrophs and these phytohormones often produce similar plant responses. Other possible mechanisms for plant growth promotion may include synthesis of enzymes that can modulate plant growth, increased nutrient uptake, enhanced stress resistance, increased nutrient availability through organic and inorganic phosphate solubilization, vitamin production, biocontrol, increases in root-adhering soil, and interactions with other microorganisms (Dobbelaere et al., 2003).

Justification and Significance of Study

This experiment attempted to address a missing piece of the puzzle regarding the ability of the prairie sites in Chapter 2 to sustain long-term N removal without fertilization. Over the course of the last 75 years (assuming the consistent yields that the county yield averages indicate), the prairie sites have exported roughly 664 kg ha⁻¹ more N than the average for wheat fields in this region (Glover et al, unpublished data). Current N removal rates between prairie sites and wheat fields are not statistically different (roughly 45 kg N ha⁻¹ yr⁻¹), despite the fact that wheat fields typically receive 70 kg N ha⁻¹ yr⁻¹ and prairie sites do not receive fertilization (Glover et al, unpublished data). In order to determine the source of this annually removed N in the hay, a N budget was adapted for our prairie sites, from data obtained at the Konza Prairie LTER site (Blair et al., 1998), approximately 70 miles from the field sites (Table 4.1).

Table 4.1. Nitrogen budget for annually hayed bottomland tallgrass prairie

Source	Minimum (kg N ha ⁻¹)	Maximum (kg N ha ⁻¹)
Inputs into System		
Atmospheric deposition	10	20
Biological N fixation	?	?
Total Inputs =	10	20
Outputs from System		
Leaching and Runoff	0.2	0.4
Denitrification	0	10
Hay removal	35	65
Total Outputs =	35	75
Min Output – Max Input = 15		
Max Output – Min Input = 65		

Biological N fixation and soil N mineralization (and source of internal transfer) are inherently difficult to measure and/or vary greatly based on site-specific environmental factors. As a result, they were treated as unknowns in this budget. These data show that when hay removal is taken into account, rates of N removal, there is a range of 15 – 65 kg N ha⁻¹ yr⁻¹ that can be attributed to soil N mineralization and/or BNF.

Although it's unknown if soil N mineralization is a major source of N removed from these fields, it seems unlikely that mineralization alone is accounting for this export. The amount of N removed from these landscapes over the past 75 years (approximately 3.6 Mg N ha⁻¹) is roughly ¼ of the total N in the first 1 meter of soil (15.4 Mg N ha⁻¹). If all N was coming from mineralized organic matter, yield declines would be expected. However, county hay yields in this region over the past century, show no evidence of yield decline. In addition, previous reports of long-term hay removal from unfertilized grasslands have shown no detectable declines in total soil N (Jenkinson et al., 2004, Mikhailova et al., 2000, Mikhailova and Post, 2006), suggesting that soil N mineralization might not be the primary source of N in the hayed biomass. Although the extent is unknown, it seems likely that biological N fixation could account for a large fraction of this removed N.

The annual contribution of legume N from symbiotic N fixation was estimated for each of the five sites. Legume biomass was measured in June and October of 2006 across all five sites. Since legumes commonly fix around 20-25kg of shoot N for every tonne of shoot dry matter (Peoples and Baldock, 2001), legume N contribution in our systems likely range from 0.2 – 9.2 kg N ha⁻¹ yr⁻¹, with an average yearly contribution of 3.5 N ha⁻¹ yr⁻¹ (Mangan and Crews, unpublished data). The relatively small amounts of N potentially coming from symbiotic N fixation, directed our research interest to associative N fixation in grasses.

This study attempted to quantify the amount of N-fixation that can occur associatively (either in the rhizosphere or endophytically) in the three dominant C4 perennial prairie grasses at the field sites—big bluestem (*Andropogon gerardii*), indiagrass (*Sorghastrum nutans*), and switchgrass (*P. virgatum*). A second experiment looked at the effect of clipping on the amount of N-fixation that can occur in big bluestem. The objectives of this study were to: i) document associative N fixation in three C4 prairie grasses, ii) quantify the relative contribution of associative N fixation to total plant N in these grasses, and iii) measure the effect of clipping on atom ^{15}N % excess in big bluestem shoots.

II. Materials and Methods

Experiment 1. Seeds of big bluestem, indiagrass and switchgrass were collected from the grasses at the Niles field site, in Ottawa County, Kansas (N^o 38.58.145, W^o 97.28.616) in August 2007. These seeds were tested for germination rates, and all three grasses exhibited < 5% germination. As a result, seeds of big bluestem (Kaw cultivar), indiagrass (Cheyenne cultivar) and switchgrass (Blackwell cultivar) with known high rates of germination were purchased from a commercial source (Sharp Brothers Seed, Healy, Kansas) and used instead of seeds collected from the field site.

Three separate patches of each grass species were located in the Niles site. On October 31, 2007, a large plant from each patch was sampled by digging around the plant with a spade to a depth of 40cm. The plant and rooting structure were removed and bulk soil was shaken off the roots. Roots from the same species were composited and all samples were stored at 4°C until processed.

The greenhouse experiment was conducted at University of California, Davis from December 5, 2007 – February 21, 2008. See Figures 4.1 and 4.2 for graphic representations of the methods. The experiment contained 3 species of grass, 3 levels of N, 2 inoculum treatments and 5 replicates, totaling 90 pots. Leonard jars (magenta units) were used in place of standard pots in order to minimize contamination from watering. The jars were made from plastic culture vessels (Phytotechnology Labs, Shawnee Mission, KS) which relied on a cotton-nylon wick to passively transfer nutrient solution from a lower holding container to the potting media (Figure 4.2a). New nutrient solution was added via a tube which bypassed the upper media compartment to the lower nutrient solution holding compartment (Figure 4.2a and 4.2b). This approach sought to avoid the flushing of microorganisms that have settled on the surface of the media into the rooting zone.

A microbial inoculum for each grass species was prepared by blending roots and rhizosphere soil with 700 ml of sterile 0.05 M K_2SO_4 solution in a Warring laboratory blender (New Hartford, CN) at 1300 rpm for 60 s to dislodge rhizosphere and endophytic microorganisms from roots and soil. The resulting slurry (microbial inoculum) was filtered through a sterilized #20 mesh (0.85 mm) filter to remove root segments and then divided evenly into two sterilized containers, one for the inoculum, the second for the control. The control containers were autoclaved for 2 hours to sterilize, and then allowed to cool to room temperature.

Seeds were surface sterilized using a 0.07M (10%) NaOCl for 15 minutes and then rinsed twice with sterile distilled water and blotted on sterile paper towels to dry. Approximately ten seeds were planted in each 350 ml sterilized pots containing a sterilized 50:50 mix of vermiculite and sand. One half of all pots were inoculated by applying 10 ml of the microbial inoculum to the surface of the media above the planted seeds in each pot. The remaining pots received 10 ml of the sterilized (control)

inoculum to account for added nutrients in the (non-sterile) inoculum. Pots were wrapped in aluminum foil to keep light out of media and the nutrient holding compartments. Sodium vapor lights were used to provide a 14 hour photoperiod and the greenhouse temperature was maintained between 16 and 32°C.

Plants were watered with one of three different ½ strength Hoagland's Solution #1 (Hoagland and Arnon, 1950), modified to three different levels of N. Ammonium nitrate ($^{15}\text{NH}_4^{15}\text{NO}_3$) with ^{15}N at 5 atom % excess, was used at 0.25mM, 1mM, and 4mM in order to test of the effects of a N gradient on N fixation rates. Watering took place before planting (Dec 3rd) and periodically throughout the experiment (Dec 21st, Jan 17th, Feb 2nd), whenever the nutrient solution was running low in the lower compartment. A total of 850ml of nutrient solution was administered to each pot, equivalent to 5.95 mg, 23.8 mg and 95.2 mg of N for the 0.25 mM, 1 mM, and 4 mM of N solutions, respectively. All nutrient solution was sterilized before adding to nutrient holding compartments and aseptic techniques were used during watering.

Twenty-nine days after planting (Jan 3rd), seedlings were thinned to one plant per pot, by clipping and removing every seedling except for the most vigorous. Plants were harvested in the pre-boot stage, 78 days after planting, on February 21, 2008. Plant shoots were separated from the roots, oven dried at 70°C for 48 hours and weighed to determine biomass. Before roots were oven-dried and weighed, they were examined for mycorrhizal infection. This was done in order to assess the effectiveness of the sterilization treatment. (It was assumed that if mycorrhizal infection was present in the control pots, then other contaminating microorganisms were likely also present in the rhizosphere of the control plants.) Roots were cleared by placing roots in glass vials with 10% KOH and incubating at 90°C for 1 hour. Roots were then removed and rinsed 5 times with distilled water and acidified by submersing them in 1M HCl for 5 minutes. HCl was drained and the roots were stained with a 1:1:1 mix of glycerol:

water: lactic acid with 0.05% trypan blue. Samples were incubated at 90°C for 45 minutes and rinsed with distilled water 5 times. Stained roots were mounted on slides and examined for mycorrhizal infection with a Nikon dissecting scope. Infection was assessed with one half of all plant roots in the experiment and was scored for presence/absence only.

Experiment 2. The second experiment was conducted on big bluestem grown with the 1 mM level of N (supplied as $^{15}\text{NH}_4^{15}\text{NO}_3$). Ten pots (5 inoculated, 5 control) of big bluestem were grown for a clipping treatment and compared with ten (5 inoculated, 5 control) pots for an unclipped control. Plants and experimental conditions were identical to those outlined in experiment 1. On February 4, 2008, each plant in the clipped treatment was clipped with scissors to simulate a haying event. Beginning with the second leaf, every other leaf was fully clipped. A maximum of three leaves were clipped, as no plant had more than 6 leaves total. Because of the pre-existing N stress observed in the plants, it was unclear if clipping the entire plant (similar to what would occur in an actual haying event) would result in the death of the plant. Instead, only half of the plants' leaves were clipped to ensure the experiment would produce results. Clipped and unclipped plants were harvested on February 21, 2008 with the plants in experiment 1.

All plant shoots were ground to powder by placing samples in 25ml scintillation vials with two, 8 mm stainless steel balls (Winsted Precision Ball Company, Winsted, CT) and shaking vigorously on a modified paint shaker. Ground samples were weighed into 5 x 9 mm tin capsules (Costech Analytical Technologies, Valencia, CA). Total N and atom % ^{15}N excess values were determined with a PDZ Europa ANCA-GSL elemental analyzer interfaced to a PDZ Europa 20-20 isotope

ratio mass spectrometer (Sercon Ltd., Cheshire, UK) at the University of California Davis Stable Isotope Facility.

Analysis of variance was used to test the effects of N level (N), inoculation (I), and $N \times I$ in the first experiment and inoculation (I), clipping (C), and $C \times I$ in the second experiment using a mixed model in SAS v.9 (Cary, NC). In addition, the effects of time (T), inoculation (I), and $T \times I$ on the clipped plants were tested using the same method. All factors were treated as fixed effects, except replicated plots which were treated as random effects. Graphs were plotted in SigmaPlot v.9 (San Jose, CA).

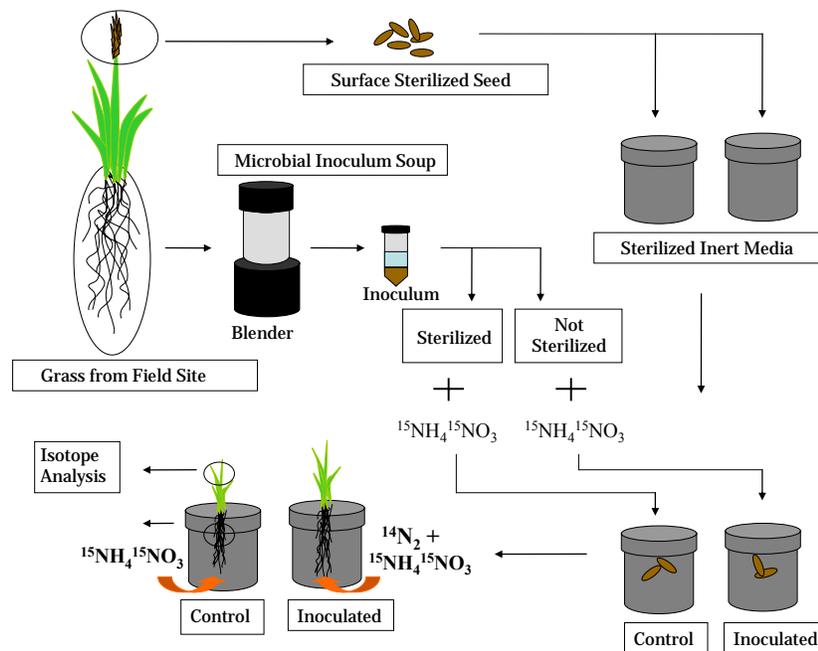


Figure 4.1. Graphic depiction of methods.

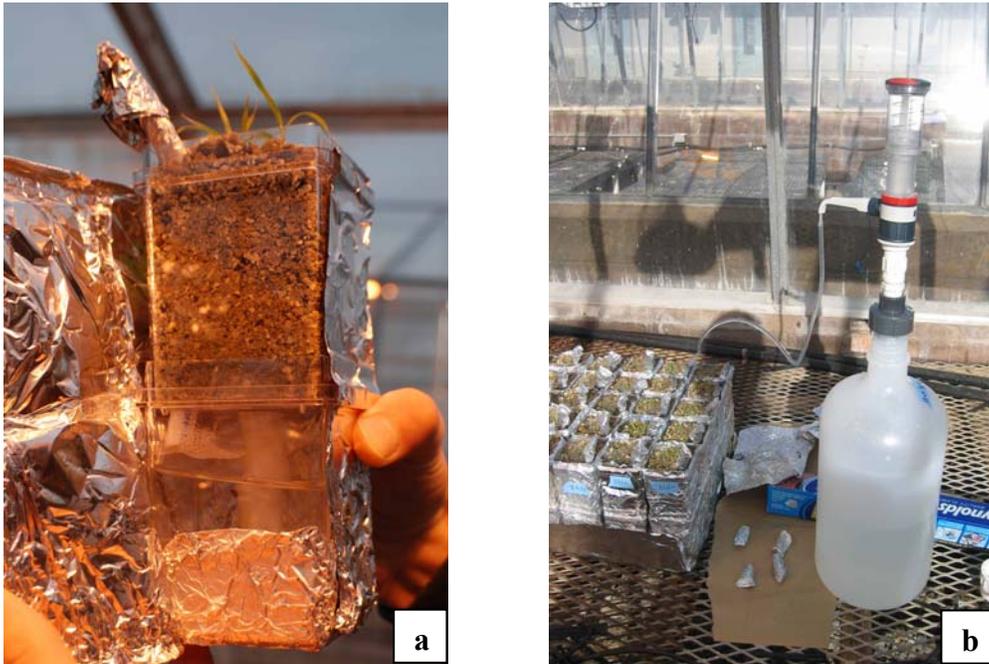


Figure 4.2. Experimental pot (a) and nutrient solution watering set up (b).

III. Results

Twenty-three days after planting (Dec 28th), 100% of the big bluestem pots, 63.3% of the indiangrass pots, and 100% of the switchgrass pots contained germinated seeds. However, throughout the experiment, all plants appeared N-stressed and some died. At harvest, only 83.3% of the big bluestem pots, 30% of the indiangrass pots, and 96.7% of the switchgrass pots contained viable plants. Because of the poor growth of indiangrass plants (9 plants out of 30 possible), these data should be interpreted with caution. Nevertheless, these data are still presented here.

Mycorrhizal infection was found in the roots of both control and inoculated plots in switchgrass. There was no evidence of infection found in indiagrass or big bluestem.

Experiment 1. Table 4.2 shows the summary of ANOVA results for all of experiment 1. Significant differences were detected in the inoculated (I) treatment, the N level (N), and the interaction between I and N ($I \times N$). However, most of these differences did not follow trends that were consistent across all species.

Table 4.2. Summary of significant differences found in experiment 1. Values represent p-values from ANOVA.

	Percent N in Shoots	Shoot Biomass	Total N in Shoots	Atom ¹⁵ N % excess
Bluestem				
Inoculated (I)	0.088	--	--	--
N level (N)	<0.001	--	--	0.002
$I \times N$	--	--	--	--
Indiagrass				
Inoculated (I)	--	--	--	--
N level (N)	--	0.063	0.011	0.031
$I \times N$	--	0.052	0.047	--
Switchgrass				
Inoculated (I)	<0.001	0.040	0.006	0.011
N level (N)	<0.001	--	0.012	<0.001
$I \times N$	--	--	0.011	<0.001

-- = no significant differences detected

In big bluestem, the percent of N found in shoots was greater in the control plants relative to inoculated plants and was also greater in the plants receiving higher levels of N (Figure 4.3a). However, no differences were found in shoot biomass or in

the total amount of N found in the shoots (calculated as percent N in shoots \times shoot biomass). The level of N in the nutrient solution, but not the inoculation treatment, significantly influenced the atom ^{15}N % excess, with lower values being reported at lower nutrient concentrations (Figure 4.3d).

No measured variables in indiangrass were influenced by the inoculation treatment (Table 4.2, Figure 4.4). However, the N level significantly influenced the shoot biomass, the total N in the shoots, and the atom ^{15}N % excess.

All variables measured in switchgrass were significantly affected by the inoculation treatment (Table 4.2, Figure 4.5). The percent N was greater in the inoculated plants, but the shoot biomass, the total N in the shoots, and the atom ^{15}N % excess, were all greater in the control plots. The level of N significantly impacted all variables, except shoot biomass (Table 4.2, Figure 4.5). In most cases, greater N levels in the nutrient solution increased the measured response variables. Significant interactions between N and I were found in indiangrass and switchgrass and resulted from smaller rates of increase in the response variable in inoculated plants relative to the control plants (Figures 4.4 – 4.5).

The only consistent trend observed across all three grasses was with the effect of N level on atom ^{15}N % excess (Table 4.2). Plants grown with nutrient solutions with lower concentrations of N had shoots with consistently lower values of atom ^{15}N % excess (Figures 4.3d, 4.4d and 4.5d).

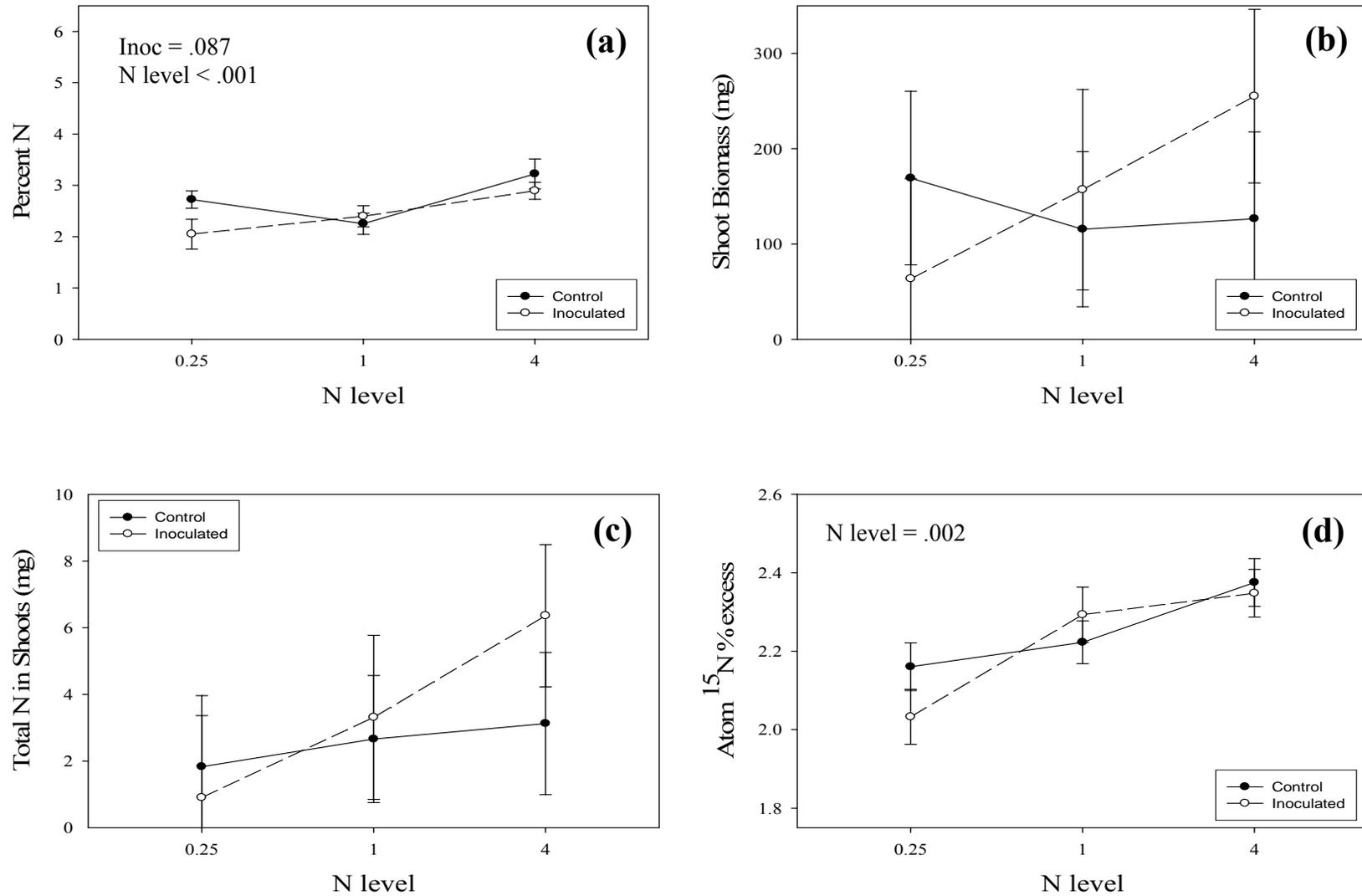


Figure 4.3. Effects of N level and inoculation on percent N (a), shoot biomass (b), total N in shoots at harvest (c) and atom ^{15}N percent excess (d) in big bluestem.

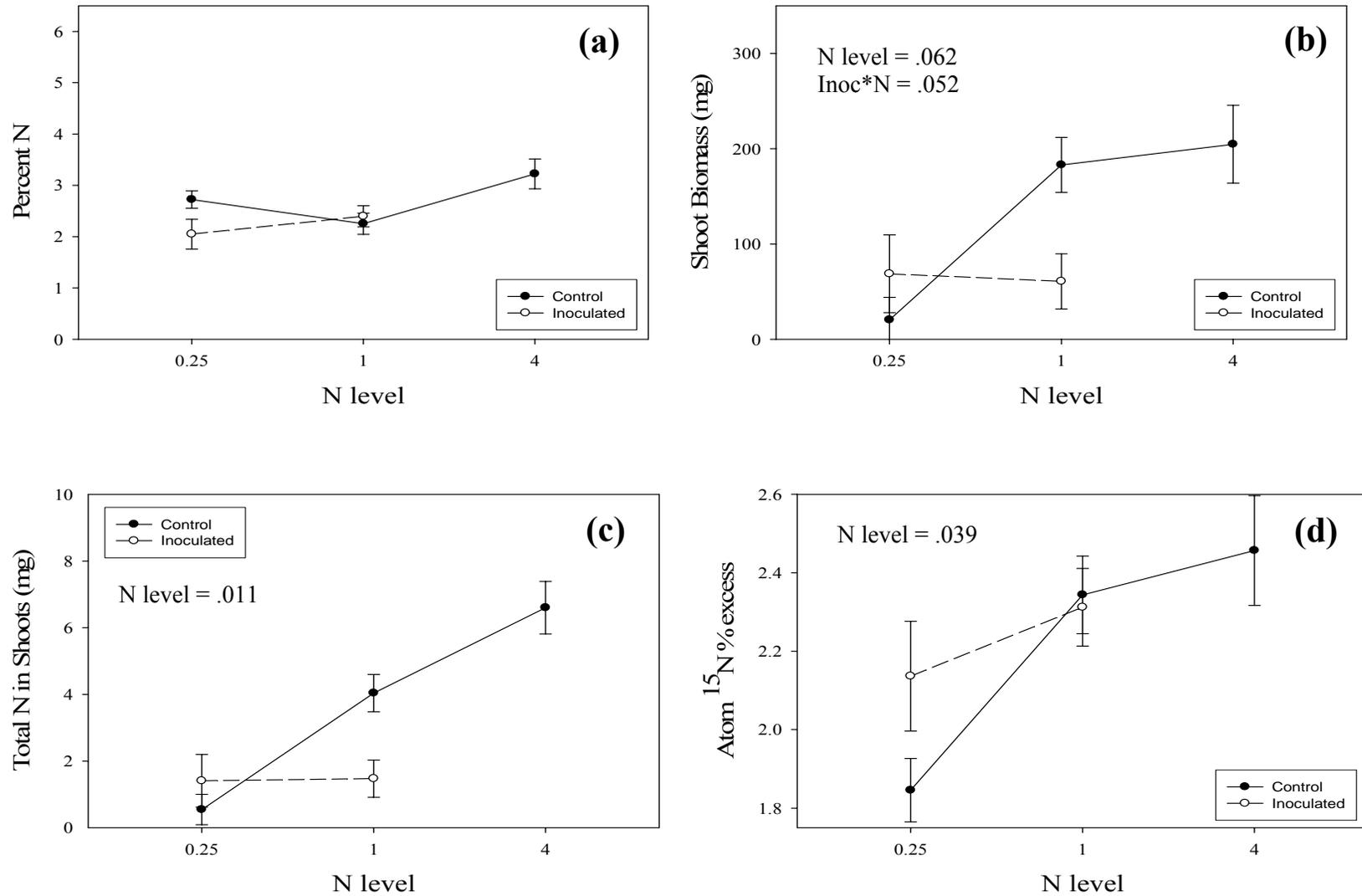


Figure 4.4. Effects of N level and inoculation on percent N (a), shoot biomass (b), total N in shoots at harvest (c) and atom ^{15}N percent excess (d) in indiagrass.

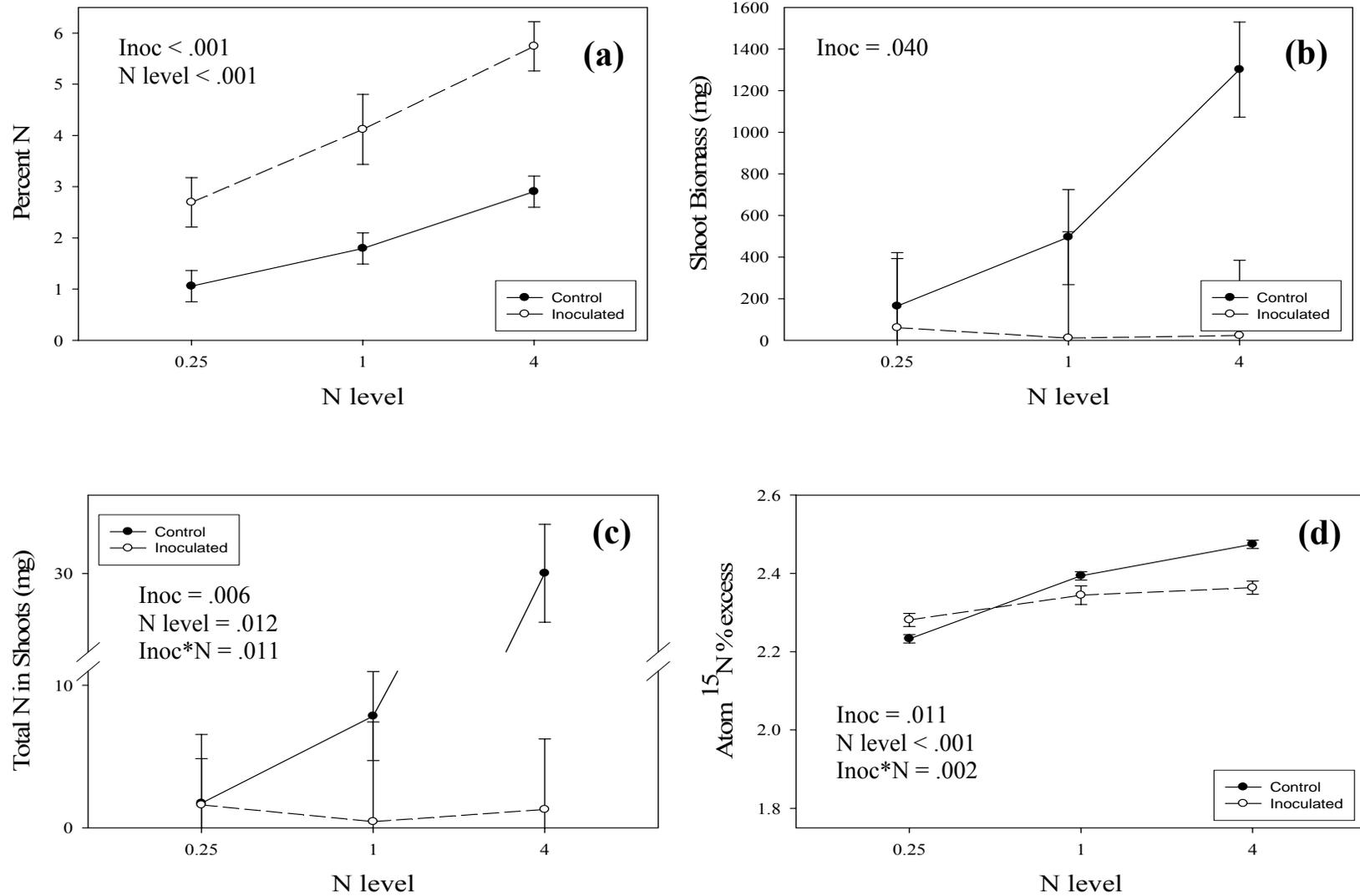


Figure 4.5. Effects of N level and inoculation on percent N (a), shoot biomass (b), total N in shoots at harvest (c) and atom ^{15}N percent excess (d) in switchgrass.

Experiment 2. Clipping big bluestem at 1 mM ammonium nitrate reduced the percent of N in the shoots at $\alpha = 0.10$ (Figure 4.6a). The interaction of inoculation and clipping ($I \times C$) was significant with shoot biomass and total N in the shoots, as clipped, control plants yield greater shoot biomass and total N (Figure 4.6b and 4.6c). Neither clipping nor inoculation had a significant effect on atom ^{15}N % excess (Figure 4.6d).

Shoot material that was clipped (i.e., clipping) to simulate the haying event was compared with the regrown shoot material collected at final harvest (labeled 'harvest'). This comparison was with material that came from the same ten clipped big bluestem plants. The percent N in the clipping and harvest material did not change, nor was it affected by the inoculation treatment (Figure 4.7a). However, the atom ^{15}N % excess was significantly lower in the inoculated plants (Figure 4.7b).

IV. Discussion

This study was conducted to quantify the amount of associative N fixation that occurred in three C4 prairie grasses and to test if clipping influenced this fixation. The evidence of mycorrhizal infection in the control plants of at least one of the grasses suggests that efforts to keep the rooting system free from microorganisms were marginally successful at best. It is likely that since mycorrhizal infection was found in some of the plants, the rooting system of the control plants were not in true sterile environments. It is unclear whether the mycorrhizal infection resulted from a lack of effective seed or inoculum sterilization, or if they were introduced after planting. In any event, these findings compromise the integrity of the control vs. inoculation treatment.

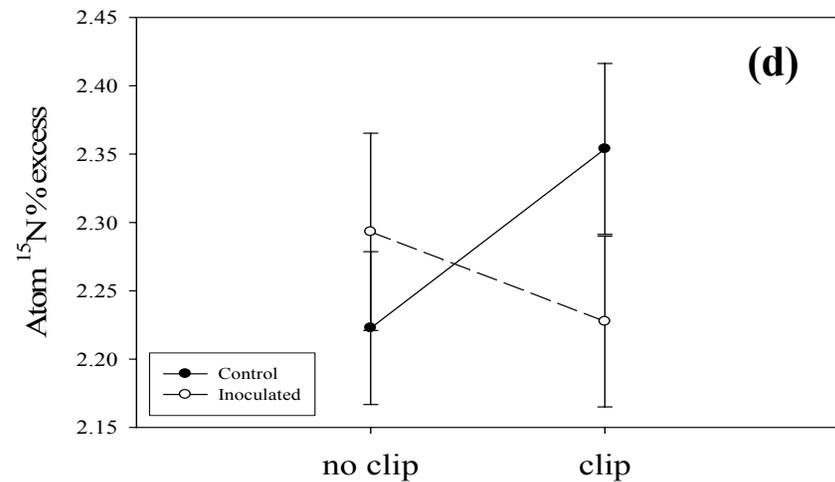
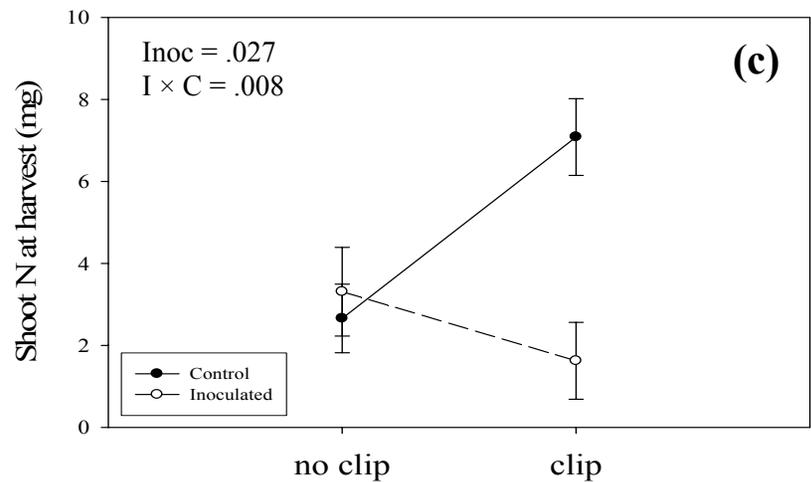
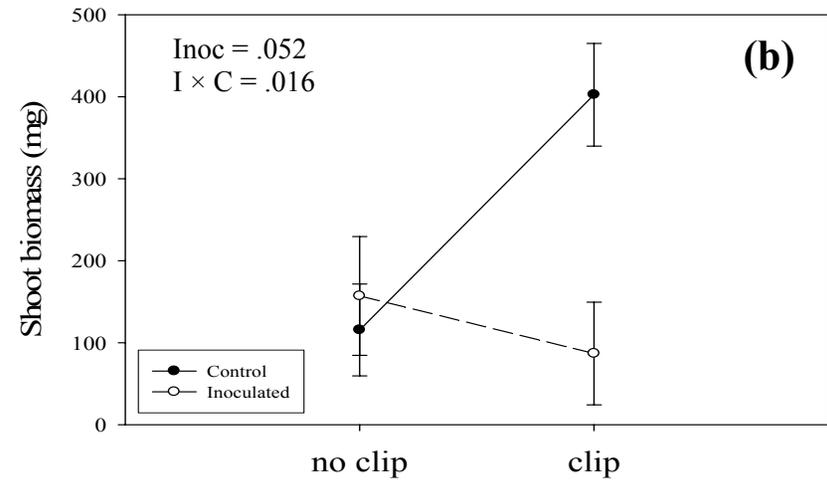
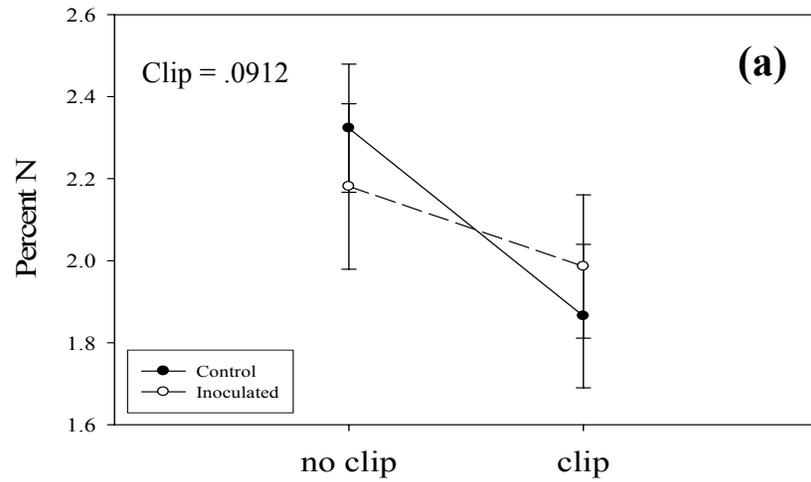


Figure 4.6. Effects of clipping and inoculation on percent N (a), shoot biomass (b), total N in shoots at harvest (c) and atom ¹⁵N percent excess (d) in big bluestem grown with 1mM N.

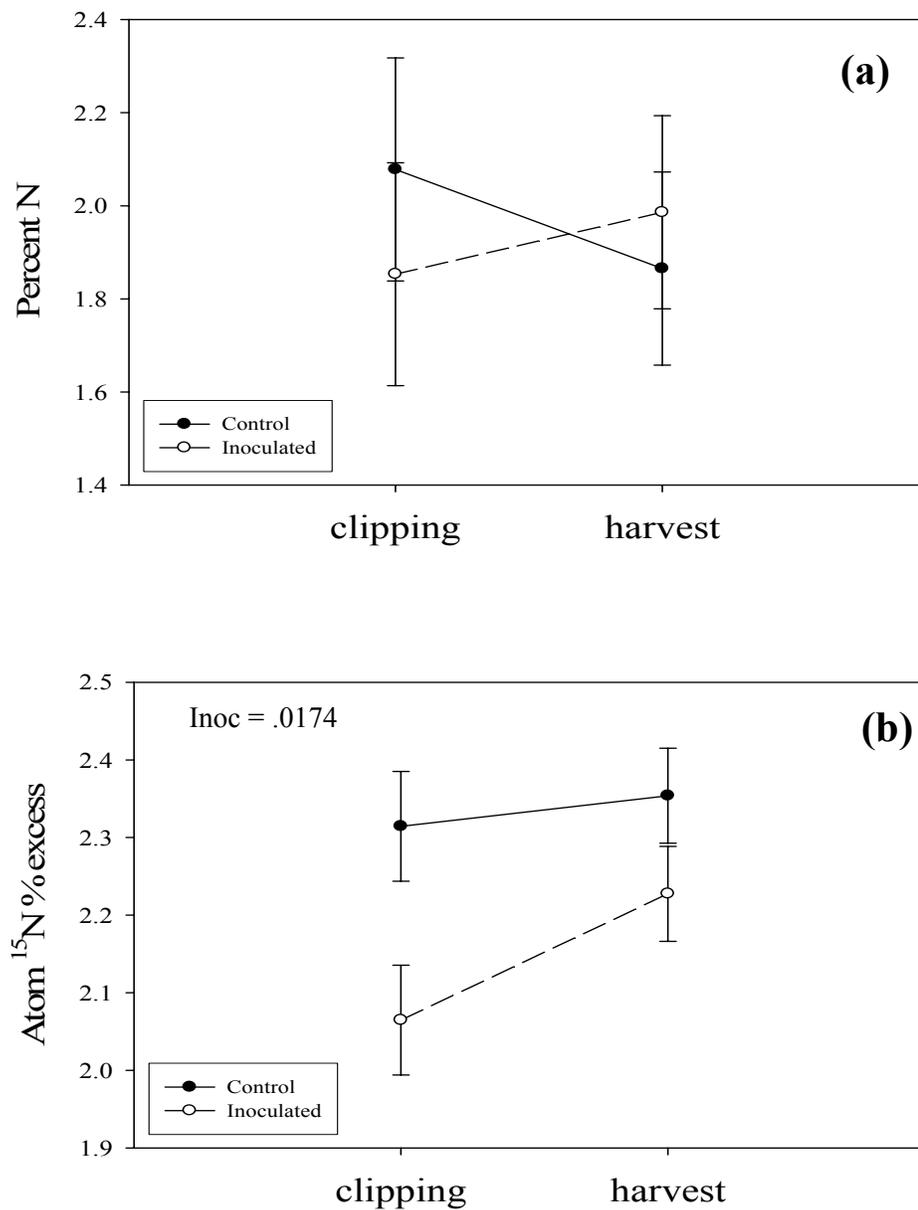


Figure 4.7. Changes with time and inoculation on percent N (a) and atom ¹⁵N percent excess (b) in clipped big bluestem grown with 1mM N. Clipping indicates analysis of clipped material; harvest indicates analysis of shoot biomass at harvest (after regrowth).

Although the inoculation treatment was likely compromised, there were still significant differences detected between control and inoculated plants with percent N in big bluestem and between all measured variables in switchgrass. This demonstrates that despite possible contamination, the microbial communities likely differed enough to influence plant growth. In switchgrass (where mycorrhizal infection was observed), inoculated plants yielded much lower amounts of shoot biomass and total N in the shoots. This suggests a parasitic role of the mycorrhizal fungi, which has been shown previously (Hendrix et al., 1992, Klironomos, 2003, Modjo and Hendrix, 1986) or possibly some other pathology due to microorganisms in the inoculum.

The significant influence of N level on percent N, shoot biomass, and total N in the shoots observed in these grasses indicates that they were N limited during growth, as more N in nutrient concentrations generally yielded greater values in these measured variables.

The only consistent trend across all three grasses was the increase in atom ^{15}N % excess with increase in N concentration in the nutrient solution. Since the nutrient solution was labeled with ^{15}N ammonium nitrate, lower atom ^{15}N % excess values indicate a dilution of the isotopic ^{15}N via N fixation. Lower atom ^{15}N % excess values at lower N levels indicate that more N fixation occurred with these plants, relative to plants grown with higher concentrations of N. Lower levels of available N has corresponded to higher levels of N fixation in many studies on symbiotic N fixation (Crews et al., 2004, Peoples and Herridge, 1990, Streeter, 1998), but to the author's knowledge, this is the first study that demonstrates this phenomenon with associative N fixation. Much of the previously reported work on associative N fixation has involved field-based studies and has shown the variability associated with rates of N fixation are due mainly to genotype (site) differences and soil moisture levels (Morris et al., 1985, Reis et al., 2001 Tjepkema and Burris, 1976). However, assessing effects

of soil N levels on fixation in these studies was not entirely possible, since variability due to differences between field sites was confounding. Here, the effect of available N levels on N fixation was made possible since soil moisture levels were held nearly constant and site conditions were controlled.

In the second experiment, clipping had a stimulatory effect on shoot biomass and total N in shoots with the control plants, but not with the inoculated plants. These results were unexpected, as previous studies have shown that clipping grasses can increase soil N pools and positively feedback into plant N uptake (Hamilton and Frank, 2001). Although it is unclear, it is possible that more N was immobilized into microbial biomass after clipping and less available for plant uptake in the inoculated plants due to differences in microbial communities between inoculated and control plants.

There was not consistent evidence that inoculation with native microorganisms influenced N fixation across all species, but there was evidence of this phenomenon in switchgrass (at 1mM and 4mM NH_4NO_3) and with clipped big bluestem. In these examples, inoculated plants had significantly lower atom ^{15}N % excess values indicating that more N was fixed associatively with these plants than with the control plants. In big bluestem, the influence of inoculation on N fixation was only apparent when considering clipped plants, suggesting an interaction between these factors.

This study was an initial attempt to document associative N fixation. The likely contamination of the control plants inhibits our ability to extrapolate how much N in the plant shoot was derived from the atmosphere, since in many comparisons, inoculated and control plants had similar atom ^{15}N % excess values, indicating no difference in N fixation between these treatments. Although the results are mixed, it is suggestive that both N levels and microbial communities interacted with the plants to determine N fixation rates. There are still many unknowns at this point, such as, how

these N fixation rates change with plant phenology, the effect of plant diversity on N fixation (i.e., interactions with neighboring plants of different species), and long term effects of clipping on associative N fixing dynamics. The results of this study show that associative N fixation in these species in the field is likely, and the rates of fixation are likely influenced to a large extent by available soil N. To what extent that associative N fixation can contribute to plant total N and to how much N is removed from the hayed prairie landscapes remains to be determined. Future work to replicate the results of this study and possible microplot field studies will be needed in the future to answer these questions.

REFERENCES

- Baldani, J.I., Baldani, V.L.D., 2005. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *Anais da Academia Brasileira de Ciencias*. 77, 549-579.
- Bashan, Y., Singh, M., Levanony, H., 1989. Contribution of *Azospirillum brasilense* Cd to growth of tomato seedlings is not through nitrogen fixation. *Canadian Journal of Botany*. 67, 2429-2434.
- Bastian, F., Cohen, A., Piccoli, P., Luna, V., Baraldi, R., Bottini, R., 1998. Production of indole-3-acetic acid and gibberellins A1 and A3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regulation*. 24, 7-11.
- Blair, J.M., Seastedt, T.R., Rice, C.W., Ramundo, R.A., 1998. Terrestrial nutrient cycling in tallgrass prairie. In: A. K. Knapp, J. M. Briggs, D. C. Hartnett, S. L. Collins (Eds.), *Grassland dynamics: longer-term ecological research in tallgrass prairie*, Oxford University Press, New York, NY.
- Boddey, R.-M., Döbereiner, J., 1995. Nitrogen fixation associated with grasses and cereals: Recent progress and perspectives for the future. *Fertilizer Research*. 42, 241-250.
- Boddey, R.M., Urquiaga, S., Reis, V., Döbereiner, J., 1991. Biological Nitrogen Fixation Associated with Sugar Cane. *Plant and Soil*. 137, 111-118.
- Cojho, E.H., Reis, V., Schenberg, A.C.G., Döbereiner, J., 1993. Interactions of *Acetobacter diazotrophicus* with an amolytic yeast in nitrogen-free batch culture. *FEMS Microbiology Ecology*. 106, 341-346.
- Crews, T.E., Brockwell, J., Peoples, M.B., 2004. Host–rhizobia interaction for effective inoculation: evaluation of the potential use of the ureide assay to monitor the symbiotic performance of tepary bean (*Phaseolus acutifolius*). *Soil Biology and Biochemistry*. 36, 1223-1228.
- Dalton, D.A., Kramer, S., Azios, N., Fusaro, S., Cahill, E., Kennedy, C., 2004. Endophytic nitrogen fixation in dune grasses (*Ammophila arenaria* and *Elymus mollis*) from Oregon. *FEMS Microbiology Ecology*. 49, 469-479.
- Day, J.M., Neves, M.C.P., Döbereiner, J., 1975. Nitrogenase activity on the roots of tropical forage grasses. *Soil Biology and Biochemistry*. 7, 107-112.

Dobbelaere, S., Vanderleyden, J., Okon, Y., 2003. Plant growth-promoting effects of diazotrophs in the rhizosphere. *Critical Reviews in Plant Sciences*. 22, 107-149.

Dobereiner, J., Day, J.M., Dart, P.J., 1972. Nitrogenase activity and oxygen sensitivity of the *Paspalum notatum*-*Azotobacter paspali* association. *Journal of General Microbiology*. 71, 103-116.

Dobereiner, J., Reis, V.-M., Paula, M.-A., Olivares, F., 1993. Endophytic diazotrophs in sugar cane, cereals and tuber plants. Palacios, R., Mora, J., Newton, W. E. (Eds.), *Current Plant Science and Biotechnology in Agriculture; New horizons in nitrogen fixation*. Kluwer Academic Publishers, Netherlands.

Fuentes-Ramirez, L.E., Jimenez-Salgado, T., Abarca-Ocampo, I.R., Caballero-Mellado, J., 1993. *Acetobacter diazotrophicus*, an indoleacetic acid producing bacterium isolated from sugarcane cultivars of Mexico. *Plant and Soil*. 154, 145-150.

Hamelin, J., Fromin, N., Tarnawski, S., Teyssier-Cuvelle, S., Aragno, M., 2002. *nifH* gene diversity in the bacterial community associated with the rhizosphere of *Molina coerulea*, an oligonitrophilic perennial grass. *Environmental Microbiology*. 4, 447 – 481.

Hamilton, E.W., Frank, D.A., 2001. Can plants stimulate soil microbes and their own nutrient supply? Evidence from a grazing tolerant grass. *Ecology*. 82, 2397–2402.

Hendrix, J.W., Jones, K.J., Nesmith, W.C., 1992. Control of pathogenic mycorrhizal fungi in maintenance of soil productivity by crop rotation. *Journal of Production Agriculture*. 5, 383-386.

Hurek, T., Handley, L.L., Reinhold-Hurek, B., Piche, Y., 2002. *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Molecular Plant Microbe Interactions*. 15, 233-242.

Hurek, T., Reinhold-Hurek, B., 1995. Identification of grass-associated and toluene-degrading diazotrophs, *Azoarcus spp.*, by analyses of partial 16S ribosomal DNA sequences. *Applied and Environmental Microbiology*. 61, 2257-2261.

Hurek, T., Reinhold-Hurek, B., Turner, G.L., Bergersen, F.J., 1994a. Augmented rates of respiration and efficient nitrogen fixation at nanomolar concentrations of dissolved O₂ in hyperinduced *Azoarcus sp.* strain BH72. *Journal of Bacteriology*. 176, 4726-4733.

Hurek, T., Reinhold-Hurek, B., Van-Montagu, M., Kellenberger, E., 1994b. Root colonization and systemic spreading of *Azoarcus sp.* strain BH72 in grasses. *Journal of Bacteriology*. 176, 1913-1923.

- Hoagland, D.R. and Arnon, D.I. 1950. The water-culture method of growing plants without soil. California Agricultural Experiment Station Circular. 347, 1950.
- James, E.K., 2000. Nitrogen fixation in endophytic and associative symbiosis. *Field Crops Research*. 65, 197-209.
- Jenkinson, D. S., P. R. Poulton, A. E. Johnston, and D. S. Powlson. 2004. Turnover of nitrogen-15-labeled fertilizer in old grassland. *Soil Science Society of America Journal*. 68, 865-875.
- Klironomos, J.N., 2003. Variation in plant response to native and exotic arbuscular mycorrhizal fungi. *Ecology*. 84, 2292–2301.
- Lifshitz, R., Klopper, J.W., Scher, F.M., Tipping, E.M., Laliberte, M., 1986. Nitrogen-fixing *Pseudomonads* isolated from roots of plants grown in the Canadian high arctic. *Applied and Environmental Microbiology*. 51, 251-255.
- Lima, E., Boddey, R.M., Dobereiner, J., 1987. Quantification of biological nitrogen fixation associated with sugar cane using a nitrogen-15 aided nitrogen balance. *Soil Biology and Biochemistry*. 19, 165-170.
- Maasdorp, B.V., 1987. Contribution of Associative Nitrogen-Fixation Acetylene Reduction in Some Grassland Ecosystems in Zimbabwe. *Soil Biology and Biochemistry*. 19, 7-12.
- Malik, K.A., Bilal, R., Mehnaz, S., Rasul, G., Mirza, M.S., Ali, S., 1997. Association of nitrogen-fixing, plant-growth-promoting rhizobacteria (PGPR) with kallar grass and rice. *Plant and Soil*. 194, 37-44.
- Mikhailova, E. A., R. B. Bryant, Vassenev, II, S. J. Schwager, and C. J. Post. 2000. Cultivation effects on soil carbon and nitrogen contents at depth in the Russian Chernozem. *Soil Science Society of America Journal*. 64, 738-745.
- Mikhailova, E. A., and C. J. Post. 2006. Organic carbon stocks in the Russian Chernozem. *European Journal of Soil Science*. 57, 330-336.
- Modjo, H.S., Hendrix, J.W., 1986. The mycorrhizal fungus *Glomus macrocarpum* as a cause of tobacco stunt disease. *Phytopathology*. 76, 688–691.
- Morris, D.R., Zuberer, D.A., Weaver, R.W., 1985. Nitrogen fixation by intact grass-soil cores using nitrogen-15 and acetylene reduction. *Soil Biology and Biochemistry*. 17, 87-92.

- Muthukumarasamy, R., Revathi, G., Lakshminarasimhan, C., 1999. Diazotrophic associations in sugar cane cultivation in South Asia. *Tropical Agriculture*. 76, 171-178.
- Myrold, D.D., Ruess, R.W., Klug, M., 1999. Dinitrogen fixation. In: G. P. Robertson, C. S. Bledsoi, D. C. Coleman, P. Sollins (Eds.), *Standard soil methods for long-term ecological research*, Oxford University Press, N.Y., pp. 241-257.
- Peoples, M.B., Baldock, J.A. 2001. Nitrogen dynamics of pastures: nitrogen fixation inputs, the impact of legumes on soil nitrogen fertility, and the contributions of fixed nitrogen to Australian farming systems. *Australian Journal of Experimental Agriculture* 41:327-346.
- Peoples, M.B., Herridge, D.F., 1990. Nitrogen fixation by legumes in tropical and subtropical agriculture. *Advances in Agronomy*. 44, 155–223.
- Poly, F., Ranjard, L., Nazaret, S., Gourbiere, F., Monrozier, L.J., 2001. Comparison of nifH gene pools in soils and soil microenvironments with contrasting properties. *Applied and Environmental Microbiology*. 67, 2225 – 2262.
- Rao, V.R., Ramakrishnan, B., Adhya, T.K., Kanungo, P.K., Nayak, D.N., 1998. Review: Current status and future prospects of associative nitrogen fixation in rice. *World Journal of Microbiology and Biotechnology*. 14, 621-633.
- Reinhold-Hurek, B., Hurek, T., Gillis, M., Hoste, B., Vancanneyt, M., Kersters, K., DeLey, J., 1993. *Azoarcus*, new genus, nitrogen-fixing Proteobacteria associated with roots of Kallar grass (*Leptochloa fusca* (L.) Kunth), and description of two species, *Azoarcus indigenus*, new species and *Azoarcus communis*, new species. *International Journal of Systematic Bacteriology*. 43, 574-584.
- Reinhold, B., Hurek, T., Niemann, E.G., Fendrik, I., 1986. Close association of *Azospirillum* and diazotrophic rods with different root zones of kallar grass. *Applied and Environmental Microbiology*. 52, 520-526.
- Reis, V.M., dos-Reis-Junior, F.B., Quesada, D.M., de-Oliveira, O.C.A., Alves, B.J.R., Urquiaga, S., Boddey, R.M., 2001. Biological nitrogen fixation associated with tropical pasture grasses. *Australian Journal of Plant Physiology*. 28, 837-844.
- Sevilla, M., Burris, R.H., Gunapala, N., Kennedy, C., 2001. Comparison of benefit to sugarcane plant growth and ¹⁵N₂ incorporation following inoculation of sterile plants with *Acetobacter diazotrophicus* wild-type and Nif- mutant strains. *Molecular Plant Microbe Interactions*. 14, 358-366.
- Streeter, J., 1998. Inhibition of legume nodule formation and N₂ fixation by nitrate. *CRC Critical Reviews of Plant Science*. 7, 1-23.

- Tan, X.Y., Hurek, T., Reinhold-Hurek, B., 2003. Effect of N-fertilization, plant genotype and environmental conditions on nifH gene pools in roots of rice. *Environmental Microbiology*. 5, 1009 – 1015.
- Tjepkema, J.D., Burris, R.H., 1976. Nitrogenase activity associated with some Wisconsin prairie grasses. *Plant and Soil*. 45, 81-94.
- Triplett, E.W., 1996. Diazotrophic endophytes: Progress and prospects for nitrogen fixation in monocots. *Plant and Soil*. 186, 29-38.
- Urquiaga, S., Cruz, K.H.S., Boddey, R.M., 1992. Contribution of nitrogen fixation to sugar cane nitrogen-15 and nitrogen-balance estimates. *Soil Science Society of America Journal*. 56, 105-114.
- Von Bulow, J.F.W., Dobereiner, J., 1975. Potential for nitrogen fixation in maize genotypes in Brazil. *Proceedings of the National Academy of Sciences of the United States of America*. 72, 2389-2393.
- Yoshida, T., Ancajas, R.R., 1971. Nitrogen fixation by bacteria in the root zone of rice. *Soil Science Society of America Proceedings*. 35, 156-158.
- Yoshida, T., Ancajas, R.R., 1973. Nitrogen fixing activity in upland and flooded rice fields. *Soil Science Society of America Proceedings*. 37, 42-46.
- Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology* 5, 539-554.

CHAPTER 5
ANALYSIS OF T-RFLP DATA USING ANALYSIS OF VARIANCE AND
ORDINATION METHODS: A COMPARATIVE STUDY⁴

Abstract

The analysis of T-RFLP data has developed considerably over the last decade, but there remains a lack of consensus about which statistical analyses offer the best means for finding trends in these data. In this study, we empirically tested and theoretically compared ten diverse T-RFLP datasets derived from soil microbial communities using the more common ordination methods in the literature: principal component analysis (PCA), non-metric multidimensional scaling (NMS) with Sørensen, Jaccard and Euclidean distance measures, correspondence analysis (CA), detrended correspondence analysis (DCA), and a technique new to T-RFLP data analysis, the Additive Main Effects and Multiplicative Interaction (AMMI) model. Our objectives were i) to determine the distribution of variation in T-RFLP datasets using analysis of variance (ANOVA), ii) to determine the more robust and informative multivariate ordination methods for analyzing T-RFLP data, and iii) compare the methods based on theoretical considerations. For the 10 datasets examined in this study, ANOVA revealed the variation from Environment main effects was always small, variation from T-RFs main effects was large, and variation from T-RF \times Environment (T \times E) interactions was intermediate. Larger variation due to T \times E indicated larger differences in microbial communities between environments/ treatments and thus demonstrated the utility of ANOVA to provide an objective assessment of community dissimilarity. The comparison of statistical methods

⁴ The chapter is currently in press in the Journal of Microbiological Methods.

typically yielded similar empirical results. AMMI, T-RF-centered PCA, and DCA were the most robust methods in terms of producing ordinations that consistently reached a consensus with other methods. In datasets with high sample heterogeneity, NMS analyses with Sørensen and Jaccard distance were the most sensitive for recovery of complex gradients. The theoretical comparison showed that some methods hold distinct advantages for T-RFLP analysis, such as estimations of variation captured, realistic or minimal assumptions about the data, reduced weight placed on rare T-RFs, and uniqueness of solutions. Our results lead us to recommend that method selection be guided by T-RFLP dataset complexity and the outlined theoretical criteria. Finally, we recommend using binary or relativized peak height data with soil-based T-RFLP data for ordination-based exploratory microbial analyses.

1. Introduction

Terminal restriction fragment length polymorphism (T-RFLP) analysis is a robust and effective DNA-fingerprinting technique commonly used to compare microbial communities (Clement et al., 1998, Liu et al., 1997, Osborn et al., 2000, Thies, 2007, Tiedje et al., 1999). Although the analysis of T-RFLP data has developed considerably over the last decade, there remains a lack of consensus about which statistical analyses offer the best means for finding trends in these data. Researchers surveying recent literature on T-RFLP analyses will find publications with common research objectives that use a wide range of statistical techniques, often with no justification of their selected method.

In this study, we aimed to address this lack of consensus by comparing the more common ordination methods used in the literature—Principal Components Analysis (PCA), Nonmetric Multidimensional Scaling (NMS, MDS, NMDS) with

either Sørensen, Jaccard or Euclidean distance, Correspondence Analysis/Reciprocal Averaging (CA) and Detrended Correspondence Analysis (DCA). The utility of a technique new to T-RFLP data analysis, the Additive Main Effects and Multiplicative Interaction (AMMI) model (Gauch, 1992), was also examined.

Blackwood et al. (2003) compared several T-RFLP datasets using two classification methods with several distance measures. Here, we explore another class of methods commonly employed to analyze T-RFLP data. We focused on ordination methods used for exploratory purposes only, not including analyses which test specific hypotheses (e.g. that two microbial communities are significantly different), or relate microbial community data to environmental variables (i.e., constrained ordinations, such as CCA). The objectives of this study were i) to determine the distribution of variation in a variety of T-RFLP datasets using analysis of variance (ANOVA), ii) to determine the more robust and informative multivariate ordination methods for exploratory analysis of T-RFLP data, and iii) compare the methods based on theoretical considerations.

2. Materials and methods

2.1. T-RFLP datasets

Ten T-RFLP datasets were used in this study (Table 5.1). Here we use the word ‘dataset’ to define a particular microbial community characterized. Each dataset consisted of multiple data matrices which were derived from the same template DNA and reflected the same community. Multiple data matrices resulted from 1) the three ways to represent T-RFLP data (binary [i.e., presence/ absence], peak height, peak area; called ‘types of data’ henceforth) and 2) use of multiple restriction enzymes. We examined 46 data matrices in all, derived from these 10 datasets.

Table 5.1. T-RFLP datasets used in this study

Dataset	System	Site Location	Targeted Community (Primers)	Restriction Enzyme(s)	References
Bioreactor	Fluidized-bed reactor	Milan, Tennessee	Bacteria (27f – 1492r)	HhaI, MspI, RsaI	Blackwood et al., 2003
Khum Bacteria, Ramp Bacteria	Soil solarization in paddy, lowland rice	Khumaltar and Rampur, Nepal	Bacteria (27f – 1492r)	HhaI, Sau96I	Culman et al., 2006
Fractionation '98 Fractionation '99	Conventional and organic corn; Alfalfa	Kellogg Biological Station, Michigan	Bacteria (27f – 1392r)	RsaI	Blackwood and Paul 2003
Prairie	Native tallgrass prairie	Niles, Kansas	Bacteria (27f – 1492r)	HhaI, Sau96I	Culman et al., unpublished data
AWD	Alternating wetting and drying in paddy lowland rice	Chiang Mai, Thailand	Archaea (Ar109f – Ar912r)	HhaI, Sau96I	Sooksa-nguan et al., unpub. Ph.D. Thesis, Suranaree University of Technology, 2007
Khum Fungi, Ramp Fungi	Soil solarization in paddy, lowland rice	Khumaltar and Rampur, Nepal	Fungi (ITS1F – ITS4)	Sau96I	Culman et al., 2006
Multiregional	Soils from wide- ranging environments	Michigan, Tennessee, Nevada	Bacteria (27f – 1492r)	HhaI, MspI, RsaI	Blackwood et al., 2003

The prairie dataset (Culman, unpublished data) was generated with the same procedures (PCR reactions, restriction enzyme digests, 3730 ABI capillary sequencer) as described in Culman et al. (2006); the alternating wetting and drying (AWD) dataset (Sooksa-nguan et al., unpub. Ph.D. Thesis, Suranaree University of Technology, 2007) was generated with the same procedures described by Lueders and Friedrich (2000), with minor modifications. Other datasets have been previously described (Table 5.1). These soil microbial community-based datasets were selected to represent substantial diversity in dataset sizes, targeted microbial communities, imposed treatments and geographical regions. Table 5.2 provides additional detail about the dataset properties and characteristics.

2.2. Data processing

Data processing (determining ‘true peaks’ from noise, manual alignment of peaks, data transformations, etc.) was performed in numerous ways based on the researcher’s original methods. Although the effects of processing raw data were not examined here, Adbo et al (2006) offer a number of good suggestions to consider in this regard.

All T-RFLP electropherograms were visually inspected to ensure quality runs. Electropherograms were tabulated in either GeneScan or Genemapper v3.5 (Applied Biosystems, Foster City, CA) using the Local Southern method as the size-calling algorithm. A baseline threshold of 50 fluorescence units was used to determine ‘true peaks’ from background noise with all datasets, except from the solarization study (Khum and Rampur Bacteria and Fungi), in which the baseline was set at 200. Terminal restriction fragments (T-RFs) less than 50 base pairs (bp) and greater than 500 bp were eliminated from all datasets. Manual alignment of peaks (often called

Table 5.2. Characteristics of T-RFLP datasets used in this study ^a

Dataset	Number of Environments (Treatments, Replicates)	Total number of T-RFs in Data Matrix	Richness^b (Minimum, Maximum T-RFs)	Evenness^c	Percent Empty Cells in Matrix	 Skew ^d	Beta Diversity^e	NMS Stress^f
Bioreactor	12 (3, 4)	66	26.0 (41, 20)	0.95	60.2	1.51	1.54	11.2
Khum Bacteria	64 (16, 4)	217	116.3 (140, 95)	0.87	46.3	3.26	0.86	7.3
Ramp Bacteria	64 (16, 4)	218	114.6 (126, 101)	0.88	47.3	3.00	0.90	7.2
Fractionation '98	48 (12, 4)	120	53.2 (68, 46)	0.92	55.7	1.95	1.26	13.1
Fractionation '99	48 (12, 4)	103	52.6 (61, 46)	0.91	48.9	1.82	0.96	9.8
Prairie	153 (51, 3)	221	88.9 (62, 111)	0.87	59.7	4.92	1.49	9.4
AWD	160 (20, 8)	356	190.5 (168, 221)	0.71	46.5	7.10	0.87	9.0
Khum Fungi	64 (16, 4)	393	168.7 (142, 204)	0.86	57.1	4.72	1.33	9.9
Ramp Fungi	64 (16, 4)	417	175.9 (123, 261)	0.85	57.8	5.29	1.37	9.1
Multiregional	12 (3, 4)	97	36.8 (29, 46)	0.95	62.2	2.06	1.64	0.1

^a For experiments with multiple restriction enzymes, values were averaged across those enzymes. Since skew and NMS stress varied between types of data, these values were averaged over all types of data and enzymes.

^b Defined as the average number of T-RFs in a dataset

^c Pielou's J

^d |Skew| values averaged over binary data, relativized peak height and relativized peak area were 0.8, 4.8, and 6.8, respectively.

^e Defined as: [(total number of T-RFs in a dataset) / (average T-RF richness in the environments)] – 1

^f Sorensen distance measure, 2 axes selected, 50 runs with real data, 0.0001 stability criterion, 50 iterations to evaluate stability; NMS stress values averaged over binary data, relativized peak height and relativized peak area were 8.6, 8.3, and 8.7, respectively.

‘binning’) is often used to account for T-RF drift (improperly sized T-RFs due to differences in fragment migration and purine content [Kaplan and Kitts, 2003, Marsch, 2005]). The Bioreactor, Fractionation ’98, Fractionation ’99 and Multiregional datasets were all manually aligned (with sample identities concealed). The peaks of the remaining datasets were aligned by rounding to the nearest integer (nucleotide) size.

Raw peak height and raw peak area were relativized to account for uncontrolled differences in the quantity of DNA between samples. Relativized peak height was calculated by dividing each raw peak height by the cumulative peak height of that sample. This is analogous to making each peak height a percentage of the total peak height of a sample. Likewise, relativized peak area was calculated by dividing each raw peak area by the cumulative peak area of that sample. Peak areas from four of the ten datasets could not be obtained.

2.3. Sources of variation in T-RFLP datasets

Multivariate T-RFLP data analysis often begins by organizing the data into a species (T-RF) \times samples matrix, analogous to those found in many other applications in community ecology. This matrix will contain three distinct sources of variation: i) main effects for **T-RFs**, also called operational taxonomic units (OTUs); ii) main effects for **Environments**, also called treatments, plots, samples, and iii) interaction effects for **T-RF \times Environment (T \times E)**. When this matrix is subjected to a dimensionality-reduction method, the selected method will generally analyze all three sources of variation simultaneously. However, here we argue that when T-RFLP is used as a tool for exploratory microbial community analysis, the scientifically interesting question commonly being asked is, “*How are the T-RFs differentially*

responding to (i.e. interacting with) the environments?” It is therefore important to distinguish between these sources of variability, and to focus on the T×E interactions.

T-RF variation arises from the fact that some T-RFs occur commonly across all samples in a dataset, while others occur only rarely. In other words, it reflects variability in the means of different T-RFs. When studying macroorganisms, species have a real and tangible meaning to the researcher. On the contrary, for T-RFs, the represented ‘species’, are often considered somewhat of a ‘black-box’ during the analysis of T-RFLP datasets. Microbial ecologists must use caution when conceptualizing T-RFs as unique species or even unique OTUs, as Clement et al. (1998) and others have shown that multiple organisms can share similar or identically-sized T-RFs. Except when working with a specific organism, group-specific primers or with a locally-constructed clone library, individual T-RFs in a T-RFLP dataset are often ambiguous, limiting the researcher from drawing information from them. T-RF variation constitutes inherently simple information on commonness or rareness that can distract a multivariate analysis from capturing the truly complex interaction information.

The logic outlined above for T-RF variation does not apply to the environmental variation in a T-RFLP data matrix, which arises from differences in numbers of peaks or overall signal strength in T-RFLP profiles representing different environments. Environments represent a very real and tangible concept to the researcher. However, true environmental variation (e.g., due to microbial biomass) is masked in T-RFLP analyses by analytical variability related to, for example, DNA purification efficiency, pipetting error, and community structure (Blackwood et al., 2003, Dunbar et al., 2000). As a result, this known source of analytical noise is commonly removed in peak height and area with the relativization process. Hence, when using T-RFLP as a method of exploratory data analysis on microbial community

structure, the $T \times E$ variation is often most relevant to the researcher. In this study, we discuss sources of T-RFLP variation in this context, assuming that $T \times E$ variation is of primary interest.

2.4. Analysis of variance

Two-way ANOVA was performed on all T-RFLP datasets in Table 5.1, using MATMODEL software (Gauch, 2007, Gauch and Furnas, 1991). The percent of variation from each source in the ANOVA (T-RF, Environment, and $T \times E$) was calculated by dividing that source's sum of squares (SS) by the treatment SS and multiplying by 100. The interaction SS was further decomposed into interaction signal SS and interaction noise SS. The interaction noise SS was estimated by multiplying the interaction degrees of freedom (df) by the mean squared error (MSE). The interaction signal SS was estimated by subtracting the interaction noise SS from the interaction (total) SS. The interaction signal SS and interaction noise SS were then divided by the treatment SS to calculate the percent variation in the dataset due to these sources. See Gauch (1992) for more details on these calculations.

2.5. Empirical testing of ordination methods

In order to determine the more robust and informative methods, several of the more common multivariate statistical ordination analyses in the literature were compared: (i) PCA, (ii) CA, (iii) DCA and (iv) NMS using the Sørensen (Bray-Curtis) distance measure, (v) NMS using the Jaccard distance measure, (vi) NMS using the Euclidean distance measure, and (vii) the Additive Main effects and Multiplicative Interaction (AMMI) model. The AMMI model, also known as 'doubly-centered PCA', has been used extensively in agriculture research, particularly in analyses of yield trials. AMMI uses ANOVA to first partition the variation into main effects and

interactions, and then applies PCA to the interactions to create interaction principal components (IPCs) (Gauch, 1992). Therefore, instead of examining overall variability of the data, AMMI can focus on the differential responses of T-RFs to the treatments, that is, T×E. Here, the ‘main effects’ are defined as the T-RF and environment variation. Including the AMMI model, there were 7 separate analyses performed on each of the 46 T-RFLP data matrices, totaling 322 graphs evaluated.

Environments (E) were replicated for all experiments in this study. Initially, two analyses of each statistical method were run on each dataset, one analysis with the original replicated dataset and a second analysis with the averages over replicates, using *T-RF Manager* software (Culman et al., unpublished; <http://cbsusrv02.tc.cornell.edu/TRF/index.aspx>). The two ordinations produced from these analyses were very similar, and if the two graphs were overlaid, the individual replicates would simply scatter somewhat around the averaged value. The two ordinations were equally discriminatory with respect to our criteria for evaluating statistical methods (see below), so we subsequently focused the comparison on the simplified datasets with averages over replicates.

2.6. Statistical software and parameters

- PCA was performed with PC-ORD v4 (MjM Software Design, Gleneden Beach, OR; McCune and Mefford, 1999) with the VARIANCE/COVARIANCE (CENTERED) option selected. This option centers the T-RFs by subtracting the average for each T-RF over environments from each matrix entry for that T-RF. This produces a variance-covariance matrix. By contrast, use of the CORRELATION (STANDARDIZED) option first centers the T-RFs and then divides each matrix entry by the standard deviation for each T-RF, thus producing a correlation matrix. Our reason for this selection is discussed section 3.3.

- CA was performed with PC-ORD, with DOWNWEIGHT RARE SPECIES not selected.
- DCA was performed with PC-ORD using the default settings: (i) DOWNWEIGHT RARE SPECIES was not selected, (ii) RESCALE AXES was selected, (iii) RESCALING THRESHOLD = 0, and (iv) NUMBER OF SEGMENTS = 26.
- NMS was performed with PC-ORD, using the MEDIUM AUTOPILOT mode. This mode specifies: (i) MAXIMUM NUMBER OF ITERATIONS = 200, (ii) INSTABILITY CRITERION = 0.0001, (iii) STARTING NUMBER OF AXES = 4, (iv) NUMBER OF REAL RUNS = 15, and (v) NUMBER OF RANDOMIZED RUNS = 30. Three separate distance measures were selected for the MDS analysis: Sørensen, Jaccard and Euclidean. When the AUTOPILOT mode recommended a final solution that was more or less than 2 axes, the analysis was re-run with the same parameters, but forcing a 2-dimensional solution (AUTOPILOT mode deselected).
- The AMMI analysis was performed with MATMODEL.

Scatterplots of the first two axes of each ordination were graphed with Minitab v.14.1 (State College, PA).

2.7. Criteria for empirical comparisons

Evaluating the ability of an ordination to reveal the true structure of a dataset is problematic, because it is only with simulated data that we know the true and exact structure of a particular dataset. However, with field data, we do have two considerations to aid our evaluation of the accuracy and effectiveness of an ordination: (i) *a priori* information about the experiment's treatment design and microbial community dynamics and (ii) consensus among the ordinations performed. In this study we used these two criteria as a surrogate for the true structure of the data. For each of the 46 data matrices, we compared the scatterplots of the first two axes

produced by the seven ordination methods and evaluated each method's ability to demonstrate known gradients or treatments, such as time, crop phenology, soil depth, etc. We also examined each method's performance against one another within each data matrix, evaluating if the method complied with the consensus reached by the majority of analyses. For each data matrix, every method was scored as demonstrating the expected gradient/s either i) very well, ii) reasonably well or iii) poorly/not at all. In addition, the method was scored as reaching a general consensus with the other methods or not. Consensus could be viewed as a more robust ranking scheme, indicating if the overall interpretation of a particular ordination was similar to the majority of analyses. The ranking of ordination methods was judged for consistency across all datasets to minimize any anomalous results specific to a given dataset. The large number of datasets examined here mitigated the subjectivity of this coarse scoring system, making this empirical assessment more robust than in any previous study.

2.8. Most informative type of data

The most informative type of data (binary, relativized peak height or relativized peak area) was evaluated based on the number of times a consensus with the majority of the methods was reached. The total number of times a consensus was reached was summed over every data matrix for that type of data.

2.9. Theoretical criteria for evaluating methods

The ordination methods in this study were not only compared empirically, but also theoretically. Although these criteria are not exhaustive, the more relevant theoretical aspects when analyzing T-RFLP data are listed below:

- 1) **Assumptions**—What assumptions does the analysis make about the data? Are these assumptions appropriate for microbial community data?
- 2) **Proportion of Variation Represented**—Does the analysis quantify the amount of variation captured in the first 2 (or 3) axes?
- 3) **Integrated Dual Analysis**—Is there an integrated analysis of T-RFs and treatments, or only an analysis of one of these?
- 4) **Uniqueness of Solution**—Given the same data, would multiple users come to similar conclusions? Or does this method suffer from ‘optionitis’? (‘Optionitis’ = an excessive number of choices not determined by objective criteria).
- 5) **Weight of Rare T-RFs**—How much importance does the analysis give rare T-RFs?

3. Results

3.1. Analysis of variance

Table 5.3 shows the distribution of variation within the T-RFLP datasets from ANOVA, arranged in descending order of variation due to main effects (T-RF and E variation). Main effects variation ranged from 89.4% (Bioreactor, relativized peak height) to 39.4% (Multiregional soil, binary). The vast majority of the main effects variation was made up of T-RF variation, with E variation contributing very little. Note that E main effects with relativized peak height and area will be exactly zero, as a result of the relativization process. However, E main effects of binary data, which were not relativized, still contributed no more than 2.0% (Bioreactor) and as little as 0.1% (Ramp Bacteria and Fractionation '98) of the total variation. This indicates that the total number of T-RFs in each sample was nearly constant.

Table 5.3. Percent of variation in T-RFLP datasets from Analysis of Variance ^a

Study	Main	T-RFs	Environments	Interaction	Signal	Noise
	Effects			Effects		
----- % -----						
Bioreactor						
Binary	83.7	81.7	2.0	16.3	6.3	10.0
Height	89.4	89.4	0.0	10.6	6.2	4.4
Khum Bacteria						
Binary	73.7	73.4	0.3	26.3	12.2	14.1
Height	83.4	83.4	0.0	16.6	6.8	9.8
Area	84.8	84.8	0.0	15.2	6.9	8.3
Ramp Bacteria						
Binary	72.6	72.5	0.1	27.4	13.3	14.1
Height	74.5	74.5	0.0	25.5	16.0	9.5
Area	75.5	75.5	0.0	24.5	15.6	8.9
Fractionation '98						
Binary	78.9	78.8	0.1	21.1	11.7	9.4
Height	79.5	79.5	0.0	20.5	15.6	4.9
Fractionation '99						
Binary	75.7	75.4	0.3	24.3	12.6	11.7
Height	76.5	76.5	0.0	23.5	17.0	6.5
Prairie						
Binary	61.9	60.9	1.0	38.1	21.6	16.5
Height	78.0	78.0	0.0	22.0	14.4	7.6
Area	78.2	78.2	0.0	21.8	13.8	8.0
AWD						
Binary	68.2	67.1	1.1	31.8	19.8	12.0
Height	84.0	84.0	0.0	16.0	11.7	4.3
Area	86.7	86.7	0.0	13.3	8.7	4.6
Khum Fungi						
Binary	60.1	59.3	0.8	39.9	14.7	25.2
Height	60.3	60.3	0.0	39.7	21.6	18.1
Area	59.4	59.4	0.0	40.6	21.9	18.7
Ramp Fungi						
Binary	52.9	51.0	1.9	47.1	21.0	26.1
Height	54.4	54.4	0.0	45.6	32.0	13.6
Area	53.5	53.5	0.0	46.5	32.8	13.7
Multiregional						
Binary	39.4	39.1	0.3	60.6	56.6	4.0
Height	46.5	46.5	0.0	53.5	52.2	1.3

^a The values are percentages based on the three T-RFLP data types: binary (presence/absence), height (relativized peak height) and area (relativized peak area). Main effects and interaction effects make up the two main sources of variation within a T-RFLP dataset. The main effects are a subtotal for the T-RFs and Environments. Similarly, the interaction effects are a subtotal for both interaction signal and interaction noise. For experiments with multiple restriction enzymes, the variation across these enzymes was averaged, as this variation was insignificant.

Percentages of variation due to interaction effects and the relative proportion of interaction signal not only varied with the dataset, but this variation coincided with background knowledge of these datasets. For example, the Bioreactor dataset is composed of bacterial communities from several bioreactors that were treated identically. These bacterial communities would be expected to vary little, and the ANOVA confirms this expectation, with the interaction effects making up only 16.3% and 10.6% of the total variation in the binary and peak height data, respectively. Of this variation, just 6% is from interaction signal for both types of data. In contrast, the Multiregional soil dataset is composed of communities from extremely different soil types. These communities appear to be very different, as the majority of the total variation in the binary and peak height data is due to interaction signal (56.6% and 53.5%, respectively). The ANOVA output given in Table 5.2 suggests that the community differences need not be that extreme in order to be detected. Relatively larger interaction variation from the Khum and Ramp fungal datasets compared to the Khum and Ramp bacterial datasets also coincides with the results of the ordinations reported elsewhere (Culman et al., 2006).

ANOVA revealed overall consistency in the distribution of variation between the three types of data: binary, relativized peak height and relativized peak area (Table 5.2). Analyzing binary data resulted in the lowest main effects variation and the highest interaction effects variation in 9 out of the 10 datasets. The ratio of interaction signal to noise was lowest in the binary data 9 out of 10 times and highest in peak height 7 out of 10 times.

3.2. Empirical ordination results

Overall, the ordinations from the seven analyses generally yielded graphs that did not drastically deviate from one another. Consequently, the scoring differences

were limited to a relatively narrow range (Table 5.4). Individual methods are discussed below.

Table 5.4. Results of empirical comparisons between ordination methods with values representing the number of data matrices scored for each category.

Method	Demonstrated known gradient(s) very well	Demonstrated known gradient(s) reasonably well	Demonstrated known gradient(s) poorly/ not at all	Reached consensus with other ordinations
T-RF-centered PCA	25	21	0	46
AMMI	25	21	0	46
CA	21	19	6	42
DCA	23	23	0	46
NMS with Sørensen	26	18	2	44
NMS with Jaccard	25	20	1	45
NMS with Euclidean	22	20	4	42

3.2.1. PCA and AMMI

In this study, both variable-centered (T-RF-centered) PCA and AMMI (doubly-centered PCA) were performed. In T-RF-centered PCA, the T-RF main effects variation is reduced to zero in a manner similar to the relativization process with the environments. Hence, with little to no main effects variation, a T-RF-centered PCA approximates the AMMI analysis, because the environment variation is small (2% or less). As a result, the T-RF-centered PCA and AMMI analysis ordinations were nearly identical and performed equally well at recovering expected gradients, reaching a consensus with every data matrix analyzed (Table 5.4).

Variants of PCA that were not T-RF-centered or T-RF-standardized did not remove the T-RF main effects variation and as a result produced quite different ordinations that were dominated by T-RF main effects. Figure 5.1a shows the results of an environment-centered PCA, an analysis that removes variation that is already

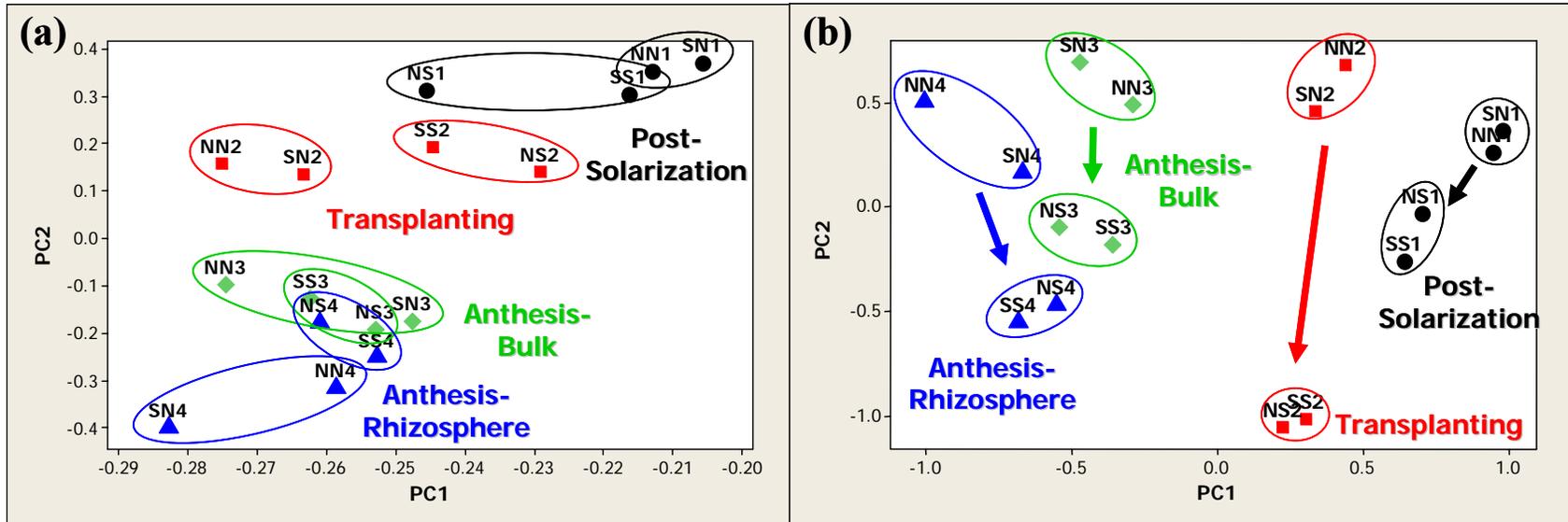


Figure 5.1. Ordinations of fungal community T-RFLP data (Khum, *Sau96I*, binary data) with (a) environment-centered PCA and (b) AMMI analysis. The AMMI analysis removes the main effects variation by applying PCA to the interaction matrix, thus focusing solely on the pattern of interest. Environment-centered PCA does not remove the variation from T-RFs (59.3% of the total variation) and therefore, the primary pattern of interest—interaction signal—is obscured by less interesting main effect variation from the T-RFs.

very small (0.8% environment main effects), but does nothing to remove the main source of variation (59.3% T-RF main effects). Figure 5.1b shows PCA (AMMI) applied to the same data, but after first removing the large T-RF variation and small environment variation. Removing these sources of variation (i.e., doubly-centering) produces the interaction matrix that AMMI ordinales. The environment-centered PCA (Figure 5.1a) is an inferior procedure, as the more subtle trends in treatment difference (indicated by circled data points) are not consistently captured. This analysis could be produced by a researcher simply entering and analyzing a mistakenly transposed T-RFLP data matrix. Note that a variable-centered PCA is the default procedure in some statistical packages (e.g., SAS, Cary, NC), while in other commonly used packages (e.g., Canoco, Microcomputer Power, Ithaca, NY), the user must select this option.

3.2.2. *CA and DCA*

CA was among the poorer methods for demonstrating expected gradients and reaching a consensus with the other methods (Table 5.4). However, when CA was rerun with rare species downweighted with four of the data matrices, the resulting ordinations were acceptable and reached a consensus with other methods, making the method more robust (data not shown). DCA performed comparable to other methods with all datasets, except the Multiregional dataset, where it scored ‘reasonably well’ in all six of the data matrices. These did not affect the overall interpretation of the dataset, enabling DCA to reach a consensus with all 46 data matrices, demonstrating its overall robustness (Table 5.4).

3.2.3. *NMS*

Overall, the ordinations produced from NMS analyses were very similar to the eigenvector-based methods evaluated above. NMS analyses with Sørensen and

Jaccard distance measures scored high in demonstrating gradients ‘very well’, with the Sørensen distance measuring scoring better than any method compared (Table 5.4). However, two NMS analyses with Sørensen distance and one analysis with the Jaccard distance measure produced ordinations that were wildly different from the consensus reached by other ordinations. These analyses were re-run several times each with different initial configurations, but produced similar idiosyncratic results. NMS with the Euclidean distance measure ranked the worst at reaching a consensus and was judged to be the poorest performing method examined (Table 5.4).

For the most complex datasets, NMS analyses with Sørensen and Jaccard distance measures performed better at demonstrating known gradients than all other analyses. For example, NMS outperformed all other analyses with the entire tallgrass prairie datasets analyzed as a whole, but when this dataset was decomposed (based on experimental design) into two separate datasets and reanalyzed, the NMS analyses were no longer more discriminatory than the other methods (data not shown). This phenomenon was also observed in the datasets from the solarization study (Figure 5.2). The combined rice and wheat dataset (Rampur, Bacteria, *Sau96I* enzyme) is shown analyzed with the AMMI model (Figure 5.2a) and with NMS with the Sørensen distance measure (Figure 5.2b). When the rice and wheat seasons were analyzed as a single dataset, the NMS analysis was more discriminatory at demonstrating the treatment differences (soil solarization) in rice than was the AMMI analysis. This is demonstrated by the circled data points and arrows in Figure 5.2b, and the lack of consistency in treatment trends represented in Figure 5.2a. However, when the rice season was decomposed and analyzed separately from the wheat season (Figure 3 from Culman et al. (2006)), the resulting ordinations from all the methods were equally informative, and were generally more informative than the NMS analysis of the combined rice and wheat dataset. Neither AMMI nor NMS revealed treatment

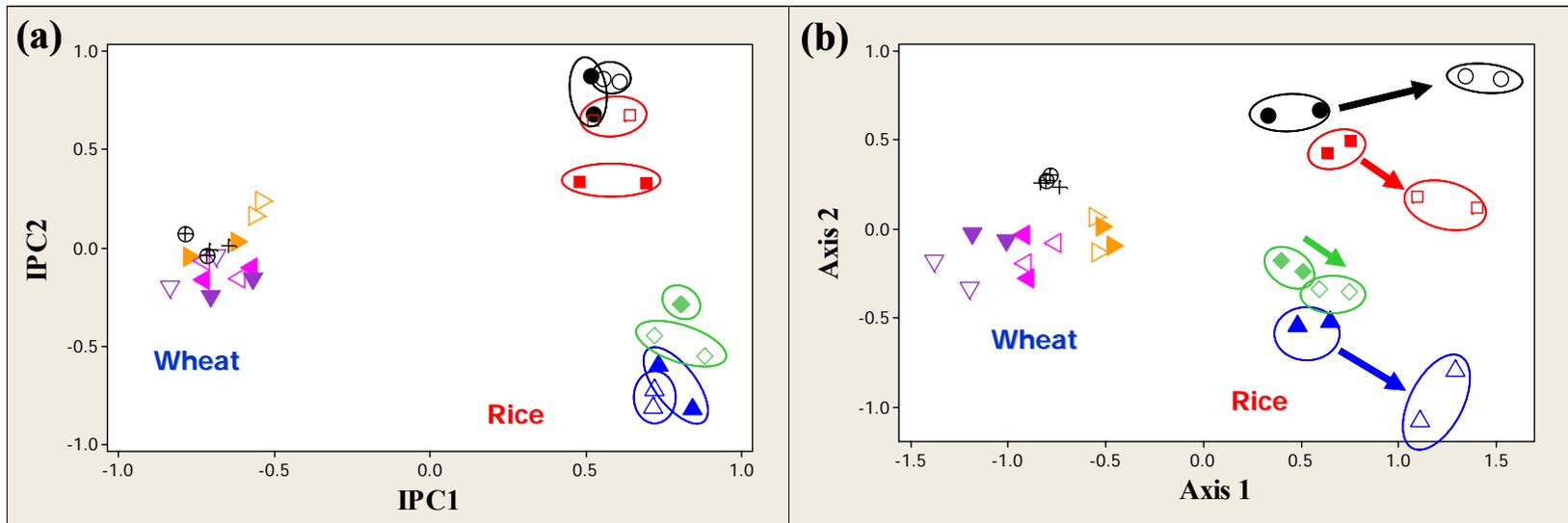


Figure 5.2. T-RFLP bacterial community data (Ramp, *Sau96I*, binary) analyzed with (a) AMMI (b) and NMS with Sørensen distance measure, demonstrating the greater discriminatory power of NMS in the rice crop with heterogeneous data (rice and wheat data analyzed together). Closed data points represent non-solarized plots; open data points represent solarized plots. Circled data points represent the rice crop; non-circled data points represent the wheat crop. Differing shapes and colors represent different sampling time periods. See Culman et al. (2006) for details of experiment and legend.

differences in the wheat sampling periods when either the combined or separated datasets were analyzed.

3.2.4. Most informative type of data—binary, peak height or peak area

Overall, there were few differences in the final ordinations derived from the three different types of data. Our analyses demonstrated that binary data were the most robust, with only one instance when a method didn't reach a consensus in the binary data matrices (18 total). A consensus was not reached five times in the relativized peak height data matrices (18 total), and five times in the relativized peak area data matrices (10 total).

3.3. Theoretical ordination results

Assumptions. PCA uses Euclidean distance and assumes a linear relationship among variables, which may not be appropriate for community ecology data (Beals, 1984, McCune and Grace, 2002). However, this issue is resolved by use of appropriate data transformations which allow PCA to be performed with a wide variety of distance metrics appropriate for community ecology data (Legendre and Gallagher 2001). CA and DCA are both eigenvector-based ordination techniques that use a chi-square distance measure and assume that T-RFs have a unimodal distribution along ecological gradients. This is a more appropriate assumption than linearity for ecological community data, as it is in agreement with the usual outcome of ecological studies (Gauch, 1982), but is still capable of being violated with community ecology datasets (Beals, 1984, Legendre and Legendre, 1998, Minchin, 1987). NMS fundamentally differs from the above techniques because it uses rank order information from a similarity matrix to ordinate the data, rather than metric information (Gauch, 1982). NMS does not assume linear relationships among

variables and it allows virtually any distance measure to be used in the construction of the similarity matrix. Rees et al. (2004) argued that these attributes make NMS more appropriate for T-RFLP community analysis over other methods.

Proportion of Variation Represented. Interpretation of ordination analyses must be guided by an objective assessment of how much variation is being captured in the first few ordination axes. PCA (including AMMI) is the only method that can precisely calculate this variance from the ratio of eigenvalues to total variation. This attribute favors PCA when an objective assessment of variation represented is desired. It should be noted that all methods can use an after-the-fact assessment which calculates a coefficient of determination (r^2) as an assessment of how much variation is captured in the first few axes (McCune and Grace, 2002). However, this approach was not evaluated here since several distance measures were compared, making these measures incommensurable.

The proportion of variation represented in the IPCs in AMMI analysis can be directly related to the predicted interaction signal given in Table 5.3. For instance, the T×E signal comprises only 14.7% of the total variation for the data represented in Figure 5.1b, but this figure focuses on T×E exclusively and AMMI's first two IPCs account for nearly all of the interaction signal. Across all datasets, the average percent of interaction variation represented in the first two IPCs was 94.5%. This directed focus can be advantageous when research interests focus on T×E, and this interaction is a small portion of the overall variation.

Integrated Dual Analysis. Biplots are scatterplots of samples *and* species (Environments and T-RFs) in the same graph. These can be extremely insightful in determining which T-RFs are most closely associated with which environments. Eigenvector techniques (PCA, AMMI, DCA, and CA) use integrated, dual analysis of the rows and columns of a data matrix, resulting in scores for both Environments (E)

and T-RFs. NMS analyses are not dual by nature, because a distance matrix is constructed based on similarities of T-RFs *or* E, but not both.

Uniqueness of Solution. Analyses that have a large number of parameters that need to be specified have potential for confusion. *Optionitis*—a term we use to describe this phenomenon—can become problematic when these parameters are not selected by objective criteria and when the results vary depending on which parameters are selected. PCA has several variants, but the most appropriate can be selected by objective criteria. CA and DCA have a few important input options that the user must select. Depending on the software used, NMS has several criteria that need to be selected (distance measure, stress level, number of iterations, starting configuration, number of axes), which can lead to a number of possible outcomes. Current computing power enables the user to be very conservative with most of these parameters, alleviating most of these concerns. However, in our experience, if these parameters are not set conservatively, the results can be quite variable.

Weight of Rare T-RFs. The distance measure that an analysis uses can impact the relative influence of rare T-RFs. The chi-square distance measure has been criticized by several authors (Faith et al., 1987, Legendre and Legendre, 1998, Minchin, 1987) for its tendency to give greater weight to rare species and less weight to commonly occurring species. This is an important issue as rare T-RFs commonly occur in T-RFLP data and can be methodologically exacerbated with mis-called electrophoretic reads of peak size (T-RF drift) or by PCR artifacts. Since both CA and DCA use the chi-square distance measure, they will potentially give greater weight to rare species—an undesirable characteristic. Although the program DECORANA (Hill, 1979), on which most DCA software is built, and various versions of CA do give the option of ‘down-weighting’ rare species, this does not fully remedy the problem (Jongman et al., 1995). Likewise, rare species are also given greater weight in

standardized (correlation-matrix) PCA (not empirically evaluated here) than in centered PCA. Because standardized variables give each T-RF equal variance, rare and common T-RFs contribute equal information to the ordination. Therefore, using T-RF-centered variables, rather than T-RF-standardized variables, is generally more desirable in T-RFLP analysis. NMS offers the greatest flexibility with regard to selecting a distance measure. Rees et al. (2004) and others (Legendre and Legendre, 1998, McCune and Grace, 2002) have recommended the Sørensen (Bray-Curtis) distance measure (with NMS) as an ecological distance measure for several reasons. Perhaps the most important is the ability of Sørensen to appropriately ignore joint absences (i.e. if two samples do not contain many of the same T-RFs, this is not measured as similarity between them). See Legendre and Legendre (1998) for further discussion of these ecological distance measures.

The potential problem of rare species receiving greater weight can be sidestepped by simply deleting rare species from the dataset. While this would not be appropriate if the researcher was interested in diversity measures, it can often reduce overall noise and improve the correlation structure in datasets. McCune and Grace (2002) suggest deleting species (T-RFs) that only occur in fewer than 5% of the samples. In our experience, deleting rare T-RFs strengthens the observed patterns from ordinations.

4. Discussion

Analysis of variance was performed on all datasets in this study in order to gain insight into the distribution of variation in these T-RFLP datasets. Main effects variation dominated the majority of the datasets and was comprised almost entirely of T-RF variation. This variation is inherently simple information on the rareness or commonness of T-RFs and is often not of primary interest. The main effects variation

from E was always small—either the total number of T-RFs across all treatments remained fairly constant (in binary data), or the variation was eliminated by the relativization step (in peak height and area) to remove a known source of analytical noise. The small contribution from E was due in part to the quality control step that eliminated poor-quality runs from the analysis, but also reflected the nature of the datasets examined. Despite often large differences between treatments, the number of T-RFs in each environment remained fairly constant (evenness values in Table 5.2; binary environment variation in Table 5.3). The nature of T-RFs and E main effects variation directs the researcher to focus entirely on the interaction variation, as nearly of all the relevant information concerning the environments, imposed treatments, samples, etc. is found in the interaction. This interaction captures how T-RFs differentially respond to the E. In this context, ANOVA can be used as a standard procedure that can be applied to any dataset in order to objectively measure microbial community dissimilarity. This in turn can provide insight into whether these differences are ecologically meaningful.

Table 5.3 demonstrates a large proportion of dataset variation from T-RFs. This variation needs to be removed by the selected analysis, or the resulting ordination will be dominated by this simple and often uninteresting information. This phenomenon is demonstrated dramatically in Figure 5.1. We were intrigued to discover that the necessity to remove the T-RF variation only seems to be germane with PCA, as all the other methods examined in this study (CA, DCA and NMS) have means to ignore or minimize this variation.

In this study, all methods produced empirical results that generally did not deviate from the overall consensus. Likewise, the theoretical criteria did not solely favor one particular analysis over the others. However all methods did not perform equally. Some methods have distinct advantages which aid in data interpretation and

tend to be more robust, while others have qualities that could be potentially problematic.

The AMMI analysis and PCA yielded nearly identical empirical results and scored favorably with all but one theoretical criterion—assumptions about the data. The AMMI analysis also measures how the variation within a dataset is distributed, including the amount of interaction signal that the IPCs capture. This can provide valuable insight into the relative strength of the observed pattern. Empirically, DCA performed well and proved to be among the more robust analyses. CA was not as robust as DCA, but performed well when rare species were downweighted. DCA and CA both scored favorably in only two out of the five theoretical criteria. NMS analyses with Sørensen and Jaccard distance measures both performed well at demonstrating expected gradients. NMS analysis with Euclidean distance ranked as being the least discriminatory. Theoretically, NMS analyses scored poorly with all but one criterion—the lack of assumptions made on the data. This attribute could potentially be very important, but we have found it is usually not germane. In our experience, soil-based T-RFLP data are generally not very complex relative to other types of ecological community data. Legendre and Gallagher (2001) refer to these types of datasets as having short gradients. Measures of dataset complexity can account for various factors, including sample heterogeneity, the number of underlying gradients, T×E interactions, and relative noise to signal. We found sample heterogeneity (also known as beta diversity) to be a satisfactory and simple measure of this complexity.

We assessed T-RFLP environment heterogeneity by measuring beta diversity, defined as $[(\text{total number of T-RFs in a dataset}) / (\text{average T-RF richness in the environments})] - 1$ (Whittaker, 1972). McCune and Grace (2002) state as a rule of thumb that beta diversity less than 1 is rather low and greater than 5 is very high. We

found the average beta diversity across all T-RFLP datasets to be 1.22, suggesting that overall heterogeneity between environments in T-RFLP datasets is relatively low. (Note that beta diversity is often highly correlated with the percent of empty cells in a data matrix.) Another observation to support the idea that soil-based T-RFLP data have relatively low complexity was the empirical findings. Overall, ordinations of the same dataset with different methods did not drastically deviate from each other, suggesting the relative ease of the ordinations by the various methods. This same phenomenon—when sample heterogeneity is low, ordinations from different methods perform comparably—has been shown in numerous ecological studies (Beals, 1984, Fasham, 1977, Gauch et al., 1977, Kenkel and Orloci, 1986, Minchin, 1987).

The lack of complexity or heterogeneity in T-RFLP datasets is likely related to the type of questions that T-RFLP is employed to ask. Microbial ecologists rarely use T-RFLP to assess community structure from two very different environments (e.g., very wet to very dry gradients, very different soil types, etc.) as most researchers will (often correctly) assume that these microbial communities will be very different. As a consequence, the length of the ecological gradient is often kept relatively short by the researcher's experimental design when employing T-RFLP.

An irony currently exists in the literature in which researchers meticulously report the concentration and reaction details of their PCR (even though T-RFLP has been shown to be robust to these variations (Osborn et al., 2000)), but often fail to report rudimentary information about the statistical analyses. A survey of recent T-RFLP literature will show many papers that fail to report any parameters about the analyses employed, even parameters as fundamental as which variant of PCA or distance measure of NMS. Omissions such as these should be avoided, as they compromise the repeatability of experiments and analyses. McCune and Grace (2002) have offered criteria to report for each analysis. Of these criteria, we judge the most

important for T-RFLP analysis to be: i) PCA and AMMI: form of cross-products matrix (correlation, variance-covariance, or doubly-centered), number of axes interpreted and the proportion of variance represented by each axis, and justification of the assumption of linear relationships among T-RFs; ii) CA and DCA: number of axes interpreted and the proportion of variance represented by each axis (after-the-fact assessment), whether downweighting was selected, and justification of the assumption of unimodal distribution of T-RFs; iii) NMS: distance measure, stress of final solution, number of dimensions in final solution, proportion of variance represented by each axis, number of runs with real data, and number of iterations for final solution.

Conclusions and Recommendations

Based on our findings there is no single method that we can recommend above all others for soil-based T-RFLP community analysis. We recommend that it become conventional for all T-RFLP data to be reported with two important dataset characteristics: beta diversity and percent of main and interaction effects. We found these two characteristics to be highly informative indicators of the nature of the datasets examined in this study. Datasets with high beta diversity (2 or greater in our experience), or otherwise greater complexity than those examined in this study, should likely employ NMS analyses with Sorensen or Jaccard (or another appropriate distance measure). Analyses of low beta diversity datasets should be more strongly guided by the outlined theoretical criteria. Both interaction effects and beta diversity are easily calculated with the free online software, T-RF Manager (<http://cbsusrv02.tc.cornell.edu/TRF/index.aspx>).

We prefer the AMMI analysis for most T-RFLP data that we have encountered, as it has performed very well empirically, and also yields an ANOVA table which provides a tool to assess microbial community dissimilarity. The ability of the analysis to focus on the differential responses to the treatments may also be amenable for use on microarray data or other approaches focusing on differential gene responses. Unless otherwise justified, DCA and CA analyses should be run with the ‘downweight rare species’ option selected. We generally do not recommend NMS with the Euclidean distance measure; it performed the worst empirically, and has no advantages over the other methods that we judge important for T-RFLP. Finally, we recommend confirming initial findings with other multivariate method/s whenever possible.

In this study, binary data were less prone to variable results than relativized peak height or area, making binary data the most robust measure. Relativized peak height had, on average, greater interaction signal to interaction noise ratios, lower |skew|, and lower NMS stress than relativized peak area. For these reasons and those outlined in Grant et al. (2003), we recommend using binary data or relativized peak height for ordination analyses over relativized peak area.

REFERENCES

- Abdo, Z., Schuette, U.M.E., Bent, S.J., Williams, C.J., Forney, L.J., Joyce, P., 2006. Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Environmental Microbiology*. 8, 929-938.
- Beals, E.W., 1984. Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. *Advances in Ecological Research*. 14, 1-55.
- Blackwood, C.B., Marsh, T., Kim, S.-H., Paul, E.A., 2003. Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Applied and Environmental Microbiology*. 69, 926-932.
- Blackwood, C. B., Paul, E.A., 2003. Eubacterial community structure and population size within the soil light fraction, rhizosphere, and heavy fraction of several agricultural systems. *Soil Biology & Biochemistry* 35:1245-1255.
- Clement, B.G., Kehl, L.E., Debord, K.L., Kitts, C.L., 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.
- Culman, S.W., Duxbury, J.M., Lauren, J.G., Thies, J.E., 2006. Microbial community response to soil solarization in Nepal's rice-wheat cropping system. *Soil Biology & Biochemistry*. 38, 3359-3371.
- Dunbar, J., Ticknor, L.O., Kuske, C.R., 2000. Phylogenetic specificity and reproducibility and a new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Applied And Environmental Microbiology*. 67, 190-197.
- Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Plant Ecology*. 69, 57-68.
- Fasham, M.J.R., 1977. A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines and coenoplanes. *Ecology*. 58, 551-561.
- Gauch, H.G., 1982. *Multivariate Analysis in Community Ecology*, Cambridge University Press, Cambridge.
- Gauch, H.G., 1992. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*, Elsevier Science Publishers, Amsterdam, The Netherlands.

- Gauch, H.G., 2007. MATMODEL Version 3.0: Open source software for AMMI and related analyses. Crop and Soil Sciences, Cornell University, Ithaca, NY.
<http://www.css.cornell.edu/staff/gauch/matmodel.html>
- Gauch, H.G., Furnas, R.E., 1991. Statistical analysis of yield trials with Matmodel. *Agronomy Journal*. 83, 916-920.
- Gauch, H.G., Whittaker, R.H., Wentworth, T.R., 1977. A comparative study of reciprocal averaging and other ordination techniques. *Ecology*. 65, 157-174.
- Grant, A., Ogilvie, L.A., Blackwood, C.B., Marsh, T.L., Sang-Hoon, K., Paul, E.A., 2003. Terminal Restriction Fragment Length Polymorphism Data Analysis. *Applied and Environmental Microbiology*. 69, 6342-6343.
- Hill, M.O., 1979. DECORANA-- a FORTRAN program for detrended correspondence analysis and reciprocal averaging, Ecology and Systematics, Cornell University, Ithaca, NY.
- Jongman, R.H.G., ter Brack, C.J.F., van Tongeren, O.F.R., 1995. Data analysis in community and landscape ecology, Cambridge University Press, Cambridge.
- Kenkel, N.C., Orloci, L., 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: Some new results. *Ecology*. 67, 919-928.
- Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia*. 129, 271–280.
- Legendre, P., Legendre, L., 1998. Numerical Ecology, Elsevier Science Publishers, Amsterdam, The Netherlands.
- Liu, W., Marsh, T.L., Cheng, H., Forney, L.J., 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology*. 63, 4516-4522.
- Lueders, T., Friedrich, M., 2000. Archaeal population dynamics during sequential reduction processes in rice field soil. *Applied and Environmental Microbiology*. 66, 2732-2742.
- Marsh, T.L., 2005. Culture-independent microbial community analysis with terminal restriction fragment length polymorphism. *Methods in Enzymology*. 397, 308–329.
- McCune, B., Grace, J.B., 2002. Analysis of Ecological Communities, MjM Software Design, Gleneden Beach, OR.
- McCune, B., Mefford, M.J., 1999. PC-ORD. Multivariate Analysis of Ecological Data. Version 4, Gleneden Beach, OR.

- Minchin, P.R., 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Plant Ecology*. 69, 89-107.
- Osborn, A.M., Moore, E.R.B., Timmis, K.N., 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.
- Rees, G.N., Baldwin, D.S., Watson, G.O., Perryman, S., Nielsen, D.L., 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. *Antonie van Leeuwenhoek*. 86, 339-347.
- Thies, J.E., 2007. Soil microbial community analysis using terminal restriction fragment length polymorphisms. *Soil Science Society of America Journal*. 71, 579-591.
- Tiedje, J.M., Asuming-Brempong, S., Nusslein, K., Marsh, T.L., Flynn, S.J., 1999. Opening the black box of soil microbial diversity. *Applied Soil Ecology*. 13, 109-122.
- Whittaker, R.H., 1972. Evolution and measurement of species diversity. *Taxon*. 21, 213-251.

CHAPTER 6
T-REX: SOFTWARE FOR THE LABELING, PROCESSING,
AND ANALYSIS OF T-RFLP DATA⁵

Background

Despite increasing popularity and improvements in terminal restriction fragment length polymorphism (T-RFLP) and other molecular-based microbial community fingerprinting techniques, there are still some formidable barriers that plague the analysis of these datasets. Many steps are required to process raw data into a format ready for analysis and interpretation. These steps can be time-intensive, error-prone, and can introduce unwanted variability into the analysis. Currently, some of greatest obstacles of T-RFLP analysis are: i) determining true peaks from noise, ii) aligning peaks between samples iii) creating two-way data matrix from tabulated raw data, iv) rapid manipulation of data matrices, and v) determining dataset complexity (sample heterogeneity and interaction effects). Each one of these issues will be discussed more thoroughly below.

Determining "true peaks" (i.e., distinguishing peaks from background fluctuations in fluorescence) is often a major challenge in T-RFLP data analysis, as the baseline threshold can dramatically affect the community fingerprint and downstream analyses. A common procedure is to apply a researcher-determined baseline threshold across all samples to delineate true peaks from noise [1-3]. However, this threshold is often subjectively determined and may not be the most appropriate approach. Since the number of spurious peaks in a sample may increase when the amount of PCR product

⁵ This chapter will be submitted to BMC Bioinformatics

analyzed increases, the amount of noise relative to signal in a sample may be a methodological artifact. When DNA concentrations vary from sample to sample, determining true peaks from noise based on variability in each sample [3], rather than a value across all samples is conceptually more appropriate.

The size (in base pairs) of every T-RF is determined by referencing the T-RF with an internal size standard. However, T-RFs can be improperly sized due to differences in fragment migration, purine content, and fluorophores [4, 5]. These analytical errors in determining fragment length (T-RF drift) are often corrected for by aligning peaks manually [2], aligning them automatically [1, 6], or simply ignored and treated as analytical error. However, to date, there have been no reports on the effect of these three approaches. Since most peak alignment software isn't integrated with downstream multivariate analyses, it is often difficult to determine the effects of this alignment on the overall interpretation of the data.

Because of the complexity associated with T-RFLP and other microbial community datasets, multivariate statistical analyses are typically performed on these data to summarize the complex relationships of the microbial communities with their environments. Raw T-RFLP data exported from Genemapper™, Peak Scanner™, or similar size-calling software is typically in a tabulated or listed format, where one column contains all the records for each variable (i.e., one column for all T-RF sizes, one column for all peak heights, etc.). However, these data need to be formatted into a two-way data matrix, which is required by most multivariate-focused statistical software packages (e.g. *Canoco*, *Primer*, *PC-ORD*). The formatting of tabulated raw data into a data matrix is generally performed in a spreadsheet software application (e.g., pivot table in *MS Excel*). In order to do this, the electrophoretic runs need to be

labeled with information about the sample, which usually pertains to the experimental design (sampling period, treatment, replicate number, etc.). Assuming the researcher employed randomization on the lab bench to minimize bias throughout the entire analysis, this labeling procedure and data matrix construction can be time-intensive and error-prone.

A thorough analysis of large T-RFLP datasets requires various data matrix manipulations, such as examining the three types of data (presence/absence, peak height, peak area), relativization of peak height or peak area, averaging replicated samples, examining specific experimental factors, deleting rare T-RFs, etc. Most spreadsheet software applications aren't always amendable to these more sophisticated or time-intensive manipulations, making an exhaustive analysis of these data difficult. In our experience, an exhaustive analysis can be quite fruitful, as manipulating datasets based on experimental design can reveal patterns in ordinations that were previously obscured by stronger patterns or by lower signal: noise ratios [7, 8].

Finally, there is a lack of consensus in the literature today about which statistical analyses are more appropriate to analyze T-RFLP data. In a comparative study of multivariate methods, Culman et al. [7] attempted to address this lack of consensus and found that most common multivariate methods yielded similar empirical results when the T-RFLP dataset exhibited low complexity, as measured by beta diversity and percent interaction effects. They reported that this complexity could be prescriptive, as datasets with greater complexity should likely employ nonmetric multidimensional scaling (NMS) analyses, and datasets with low complexity should be guided by theoretical criteria. Although beta diversity can easily be calculated in a spreadsheet application, the calculation of interaction effects is computationally-intensive. Culman

et al. [7] also demonstrated the utility of the Additive Main Effects and Multiplicative Interaction (AMMI) model as a robust and advantageous method for T-RFLP analysis. This model is found in only a few multivariate software packages offered today.

Currently, there are few options that researchers have to choose from when analyzing microbial community data. Most software that has been developed is aimed at referencing community fingerprints or profiles with a sequence database in order to predict specific taxa present in the profile (e.g. TAP-TRFLP [9, 10], MiCA [11], T-RFLP FRAGSORT [12], PAT [13]. There are, however, a few available packages that do aid with multivariate data analysis. T-Align [6] implements an algorithm to align peaks, hence removing the potential of subjective bias during peak alignment.

However, this peak alignment is limited to single or duplicate samples, and fails to simultaneously align peaks in samples containing more than two replicates. Another package, T-RFLP Stats [3] allows users to align peaks (as does T-Align), group samples based on various classification procedures and then references these profiles to a clone library. This software shows promise, but is written in three separate languages (R, Perl and SAS) requiring three separate platforms. These platforms are primarily command driven and are all potentially cumbersome to inexperienced users. SAS also requires a purchased license for use. In addition, T-RFLP Stats offers no labeling procedure to designate and format raw data, nor does it perform any ordination analyses, argued by some to be superior to classification procedures for the analysis of microbial community data [14]. More recently, a few commercial software packages have become available that offer a range of features regarding electropherogram manipulation, with some limited multivariate procedures, most notably GelQuest (SequentiX, Germany), Genemarker (SoftGenetics, USA), and Torast (Dresden, Germany). However, the high costs of these programs (for a single

license, GelQuest costs €1,299, with an additional €669 for ClusterVis, a program for clustering procedures; Genemarker costs \$3,500; Torast costs €249) make them inaccessible to some research labs. In addition, features and functions vary widely between these programs, as T-RFLP community analysis was not the primary motive for the development of these software packages.

Here, we developed *T-REX* (**T-R**F_{LP} analysis **EX**pedited), a free, web-based tool that was developed to address current obstacles of T-RFLP analysis. *T-REX* allows users to i) label raw data with attributes related the experimental design of the samples, ii) determine a baseline threshold for identification of true peaks over noise, iii) align T-RFs in all samples (bin T-RFs), iv) construct a two-way data matrix from labeled data and manipulate the matrix in a variety of ways, v) produce several measures of data matrix complexity, including the distribution of variance between main and interaction effects and sample heterogeneity, and vi) analyze a data matrix with the additive main effects and multiplicative interaction model (AMMI). *T-REX* offers users a consolidated and rapid analysis of T-RFLP data with great flexibility of functions performed on the data.

Implementation

T-REX can be found at the web address: <http://trex.biohpc.org/>. The program is free, platform independent, and requires only a web browser and an internet access to use. The **Home** page outlines the program's features and introduces the user to the typical flow of analysis (Figure 6.1). Tabs on the left of the page direct the user to different functions of the program.

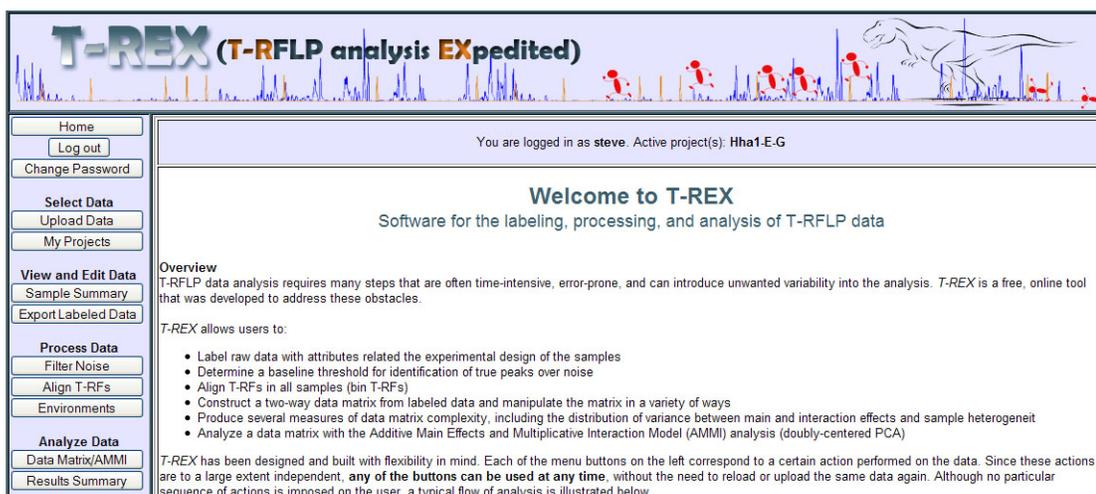


Figure 6.1. Screenshot of T-REX Homepage.

User Profile and Project Management (*Log In*). Users can work as a guest or may become a registered user. Registered users can store multiple projects (up to 25 unique projects) on the server at any time. Saved projects can later be concatenated, renamed, or deleted. Guests are allowed everything except storing their data on the server. New users can self-register and registered users log into their account via the **Log in** page.

Uploading Data and Labeling Procedure (*Upload Data and My Projects*). The first step in using T-REX is to upload and label the data, which happens simultaneously and requires two files: i) the raw data file and ii) the label file. The raw data file is the tabulated file that is exported in Genemapper™ or similar size-calling software that contains the peak information for a set of samples. The label file contains a set of labels/attributes that describe each sample and often correspond to factors in the experimental design. Both files should be a simple text file in tab-delimited format.

Please see the *T-REX*'s program documentation for specific guidelines on raw data and label file formats.

A new project is created when a user uploads and labels data. Registered users should specify a unique name for this project as it will be stored on the server, and can later be modified. Alternatively, a registered user can upload new raw data to a specified existing project. The name of the active project and the registered user is displayed in the blue box under the *T-REX* header icon.

T-REX performs several functions that take advantage of information provided by replicated data (i.e., samples that are conceptually identical, or belong to the same environment/treatment). However, data need not be replicated for the majority of the program's functions. If replication is part of the experimental design, users can define what samples are replicates during data upload, or manually in the **Sample Summary** page. Replicated samples are organized into groups called environments. Each environment has a unique identifier—a positive integer. At any given time, each sample belongs to only one environment and the corresponding environment identifier is displayed in the Env column on **Sample Summary** page. *T-REX*'s program documentation outlines specific guidelines on defining replicated data into environments.

Missing data occurs when one or more samples are omitted from the analysis. Missing data can result from multiple scenarios. First, a researcher could throw a sample out after the visual inspection of the electropherogram showed that the run was of too poor quality to be meaningful. Genemapper™ software can also omit poor-quality samples from the exported Genemapper™ file. (In this scenario, the sample will not be present

in the raw data file.) Missing data can also arise from filtering procedures or other parameters applied to the dataset in *T-REX* that eliminate samples from down-stream analyses.

T-REX is able to appropriately deal with all possible cases of missing data. In scenarios when there is a discrepancy regarding samples between the raw data file and label file the program will enter and record these non-matched samples, but mark them as missing data. More specifically, if the label file contains a file name that is not represented in the raw data file, the corresponding sample will be entered into the system and marked as missing data. Likewise, if the raw data file contains file names not in the label file, the samples corresponding to these missing names will be marked as missing data and zeroes will be added as labels. In an extreme case when a label file is not supplied, all samples in the raw data file will be marked as missing data. Users have the option of manually supplying the labels and removing the missing data mark from the **Sample Summary** page. Missing data are stored in the active project, but do not affect any procedures or manipulations of data. They are marked in red and can be viewed in the **Sample Summary** page.

Once a project is created, it can be renamed, merged, or deleted in the **My Projects** page. Users can also come back to pre-existing projects and load them in this page for further manipulation.

Viewing and Editing Individual Samples (*Sample Summary*). The **Sample Summary** page is synonymous to the home page of a particular project (Figure 6.2). All samples are consolidated to show the total number of peaks, total peak height and peak area, as well as the properties relating to the experimental factors assigned in the

labeling procedure. If data are replicated, users can view which environment they are grouped with in the **Env** column. The **Sample Summary** page also shows users how data processing procedures (such as noise filtering or T-RF aligning) have removed peaks originally found in the raw data file.

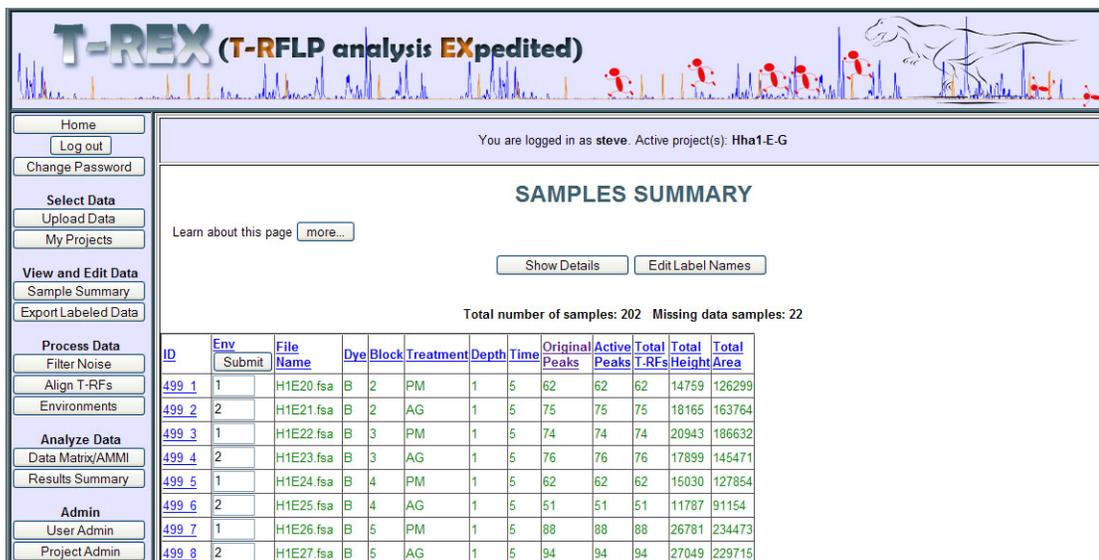


Figure 6.2. Screenshot of T-REX Samples Summary page.

Individual samples can be viewed, edited, and even removed from the analysis in the **Sample Details** page, accessible by selecting the sample **ID** in the **Sample Summary** page. Once viewing an individual sample, the user will see individual peak properties and will be able to manipulate labels, remove individual peaks of that sample, or mark the entire sample as missing data within the project.

Export Labeled Data to Use Elsewhere (Export Labeled Data). The **Export Labeled Data** page was designed for users who want to take advantage of T-REX's rapid labeling procedure, but analyze their data with another software program. Users

can also manipulate data via the **Filter Noise** and/or **Align T-RFs** pages before exporting the labeled data. Labeled data is exported as a simple text file with columns separated by a specified separator (tab-delimited by default).

Procedure to Filter out Noise from True Peaks (*Filter Noise*). *T-REX* uses the approach outlined by Abdo et al. [3] to find true peaks and eliminate background noise. True peaks are identified as those whose height (or area) exceeds the standard deviation (assuming zero mean) computed over all peaks and multiplied by the factor specified in the box provided. The procedure is then reiterated with the peaks which were not identified as true ones. The iterations continue until no new true peaks are found.

The filtering of peaks can be based on standard deviations of peak height or area and may be applied to all samples or just selected samples in the active project. Users should select an appropriate standard deviation multiplier based on the original electropherograms and results of the filtering procedure. The program allows for rapid manipulation of the multiplier and subsequent reviewing of results in the **Samples Summary** page if a user wants to determine an appropriate multiplier empirically. At any time the filtering procedure can be cleared and the data reverted to their original state with the ‘Clear filtering’ button.

Automated Alignment of Peaks (*Align T-RFs*). An automated alignment of peaks across all samples is possible in the **Align T-RFs** page. This function models the approach taken by the software program *T-Align* [6]. Briefly the smallest peak across all samples is identified and tagged. Peaks within the range specified by the clustering threshold are then identified and grouped into a T-RF. The next smallest peak across

all samples not falling into the first T-RF is identified and tagged. Peaks within the specified clustering threshold are identified and grouped with the second T-RF. This process continues until all peaks are grouped into a T-RF. Alternatively, peaks can simply be rounded to the nearest nucleotide (integer) size with no real clustering across samples by using the ‘Round to the nearest integer’ function.

Grouping Samples into Environments (*Environments*). The **Environments** page allows users to rapidly classify samples into environments based on the given labels. This approach is especially useful when replication in an experiment occurred at multiple scales (e.g., analytical, field) and a user wants to compare results based these different ways of defining replication. Users can assign and/or reassign replicated samples into environments by using the provided checkboxes to define the set of labels that determine an environment. Samples will be considered replicates (i.e., belonging to the same environment) if they have identical sets of checked label values.

Data Matrix Construction and AMMI analysis (*Data Matrix/ AMMI*). The **Data Matrix/ AMMI** page allows users to first construct a two-way data matrix and second run the AMMI model on this data matrix. Data matrix construction involves six steps. The first two steps require that all peaks be assigned to a particular T-RF via the **Align T-RFs** function and that each sample be associated with an environment. Note that a data matrix can be constructed when samples are only rounded to the nearest base pair and when data are not replicated (i.e., each sample is designated into a unique environment). The third step allows users to specify which type of data (presence/absence, peak height, or peak area) to use for data matrix construction, and if these data should be averaged across replicates and/or relativized. The fourth step

allows users to select which experimental factors should be included in the data matrix. Users have the option of selecting all, or just specific fluors and/or specific factors to be included in the data matrix and subsequent analysis. The fifth step allows users to omit rare T-RFs or samples with poor peak representation. T-RFs or entire samples can be omitted based on number or percentage of occurrences across samples, or based on total T-RFs in samples or the cumulative peak height or area within a sample. This step represents a final quality control process to be placed a final data matrix. Selecting ‘Create Data Matrix’ in the sixth step will take the user to another page where a data matrix in tab-delimited format is ready, as well as output on basic data matrix properties, such as total samples and T-RFs present, maximum and minimum, and average number (richness) of T-RFs across samples, and sample heterogeneity (Figure 6.3).

You are logged in as **steve**. Active project(s): **Hha1-E-G**

DATA ANALYSIS: run AMMI

Data Matrix created successfully !

Basic Data Matrix Properties [\(explain\)](#)

Samples present	164 out of 202
T-RFs present	335 out of 335
Number of column blocks (environments)	36
Numbers of replicates in consecutive environments	5,5,4,4,4,5,5,5,5,5,4,3,5,3,5,5,5,5,5,3,4,5,5,5,5,5,4,5,4,4,5,5,3,5
Missing data represented by	-1
Minimum number of T-RFs	20 in sample 499_89 (H10F20.fsa)
Maximum number of T-RFs	112 in sample 499_16 (H1F27.fsa)
Average T-RF richness	48.21
Sample Heterogeneity	5.95
Percent of empty cells in data matrix	0.14

The tab-delimited data matrix is ready for download [here](#)

If you wish, you can now click on the button below to run AMMI analysis of the data matrix just created using MATMODEL 3.0.

[View AMMI analysis](#) [Back to Data Matrix](#)

Figure 6.3. Screenshot of T-REX Create Data Matrix/ Run AMMI page.

At this point the user is able to export this data matrix for analysis with another software package, or continue with the AMMI analysis by clicking ‘View AMMI Analysis’. Choosing the latter will take the user to another page where four output tables summarize the ANOVA results and there are a number of files containing output available to download (Figures 6.4 and 6.5). The first table reports the degrees of freedom, sum of squares (SS) and mean squares for the sources of variance; the second table reports the estimations of interaction SS for pattern and noise, if the data are replicated. The third table reports the percentages of variation from each source of the main effects and interaction effects and the fourth table reports the percentage of interaction signal variation that is captured by the first two interaction principal components axes (IPCA) (Figure 6.5). These tables are written to a text file called “AMMI Summary” and the graphing scores are written to the “AMMI Graphing Data” file.

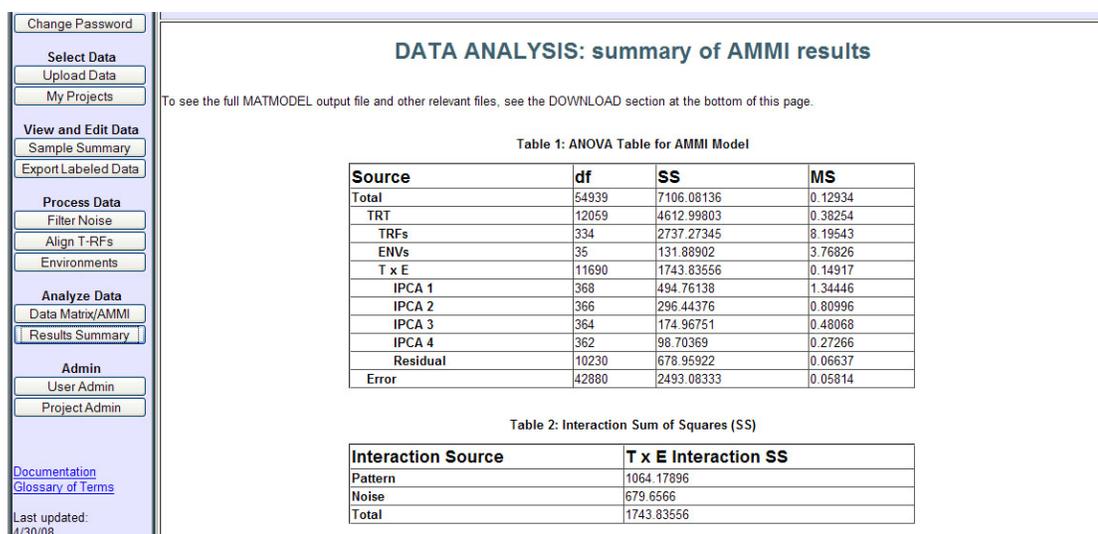


Figure 6.4. Screenshot of tables 3 and 4 produced with output from the AMMI analysis in *T-REX*

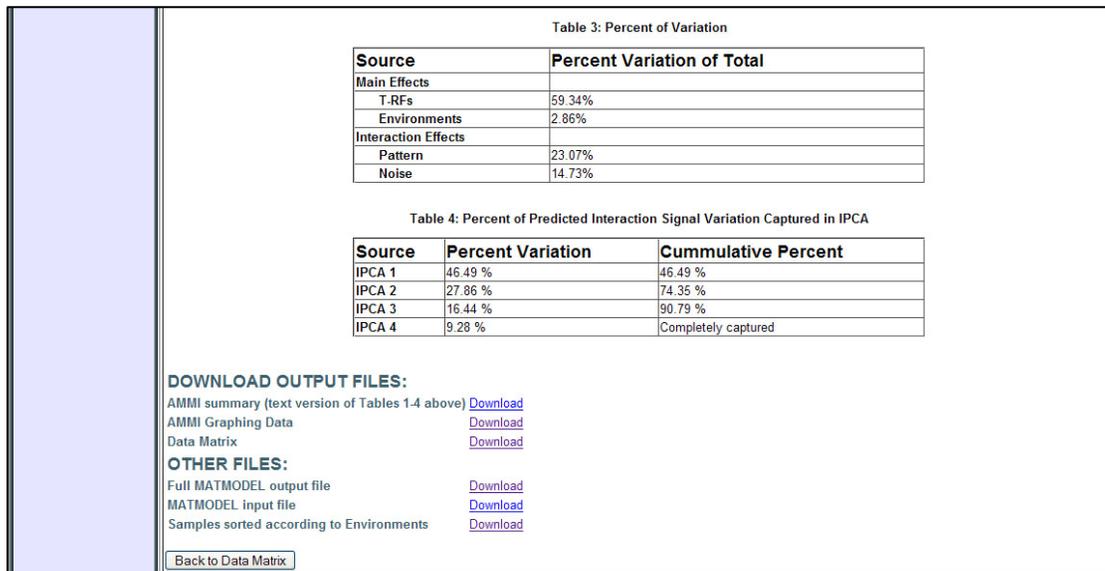


Figure 6.5. Screenshot of the tables 3 and 4 produced with output from the AMMI analysis in *T-REX*

Summary of Results and Output (*Results Summary*). The **Results Summary** page reports the results of relevant basic data matrix properties and summarizes the results of the AMMI analysis in one place. The ‘T-RF Abundance table’ reports the number of samples (samples present) and percentage of samples (% of samples present) that each T-RF occurs. All generated output files are also available for download at this page.

Example Dataset. We used *T-REX* to analyze 16S T-RFLP data from Chapter 2, generated from soils under native tallgrass prairies and annual wheat fields from five different sites across north central Kansas. Soil was sampled at 6 different depth intervals on three separate dates. This experiment created a complex factorial design

(3 sampling dates × 5 field sites × 6 sampling depths × 2 management histories), totaling 180 samples.

Results and Discussion

Analysis of the all samples in the dataset revealed that sampling date and depth were the two primary drivers of bacterial community structure, while differences in management history were not observed (Figure 2.4, in Chapter 2). Sample heterogeneity was extremely high for previously encountered soil T-RFLP datasets (Table 5.2 in Chapter 5) with a beta diversity of 4.82. However, differences in management history were of primary interest in this study, and attempts to minimize or eliminate variance from sampling date and depth were taken in order to reveal possible, more subtle trends of management history. Sampling date variance was easily removed by selecting only the last sampling date (June 2007) during the data matrix construction in *T-REX*. Removing the first two sampling dates resulted in a dataset with a smaller beta diversity (3.46), but still large enough to indicate that NMS analyses would produce more discriminatory results compared to parametric ordinations (AMMI, principal components analysis, etc.).

Eliminating sampling date variance revealed more subtle differences in management history (Figure 2.7 in Chapter 2). Since the relationship between management history and depth was of interest in this study, the June 2007 dataset was decomposed down in *T-REX* by looking at the surface three depths together (0 – 10 cm, 10 – 20 cm, 20 – 40 cm), and then, the bottom three depths (40 – 60 cm, 60 – 80 cm, 80 – 100 cm). Looking at a subset of this dataset revealed that management history and depth drove

microbial community composition in the surface depths, while differences due to site largely shaped the microbial communities at the lower depths (Figure 2.8 from Chapter 2). Analyzing the June 2007 data matrix in two subsets lowered the beta diversity measures in the surface depths to 2.48 and in the lower depths to 2.88.

A final approach was taken to minimize both sources of variation due to differences in site and sampling date—averaging the samples across both site and sampling date. This approach proved to be fruitful, as ordinations revealed that the bacterial communities were different between prairie and annual wheat sites down to 60cm depth (Figure 2.11 from Chapter 2), aligning with similar findings in SOC and soil N pools (Chapter 2).

The T-RFLP dataset in this study was complex, as they were many strong drivers of community structure present. However, some signals were so strong (sampling date, depth) that they obscured more subtle signals of interest (management history). Hence, exploratory data analysis and data matrix manipulation was required to elucidate which factors exerted the greatest influence on bacterial community structure and how those factors changed through depth. These data matrix manipulations would have been time intensive to perform manually and prone to error for even simple manipulations, such as averaging of samples. However, with *T-REX*, a rapid, robust, and thorough analysis of this dataset was made possible.

In addition to rapid data matrix manipulation, *T-REX* also produced a more robust dataset, as prior to data matrix construction, the data were subjected to noise filtering and T-RF alignment. The calculations of beta diversity and interaction effects that *T-*

REX output were also used as prescriptive indicators that the data were complex and non-parametric analyses, such as NMS should be employed [7].

Conclusions

T-REX facilitates a streamlined, more reliable and less biased analysis of microbial community data with a suite of flexible functions that allows researchers to choose the most appropriate data manipulations based on research objectives. *T-REX* also enables researchers to implement the AMMI analysis, a method which holds many advantages for microbial community data analysis. In addition, this software provides a tool to the research community to rapidly and robustly test the effects of various data processing methods on the overall results of datasets. Many of these processing methods are known sources of analytical variability, but there is no consensus in the literature of how to most appropriately minimize this variability. *T-REX* will allow microbial community analyses to continue to develop as an important tool in understanding microbial community dynamics and their effects on ecosystem processes.

Availability and Requirements

- **Project name:** T-REX
- **Project home page:** <http://trex.biohpc.org>
- **Operating system(s):** Platform independent
- **Programming language:** Sequel
- **License:** GNU GPL
- **Any restrictions to use by non-academics:** none

REFERENCES

1. Dunbar J, Ticknor LO, Kuske CR: Phylogenetic Specificity and Reproducibility and New Method for Analysis of Terminal Restriction Fragment Profiles of 16S rRNA Genes from Bacterial Communities. *Applied and Environmental Microbiology* 2001, 67:190-197.
2. Blackwood CB, Marsh T, Kim S-H, Paul EA: Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Applied and Environmental Microbiology* 2003, 69(2):926-932.
3. Abdo Z, Schuette UME, Bent SJ, Williams CJ, Forney LJ, Joyce P: Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Environmental Microbiology* 2006, 8:929-938.
4. Kaplan CW, Kitts CL: Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content. *Journal of Microbiological Methods* 2003, 54:121-125.
5. Marsh TL: Culture-independent microbial community analysis with terminal restriction fragment length polymorphism. *Methods in Enzymology* 2005, 397:308-329.
6. Smith CJ, Danilowicz BS, Clear AK, Costello FJ, Wilson B, Meijer WG: T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiology Ecology* 2005, 54:375-380.
7. Culman SW, Gauch HG, Blackwood CB, Thies JE: Analysis of T-RFLP data using Analysis of Variance and Ordination Methods: A Comparative Study. *Journal of Microbiological Methods* 2008, In press.
8. Culman SW, Duxbury JM, Lauren JG, Thies JE: Microbial community response to soil solarization in Nepal's rice-wheat cropping system. *Soil Biology and Biochemistry* 2006, 38:3359-3371.
9. Marsh TL, Saxman P, Cole JR, Tiedje JM: Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Applied and Environmental Microbiology* 2000, 66(8):3616-3620.
10. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM *et al*: The Ribosomal Database

Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research* 2003, 31(1):442-443.

11. Liao K, Bent SJ, Shyu C, Sale KL, Soule T, Forney LJ, Foster JA: Microbial Community Analysis (MiCA): Web-based computational tools for the analysis of microbial community structure and composition based on terminal restriction fragment length polymorphisms (T-RFLP) of 16S rDNA genes. *Abstracts of the General Meeting of the American Society for Microbiology* 2002, 102:462.
12. Michel FC, Jr., Sciarini S: T-RFLP FRAGSORT; a computer program to correlate multiple 16S rRNA gene T-RFLP profiles with in silico amplification and digestions of ribosomal database project II alignments. *Abstracts of the General Meeting of the American Society for Microbiology* 2003, 103:N-289.
13. Kent AD, Smith DJ, Benson BJ, Triplett EW: Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. *Applied and Environmental Microbiology* 2003, 69(11):6768-6776.
14. Grant A, Ogilvie LA, Blackwood CB, Marsh TL, Sang-Hoon K, Paul EA: Terminal Restriction Fragment Length Polymorphism Data Analysis. *Applied and Environmental Microbiology* 2003, 69:6342-6343.