This is to certify that the

dissertation entitled

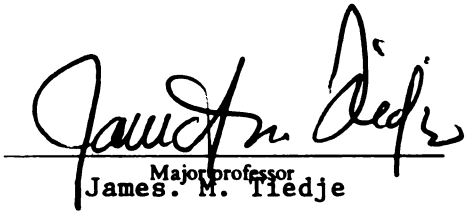
INFLUENCE OF SOIL AGE AND VEGETATIVE COVER ON MICROBIAL COMMUNITY COMPOSITION: A RIBOSOMAL DNA ANALYSIS OF HAWAIIAN SOILS

presented by

Klaus R.L. Nüsslein

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Microbiology



Major professor
James. M. Tiedje

Date May 12, 1998

INFLUENCE OF SOIL AGE AND VEGETATIVE COVER ON MICROBIAL COMMUNITY COMPOSITION: A RIBOSOMAL DNA ANALYSIS OF HAWAIIAN SOILS

By

Klaus R. L. Nüsslein

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology

February 1998

ABSTRACT

The influence of two environmental factors, soil age and vegetative cover, on the structure and composition of soil microbial communities was investigated using Hawaiian soils and DNA-based methods. Hawaiian island soils were expected to have lower diversity than continental soils because they are young, geographically isolated, harbor low plant diversity, and the islands are of relatively constant climate. Molecular methods were used to identify some of the dominant and rare bacterial community members, and yielded reproducible community profiles within sites but distinguishable between sites. The DNA was first fractionated based on its guanine and cytosine content (G+C), and then rDNA amplified from a fraction with high biomass (63% G+C) and a fraction of low biomass (35% G+C). The rDNA clone libraries derived from these fractions were screened by amplified rDNA restriction analysis (ARDRA) to determine phylotype distribution for each soil.

For the 63% G+C fractions the two major changes in community structure were seen for a chronosequence of soils of ages 200, 2,100, 20,000, and 150,000 years. The maximum in phylotype diversity at the 20,000 year old site is correlated with maxima in plant available nitrogen and soil organic matter, while minimum diversity at the 150,000 year old site corresponded to a maximum rate of litter decomposition. The

findings are consistent with the macro-ecological principle of diversity responses to nutrient availability. The rDNA method was reproducible within sites, however, dominant phylotypes were different for each site.

Comparison of the oldest rainforest soil and a soil of an adjacent pasture created by clear cutting eight decades ago revealed a major shift in the microbial community structure and composition. The G+C profile showed a shift to organisms with higher G+C content in the pasture. None of the nucleotide sequences analyzed matched sequences in the databases, and hence reflect novel bacteria.

The two step rDNA analysis used here uncovered new rDNA diversity and provided evidence that soil microbial diversity is much higher than can be revealed in eubacterial clone libraries from total community DNA. This work provides the first evaluation of the response of dominant and rare members within a soil microbial community to environmental factors, and revealed that vegetation was a particularly strong factor in shaping microbial community composition. Understanding what parameters and principles influence soil microbial community composition may aid prediction and management of soil biological processes.

This dissertation is dedicated to my parents Hildegard and Rudolf,
and to my wife Christina

ACKNOWLEDGMENTS

I would like to express gratitude to my major professor and research advisor, James M. Tiedje, for his instruction, guidance and encouragement during my Ph.D study. I have learned much by his example, and I will remain grateful for all the opportunities he made possible for me. Special thanks are extended to my advisory committee, Mike Klug, Eldor Paul, and Loren Snyder for their instruction and encouragement throughout this research work, for the time and effort spent on reviewing this dissertation, and for their valuable comments and suggestions.

Special thanks are extended to Mary Ann Bruns, Larry Forney, Joyce Wildenthal, Bernard Schroeter, Dave Harris, Tom Schmidt, John Urbance, Jizhong Zhou, and Dave Odelson for their invaluable help and assistance during this work. I also thank my fellow students and coworkers for their friendship, support and stimulating discussions during my years at Michigan State University.

This study was funded by a grant from the National Science Foundation Center for Microbial Ecology at Michigan State University under Microbial Ecology Grant NSF DEB #9120006.

TABLE OF CONTENTS

Table of contents, continued

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

List of Tables, continued

APPENDIX

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

List of Figures, continued

List of Figures, continued

CHAPTER 3

List of Figures, continued

CHAPTER 4

List of Figures, continued

List of Figures, continued

APPENDIX

List of Figures, continued

LIST OF ABBREVIATIONS

ARDRA ..... Amplified ribosomal DNA restriction analysis

DGGE ..... Denaturing gradient gel electrophoresis

DTAF ..... 5-(4,6-dichlorotriazin-2-yl)amino fluorescein

G+C ..... guanine and cytosine

HRS-1 ..... Hawaiian rainforest soil clone number one

HPS-2 ..... Hawaiian pasture soil clone number two

NPP ..... Net primary production

PCR ..... Polymerase chain reaction

rDNA ..... ribosomal DNA

R_f ..... refractive index

RFLP ..... restriction fragment length polymorphism

SSU ..... small-subunit

CHAPTER 1

INTRODUCTION AND RATIONALE

BACKGROUND

1. Soil microbial communities

Soil microbial communities are some of the most complex, diverse, and important assemblages of organisms in the biosphere, yet little is known about their structure and their response to environmental influences. The complexity of their environment results in numerous ecological niches. Niche availability and diversity is given among other factors by mineral composition of the soil, salinity, pH value, nutrient availability, organic input, temperature, and water content. This sum of basic variables is overlaid by a high spatial heterogeneity caused by low mixing rates, high surface area, the formation of macro- and microaggregates, and the influence of plants and soil animals. In addition to new niches through the input or the metabolism of organic matter, plants and animals also create new habitats of increased metabolic activity and complexity such as the rhizosphere or animal digestive tracks. Finally, seasonal and vegetative changes

superimpose a layer of temporal heterogeneity, resulting in the common notion that no two soil samples have the same microbial community (Liesack *et al.*, 1997).

As a result of the high niche variability of the soil habitat microbial diversity is thought to be extremely high, however the extent of that diversity is unknown. Furthermore, the influence of soil environmental parameters on this high diversity remains largely unexplored. Diversity and community structure currently are the focal points of soil microbial community studies. In this study diversity is defined as the number of different species present in a soil habitat, or, in molecular terms, the number of different phylotypes (fingerprint patterns or nucleotide sequences) present in a habitat. Community structure includes quantitative information on the number of individuals of different phylotypes or otherwise defined groups.

Analysis of soil microbial diversity presents a major challenge. Molecular techniques are reinforcing the evidence for extremely high below ground diversity among the prokaryotes. Torsvik *et al.* (1990) used reassociation kinetics of soil extracted DNA to yield an estimate of approximately 4,000 different microbial genomes in one gram of soil, representing perhaps as many as 13,000 different species (Torsvik *et al.*, 1994). Diversity data need to be accompanied by data for species or group abundance, since each strain can vary from one cell per gram of soil to maybe 10^8.

2. Why Hawaii

Consequently, Hawaii was chosen for the study of soil microbial community structure and development based on the notion that a young and isolated land mass is lesser developed than old continental systems, and hence exhibits a natural soil community but at reduced microbial diversity. The selected sites were some of the most pristine and simple sites available for addressing questions about whether the particular factors of age and vegetative type influence the microbial community structure and composition.

The Hawaiian Islands are ideal natural laboratories for ecological research. A series of features of the Hawaiian Islands make them particularly suitable for studies of long term soil and ecosystem development. First, the Hawaiian Islands are the most geographically isolated archipelago on Earth, yet, they are very accessible. They exhibit among the youngest soils in the world, with rapid biological development. Macrospecies richness is at a low level. The few species that colonized have radiated to occupy an extremely broad range of environments and soils (Carlquist, 1980). Second, other variables can be held nearly constant. The selected sites have a relatively constant climate without seasonal changes, and the high mean annual precipitation and temperature are nearly constant throughout the year. All sites are at about the same elevation, hence same temperature, are about 1 meter deep, and, with one exception, support intact rain forest vegetation dominated by the same tree species. The mineral soil substrate shows no or little chemical variation. The maritime tropical environment

reduces the impact of Pleistocene climate change. Glaciation, which would reset soil and ecosystem development, was precluded. Third, the Hawaiian Islands result from the movement of the Pacific tectonic plate over a stationary "hot-spot" in the mantle (Hawaii Scientific Project Drilling Team, 1996). The distance from the currently active volcanoes correspond to substrate age. Fourth, the Hawaiian Islands are one of the best studied areas by ecologists. Much background information is supplied on these well selected, representative sites by other research groups working at these sites.

Several critical aspects must not be overlooked. For example, Pleistocene climatic variations occurred also in Hawaii, if in a dampened manner compared to continental and temperate areas. Therefore, sites over 14,000 y of age developed under conditions different from the present (Hotchkiss *et al.*, 1993). Also, subtle effects of surface erosion, varying exogenous inputs of atmospheric dust from Asia ((Jackson *et al.*, 1971; Betzer *et al.*, 1988), and the recent introduction of plant and animal species from outside Hawaii have changed the environment (Cuddihy *et al.*, 1990). Nevertheless, environmental variation that could affect soils and ecosystems have been constrained in Hawaii, to an extent that cannot be matched elsewhere.

3. Ribosomal RNA as a tool in microbial ecology

Knowledge of bacterial diversity obtained after more than 100 years of pure culture study is incomplete, and very few of the total number of microbial species are in culture (Torsvik *et al.*, 1990; Amann *et al.*, 1995). Until the middle of the last decade enumeration and identification of soil microorganisms had to rely on phenotypic methods. However, phenotypic methods are restricted to only those bacteria that can be isolated and cultured. Other, maybe completely unsuspected groups which may be abundant or very active will not be considered, rendering the emerging picture of the soil microbial community false. Only a small proportion of all prokaryotes have so far been cultivated, and the majority of soil bacteria observed microscopically cannot be cultivated. In addition, the selective enrichment culture has severe limitations as an approach to study the community composition of naturally occurring microorganisms (Ward *et al.*, 1992; Amann *et al.*, 1995). Results are not always reproducible due to the variability of phenotypic properties in relation to culture conditions. Furthermore, laboratory cultivation introduces serious bias to community analysis (Boivin-Jahns *et al.*, 1995; Ferris *et al.*, 1996; Rheims *et al.*, 1996), since the bacterial populations obtained through plating are mainly dependent on the isolation media used (Sørheim *et al.*, 1989), as well as on purification and maintenance procedures. Since nutrient-rich media are used the selection might be biased towards copiotrophic bacteria rather than dominant community members. In addition. properties of microbes *ex situ* may differ, because they lack the interaction among populations or with the natural environment, or

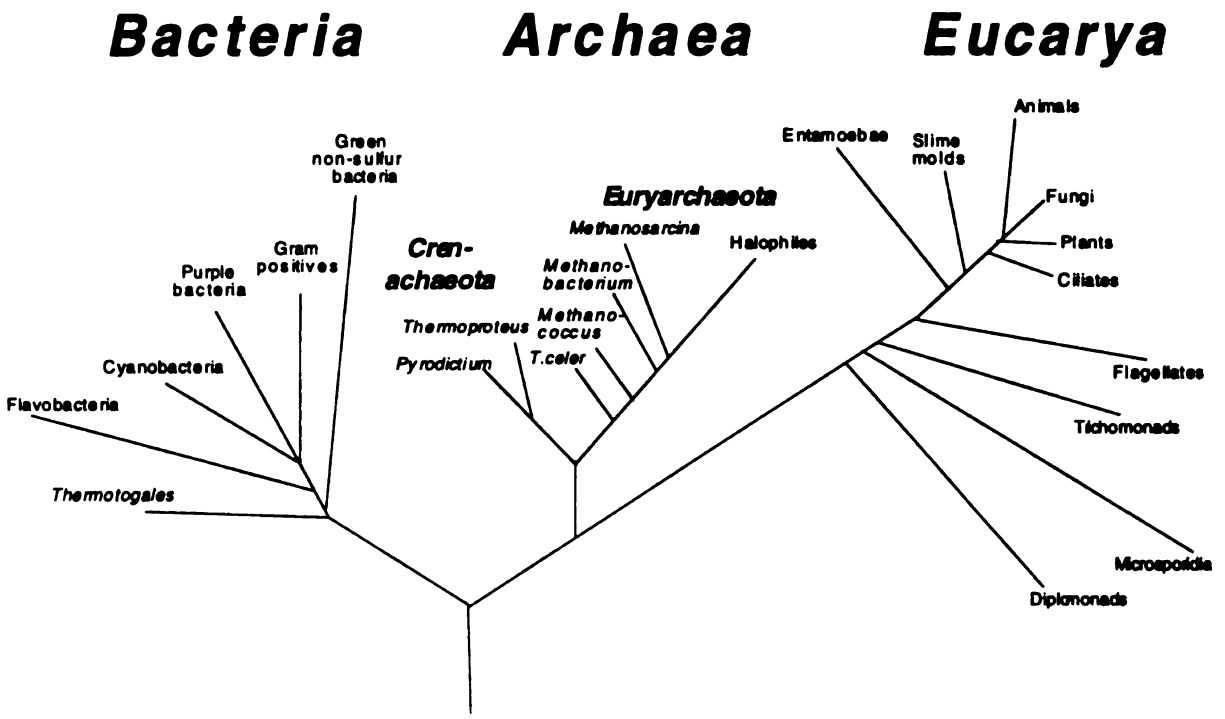


Figure 1. Rooted universal phylogenetic tree showing the three domains, *Archaea*, *Bacteria*, and *Eucarya* based upon prokaryotic SSU rRNA sequences (adapted from Woese, 1994).

change their physiological state as a result of adaptation to the changed environment. Consequently, *in situ* diversity and community structure are unlikely to be represented by collections of natural isolates.

Alternatively, molecular approaches based on phylogenetic analyses of biological markers amongst cellular components are capable of better encompassing the diversity of all well represented members of the entire community. Since the mid-1980s, the use of small-subunit ribosomal ribonucleic acid (SSU rRNA) based techniques has facilitated a culture-independent approach of investigating microorganisms as they occur in nature (Olsen *et al.*, 1986; Ward *et al.*, 1992; Amman *et al.*, 1995). The comparison of these molecular "signature" sequences transformed microbial taxonomy from a pure identification system to an evolutionarily-based framework (Gray *et* al., 1984; Woese, 1987; Olsen *et al.*, 1994). Based on these studies all forms of life are separated into three major evolutionary lines, the three so-called domains: *Bacteria*, *Archaea*, and *Eucarya* (Fig. 1; Fox *et al.*, 1980; Woese *et al.*, 1990). Ribosomal RNA and the corresponding genes (*rrn*) are now widely used as powerful evolutionary and investigative biomarkers for the following reasons (Olsen *et al.*, 1986): (i) Ribosomal RNAs are essential to protein synthesis, and therefore are ubiquitous to all organisms, and structurally and functionally conserved ; (ii) ribosomal RNAs are readily isolated and identified, (iii) they contain variable and highly conserved regions in both primary and secondary structure, (iv) and they appear to change in sequence very slowly, and they do not exhibit horizontal gene transfer found with many other prokayotic genes; therefore relationships between rRNAs reflect evolutionary relationships. These traits

make rRNAs not only the most widely used biomarker, but also a powerful tool for microbial ecology studies, particularly for complex terrestrial environments such as the soil with enormous and undiscovered diversity.

Amongst a variety of cellular biomarkers studied, the SSU rRNA gene provides certain aspects of information that makes it an extremely versatile tool and the best culture-independent biomarker to study microorganisms. Each SSU rRNA gene contains highly conserved regions found among all living organisms as well as unique variable regions and hence diagnostic to certain organisms or related groups. Furthermore, the primary structure of the approximately 1,500 base SSU rRNA gene allows the inference of phylogenies based on comparative sequence analysis. By estimating the phylogenetic relatedness to known microorganisms based on the homology of the gene sequence, the closest affiliation of a newly isolated or molecularly detected microorganism can be established. In combination with a large and growing SSU rRNA database, microorganisms can be sorted according to their phylogenetic affiliation, and conversely, gene probes can be constructed at different levels of specificity. To date, over 6,000 SSU rRNA sequences have been made available for comparison (Maidak *et al.*, 1996). As these databases rapidly expand, they constantly improve the process of matching new sequences to known microorganisms.

4. Molecular microbial diversity of soils

Since the development and widespread application of the polymerase chain reaction (PCR; Saiki *et al.*, 1988), rRNA sequences can directly be obtained from lysed cells, which has contributed to the exponential increase in known prokaryotic SSU rRNA sequences in recent years (Fig. 2). Results from molecular ecological studies within the last seven years from marine (Giovannoni *et al.*, 1990; Schmidt *et al.*, 1991; Moyer *et al.*, 1994), thermophilic (Weller *et al.*, 1991; Ferris *et al.*, 1996), terrestrial (Liesack and Stackebrandt, 1992; Borneman *et al.*, 1996) environments and on symbionts (Amann *et al.*, 1991) documents the success of this strategy.

Several studies explored the microbial diversity of soil bacterial communities by sequencing SSU rDNA clone libraries (Liesack *et al.*, 1992; Ueda *et al.*, 1995; Lee *et al.*, 1996; Borneman *et al.*, 1996, 1997; Kuske *et al.*, 1997; Zhou *et al.*, 1997; Ludwig *et al.*,1997). In all cases, libraries were prepared by polymerase chain reaction amplification (PCR) of rRNA genes directly from total soil genomic DNA. These studies reveal extensive diversity of rRNA operons, and many examples of novel sequences that are only distantly related to those known from cultivated species. Differences between these studies are summarized in Table 1. None of the studies reported duplicate sequences in the analyzed clone libraries, which supports the notion of high diversity. While all the studies analyzed diversity of a particular soil site they cannot readily be compared, because they either use PCR primer sets of unique

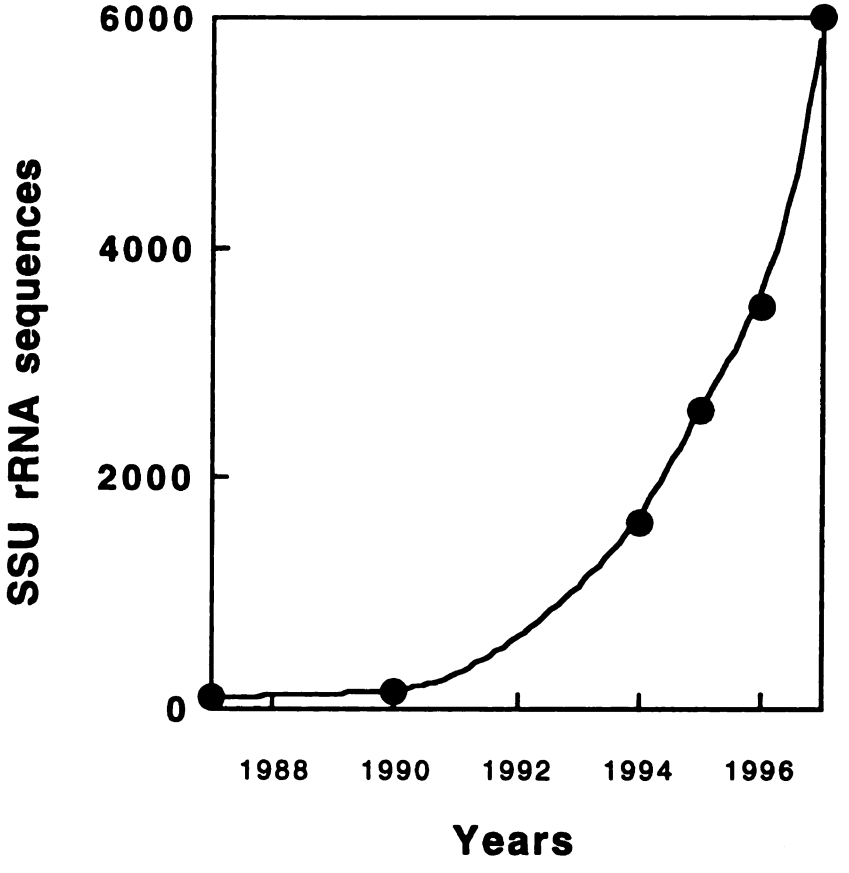


Figure 2. Increase of the number of SSU rRNA genes sequenced since 1987

specificity (Liesack *et al.*, 1992; Rheims *et al.*, 1996) or they analyze sequences in different portions of the SSU rDNA.

The studies of soil community structure found to date tend to concentrate on the relatively few most abundant species, but do not explore the rich variety of organisms anticipated to occur in smaller numbers. These microorganisms may be important in acting as a reservoir of different physiological types, which may respond by population size increase following changes in the habitat. This would also imply a degree of builtin redundancy in the microbial community. Some organisms likely are also remainders of past conditions. Knowledge of the composition of this reservoir will allow us to distinguish between a change in only the dominant community members and a shift of the whole community. In the present study a different, more sensitive path to investigate bacterial diversity of the soil was used, namely the application of a guanine and cytosine (G+C) based fractionation of total soil DNA, and the use of only one G+C fraction out of approximately 80 as the source of community DNA for SSU rDNA studies (Fig.3). Since an underlying community structure was not revealed by analyzing the total eubacterial rDNA, as was shown in this and all of the above mentioned SSU rDNA library studies, total soil extracted DNA was first fractionated by its G+C content using bis-benzimidazole and equilibrium centrifugation (Holben *et al.*, 1994). Thereafter, bacterial rDNA was amplified by PCR using universal eubacterial primers from a fraction with high biomass (63% G+C), and a fraction of low biomass (35% G+C). The rDNA clone libraries were screened by amplified ribosomal DNA restriction analysis (ARDRA) to determine phylotype distribution. Similar restriction patterns

Table 1. Characteristics and differences of seven published rDNA sequenced-based analyses of soil bacterial communities.

Geographic location	Soil environment	16S rRNA Primer used	Primer location	Clones analyzed	Major taxon most commonly found	Reference
Brisbane, Australia	acidic forest mountains	5'-Universal 3'- Streptomyces specific	27F 1224R	113	α-Proteobacteria Actinomycetes Planctomycetes Novel lineages[a]	Liesack *et al.*, 1992; Stackebrandt *et al.*, 1993
Kyushu, Japan	soybean field	5'-Universal 3'-Universal	1100F 1400R	17	Proteobacteria Gram positive groups Green sulfur division Archaea	Ueda *et al.*, 1995
Eastern Washington, U.S.A.	grassland	5'-Universal 3'-Universal	68F 1406R	7	Proteobacteria Planctomycetes Flexibacter Novel lineages[a] (43%)	Lee *et al.*, 1996
Gifhorn, Germany	peat bog	5'-Universal 3'-Gram positive specific	N/A	92	Proteobacteria Gram positives Novel lineages (38%)	Rheims *et al.*, 1996
Wisconsin, U.S.A.	clover pasture	5'-Universal 3'-Universal	530F 1492R	124	Proteobacteria Gram positive groups Cytophaga-group Novel lineages (39%)	Borneman *et al.*, 1996

Table 1. (continued)

Geographic location	Soil environment	16S rRNA Primer used	Primer location	Clones analyzed	Major taxon most commonly found	Reference
Paragominas, Brazil	tropical rainforest, and pasture after clear cutting (29yr)	5'-Universal 3'-Universal	530F 1492R	100	Proteobacteria Gram positive groups Planctomycetes Fibrobacter Cytophaga-group Novel lineages (18%)	Borneman *et al.*, 1997
Kolyma Lowland, Siberia, Russia	arctic tundra	5'-Universal 3'-Universal	27F 1492R	43	Proteobacteria Gram positive groups Fibrobacter	Zhou *et al.*, 1997
Munich, Germany	genetically modified plants	5'-Universal 3'-Universal	27F 1522R	144	N/A	Ludwig *et al.*, 1997
Northern Arizona, U.S.A.	piyon-juniper woodlands on volcanic cinder field	5'-Universal 3'-Universal	27F 1492R (partials only)	60	Proteobacteria Gram positive groups Planctomycetes Flexibacteria Novel lineages (64%)	Kuske *et al.*, 1997

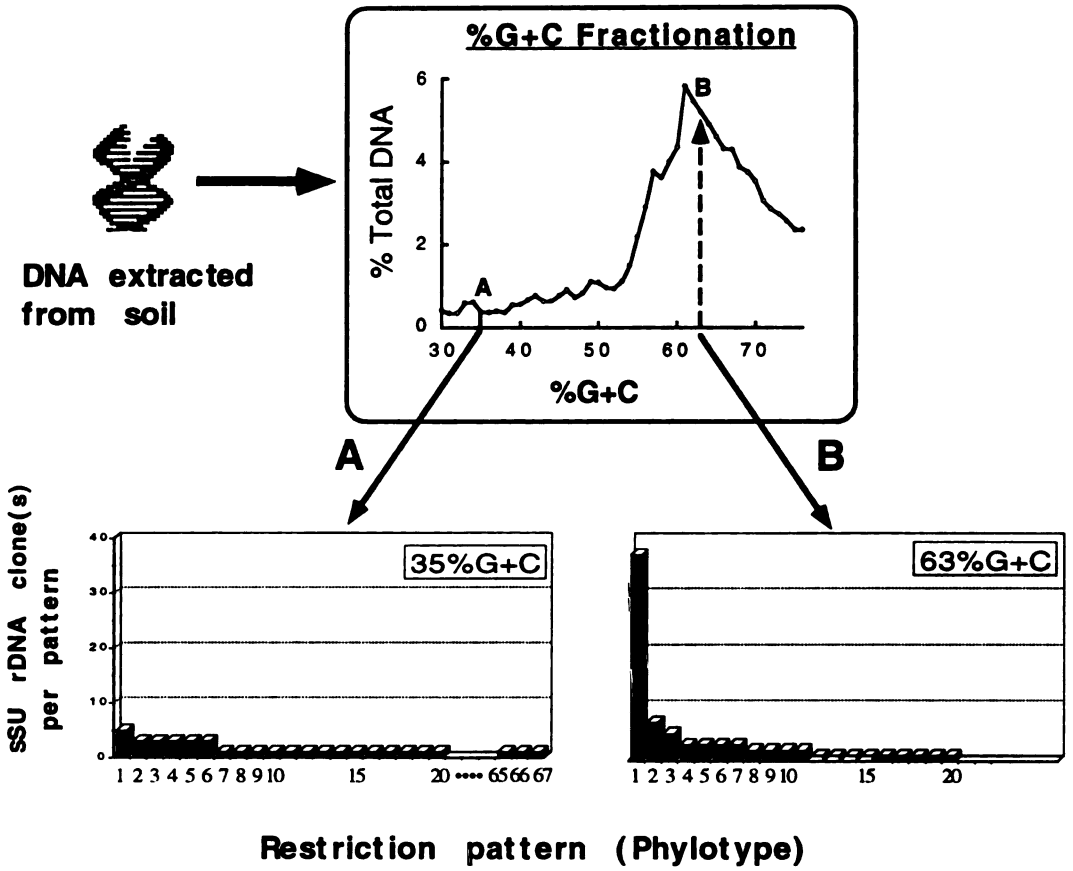


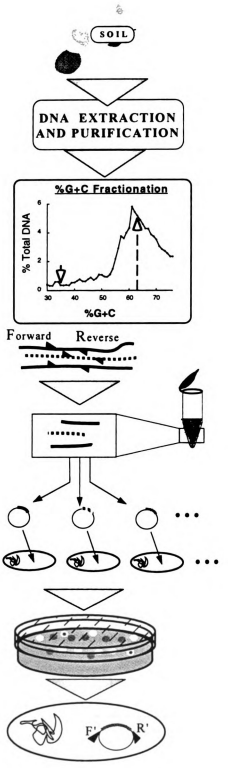
Figure 3. Total soil DNA is fractionated according to its G+C content, and two single fractions are analyzed for their phylotype rank-abundance patterns after restriction digestion with two pairs of restriction enzymes

were grouped into phylotypes by computer assisted cluster analysis to create rank-abundance plots (Fig. 4). Only the nucleotide sequences of the three most dominant and some of the rare clones were determined, and their phylogenetic affiliation identified by comparative sequence analysis. The term "rare" is used here relative to dominant community members, although we cannot speak of rarity in an ecosystem at a sensitivity of $\geq 10^4$ cells per phylotype.

In addition to lowering diversity to a more manageable level, the G+C-based fractionation of total soil DNA yields community profiles specific for each soil site. Evidence from ^{3}H-labeling experiments and G+C based fractionation of soil community DNA indicates that activity and growth are not confined to a small fraction of the bacterial biomass but are widespread amongst the genera represented in the profile (Harris, 1994). Consequently, the two step rDNA analysis used here uncovered more diversity than can be detected by the direct rDNA analysis of total community DNA. The G+C separation step is also a means to detect less dominant organisms in a community.

Since the structure of microbial communities in soil is made up of a high level of prokaryotic diversity, sequencing large numbers of clones would be too costly and time-consuming. The restriction analysis of amplified rDNA (ARDRA) is a tool for the fast and reliable SSU-rDNA-based differentiation of a wide diversity of microorganisms without sequencing. The generation of restriction patterns as fingerprints provides a means for the stepwise exclusion of redundant clones, or for the quantification of repeated patterns. Internet services exist (Hermjakob, 1997; Kim *et al.*, 1997) that

Figure 4. Flowchart of the analytical procedure employed in this study: Soil DNA extraction, fractionation and bacterial community fingerprinting of amplified 16S ribosomal DNA gene sequences and restriction endonuclease analysis (ARDRA) of a clone library



Environmental sample: soil, sediment, water, gut liquid, faeces, etc.

Direct Lysis Method: mechanical, high temperature, and detergents

Fractionation Method: DNA fractionation based on G+C content of genomic DNA; fractions of 35% and 63% G+C are indicated

PCR 1: PCR amplification of 16S rRNA from soil DNA with universal eubacterial primers ( )

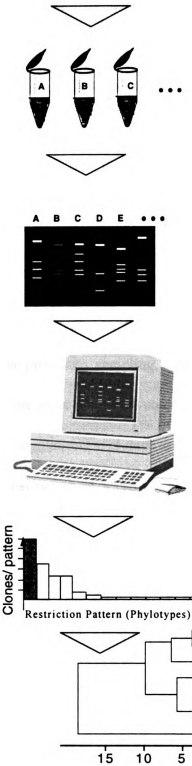
PCR1 product is a mixture of copies of soil bacterial 16S rDNA fragments

Ligation of 16S rDNA pieces into vector p*CR II*

Transformation of vector p*CRII* containing the 16S rDNA insert into *E. coli* Top10' cells

Raising transformants: white colonies indicate the vector contains a 16S rDNA fragment

Figure 4. (continued)



PCR 2 :
PCR amplification of each clone 16S rDNA insert from whole cells

Digestion with two pairs of restriction endonucleases

Agarose gel analysis of restriction length polymorphisms

Computer assisted cluster analysis of similarities between restriction patterns (phylotypes)

Community profile
Group identical restriction patterns and sort by abundance into rank-abundance pattern

Phylogenetic SSU rDNA sequence analysis

compare restriction patterns against a database of theoretical restriction patterns generated form a 16S rDNA database (e.g. RPD database; Maidak *et al.*, 1996). These programs return a list of microorganisms whose computed theoretical restriction patterns match the laboratory derived patterns.

5. Potential biases of SSU rDNA-based studies

The constructed libraries do not always reflect the existing total microbial community, because, as with every method, bias might occur during its construction. The intial step, sampling, could be biased. The clone library might not be representative of the most successful community in this soil due to the potentially patchy distribution of populations in clumps and aggregates in the soil sample. The most critical step in the construction of a clone library is the extraction of nucleic acids from the ecosystem. Isolated DNA should reflect the existing genetic diversity. But microbial cells do not lyse equally well. Especially Gram positive cells might be underrepresentated in the clone library due to their resistence to lysis. Some DNA might get lost during purification treatment. Since previous research suggests (Holben, 1994) that direct lysis recovers a more representative fraction of the genetic diversity than the alternative cell extraction technique, this approach was applied in this study.

Another potential source of bias is the PCR reaction, where certain SSU rRNA genes might preferentially be amplified (Amann *et al.*, 1995). The primer sequences

used in the PCR-mediated amplification of SSU rDNA should cover the phylogenetic diversity present in the sample, but primers can only be as good as the databases on which they are based. Our databases, however, are focused on economically important strains, *i.e.* medical isolates, food and industry strains, and do not reflect the high diversity of terrestrial and other environmental samples.

A third potential source of bias is the PCR coamplification of mixed genomes which leads to the formation of chimeric SSU rDNA molecules (Liesack *et al.*, 1991; Wang *et al.*, 1997). PCR-induced chimeras are formed by annealing of partial-length fragments of different SSU rDNA genes via highly conserved regions. The following primer independent elongation phase then creates full fragments, thereby leading to reports of nonexisting organisms.

A fourth factor that could influence selective recovery is copy number variation of the *rrn* gene between bacterial species which could result in a quantitative misrepresentation of the community profile. This renders clone numbers of low abundance per phylotype unreliable for quantitative evaluation. Other, less important sources of bias in DNA sequence-based analyses of natural communities have been discussed extensively elsewhere (Ward *et al.*, 1992; Stackebrandt *et al.*, 1993; Liesack *et al.*, 1997).

In this study the following seven measures were implemented to reduce potential bias formation: (i) Culturing biases were omitted by using a molecular approach. (ii) Using the direct lysis method we were capable of extracting high yields of non-sheared DNA. The fact that Gram negative as well as many Gram positive organisms were

detected supports the notion that at least some of the reputed lysis resistant bacteria are not underrepresented in the extracted DNA. The high molecular weight of the extracted DNA reduces the probability of chimera formation. (iii) Since primers for the PCR-based amplification of SSU rDNA were universal for conserved regions throughout the entire (known) bacterial domain in the RDP database it is inferred that the mixture of amplified SSU rDNA fragments is as diverse as the genomic DNA extracted. (iv) PCR products were cloned directly in the vector to avoid the introduction of bias from uneven vector to insert ratios needed in other cloning techniques. (v) Only 22 to 25 PCR cycles were used to stay within the quantifiable linear range of the amplification reaction series (Suzuki and Giovannoni, 1996). (vi) Soil DNA templates were denatured immediately before adding them into the PCR reaction assay to free the target DNA from potential contamination with persistent humic compounds, and to optimally expose all target sites.

Despite the care taken in the design of this study, there are still method limitations and biases inherent to each step in community analysis that will influence the results and conclusions. Nevertheless, this study used currently available methods as critically as possible with the present knowledge to minimize the introduction of potential biases.

PURPOSE AND GOAL OF THIS STUDY

This dissertation focuses on the analysis of microbial community composition in a sequence of comparable Hawaiian soils except for differences of age and different vegetative cover (Fig. 5). The goals of the present study were to (i) determine how a major subsample of the soil microbial community changes with soil age, and (ii) to assess the influence of environmental parameters soil age and vegetation on species diversity and dominance structure. Nucleic acid based techniques were used to achieve these goals. The G+C based fractionation of soil extracted DNA yielded patterns of the entire soil microbial community that can be compared, and allowed the selective analysis of soil community subsets. SSU rRNA genes were used as interpreters to distinguish community diversity and community structure, and to describe dominant and rare community members under the influence of soil age and vegetative cover. This study does not attempt the complete comprehensive representation of community structure in a soil, but rather uses around one hundred of the most successful members to be compared between soils. The degree to which environmental parameters affect soil microbial community structure has implications for both our understanding of soil ecosystem structure and our management of soil as an ecosystem resource.

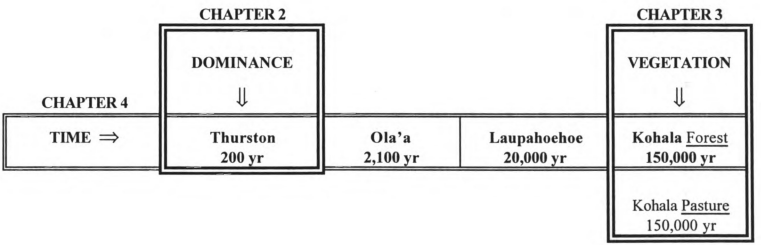


Figure 5. The three projects described in this thesis, and their underlying variables at four soils of a chronosequence in Hawaii

OUTLINE AND THESIS DESIGN

This dissertation is divided into five chapters. This introduction constitutes the first chapter. Chapter 2 describes the analysis of a young and geographically isolated soil for its dominant and rare community members. This study advances the use of G+C based fractionation of the total soil extracted DNA. By concentrating the community structure analysis to two fractions, one within an area of low biomass the other in the G+C range typically high in biomass, the level of sensitivity could be lowered to determine dominance structures within subsets of the soil community. The dominant biomass reflected by the 63% G+C fraction contained several dominant phylotypes, while the community members that were less successful (35% G+C fraction) did not show dominance but a very high diversity of phylotypes. Nucleotide sequence analysis revealed taxa of the groups expected for the particular G+C contents, and important validation for the method. Dominant phylotypes in the 63% fraction were of the *Pseudomonas, Rhizobium-Agrobacterium*, and *Rhodospirillum* assemblages, while all clones sequenced from the 35% fraction were affiliated with several *Clostridium* assemblages. The two step rDNA analysis used here uncovered more diversity than can be detected by the direct rDNA analysis of total community DNA. The G+C separation step is also a means to detect less dominant organisms in a community. Chapter 2 represents a manuscript published in *Applied and Environmental Microbiology* (1998) under the authorship of K. Nüsslein and J. M. Tiedje.

Chapter 3 is a comparison of two adjacent soils with the same long term history but which now differ in their vegetative cover. The reliability of the G+C fractionation

technique could be shown through statistical comparisons of repetitive experiments. A strong shift in soil bacterial community composition was related to change in vegetation cover. Sequence analysis of the three most dominant members revealed that their taxa changed from *Fibrobacter* and *Syntrophomonas* assemblages to *Burkholderia* and *Rhizobium-Agrobacterium* assemblages. The rare community members overlapped by only 5% as determined by ARDRA analysis. This work represents the first evaluation of the response of dominant and rare members of a soil microbial community structure to a major environmental factor, plant type. Chapter 3 is currently a manuscript in preparation for submission as a note to *Applied and Environmental Microbiology* under the authorship of K. Nüsslein and J. M. Tiedje.

Chapter 4 constitutes a study of the influence of soil age on the community structure of a chronoseries of four Hawaiian rainforest soils ranging in age of parent material from 200 y to 150,000 y. The analysis of the 63% G+C fractions demonstrated that while dominance structures intensified with soil age, bacterial diversity decreased. The most abundant phylotype at the oldest soil comprised 45(±2)% of the bacterial clone library. Microbial diversity was found maximal at an intermediate soil age, which correlated with extreme values in a variety of soil physical and chemical parameters. The reproducibility of the technique to screen and analyze clone libraries by their ARDRA pattern was shown by comparison of the identity of dominant phylotypes in replicate samples. Chapter 4 is currently a manuscript in preparation for submission to *Ecology* under the authorship of K. Nüsslein and J. M. Tiedje.

CHAPTER 2

CHARACTERIZATION OF THE DOMINANT AND RARE MEMBERS OF A YOUNG HAWAIIAN SOIL BACTERIAL COMMUNITY WITH SMALL-SUBUNIT RIBOSOMAL DNA AMPLIFIED FROM DNA FRACTIONATED ON THE BASIS OF ITS GUANINE AND CYTOSINE COMPOSITION

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 1998, p. 1283–1289 Vol. 64, No. 4
0099-2240/98/$04.00+0

Characterization of the Dominant and Rare Members of a Young Hawaiian Soil Bacterial Community with Small-Subunit Ribosomal DNA Amplified from DNA Fractionated on the Basis of Its Guanine and Cytosine Composition

KLAUS NÜSSLEIN[1] AND JAMES M. TIEDJE[1,2]*

Center for Microbial Ecology,[1] and Department of Crop and Soil Sciences,[2] Michigan State University, East Lansing, Michigan 48824-1325

Received 18 August 1997/Accepted 12 January 1998

The small-subunit ribosomal DNA (rDNA) diversity was found to be very high in a Hawaiian soil community that might be expected to have lower diversity than the communities in continental soils because the Hawaiian soil is geographically isolated and only 200 years old, is subjected to a constant climate, and harbors low plant diversity. Since an underlying community structure could not be revealed by analyzing the total eubacterial rDNA, we first fractionated the DNA on the basis of guanine-plus-cytosine (G+C) content by using bis-benzimidazole and equilibrium centrifugation and then analyzed the bacterial rDNA amplified from a fraction with a high biomass (63% G+C fraction) and a fraction with a low biomass (35% G+C fraction). The rDNA clone libraries were screened by amplified rDNA restriction analysis to determine phylotype distribution. The dominant biomass reflected by the 63% G+C fraction contained several dominant phylotypes, while the community members that were less successful (35% G+C fraction) did not show dominance but there was a very high diversity of phylotypes. Nucleotide sequence analysis revealed taxa belonging to the groups expected for the G+C contents used. The dominant phylotypes in the 63% G+C fraction were members of the *Pseudomonas*, *Rhizobium-Agrobacterium*, and *Rhodospirillum* assemblages, while all of the clones sequenced from the 35% G+C fraction were affiliated with several *Clostridium* assemblages. The two-step rDNA analysis used here uncovered more diversity than can be detected by direct rDNA analysis of total community DNA. The G+C separation step is also a way to detect some of the less dominant organisms in a community.

Soil microbial communities remain some of the most difficult communities to characterize due to their extreme phenotypic and genotypic diversity. Estimates of the genotypic diversity in these communities based on DNA renaturation experiments suggest that there are 4×10^3 to 7×10^3 different genome equivalents per g of soil (36), which, if extrapolated to species diversity, suggests that there are perhaps 10^3 or even more species per g of soil. Data from culture-based methods also suggest that there is high microbial diversity in soil, but these methods are extremely biased (25, 32) and recover less than 1% of the viable community (3, 20, 36, 39). Molecular approaches in which rRNA sequences are used to determine the composition of natural communities identify more of the entire community. While these approaches also suffer from some biases and lack resolution at the species level, previous rRNA characterizations have confirmed that there is a high level of bacterial diversity in soil communities (4, 20, 35, 38).

To ask meaningful questions about soil community composition, a more manageable level of diversity (lower diversity) is needed. We sought to study a community with lower diversity by focusing on a geographically isolated, young soil, namely, soil formed from volcanic ash deposited 200 years ago on the island of Hawaii. Due to the geographic isolation of the Hawaiian Islands, the diversity of the native fauna and flora is low (6, 37). Furthermore, we used the site studied because it experiences a constant annual climate, which could also lessen selection for diversity. The low level of diversity in the flora and fauna has made the Hawaiian Islands an attractive site for studies on radiation of species and invasion of alien species. If the soil bacterial community is also less diverse for the same reasons, not only would it be less complex to analyze, but it would allow questions about soil community development to be addressed. An additional advantage of the site selected was the large amount of previously collected data on plant composition, ecosystem processes, and soil characteristics which was available (6, 37).

We analyzed the soil bacterial community diversity of this 200-year-old site by performing an amplified ribosomal DNA (rDNA) restriction analysis of a PCR-amplified rDNA clone library. The initial analysis revealed a diversity too great to be captured in a reasonable number of clones analyzed (80 to 90 clones per soil sample). To lower the soil bacterial diversity to a manageable level, we fractionated the soil DNA on the basis of G+C content as described by Holben and Harris (12) and analyzed rDNA clones in two discrete fractions, a fraction with a large amount of DNA (the 63% G+C fraction) and a fraction with a small amount of DNA (the 35% G+C fraction).

MATERIALS AND METHODS

Soil origin and soil sampling. Soil was collected from an undisturbed montane rainforest on the island of Hawaii near Thurston Lava Tube on Kilauea Volcano, within Volcano National Park (19°25'N, 155°15'W). The site is dominated by the native tree species *Metrosideros polymorpha* (which accounts for 91% of the cover) and harbors a total of 34 vascular plant species (6). Several tree ferns form the dominant native understory, which also includes *Citobium* spp., *Coprosoma* spp., and *Vaccinium claycinum*. The site is fenced to keep feral pigs (*Sus scrofa*) out of the park. The sampling site is located at an elevation of 1,200 m, has a mean annual air temperature of 16°C with very little variation (8), and has a mean annual rainfall of 2,500 mm (10) which is well distributed throughout the

* Corresponding author. Mailing address: Center for Microbial Ecology, Plant and Soil Sciences Building, Michigan State University, East Lansing, MI 48824-1325. Phone: (517) 353-9021. Fax: (517) 353-2917. E-mail: tiedjej@pilot.msu.edu.

year due to the relatively constant northeast trade winds (5). The soil is a Hydric Dystrandept developed on several tephra (volcanic ash) depositions ranging in age from 200 to 400 years and is approximately 38 cm deep (31). We removed the litter and the first 1 cm of soil and sampled 7.5 cm of the upper layer, which was deposited 200 years ago. The samples were collected on the perimeter of the Vitousek group's main study site (6). Duplicate soil samples (750 g each) were packaged on site in sterile polypropylene bags and immediately put on ice. The next day they were placed in dry ice coolers and shipped by express mail to Michigan, where they were stored at −20°C. The soil moisture content was determined by drying soil overnight at 100°C. Soil mechanical and chemical analyses were done at the Soil Analysis Laboratory, Michigan State University, by using the methods described by Peck et al. (27).

DNA extraction and purification. Soil microbial DNA was extracted from 10 g of soil by the direct lysis method of Holben (11), except that EDTA was not included to reduce coextraction of humic compounds, a low shaker speed was used to prevent extensive DNA shearing, the shaking time was extended to 45 min, and the phosphate concentration in the lysis buffer was adjusted to 100 mM to overcome the high phosphate absorption capacity of the young minerals. The subsequent DNA purification was also modified to include agarose gel purification (0.4% agarose) and a single Microcon-100 microcolumn (Amicon Corp., Beverly, Mass.) passage of the excised and melted gel piece, followed by repeated washing steps. The extraction efficiency was determined by comparing the amount of DNA extracted with the amount of DNA expected, as calculated from the difference between the direct microscopic counts of bacterial cells before and after lysis. Bacterial cells were counted directly by computer-aided microscopic counting procedures (43). For consistency, all counts were obtained by a single investigator.

DNA was quantified by fluorometry (18) with a model TK 100 fluorometer (Hoefer Scientific Instruments, San Francisco, Calif.) by using the extended assay protocol of the manufacturer. Five replicates were used to estimate DNA yields. Known amounts of lambda phage DNA (Boehringer Mannheim, Indianapolis, Ind.) were used for all calibrations. The fluorescence intensity of DNA was also estimated based on the relative intensities in agarose gels of PCR amplification products and restriction digests of known mass.

G+C fractionation technique. DNA fragments were separated on the basis of G+C content by the procedure of Holben and Harris (12). Briefly, DNA was mixed with bis-benzimidazole (Hoechst reagent no. 33258), which binds to adenine and thymidine and changes the buoyant density of DNA in proportion to its G+C content (40). A gradient of G+C concentrations was then established by equilibrium density gradient ultracentrifugation, and 0.2-ml fractions were collected with a fraction collector. The DNA in each fraction was quantified by spectrophotometry, and its G+C content was established by using a standard curve relating G+C content to density, which was measured with a Bausch & Lomb refractometer. To make PCR amplification possible, bis-benzimide and CsCl were removed from DNA fractions by five repeated extractions in CsCl-saturated isopropanol, followed by spin column chromatography (Wizard PCR Preps; Promega, Madison, Wis.) with two washing steps. A_{260} was determined before and after purification to monitor for potential losses of DNA during the purification procedure.

PCR amplification of SSU rRNA genes from soil DNA. Small-subunit (SSU) rRNA genes were PCR amplified from purified soil DNA by using eubacterium-specific primers fD1 and rP2 of Weisburg et al. (41). PCR were performed with *Taq* DNA polymerase (Boehringer Mannheim) by using the manufacturer's protocol, an additional 400 ng of bovine serum albumin (Sigma Chemical Co., St. Louis, Mo.) per μl, and a model 9600 thermal cycler (Perkin-Elmer Cetus, Norwalk, Conn.). The protocol used consisted of an initial denaturation step (94°C for 130 s) followed by 25 cycles consisting of denaturation at 94°C for 60 s, primer annealing at 55°C for 30 s, and elongation at 72°C for 120 s plus an additional 7-min cycle to finalize the chain reaction. Negative controls without added DNA, as well as positive controls with pure culture genomic DNA, were included in all PCR. Aliquots (3 μl) of the amplified products were separated in a 0.9% agarose gel by electrophoresis in 1× TAE buffer (22), the gel was stained with ethidium bromide (500 ng/μl), and the bands were visualized by UV excitation. The PCR products were stored at −20°C.

To ensure that only soil bacterial rRNA genes were amplified, the following four quality control steps were used: primer purity was established by high-performance liquid chromatography; oligonucleotide primers were prepared fresh from lyophilized stocks for each use; preamplification heating was used to maximize PCR sensitivity and specificity (7); and the bovine serum albumin stock solution and the commercial *Taq* DNA polymerase were tested for potential contamination with bacterial nucleic acids (29).

Analysis of SSU rDNA clone library. The concentrations of PCR-amplified SSU rRNA genes were determined by comparing the fluorescent-band intensities on agarose gels to the fluorescent-band intensities of known concentrations of standard lambda DNA. Prior to cloning, the amplified SSU rDNA fragments were purified by spin column chromatography (Wizard PCR Miniprep; Promega). An equimolar amount of amplified PCR products was ligated to the vector pCR II (Invitrogen, San Diego, Calif.). Ligation and transformation into *Escherichia coli* Top-10F' competent cells were carried out according to the manufacturer's protocol. A primer pair specifically designed to complement the polylinker of the vector pCR II (44) was used to amplify plasmid inserts directly from the transformant cells for SSU rDNA gene screening. The following two types of negative controls were used in the amplification of clone inserts: controls without target DNA added and controls in which untransformed cells were used as a target. The PCR were performed as described above, except that the primer annealing temperature was higher (68°C). To screen for SSU rDNA diversity, the amplified inserts were first digested overnight at 37°C by adding 0.2 U of *Hha*I and 0.2 U of *Hae*III (Gibco BRL, Gaithersburg, Md.) to 5 μl of the PCR product. The resulting fragments were separated by gel electrophoresis in 3.5% Metaphor agarose (FMC Bioproducts, Rockland, Maine) in the presence of ethidium bromide and fresh 1× TBE buffer (22) at 4°C and 5 V/cm for 6 h. Clones with identical restriction patterns were digested with two additional tetrameric restriction endonucleases (0.2 U of *Msp*I and 0.2 U of *Rsa*I). The similarities between the electrophoretic patterns of restriction fragments were analyzed with GelCompar software (Applied Maths, Kortrijk, Belgium). The cluster analysis method used was comparative numerical analysis with the unweighted pair group method using arithmetic averages (UPGMA). Individual clones were grouped by using a cutoff of 97% similarity and a 5% error rate for the band position. The diversities of the phylotypes in different samples were compared by rarefaction analysis (13, 30).

DGGE. Denaturing gradient gel electrophoresis (DGGE) was performed as previously described by Muyzer (24) with eubacterial PCR primers F-968 and R-1401 of Nübel et al. (26). The parallel denaturing gradient was cast with denaturing agent concentrations ranging from 0 to 60%. The fragments were made visible by acidic silver staining.

Determination of nucleotide sequences and phylogenetic analysis. Amplified SSU rDNA clone inserts were purified (Wizard PCR Preps; Promega) and partially sequenced. Nucleotide sequences were determined by the fluorescent DiDeoxy termination method by performing automated fluorescent *Taq* cycle sequencing with the ABI Catalyst 800-ABI 373A sequencing system (Applied Biosystems, Foster City, Calif.). To ensure accuracy and to aid in chimera detection, both ends of the SSU rDNA molecule were sequenced with reverse primers J529R (5'-CGCGGCTGCTGGCAC-3') and rP2 (41). All sequences were about 400 bases long and were aligned manually with sequences in the SSU database of the Ribosomal Database Project (RDP) (21) based on primary- and secondary-structure considerations. The results obtained were compared to alignments obtained with Align Sequence, version 2.0, from the ARB sequence analysis software package (33). Phylogenetic relationships were inferred by using the neighbor-joining method (28) and the modified Jukes-Cantor algorithm (16, 42). The robustness of the final topology was tested with the tree-building methods PAUP (34) and fastDNAml (21). All phylogenetic assignments were made and phylogenetic trees were constructed within the ARB software package (33) by using version 5.0 of the RDP database (21). Only unambiguously aligned regions were used for the sequence analysis (Table 1).

To detect potential chimeric artifacts in the partial sequences of the 3' end and the 5' end, as well as the entire SSU rDNA gene, two strategies were used. The partial sequences that were around 400 bases long were (i) examined with the CHECK_CHIMERA program offered by RDP (21) and, for comparison, (ii) examined with the mglobalCHI program offered by the USC Computational Biology web site (17). To detect potential chimeric artifacts in a complete SSU rDNA gene, the phylogenies determined from the sequences of the 3' ends and the 5' ends were compared.

Nucleotide sequence accession numbers. The sequence data for Hawaiian rainforest soil clones HRS-1 through HRS-23 have been deposited in the GenBank database under accession no. AF016514 to AF016533.

RESULTS

Soil analysis. The chemical analysis of the soil revealed a composition quite typical for a very young soil compared to similar but older forest soils. The soil organic matter content was 3.1%, and the pH was 5.4. The soil nitrate N and ammonia N contents were determined to be 0.2 and 3.6 μg/g, respectively, while the total nitrogen content was 0.08%. The sand particles had sharp faces, which indicated that there had been little weathering and resulted in rapid water infiltration. The humic material had to be less than 200 years old, and hence its chemical structure was different than that of typical soil humic acid.

In this young tropical rainforest soil the C/N ratio (a measure of biological activity in soils) was 16:1 (compared with a global average of 14:1), suggesting that an active bacterial community was present. This assumption was supported by high leaf litter decomposition rates (37), as well as elevated in situ mineralization and nitrification rates (6). The cation exchange capacity was low (10.8 meq/100 g) due to primary minerals.

TABLE 1. Phylogenetic affiliations based on SSU rDNA genes of members of a Hawaiian rainforest soil bacterial community[a]

Phylotype	Relative abundance (%)[b]	Phylogenetic affiliation	Most closely related organism in RDP database		% Similarity to most closely related organism
			Taxon	RDP subdivision	
35% G+C fraction					
HRS-1	6.8	Clostridia and their relatives	[*Clostridium tetani*][c]	*Clostridium novyi* subgroup	80.6
HRS-2	6.8	Clostridia and their relatives	[*Clostridium tetani*]	*Clostridium novyi* subgroup	76.1
HRS-3	6.8	Clostridia and their relatives	*Clostridium fallax*	*Clostridium perfringens* assemblage	92.9
HRS-4	6.8	Clostridia and their relatives	*Clostridium fallax*	*Clostridium perfringens* assemblage	94.0
HRS-6	5.1	Clostridia and their relatives	*Clostridium puniceum*	*Clostridium butyricum* subgroup	95.3
HRS-7	1.7	Clostridia and their relatives	*Clostridium butyricum*	*Clostridium butyricum* subgroup	94.7
HRS-8	1.7	Clostridia and their relatives	*Clostridium puniceum*	*Clostridium butyricum* subgroup	98.0
HRS-10	1.7	Clostridia and their relatives	*Clostridium beijerinckii*	*Clostridium butyricum* subgroup	98.5
HRS-21	1.7	Clostridia and their relatives	*Clostridium puniceum*	*Clostridium butyricum* subgroup	98.5
HRS-22	1.7	Clostridia and their relatives	[*Clostridium tetanomorphum*]	*Clostridium butyricum* subgroup	80.0
63% G+C fraction					
HRS-12	13.2	Alpha purple bacteria	[*Rhodopseudomonas viridis*]	*Rhizobium-Agrobacterium* group	82.6
HRS-13	10.5	Gamma purple bacteria	*Pseudomonas syringae*	*Pseudomonas* subgroup	97.9
HRS-16	7.9	Alpha purple bacteria	*Azospirillum lipoferum*	*Rhodospirillum rubrum* assemblage	88.0
HRS-17	1.3	Alpha purple bacteria	*Zoogloea ramigera*	*Brucella* assemblage	90.4
HRS-18	1.3	*Fibrobacter* phylum	Environmental strain MC 103	*Acidobacterium* subdivision	93.2
HRS-19	1.3	*Bacillus-Lactobacillus* subdivision	[*Bacillus flavothermus*]	*Bacillus megaterium* group	80.3
HRS-20	1.3	Delta purple bacteria	Environmental strain FIE 20	*Myxobacterium* group	92.9

[a] Only unambiguously aligned regions were used in the analysis.
[b] Relative abundance of clones belonging to a phylotype, calculated by dividing the number of clones belonging to the phylotype by the total number of clones analyzed.
[c] Brackets indicate that the taxonomic assignment of the closest relative is uncertain (level of similarity, less than 85%).

DNA extraction and purification. Direct microscopic cell counts of soil smears before and after DNA extraction revealed a high lysis efficiency, 91% ± 3%. The lysis efficiency as estimated by DNA yield was also high; 6.4 μg of DNA/g (dry weight) of soil was recovered, compared to an expected DNA yield of 5.8 μg/g, which was calculated by multiplying the microscopic bacterial cell counts ($1.6 \times 10^9 \pm 0.3 \times 10^9$ cells/g) by an average of 4×10^{-15} g of DNA cell^{-1} (9). Coextraction of humic material prevented PCR amplification of SSU rRNA genes if the standard purification protocol of Holben was used, probably because of the unusual nature of the yellow humic acids. Additional clean-up by preparative agarose gel electrophoresis and an additional series of washing steps in a centrifugal concentrator were needed to obtain DNA clean enough for reliable amplification of the bacterial SSU rRNA genes.

Community G+C profile. A profile of the community composition was obtained based on the amounts of DNAs having different G+C contents. The G+C contents of the majority of the soil DNA were in the range from 52 to 68% (Fig. 1); this included members of genera known to dominate soil bacterial communities, including the genera *Agrobacterium* (57 to 63% G+C), *Alcaligenes* (56 to 63% G+C), *Arthrobacter* (63 to 69% G+C), and *Pseudomonas* (58 to 66% G+C) (12). A rather consistent but minor quantity of DNA was found with G+C contents ranging from 30 to 50%, a range which is found in members of soil genera like *Streptococcus* (35 to 40% G+C), *Clostridium* (24 to 54% G+C), and *Bacillus* (32 to 69% G+C). Compared with the community profiles of midwestern agricultural soils, the main peak of the Hawaiian bacterial community profile was shifted toward a lower G+C content (by approximately 4% G+C) (Fig. 1).

Phylotype abundance patterns. Agarose gel electrophoresis revealed a clean band of SSU rRNA genes selectively amplified from purified soil-extracted DNA. The clones were digested with restriction enzymes (amplified rDNA restriction analysis) (Fig. 2) and sorted by cluster analysis. Of the 81 clones obtained from unfractionated soil DNA that tested positive for alpha-complementation of β-galactosidase, 67 contained the 1.5-kb SSU rDNA insert. Primary restriction with *Hae*III and *Hha*I resulted in 64 different restriction patterns (Fig. 2 and 3A), indicating that there was a high level of diversity. Clones with similar restriction patterns were not differentiated when the preparations were further digested with a combination of *Msp*I and *Rsa*I. The two patterns that were repeated each accounted for only 3% of all of the SSU rDNA clones, while the remaining 60 patterns were each represented

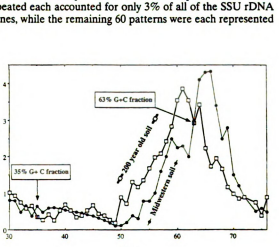


FIG. 1. Microbial community structure of a young rainforest soil determined by the G+C content of its DNA. The bacterial community profile of a mid-Michigan agricultural soil is included for comparison (12).

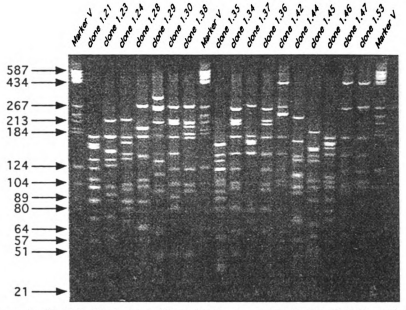


FIG. 2. Restriction patterns of amplified SSU rRNA clones in the 63% G+C fraction after restriction digestion with *Hae*III and *Hha*I. Plasmid pBR322 digested with *Hae*III (marker V) was used as a DNA size marker.

by a single clone (Fig. 3A). In order to reduce this diversity to a more manageable level, two G+C content fractions (35 and 63% G+C fractions) of the whole community DNA were used in a similar analysis. We chose the 63% G+C fraction because it was located within the major peak of DNA abundance typically found in temperate region soils. The second fraction, the 35% G+C fraction, was chosen randomly to represent a portion of the minor members of the community (Fig. 1). Table 1 shows the relative abundance of selected clones as a measure of dominance relative to the entire clone library examined. The 63% G+C fraction produced 46 different patterns for the 76 clones examined (Fig. 3C). Eleven of these patterns were represented by two or more clones, which together accounted for 54% of the clones investigated. The three most dominant restriction patterns, patterns 1, 2, and 3, accounted for 13, 11, and 8%, respectively of the SSU rDNA inserts analyzed. The 35% G+C fraction produced 47 different patterns for the 59 clones examined (Fig. 3B). Only seven of these patterns were represented by two or more clones, and they accounted for 32% of the clones examined. The two most dominant patterns each represented 7% of the SSU rDNA insert diversity in this fraction. In both fractions together only two clones with similar restriction patterns were differentiated when the preparations were digested with the second restriction endonuclease pair, *Msp*I plus *Rsa*I.

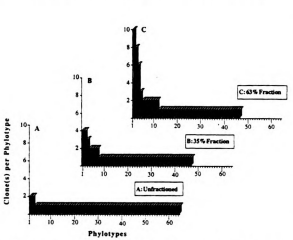


FIG. 3. Frequency distribution of SSU rDNA gene phylotypes (restriction patterns) from the total soil DNA (A), the 35% G+C fraction (B), and the 63% G+C fraction (C) of a young Hawaiian soil. The profiles are based on results obtained after digestion with tetrameric restriction endonucleases *Hae*III plus *Hha*I and *Msp*I plus *Rsa*I.

The quantitative effect of using a fraction of the community DNA instead of the total community DNA on the diversity of phylotypes was evaluated by rarefaction analysis. Given the species abundance distribution of a clone library, rarefaction gives estimates of the species richness of subsamples taken from it. This analysis verified that the unfractionated DNA contained too many phylotypes to reveal any structure but that fractionation of the DNA on the basis of G+C content did reduce the diversity to a level at which structures of dominance could be detected; the data also suggested that the phylotype sampling in the two fractions was far from complete (Fig. 4).

DGGE analysis. The abundance of particular SSU rDNA clones in our clone library may not represent the actual quantitative abundance of the clones in the soil sample due to a PCR or cloning bias. To assess this, we used DGGE to determine whether the dominant phylotypes in the clone library corresponded to well-represented phylotypes on DGGE gels of community DNA. Parallel analysis by DGGE of the two G+C fractions and dominant clones from the two fractions revealed that the intensely stained bands in the community DNA in each G+C fraction (indicating strong representation) corresponded to the bands obtained from the individual clones

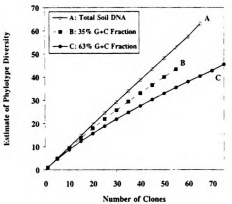


FIG. 4. Rarefaction curves for the phylotypes found in total soil DNA and in the 35 and 63% G+C fractions of a Hawaiian soil. The expected numbers of phylotypes calculated from a random sample of individuals taken from the total population of phylotypes are shown on the *y* axis.

(Fig. 5). Even though phylogenetically unrelated strains can have bands with the same R_f value due to identical melting behavior of the SSU rDNA fragment, it is unlikely that a strain other than the one detected was both strongly represented and had the same R_f value.

Sequence analysis. Two representative clones of the five dominant phylotypes from both abundance profiles (Fig. 3) and selected rare individuals were used for a partial sequencing and phylogenetic characterization analysis. This analysis revealed corresponding phylogenetic affiliations for clones belonging to the same phylotype (e.g., clones HRS-1 and HRS-2 and clones HRS-3 and HRS-4) (Table 1), although the clones varied somewhat in sequence similarity (not all matching pairs are shown in Table 1). Of the 23 clones sequenced, 20 were members of the domain *Bacteria*, while three clones were dismissed as possible chimeras. The phylogenetic affiliations and closest relatives in the RDP database are shown in Table 1. Figure 6 illustrates the phylogenetic relationships among some of the 35% G+C fraction clones. Several of the clostridial clones are closely related to each other. The phylogenetic affiliations of particular clones and their levels of abundance are summarized in Fig. 7. None of the clones exhibited an exact match with any of the SSU rDNA sequences found in the databases. In particular, clone HRS-18, which is related to the *Acidobacterium* subdivision, confirmed that novel taxa discovered previously in other molecular surveys of soil were present (20).

DISCUSSION

Geographic isolation, young age (200 years), constant climate, and low diversity of plant species did not reduce the microbial diversity in the soil studied to an extent that revealed more than two instances of resampling of the same eubacterial phylotype in a 70-clone library. Hence, the diversity was too great to reveal underlying community structure by this method. Similarly, Borneman et al. (4) found only 4% duplicates among 124 soil rDNA clones from an older continental soil. One approach to reduce complexity is to limit the study to only certain subsets of the community, an approach typically used by macroecologists. We attempted to do this by analyzing rDNA fractions having certain G+C contents since particular bacterial taxa have characteristic G+C contents. Clone libraries obtained from pooled DNA fractions (e.g., DNAs having G+C contents of 61 to 65%) still contained few repeated phylotypes (data not shown), but samples from one of these fractions (e.g., the 63% G+C fraction) contained more repeated phylotypes, indicating that more complete coverage of the rDNA types in this sample was obtained. Rarefaction analysis also indicated that there was reduced diversity in the individual G+C fractions compared to the total DNA, especially the 63% G+C fraction, but that rDNA diversity was far from exhausted in a 76-clone library of this fraction (Fig. 4). Hence, separation on the basis of G+C content revealed new diversity and provided evidence that soil rDNA diversity is much greater than the diversity that is revealed by eubacterial clone libraries of total community DNA. This estimate of greater soil rDNA diversity supports the high level of bacterial diversity estimated by Torsvik et al. (36) based on rates of soil DNA reannealing. The G+C content separation method also offers a way to enrich for rarer members of the community since DNAs having other G+C contents, especially DNAs of dominant types, can be removed by separating the DNAs into different fractions.

Dominance of phylotypes was observed in the 63% G+C fraction but not in the 35% fraction. This difference is consistent with the ecological prediction that the most dominant

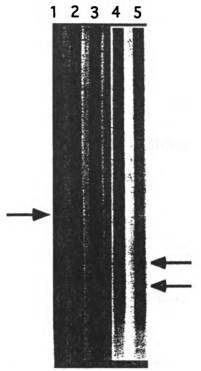


FIG. 5. DGGE analysis of SSU rDNA fragments (length, 434 bp) obtained after PCR amplification of the 35 and 63% G+C fractions of soil DNA and individual SSU rDNA clones from the clone libraries that were most frequent as determined by amplified rDNA restriction analysis. Lane 1, clone HRS-2 of the 35% G+C fraction; lane 2, 35% G+C fraction; lane 4, 63% G+C fraction; lane 5, clone HRS-12 of the 63% G+C fraction. Lane 3 contained a mixture of several bacterial genomic DNAs as a marker and positive control. The figure is a negative image of a silver-stained DGGE separation pattern. The arrows indicate dominant bands of well-represented clones, which were also found in the respective soil DNA fractions. PCR products obtained from some strains (lane 5) produced more than one band due to sequence heterogeneities of 16S rRNA operons (24).

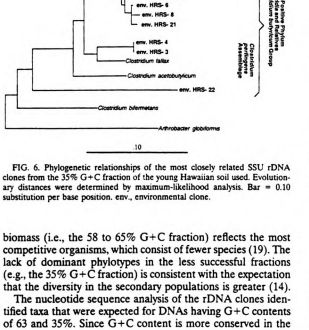


FIG. 6. Phylogenetic relationships of the most closely related SSU rDNA clones from the 35% G+C fraction of the young Hawaiian soil used. Evolutionary distances were determined by maximum-likelihood analysis. Bar = 0.10 substitution per base position. env., environmental clone.

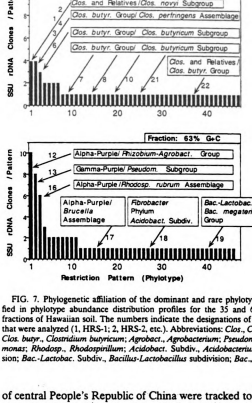


FIG. 7. Phylogenetic affiliation of the dominant and rare phylotypes identified in phylotype abundance distribution profiles for the 35 and 63% G+C fractions of Hawaiian soil. The numbers indicate the designations of the clones that were analyzed (1, HRS-1; 2, HRS-2, etc.). Abbreviations: *Clos., Clostridium*; *Clos. butyr., Clostridium butyricum*; *Agrobact., Agrobacterium*; *Pseudom., Pseudomonas*; *Rhodosp., Rhodospirillum*; *Acidobact. Subdiv., Acidobacterium* subdivision; *Bac.-Lactobac.* Subdiv., *Bacillus-Lactobacillus* subdivision; *Bac., Bacillus*.

biomass (i.e., the 58 to 65% G+C fraction) reflects the most competitive organisms, which consist of fewer species (19). The lack of dominant phylotypes in the less successful fractions (e.g., the 35% G+C fraction) is consistent with the expectation that the diversity in the secondary populations is greater (14).

The nucleotide sequence analysis of the rDNA clones identified taxa that were expected for DNAs having G+C contents of 63 and 35%. Since G+C content is more conserved in the *rrn* operon than in the genome as a whole (39), the separation method must be driven primarily by the G+C content of the flanking DNA. The DNA fragments obtained by the DNA isolation method used were usually more than 20 kb long.

The aerobic bacteria most often cultured from soil (e.g., members of the genera *Arthrobacter*, *Pseudomonas*, and *Burkholderia*) have the same G+C contents as the most dominant biomass, as determined from the DNA data. Since only a low percentage of soil bacteria can be cultivated, these data suggest that the majority of the unculturable types must have DNA G+C contents of 55 to 68%, and hence these organisms are either close relatives of the typical culturable forms or are new types but have G+C contents in this range.

Finding a large diversity of clostridia in the 35% G+C fraction was initially unexpected since the site used is well drained with high aeration and porosity (Fig. 6). However, the high rainfall throughout the year and the young organic matter could provide anaerobic microsites for growth of the clostridial community (1), and bird feces is one of the most feasible inoculum sources for this group.

The microbial colonization of the young Hawaiian soil examined by very diverse phylotypes is not easily explained, especially considering the very high rDNA diversity in the G+C fractions. The geographic isolation of the Hawaiian Islands certainly limited colonization by plants (seed dispersal), insects, and other organisms, including humans (until 500 AD). Bacterial colonization of soil could have resulted from avian transport, including avian fecal introductions, from seawater spray, or from aerial transport. Aerial transport has been considered very inefficient due to poor microbial survival caused by long exposure to UV light, desiccation, and the likely fallout of any microbe-associated soil particles during transport from Asia (more than 10,000 km). Recently, however, sand grains from sandstorms in the Gobi Desert and loess plateau regions of central People's Republic of China were tracked to Hawaii, even though the density and the size of the sand grains suggested that they should have been deposited long before they reached Hawaii (2, 15, 23). Regardless of the sources of bacterial colonization of Hawaiian soil, the soil communities appear to be unexpectedly complex, but this does not mean that they display the full complement of microbial types or diversity found in montane rainforest soils of continental environments.

Studies such as this one, performed by using PCR amplification of SSU rDNA genes from community-derived DNA, should never be assumed to be comprehensive because of well-known biases (17, 39). In particular, they are likely to reflect *rrn* operons that are more readily PCR amplifiable. Nevertheless, this study supports the notion that even young terrestrial environments exhibit enormous diversity and contain novel, uncultivated organisms.

ACKNOWLEDGMENTS

We thank T. Hattori and J. Urbance for comments on the manuscript and P. Lepp for instructions on use of the ARB software. We thank D. Harris for his introduction to the G+C fractionation technique and microscopic image analysis.

This work was supported by National Science Foundation grant DEB 9120006 to the Center for Microbial Ecology.

REFERENCES

1. **Arah, J. R. M., and K. A. Smith.** 1985. Anaerobic micro-environments in soil and the occurrence of anaerobic bacteria, p. 247–261. *In* V. Jensen, A. Kjøller, and L. H. Sørensen (ed.), Microbial communities in soil microbiology. Elsevier Applied Science Publishers, London, England.
2. **Betzer, P. R., K. L. Carder, R. A. Duce, J. T. Merrill, N. W. Tindale, M. Uematsu, D. K. Costello, R. W. Young, R. A. Feely, J. A. Breland, R. E. Bernstein, and A. M. Greco.** 1988. Long-range transport of giant mineral

aerosol particles. Nature 336:568–571.
3. **Boivin-Jahns, V., A. Bianchi, R. Ruimy, J. Garcin, S. Daumas, and R. Christen.** 1995. Comparison of phenotypical and molecular methods for the identification of bacterial strains isolated from a deep subsurface environment. Appl. Environ. Microbiol. 61:3400–3406.
4. **Borneman, J., P. W. Skroch, K. M. O'Sullivan, J. A. Paulus, N. G. Rumjanek, J. L. Jansen, J. Nienhus, and E. W. Triplett.** 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. Appl. Environ. Microbiol. 62:1935–1943.
5. **Carlquist, S. C.** 1980. Hawaii, a natural history. Pacific Tropical Botanical Garden, Honolulu, Hawaii.
6. **Crews, T. E., K. Kitayama, J. H. Fownes, R. H. Riley, D. A. Herbert, D. Mueller-Dombois, and P. M. Vitousek.** 1995. Changes in soil phosphorous fractions and ecosystem dynamics across a long chronosequence in Hawaii. Ecology 76:1407–1424.
7. **D'Aquila, R. T., L. J. Bechtel, J. A. Videler, J. J. Eron, P. Gorczca, and J. C. Kaplan.** 1991. Maximizing sensitivity and specificity of PCR by preamplification heating. Nucleic Acids Res. 19:3749.
8. **Department of Geography, University of Hawaii.** 1983. Atlas of Hawaii, 2nd ed. University of Hawaii Press, Honolulu.
9. **Ellenbroek, F. M., and T. E. Cappenberg.** 1991. DNA synthesis and tritiated thymidine incorporation by heterotrophic freshwater bacteria in continuous culture. Appl. Environ. Microbiol. 57:1675–1682.
10. **Giambelluca, T. W., M. A. Nullet, and T. A. Schroeder.** 1986. Rainfall atlas of Hawaii. Department of Land and Natural Resources, Honolulu, Hawaii.
11. **Holben, W. E.** 1994. Isolation and purification of bacterial DNA from soil, p. 727–751. *In* R. W. Weaver et al. (ed.), Methods of soil analysis, part 2. Microbiological and biochemical properties. Soil Science Society of America Book Series no. 5. Soil Science Society of America, Madison, Wis.
12. **Holben, W. E., and D. Harris.** 1995. DNA-based monitoring of total bacterial community structure in environmental samples. Mol. Ecol. 4:627–631.
13. **Hurlbert, S. H.** 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586.
14. **Huston, M. A.** 1994. Biological diversity. Cambridge University Press, Cambridge, United Kingdom.
15. **Jackson, M. L., T. W. M. Levelt, J. K. Syers, R. W. Rex, R. N. Clayton, G. D. Sherman, and G. Uehara.** 1971. Geomorphological relationships of tropospherically derived quartz in the soils of the Hawaiian Islands. Soil Sci. Soc. Am. Proc. 35:515–525.
16. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, New York, N.Y.
17. **Komatsoulis, G. A., and M. S. Waterman.** 1997. A new computational method for detection of chimeric 16S rRNA artifacts generated by PCR amplification from mixed bacterial populations. Appl. Environ. Microbiol. 63:2338–2346.
18. **Labarca, C., and K. Paigen.** 1980. A simple, rapid and sensitive DNA assay procedure. Anal. Biochem. 102:344–352.
19. **Levine, S. H.** 1976. Competitive interactions in ecosystems. Am. Nat. 110: 903–910.
20. **Liesack, W., and E. Stackebrandt.** 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. 174:5072–5078.
21. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The Ribosomal Database Project. Nucleic Acids Res. 24:82–85.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Merrill, J. T.** 1989. Atmospheric long-range transport to the Pacific Ocean, p. 15–50. *In* J. P. Riley and R. Chester (ed.), Chemical oceanography, vol. 10. Academic Press, New York, N.Y.
24. **Muyzer, G.** 1995. Protocols for the DGGE, p. 8–14. *In* K. Smalla and G. Muyzer (ed.), EU-Workshop on the Application of DGGE and TGGE in Microbial Ecology, Braunschweig, Germany. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig, Germany.
25. **Nannipieri, P., S. Greco, and B. Ceccanti.** 1990. Ecological significance of the biological activity in soil, p. 293–356. In J.-M. Bollag and G. Stotzky (ed.), Soil biochemistry, vol. 6. Marcel Dekker, Inc., New York, N.Y.
26. **Nübel, U., B. Engelen, A. Felske, J. Snaidr, A. Wieshuber, R. I. Amann, W. Ludwig, and H. Backhaus.** 1996. Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. J. Bacteriol. 178:5636–5643.
27. **Peck, T. R., D. Megel, K. Eik, D. A. Whitney, D. Warncke, R. C. Munter, J. R. Brown, D. Knudsen, W. C. Dahnke, D. J. Eckert, D. Beegle, P. E. Fixen, and E. E. Schulte.** 1988. Recommended chemical soil test procedure for the north central region. North Dakota Agricultural Experiment Station, North Dakota State University, Fargo.
28. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.
29. **Schmidt, T. M., B. Pace, and N. R. Pace.** 1991. Detection of DNA contamination in *Taq* polymerase. BioTechniques 11:176–177.
30. **Simberloff, D.** 1978. Use of rarefaction and related methods in ecology, p. 150–165. *In* K. L. Dickson, J. Cairns, Jr., and R. J. Livingston (ed.), Biological data in water pollution assessment: quantitative and statistical analyses. American Society for Testing and Materials, West Conshohocken, Pa.
31. **Soil Survey Staff.** 1972. Soil survey of the islands of Kauai, Oahu, Maui, Molokai, and Lanai, state of Hawaii. U.S. Government Printing Office, Washington, D.C.
32. **Sørheim, R., V. L. Torsvik, and J. Goksøyr.** 1989. Phenotypical divergences between populations of soil bacteria isolated on different media. Microb. Ecol. 17:181–192.
33. **Strunk, O., O. Gross, B. Reichel, M. May, S. Hermann, N. Stuckmann, B. Nonhoff, M. Lenke, A. Ginhart, T. Ludwig, A. Bode, K.-H. Schleifer, and W. Ludwig.** Arb: a software environment for sequence data. Nucleic Acids Res., in press.
34. **Swofford, D. L.** 1990. PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois Natural History Survey, Champaign.
35. **Tiedje, J. M., J.-Z. Zhou, K. Nüsslein, C. L. Moyer, and R. R. Fulthorpe.** 1997. Extent and patterns of soil microbial diversity, p. 35–41. *In* M. T. Martins et al. (ed.), Progress in microbial ecology. Brazilian Society for Microbiology, Sao Paulo, Brazil.
36. **Torsvik, V., J. Goksøyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56:782–787.
37. **Vitousek, P. M., D. R. Turner, and K. Kitayama.** 1995. Foliar nutrients during long-term soil development in Hawaiian montane rain forest. Ecology 76:712–720.
38. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1990. 16S rRNA sequences reveal numerous microorganisms in a natural community. Nature 345:63–65.
39. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature, p. 219–286. *In* K. C. Marshall (ed.), Advances in microbial ecology, vol. 12. Plenum Press, New York, N.Y.
40. **Weisblum, B., and E. Haenssler.** 1974. Fluorometric properties of the bisbenzimidazole derivative Hoechst 33258, a fluorescent probe specific for AT concentration in chromosomal DNA. Chromosoma 46:255–260.
41. **Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane.** 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173:697–703.
42. **Weisburg, W. G., J. G. Tully, D. L. Rose, J. P. Petzel, H. Oyaizu, D. Yang, L. Mandelco, J. Sechrest, T. G. Lawrence, J. Van Etten, J. Maniloff, and C. R. Woese.** 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. 171:6455–6467.
43. **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. Rapid method for the recovery of DNA from soils of diverse composition. Appl. Environ. Microbiol. 62: 316–322.
44. **Zhou, J., M. E. Davey, J. B. Figueras, E. Rivkina, D. Gilichinsky, and J. M. Tiedje.** 1997. Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. Microbiology 143:3913–3919.

CHAPTER 3

SOIL BACTERIAL COMMUNITY SHIFT LINKED TO CHANGE FROM FOREST TO PASTURE VEGETATION IN A TROPICAL SOIL

ABSTRACT

The change in vegetative cover of a Hawaiian soil from forest to pasture led to significant changes in the composition of the soil bacterial community. DNA was extracted from both soils and compared by the abundance of its guanine and cytosine (G+C) composition, by analysis of abundance of phylotypes of small-subunit ribosomal DNA (SSU rDNA) amplified from 63% and 35% G+C fractions, and by phylogenetic analysis of the dominant rDNA clones in the 63% G+C fraction. All three methods showed differences between sites that were greater than within sites, providing evidence that vegetation had a strong influence on microbial community composition at several levels of taxon resolution. The forest soil DNA had a peak in G+C content of 61% while the DNA of the pasture soil was shifted to higher G+C content with a peak at 67%. The phylotype abundance profiles of the 63% G+C fractions showed more dominance in the forest soil and greater phylotype richness at the pasture soil. None of the dominant phylotypes found in the forest soil could be detected in the pasture soil.

In the 63% G+C fraction sequence analysis of the three most dominant members revealed that their taxa changed from *Fibrobacter* and *Syntrophomonas* assemblages in the forest soil to *Burkholderia* and *Rhizobium-Agrobacterium* assemblages in the pasture soil.

INTRODUCTION

Dispersal of inoculum and selective forces for growth provided by the environment are the two main factors which determine which organisms dominate a habitat. Vegetation is one environmental factor thought to be a major determinant of the composition of the soil microbial community since it provides the primary resource for heterotrophic growth. Since different plant species are comprised of different carbon compounds, different microorganisms might be expected to grow in different plant communities. Furthermore, the plant may alter other physical and chemical features of the soil and hence favor the growth of different species. Little data exists, however, on the influence of vegetation on the composition of soil microbial communities.

We investigated the influence of vegetative cover on the structure and composition of a microbial community developed in a volcanic ash soil on the Big Island of Hawaii. Because of its young age and geographic isolation, Hawaii is depapurate in macrobiological species providing a simpler ecosystem for study. Two adjacent soils from the Kohala region were compared, one which has been continuously covered by a

native tropical forest, and the other covered by a grass pasture, which replaced the rain forest approximately 80 years ago. DNA based methods were used rather than culturing methods to compare the soil communities since they better recover the dominant prokaryotic populations. Furthermore, community DNA allows analysis at different levels of taxon resolution and provides more quantitative comparisons than culture based methods. The three levels of resolution used in order of increasing specificity and decreasing comprehensiveness are as follows. First, community DNA was fractionated by its guanine and cytosine (G+C) content and the amount of DNA of each G+C content quantified. Second, DNA fractions of two different G+C contents, one from a fraction of high biomass (63% G+C) and the other of low biomass (35% G+C) were then selected for analysis of small subunit ribosomal RNA (SSUrRNA) genes amplified by PCR. The rDNA clone libraries were screened by amplified ribosomal DNA restriction analysis (ARDRA) to determine pattern abundance profiles. Third, the dominant clones in the 63% biomass fraction from each soil were fully sequenced and analyzed phylogenetically. Together these three methods provide information at different levels of resolution on similarities and differences in the microbial communities from the two soils. Differences in the bacterial community due to vegetation were seen by all three methods.

MATERIALS & METHODS

Soil origin and sampling. The soil is located within the Kohala Forest Reserve on the Big Island of Hawaii (lat./long. 20°03'/155°41') and has never been cleared by humans. The forest site is the same as the Kohala site in Vitousek's chronosequence (Chapters 2 and 4). The pasture site was cleared in the 1920's for ranchland, and lies immediately adjacent to the Kohala rainforest. The distance between sampling sites was approximately 100m. The pasture was seeded with the African grass, *Pennisetum clandestinum,* which now dominates this species-poor site. Cattle graze the pasture.

The parent material of this Typic Placandept is a volcanic ash deposit, tephra, and is 150,000 years old. The sampling sites are at 1122 m elevation which corresponds to a mean annual air temperature of 16°C (Atlas of Hawaii, 1983). The mean annual rainfall is near 2500 mm (Giambelluca *et al.*, 1986). The pristine closed canope rainforest is mono- dominated (83% cover) by the native tree *Metrosideros polymorpha*, the most widespread tree on the Hawaiian islands (Kitayama *et al.*, 1996). The well developed understory vegetation is dominated by native tree ferns *Cibotium* spp. (25% cover). Other genera of trees and shrubs at this site are *Cheirodendron* (18% cover), the understory tree *Coprosoma* (4% cover), and the shrub *Vaccinium* (2% cover), account for most of the remaining cover (Crews *et al.*, 1995). The exotic plant species *Psidium cattleianum* and *Hedychium garnerianum* were also present in low abundance (Riley and Vitousek, 1995). The total number of vascular plant species per 0.2 ha is 53, while the

native plant species diversity is 45 (Crews *et al.*, 1995). This compares to about 140 vascular plant species per 0.2 ha in temperate forests (Huston, 1993).

Soil samples were taken from the upper 8-10 cm of the umbric A horizon after removing the overlaying organic root and litter layer, and placed immediately on wet ice. After 48 h samples were shipped on dry ice to our laboratory in Michigan and stored at -20°C. The soil moisture content was determined by drying soil overnight at 100°C. Soil mechanical and chemical analyses were done by the Soil Analysis Laboratory, Michigan State University, using the methods described in Peck *et al.* (1988).

DNA extraction and purification. DNA was extracted from 10 g of soil using the direct lysis method of Holben (1994) with the modifications previously described for the Hawaiian soils (Chapter 2). Extraction efficiency was determined by comparing the amount of DNA extracted with the amount of DNA expected calculated from the difference in direct microscopic counts of bacterial cells before and after lysis. DNA was quantified by fluorometry (Chapter 2; Glasel, 1995). The amount of extracellular DNA was determined following extraction with sodium phosphate before lysis (Chapter 4).

G+C fractionation technique. DNA fragments were separated according to G+C content following the procedure of Holben and Harris (1995). Briefly, DNA was mixed with the A-T specific dye bis-benzimidazole, and the bouyant density of the resulting DNA-bisbenzimide complex is decreased in proportion to the amount of dye bound. A gradient of %G+C concentration was then established by equilibrium density gradient ultracentrifugation. The amount of DNA in each fraction was determined by

absorption spectroscopy. DNA fractions were purified from the CsCl and the dye as previously described (Chapter 2) to allow PCR amplification. To verify the reliability of the G+C fractionation curves, repetitive analysis was done with two samples from the same soil sampling bag.

rRNA analysis. Small subunit rRNA (SSU rRNA) genes were amplified by PCR from purified soil DNA using the eubacteria specific primers fD1 and rP2 of Weisburg *et al.* (1991). PCR reactions were performed with *Taq* DNA polymerase (Boehringer Mannheim, Indianapolis, Ind.) according to the manufacturer's protocol and added quality control steps to ensure only soil bacterial rRNA genes were amplified (Chapter 2). Amplified products were separated in agarose gels, and the bands visualized by UV excitation after ethidium bromide staining. Prior to cloning, the amplified SSU rDNA fragments were purified by spin column chromatography (Wizard™ PCR Miniprep; Promega, Madison, Wis.), and an equimolar amount of amplified PCR products was ligated to the vector pCR™ II (Invitrogen Corp., Carlsbad, Cal.) and transformed into *Escherichia coli* Top-10F' competent cells. A primer pair specifically designed to complement the polylinker of the vector pCR™ II was used to amplify plasmid inserts directly from the transformant cells for SSU rDNA gene screening (Zhou *et al.*, 1997). To screen for SSU rDNA diversity, the amplified inserts were digested with two sets of tetrameric restriction endonucleases, *Hae*III and *Hha*I, *Msp*I and *Rsa*I (Chapter 2). The resulting fragments were electrophoretically resolved and the similarities between the electrophoretic patterns were analyzed using GelCompar™ (Applied Mathematics, Belgium)- cluster analyses by comparative

numerical analysis using UPGMA (Unweighted pair-group method using arithmetic averages). Individual clones were grouped using a cut off of 97% similarity and a 5% error rate for the band position. Each different pattern is termed a phylotype.

Determination of nucleotide sequences and phylogenetic analysis. Purified SSU rDNA clone inserts were sequenced in both directions with the fluorescent DiDeoxy™ termination method using automated fluorescent *Taq* cycle sequencing on the ABI Catalyst 800 and ABI 373A sequencing system (Applied Biosystems, Forester City, Calif.). The forward primer was LFP1 (Weisburg *et al.*, 1991), the reverse primer J529R (Weisburg *et al.*, 1991) which targets a conserved region (positions 515-529 in *E. coli* numbering). All sequences were aligned, and phylogenetic relationships were inferred (Chapter 2). Only unambiguously aligned nucleotide positions were used for the sequence analysis. Potential chimeric artifacts were evaluated by CHECK_CHIMERA (Maidak *et al.*, 1996), and mglobalCHI (Komatsoulis and Waterman, 1997). Furthermore, any ARDRA restriction pattern seen more than once is strong evidence against that clone pattern being from a chimera.

RESULTS

Changes in soil characteristics. The shift in vegetative cover to pasture caused some changes in soil properties including a decrease in soil acidity and organic carbon, and an increase in bulk density (Table1). The higher cation exchange capacity in the forest soil is likely due to its higher soil organic matter content.

G+C profiles. The G+C profile of soil DNA was shifted to significantly higher G+C DNA with the change from forest to pasture (Fig. 1). The majority of soil DNA was in the 55% to 70% G+C range for both soils but the major peak was 61% G+C for the forest soil DNA and 67% G+C for the pasture soil. The pasture soil community exhibited an additional peak in the 42 to 45% G+C range, and displayed a small peak of 71% G+C. DNA extracted from replicate soil samples showed repeated profiles such that within site variation was much less than differences between the two sites. The standard deviations of the mean curve differences for the forest and pasture soil communities were small, $s = 0.14$ and $s = 0.27$, respectively (Fig. 1).

Phylotype replication within and between sites. The same phylotypes were observed from replicate soil samples for the three dominant phylotypes in seven of the nine cases for the forest and the pasture soils (Fig.2, dots). Examples of ARDRA data documenting the same phylotypes in replicates of the pasture soil as well as different phylotyes are shown in Figure 3. There was no overlap in dominant phylotypes between forest and pasture samples, suggesting that the populations were different between the two sites. Non dominant phylotypes were very different even between

Table 1. Differences in soil properties between the Kohala forest and pasture soils

Soil parameter	Forest soil	Pasture soil
Soil classification	Typic Placandept	Typic Placandept
pH	4.2	5.2
Total carbon (%)	13.5	8.1
Total nitrogen (%)	1.5	ND
NO_3^--N (μg/g)	2.2	5.6
NH_4^+-N (μg/g)	83.3	22.4
CEC (meq/100 g)	89.3	51.9
Soil bulk density (g/cm^3)	0.498[a]	0.587
Dominating plant	*Metrosideros polymorpha*[b]	*Pennisetum clandestinum*[c]

[a] Data from Crews *et al.*, 1995
[b] the overstory is dominated by *Metrosideros polymorpha*, a native C-3 tree
[c] the plant *Pennisetum clandestinum* is an African C-4 grass

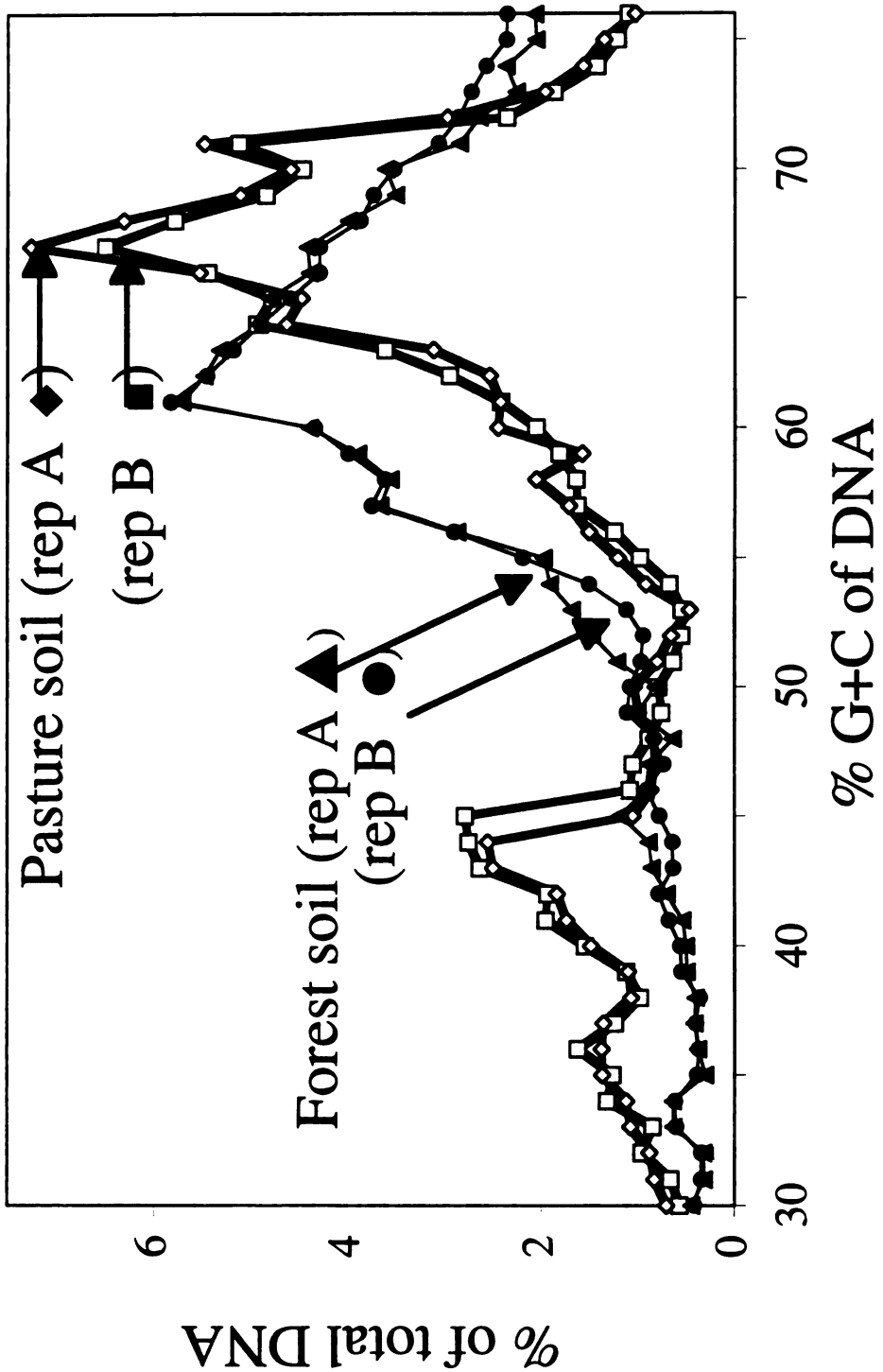


Figure 1. Comparison of G+C profiles of DNA extracted from two adjacent soils in the Kohala area. One soil supports a rainforest and the adjacent soil was converted to pasture 80 years ago. Profiles of DNA extracted from replicate samples at the same site are shown

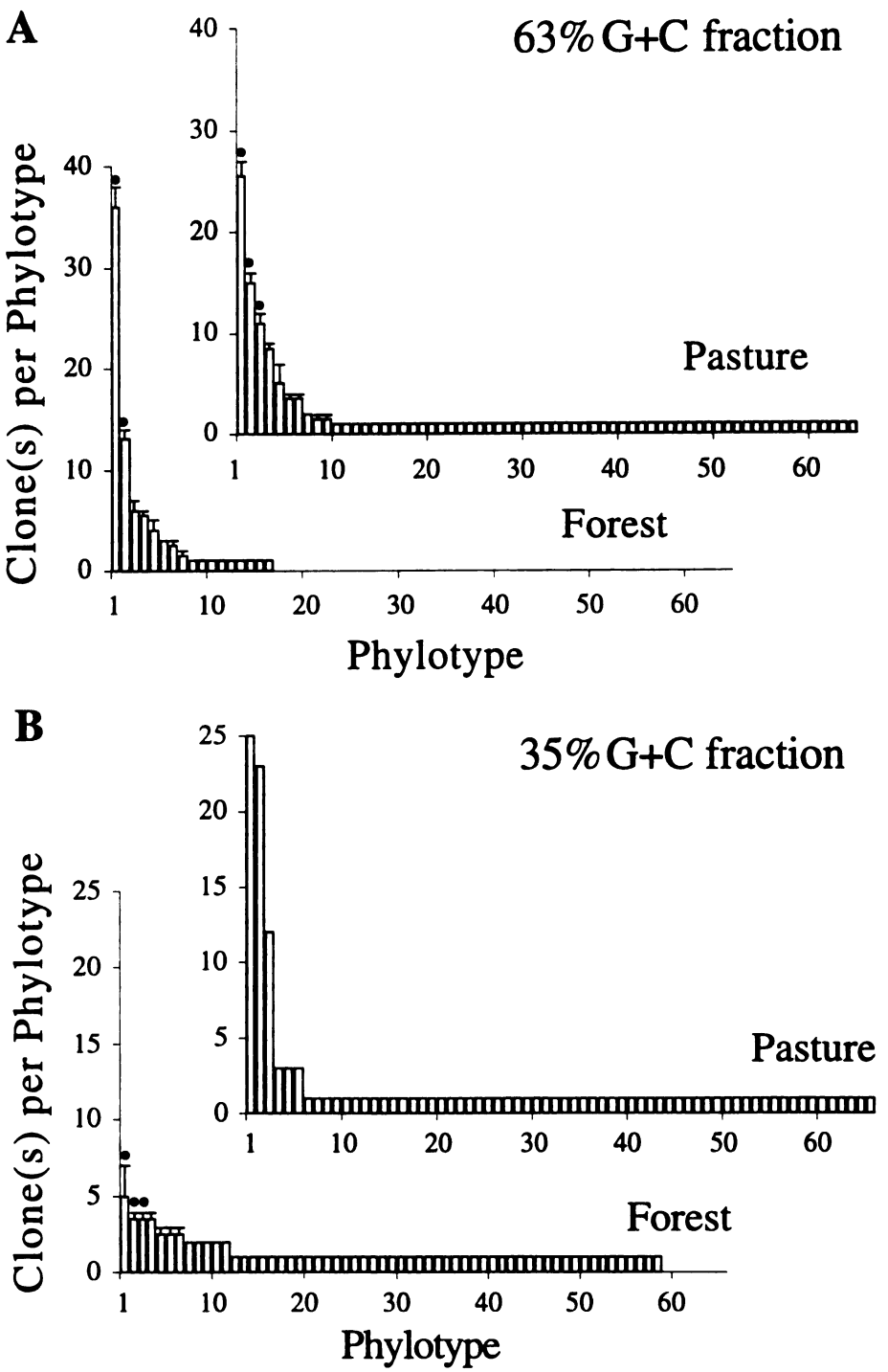


Figure 2. Frequency distribution of ssu rdna gene phylotypes from the 63% G+C fractions (A) and the 35% G+C fractions (B) from the rainforest and pasture soils. Each graph shows the median and range (bars) of data of two replicates. Solid dots (•) are placed above the phylotype which were identical in pattern and rank for the two replicates. A replicate of the 35% G+C pasture sample was not available.

replicates of the same vegetation type which lends support to the assumption that the establishment of a clone library is a random rather than a biased process. Patterns of rare clones (n=1) within the replicates of the 63% G+C pasture community overlapped by 22% while the rare clones of the corresponding forest community replicates overlapped by 12%. Only seven of the rare phylotypes overlapped between the two vegetation types. The two most dominant patterns of the 35% G+C fraction of the forest had very low frequencies of only 6.5(±2.5)% and 4.5(±0.5)%, but were found in both replicates. There was only 4% overlap for the remaining patterns within the forest replicates.

The phylotype abundance curves (Fig.2) of both the forest and pasture soil replicates could be repeated successfully with very low standard deviation. The two profiles of the 63% fraction of the forest soil were similar, *i.e.* 94% of all data fell within a single standard deviation range of ±1.17 of the mean curve difference. The corresponding numbers for the pasture profiles at 63% were 93% at ±0.77, and for the 35% G+C forest samples 98% at ±0.61.

Changes in the ARDRA community profile. Microbial rDNA diversity in the 63% G+C fraction differed between the pasture and the forest soil. The phylotype and nucleotide sequence analysis showed that the dominant strains had shifted to different, unrelated taxa. The rank abundance profiles of the 63% G+C fractions are separated by the stronger trend for dominance in the forest soil, but the much greater phylotype richness in the pasture soil (Fig. 2). The four most dominant members contribute 77%(±6) of the forest soil community while they represent about half

Table 2. Phylogenetic affiliations based on SSU rDNA genes of the three most dominant phylotypes of the two replicates from the 63% G+C fractions of the rainforest soil (HRS) and the pasture soil (HPS).

Phylotype	Replicate	Relative abundance[a] (%)	Phylogenetic affiliation[b]	Most closely related organism in RDP database	%Similarity
Forest soil					
HRS-46	A	43	*Fibrobacter* Phylum	env. str. MC 9[c]	70.2
HRS-55	A	15	*Fibrobacter* Phylum	*Acidobacterium capsulatum*[c]	65.4
HRS-60	A	9	*Syntrophomonas* assemblage	*Desulfotomaculum geothermicum*[d]	41.6
HRS-47	B	48	*Fibrobacter* Phylum	env. str. MC26[c]	73.2
HRS-56	B	18	*Fibrobacter* Phylum	*Acidobacterium capsulatum*[c]	67.0
HRS-50	B	6	*Fibrobacter* Phylum	env. str. MC 101[c]	72.6
Pasture soil					
HPS-54	A	21	Beta Purple	*Burkholderia pseudomallei*[e]	70.5
HPS-61	A	11	Beta Purple	*Burkholderia pseudomallei*[e]	68.0
HPS-49	A	8	Alpha Purple	*Bradyrhizobium japonicum*[f]	74.3
HPS-52	B	19	Beta Purple	*Burkholderia pseudomallei*[e]	77.1
HPS-45	B	13	Beta Purple	*Burkholderia pseudomallei*[e]	63.2
HPS-64	B	9	Alpha Purple	*Nitrobacter winogradskyi*[f]	68.0

[a] Relative abundance of the clones belonging to a phylotype, calculated by dividing the number of clones belonging to the phylotype by the total number of clones analyzed.

[b] Only unambiguously aligned regions were used for the analysis.

[c] RDP subdivision: *Acidobacterium* Subdivision

[d] RDP subdivision: *Syntrophomonas* Group

[e] RDP subdivision: *Burkholderia* Subgroup

[f] RDP subdivision: *Rhizobium-Agrobacterium* Group

(47%[±3]) of the pasture soil community. With 62(±4) phylotypes the pasture had a phylotype richness which was almost 2.2(±0.4) times higher than for the forest soil. The difference in abundance within the four most dominant phylotypes is more prominent in the forest soil community. Clones of the first phylotype are almost three times as abundant as the next group while the respective difference in the pasture clones is under 60% (Fig. 2A).

The pasture soil displayed a prominent dominance structure of the first three phylotypes from the 35% G+C fraction (Fig. 2B), while both replicates of the forest soil show no significant dominance pattern with no phylotype frequency over 10%. The three most dominant members represent 47% of the pasture community, and only about one third (15%[±3]) of the forest soil community. The phylotype richness of the 35% G+C fraction was about 43%(±4) higher than for the forest soil.

rDNA sequence analysis. One clone of each of the three dominant phylotypes was sequenced. Most clones sequenced were only distantly related to known sequences (Table 2). The dominant clones from the two vegetation types were very different representing two very different phyla, *Fibrobacter* and *Proteobacter*. Many of the phyla within the *Fibrobacter* group were dissimilar from each other (Table 3) while those within the *Proteobacteria* were affiliated with *Burkholderia* and more similar to each other (Tables 2 and 3).

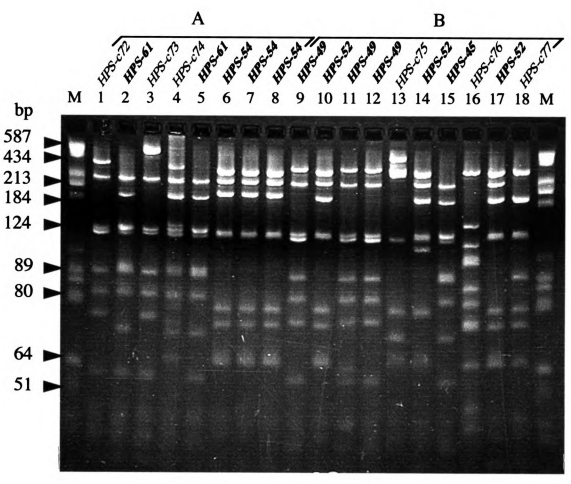


Figure 3. An ARDRA profile following double digestion with *Hae*III and *Hha*I of clones form the 63% fraction of the pasture soil. Similar patterns can be seen for three groups of clones, each found in both replicates: lanes 2,5,15; 6,7,8,10,14,17, and 9,11,12. Lanes 1 to 8 are from replicate A and lanes 9 to 18 from replicate B. Plasmid p*BR*322 digested with *Hae*III was used as a size marker (M).

Table 3. Evolutionary distances among SSU rDNA clones from dominant members of the 63%G+C soil bacterial community from the rainforest and pasture soils. The numbers represent percent SSU rDNA nucleotide sequence similarity.

A. *Fibrobacter / Acidobacterium* group

Clone[a]	Evolutionary distance for clone number[a]:				
	1	2	3	4	5
1. HRS-46	-				
2. HRS-47	89.6	-			
3. HRS-50	90.0	82.3	-		
4. HRS-55	75.7	70.0	78.5	-	
5. HRS-56	77.6	72.2	79.2	98.7	-

B. *Burkholderia* group

Clone[a]	Evolutionary distance for clone number[a]:			
	1	2	3	4
1. HPS-45	-			
2. HPS-52	99.6	-		
3. HPS-54	98.0	99.0	-	
4. HPS-61	97.0	97.0	97.2	-

[a] Organisms are represented by respective numbers in header.

detect

used

comm

betwe

analy z

showe

sites.

soil m

commu

The G

compre

but can

they fo

increase

bacteria

1920's,

the repla

DISCUSSION

The high microbial diversity even in this young, isolated Hawaiian soil makes the detection and analysis of community components challenging. The molecular methods used here, however, were able to distinguish clear differences in the microbial community between the forest and pasture soil.

Because of the high microbial diversity in soil, distinguishing statistically between inter *versus* intra sites is often very difficult. In this study replicate samples analyzed by G+C profile, ARDRA abundance patterns and rDNA sequence analysis all showed greater reproducible intra site results and significantly different results between sites. Hence, these methods may generally be useful to distinguish differences between soil microbial communities. It is also important that these methods analyze the community at different levels of resolution and different levels of comprehensiveness. The G+C method provides an analysis at a coarser level of taxon distinction but is comprehensive for *all* DNA, while the rDNA analysis is effective to nearly species level but can only be done for amplifiable and selected dominant members. Hence, together they form a complementary suite to better analyze community differences.

The change in vegetation type altered bacterial community composition and increased diversity at least as revealed by the methods used here. The shift in the soil bacterial community could be a result of either the disturbance of the clear cutting in the 1920's, or the switch from forest to pasture, or both. A radical change in vegetation like the replacement of a rainforest with a pasture certainly constitutes a major shift of the

soil en

charge

(Table

et al.

forest s

vegetat

commu

Hawaii

4). Thi

but not

in the 6

develop

the char

not dis

although

harmon

commu

young

develop

unusual

competi

of phylo

soil environment, especially in carbon sources for the microbes, and brings along a change in soil properties like a decrease in soil acidity and an increase in bulk density (Table 1) similar to those found in other ecological studies (Reiners *et al.*, 1994; Spaans *et al.*, 1989). Furthermore, cattle graze in the pasture site while wild pigs disturb the forest soil site. Hence there are both primary and secondary factors associated with the vegetation shift that could be important determinants of the current microbial communities. For my study of the chronosequence of similar soils on the big island of Hawaii, I found that strong rDNA dominance patterns developed with soil age (Chapter 4). This dominance pattern was found in the 150,000 year old forest soil investigated, but not in the 20,000 year or younger forest soils. However, the pattern of dominance in the 63% G+C fraction of the pasture, which was established only eighty years ago, developed a shape similar to the forest community (Fig. 2A). I therefore conclude that the change in vegetation is most likely the factor responsible for the community to shift, not disturbance. The dominance pattern in these fractions retain a similar shape, although the decline in the first seven phylotypes of the pasture community is more harmonic than the respective group from the forest profile. Perhaps the older forest community has a more established system of specialists with defined niches, while the young pasture soil has a successional competition profile which is typical for lesser developed systems. However, the number of rare members in the pasture community is unusually high compared to macroecological systems, which indicates no or limited competition among soil bacteria (Zhou *et al.* , 1997). The significantly increased number of phylotypes in the 63% G+C fraction of the pasture soil suggests that the change in

vegetation type limited competition between the community members by opening more niches to an already highly diverse community (Fig. 2A). Hence, I reject the assumption of successional competition.

This proposed change in niches conceivably caused the observed change of dominant phyla as inferred from rDNA sequences in the 63% G+C fraction. The dominant taxa were from the *Fibrobacter* phylum in the forest soil and from the *Proteobacteria* in the pasture soil. A statement about the change in the phylotypes amongst the rare members cannot be made due to the extremely high bacterial diversity found in soils. I assume a random selection as the underlying principle of our clone libraries, hence any mixture of soil bacterial SSU rRNA clones can make up this selection. A supporting argument to the niche addition theory is that community diversity increases even though the relative amount of contributing community DNA is lowered from the forest to the pasture soil (Fig. 1).

The application of molecular tools to analyze the soil bacterial community described here enabled us not only to look at bacterial community profiles at several levels of resolution, but also to investigate the dominant as well as some rare members within these communities. The results support the common notion that terrestrial environments exhibit enormous diversity, and contain novel, uncultivated organisms. A more refined picture of the phylotype abundance distribution within this community could be gained by complete sequence analysis of all SSU rDNA inserts in the clone libraries, but this is too costly. Furthermore, potential biases would always limit the interpretation. To minimize bias, I have used the best available methods at present as critically as possible.

CHAPTER 4

SOIL MICROBIAL COMMUNITY DEVELOPMENT AS INFLUENCED BY SOIL AGE

ABSTRACT

Hawaiian soil sites were studied since these soils and ecosystems were reset by volcanic eruptions to early stages of development, and because these soils of ages 200 to 150,000 years are underlaid by similar substrata, and influenced by similar climates and biotic communities. Total microbial community DNA was extracted and fractionated based on its G+C content. G+C profiles were similar for soils of different age except for a slight decrease in low G+C content DNA in the oldest two soils. SSU rDNA gene libraries were prepared from the 63% G+C fraction, which represents a dominant biomass fraction, and from a 35% G+C fraction, which represents a minor biomass fraction. The two major changes in community structure along the chronosequence are the development of a dominance structure, and the decrease in phylotype diversity from the youngest soils (200, 2,100, 20,000 yrs) to the oldest site (150,000 yrs). Maximum phylotype richness within the dominant biomass fraction occurs at the mid-aged Laupahoehoe site (20,000 yrs), which was confirmed by both the maximum contribution

of high biomass DNA in the total DNA community profile at this site and the maximum in phylotype richness in the 63% G+C rank-abundance pattern which was even more evident by their rarefaction curves. This maximal microbial diversity at intermediate soil age is correlated with extreme values in several soil physical and chemical parameters, especially plant available nitrogen, and soil bulk density. The findings are consistent with the macroecological principle of diversity responses to nutrient availability. All major changes with age described occurred in the fraction of high biomass (63% G+C), whereas no statistically significant changes could be found in the 35% G+C fraction. Dominant phylotypes were found again in repeated experiments of the same soil sample. However, dominant phylotypes were generally different between each site along the chronosequence. These results indicate that variability within a soil site of a particular age is not greater than between different soils. All of the soil clones sequenced affiliate with the *Eubacteria* as was expected due to the specificity of the primer pair used. None of the nucleotide sequences analyzed in this study matched known sequences from any cultured species, or were identical to sequences in the databases. Several of the most dominant clones belong to the clone cluster VI (*Fibrobacter* phylum) defined by Stackebrandt *et al.* (1993), which forms an individual line of descent and is now part of a new phylogenetic division. This novel group was more frequent at sites older than 2,100 yrs.

INTRODUCTION

Soil microbial diversity is thought to be extremely high even at a small scale (Torsvik *et al.*, 1990). However, the magnitude of this diversity remains unknown (Liesack and Stackebrandt, 1992; Borneman *et al.* 1997). In addition, very little is known about the environmental parameters that link ecological processes to that diversity, and could influence the composition of the microbial community. Although important microbial processes might occur at the small scale of microbial communities, the great abundance of these communities in the soil extends their importance to the larger ecosystem scale. Conversely, changes in parameters important at the ecosystem level (*e.g.*, soil age, temperature, moisture, vegetation) may have significant influences on microbial communities, in particular on community structure and composition. These microbial community changes may in turn lead to changes in diversity and activity of larger organisms. Without insight into soil biodiversity, models of ecosystem functioning and prediction of human impacts on changing ecosystem and global patterns may not be understood.

Recent studies of soil bacterial diversity not only confirmed the enormous variety of microbes in soil but also corroborated the limitations of common diversity studies. Use of the entire soil bacterial DNA as the target of study will only reveal the most well represented species (Liesack *et al.*, 1992; Borneman *et al.*, 1996). Macroecologocial studies have long focused on certain groups of organisms, and not the entire macroecological community. Consistent with this practice, I limited the target

group in this microecological project to only parts of the soil bacterial community, namely a fraction of soil extracted DNA with high biomass, and a fraction of low biomass. Since cultivating methods only recover about 1% of all soil microorganisms, molecular approaches were employed because they more comprehensively sample the soil community and identify its most dominant members.

The DNA was separated according to its guanine and cytosine content (%G+C) which provides a coarse level of separation of prokaryotic taxa (Holben and Harris, 1995). The fraction of high bacterial biomass selected had 63% G+C, and the fraction of low biomass had 35% G+C. These DNA fractions were used as a template to generate gene libraries of small subunit ribosomal DNA (SSU rDNA) by PCR amplification. Community structure was established by determining the diversity and abundance of SSU rDNA clone types which were distinguished by digestion with two pairs of tetrameric restriction endonucleases. Because the SSU rRNA gene contains diagnostic variable regions that are unique to particular organisms, the digestion of the amplified SSU rDNA gene with a pair of tetrameric restriction endonucleases results in mostly unique restriction length fragment polymorphism (RFLP) patterns. Since these RFLP patterns are indicative of the phylogeny of the SSU rDNA gene, they can be used as operational taxonomic units. For this study we use the term phylotype, which is defined as the pattern resulting from RFLP analysis of the SSU rDNA.

The purpose of this research was to determine how patterns of soil bacterial community structure and composition change with the age and development of soil. I studied the soils on the Big Island of Hawaii since they are the youngest and most

geographically isolated soils and hence changes in microbial community development might be more easily seen. The soil was built from periodic volcanic eruptions, each of which resets soil and ecosystem development. I sampled the four youngest soils of a recently described chronosequence across the Hawaiian Islands, *i.e.* soil ages ranging from 200 years to 150,000 years (Crews *et al.*, 1995). Most importantly, many other environmental variables were almost constant for the sites sampled. The parent material shows little or no chemical variation (Wright *et al.*, 1987), the precipitation and the temperature are nearly constant throughout the year, and the discontinuous climatic effects of pleistocene glaciation are minimized (Vitousek *et al.*, 1997). All sites support intact rain forest vegetation dominated by the same tree species. By choosing these young and geographically isolated Islands as my study site, I attempted to reduce the extent of microbial diversity to a more manageable level. Their -- compared to older soils -- lesser developed ecosystems are reflected in a low level of macrospecies richness (Crews *et al.*, 1995).

MATERIALS & METHODS

Soil origin and soil sampling. The chronosequence consists of four soils located on the Big Island of Hawaii, ranging in age of the volcanic parent material from 200 to 150,000 years (Table 1, Fig. 1). Since these soils are matched in elevation, precipitation, slope position, and disturbance history they form a sequence that varies

mainly in one parameter, soil age (Crews *et al.*, 1995). The parent material of all soils is basaltic tephra, a volcanic ash deposit (Clague and Dalrymple, 1987). All of the sites are on minimal slopes (<6%) near 1200 m elevation, with 16°C mean annual air temperature with little seasonal variation (Department of Geography, University of Hawaii, 1983). The sites currently receive approximately 2500 mm of precipitation annually (Giambelluca *et al.*, 1986), which is well distributed throughout the year due to the relatively constant NE trade winds (Carlquist *et al.*, 1980). Soil thickness remains constant at about 1 m. Each site is located under undisturbed, intact montane rainforest with an unusually low level of species richness. The overstory of all sites is dominated by the most widespread tree on the Hawaiian islands (Kitayama *et al.*, 1995), *Metrosideros polymorpha*, the native tree which increases in height from the youngest site to the mid-aged Laupahoehoe site, to decline by around 50% at the oldest site, Kohala (Crews *et al.*, 1995). The well developed understory vegetation is dominated by native tree ferns, *Cibotium* spp. Other trees and shrubs that account for most of the remaining cover are *Cheirodendron, Ilex, Coprosoma* and *Vaccinium* (Crews *et al.* 1995). The exotic plant species *Psidium cattleianum* and *Hedychium garnerianum* were also present in low abundance (Riley and Vitousek, 1995). The total number of vascular plant species per 0.2 ha was lowest at the youngest site with 34 species, and increased with soil age to 47, 48 and then 53 species, respectively (Crews *et al.*, 1995). As is common for these montane areas of high precipitation, no nitrogen-fixing vascular plant-microbe symbioses was found within any of the sites (Vitousek and Walker, 1989;

Table 1. Site descriptions and soil properties of sampling locations

Soil site	Approx. soil age (yr)	Location (N/W)	Soil classification	Elevation (m)[a]	Moisture content (%)	Organic matter (%)	pH ($CaCl_2$)	% N	C/N ratio	Soil bulk density[b] (g/cm^3)
Thurston	200	19°25'/155°15'	Hydric Dystrandept	1176	59	3.1	5.5	0.08	15.5	0.69
Ola'a	2,100	19°28'/155°15'	Typic Hydrandept	1200	55	7.1	5.4	0.19	14.9	0.61
Laupahoehoe	20,000	19°46'/155°16'	Typic Hydrandept	1170	65	42.7	4.8	0.94	18.2	0.37
Kohala	150,000	20°03'/155°41'	Typic Placandept	1122	67	40.3	4.4	0.77	20.9	0.50

[a] Mean Annual Temperature and Mean Annual Precipitation were the same for all sites, approximately 16°C and 2500 mm, respectively. The substrate is tephra.
[b] Calculated based on data by Crews *et al.* (1995)

Crews *et al.* 1995). None of the sites is known to have been cleared by humans (Crews *et al.*, 1995).

The youngest site is located near Thurston Lava Tube on the east slope of Kilauea Volcano within Volcano National Park (Fig. 1). The site is fenced to keep feral pigs (*Sus scrofa*) out of the Park. This soil is a Hydric Dystrandept developed on several tephra (volcanic ash) depositions ranging in age from 200 years at the top mineral layer to 400 years, and totaling approximately 38 cm in depth (Soil Survey Staff, 1973). The sample analyzed was from the 200 year old layer. The next older soil is located in Ola'a Forest on the north slope of Kilauea volcano. This soil is a Typic Hydrandept originating from tephra deposits about 2,100 yr old (Soil Survey Staff, 1973). The 20,000 yr old Laupahoehoe soil is also a Typic Hydrandept, and is located on the north east slope of Mauna Kea (Soil Survey Staff, 1973). The 150,000 year old soil is a Typic Placandept (Soil Survey Staff, 1973), and is from the east slope of Kohala volcano.

I removed the litter and the first centimeter of soil, and sampled the next 7.5 cm of soil. The samples were collected in the perimeter of Crews *et al.* (1995) main study site. Duplicate soil samples of 750 g each were packaged on site in sterile polypropylene bags and immediately put on ice. The next day they were placed in dry ice coolers and shipped by express mail to Michigan where they were stored at - 20°C. The soil moisture content was determined by drying soil overnight at 100°C. Soil mechanical and chemical analyses were done by the Soil Analysis Laboratory, Michigan State University, using the methods described in Peck *et al.* (1988). The two oldest

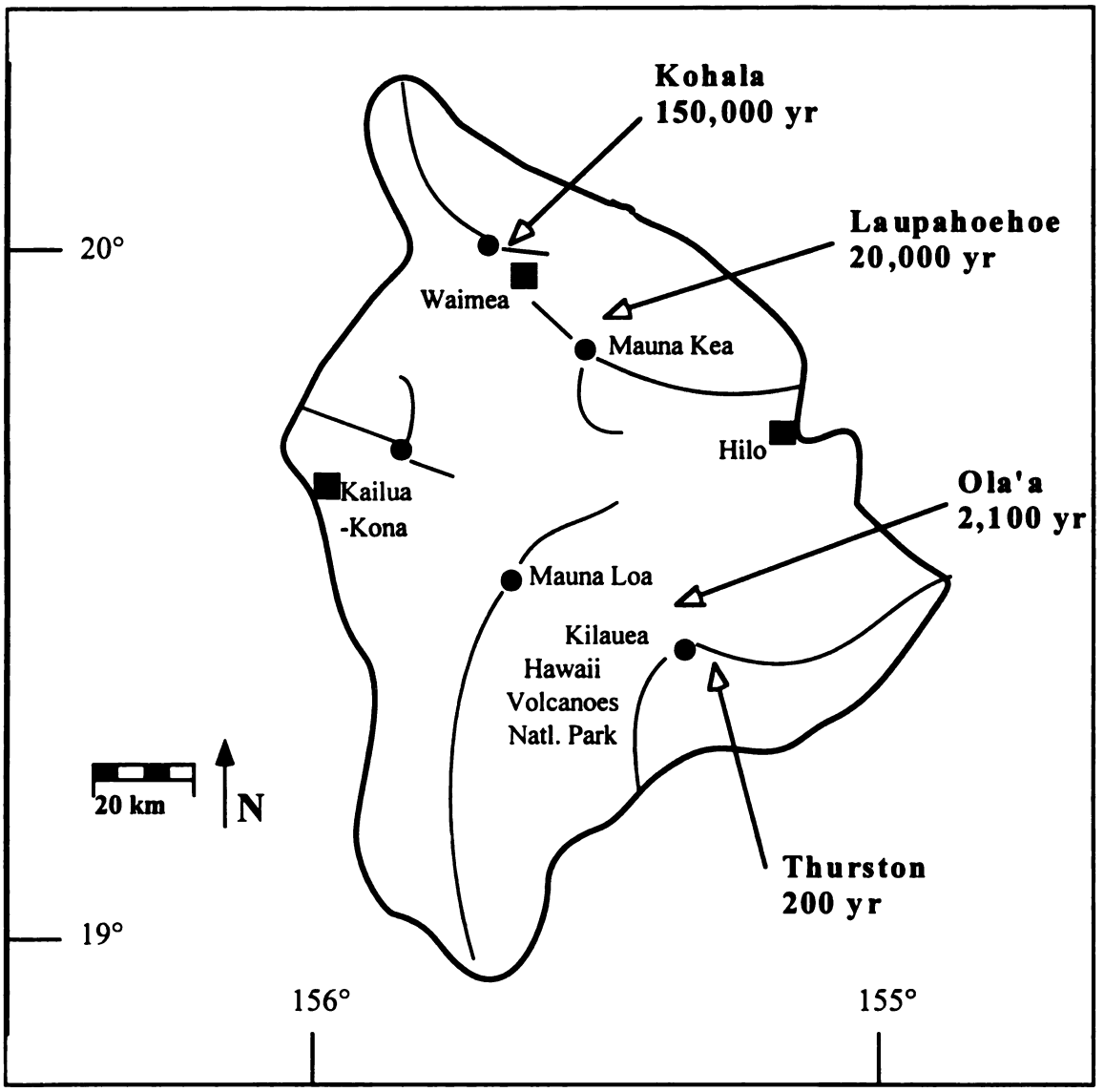


Figure 1. Map showing the location of the four soil sites of the chronosequence on the Big Island of Hawaii, HI, USA.

soils, Kohala and Laupahoehoe were sampled in 1992 and in 1994. To verify low variability of the G+C fractionation curves and of the clone library, repetitive analyses were done with two samples of the same soil sampling bag for the 1992 Kohala samples, and compared with the 1994 Kohala sample collected in the same area. To determine spatial variation within a sampling site, three samples were collected at the Laupahoehoe site that were 3 m apart from each other.

DNA extraction and purification. Soil microbial DNA was extracted from 10 g of soil using the direct lysis method of Holben (1994) with several modifications as described in Chapter 2. Subsequent DNA purification was modified to include agarose gel purification and a single microcolumn passage of the excised and melted gel piece. Extraction efficiency was determined by comparing the amount of DNA extracted with the amount of DNA expected calculated from the difference in direct microscopic counts of bacterial cells before and after lysis. Bacterial cells were counted directly using computer aided microscopic counting procedures (Zhou *et al.* 1996). For consistency, all counts were obtained by a single investigator. Details of the methods in this paragraph are described in the Material and Methods section of Chapter 2.

Extraction of extracellular DNA from soil. In order to determine the amount of extracellular DNA, 5 g of fresh soil (wet weight) was sequentially extracted four times with 200 ml of 0.12 M sodium phosphate buffer at pH 8.0. DNA was removed directly from the resulting extract by ethanol precipitation, and subsequently purified by spin columns with a pore size allowing a molecular weight cut off of over 100K (Schleicher

and Schuell). Ogram *et al.* (1987) showed that this extraction procedure removes 99.9% of previously added, radio-labeled DNA.

G+C fractionation technique. DNA fragments were separated according to G+C content following the procedure of Holben and Harris (1994). Briefly, DNA was mixed with the A-T specific binding dye bis-benzimidazole, and the bouyant density of the resulting DNA-bisbenzimide complex decreases in proportion to the amount of dye bound. A gradient of %G+C concentration was then established by equilibrium density gradient ultracentrifugation, and fractionated. To make PCR amplification possible DNA fractions were purified as described in Chapter 2.

PCR amplification of small subunit rRNA genes from soil DNA. Small subunit rRNA (SSU rRNA) genes were amplified by PCR from purified soil DNA using the eubacteria specific primers fD1 and rP2 of Weisburg *et al.* (1991). Details are described in the Material and Methods section of Chapter 2.

Analysis of SSU rDNA clone library, determination of nucleotide sequences and phylogenetic analysis. The amplified SSU rDNA fragments were purified, an equimolar amount was ligated to the vector pCR™ II (Invitrogen, San Diego, Calif.), and the vector was transformed into *Escherichia coli* Top-10F' competent cells (Invitrogen, San Diego, Calif.). A primer pair specifically designed to complement the polylinker of the vector pCR™ II (44) was used to amplify plasmid inserts directly from the transformant cells for SSU rDNA gene screening. Clones with identical restriction patterns were digested with two additional tetrameric endonuclease restriction enzymes. Cluster analyses were performed by comparative numerical

analysis and rank-abundance distribution patterns were established. The amplified SSU rDNA clone inserts of the three most dominant phylotypes in each clone library were purified, and partially sequenced at their 5' end using automated fluorescent *Taq* cycle sequencing. For the detection of potential chimeric artifacts partial sequences that were around 400 bases long were submitted to the CHECK_CHIMERA program offered by RDP (Maidak *et al.*, 1996). All sequences were aligned, and phylogenetic relationships were inferred. Details of these steps are described in the Material and Methods section of Chapter 2.

RESULTS

DNA extraction. High efficiency in DNA extraction is crucial to the representativeness of the underlying soil bacterial community. I calculated the percent of cells lysed using DTAF [5-(4,6-dichlorotriazin-2-yl)amino fluorescein)] stained cell counts, since DTAF counts commonly yield lower bacterial counts than the corresponding acridine orange count (Suzuki *et al.*, 1993). Acridine orange is more likely to nonspecifically stain organic particles, and these tropical soils are rich in organic matter (Table 1). After optimizing lysis techniques I achieved a mean DNA extraction efficiency of 88(±16)% (Table 2) based on loss of direct microscopic counts. Free (extracellular) DNA which could add bias to the depiction of the soil bacterial community does not appear to influence the amount of total community DNA extracted

Table 2. Plate counts, direct counts, cell lysis efficiency and DNA yields for individual samples of five soils of culturable heterotrophic bacteria

Site	Soil Age (yr)	Prelysis count[a] (CFU (±SD) /g dw)	Direct count[b] (cells /g dw)	Cell lysis efficiency[c] (%)	DNA yield expected[d] (µg/g dw)	DNA yield (µg/g dw)	DNA yield vs. expected yield (%)
Thurston	200	(2.4±0.2) x10^8	(1.6±0.3) x 10^9	91(±3)	5.8	6.4	110
Ola'a	2,100	(7.3±0.8) x10^8	(5.4±0.6) x 10^9	89(±4)	19.2	14.1	73
Laupahoehoe	20,000	(6.2±0.3) x10^8	(4.4±0.2) x 10^9	77(±8)	13.6	11.4	84
Kohala	150,000	(1.1±0.3) x10^8	(7.2±0.5) x 10^9	82(±3)	23.6	19.8	84

[a] Culturable bacteria on nonselective medium R2A (Difco, Detroit, Michigan) computed per gram dry weight (dw) of soil (mean counts ± standard deviation; two plates per sample)
[b] Microscopic evaluation of total bacterial numbers (mean count ± standard deviation; three smears per sample)
[c] Calculated as [1-(Direct count after lysis / Direct count before lysis)] x 100
[d] Expected DNA yield based on literature estimates of theoretical bacterial DNA content assumed to be 4 x 10-15 g DNA per cell (Ellenbroek and Cappenberg, 1991)

by the direct lysis protocol (Table 3). Initial extraction of DNA prior to lysis in the oldest and the youngest soil did not recover detectable extracellular DNA. An attempt to amplify the extract by PCR yielded no product, even after purification of the extract. In agarose gels the fragment size composition of soil extracted DNA was predominantly of high molecular weight (*i.e.*, located above the 23 Kb marker) with only a faint signal in the <23 Kb size range (data not shown).

Total community DNA profiles. The relative abundance of DNA of different base composition (% G+C) was used to produce profiles of the community composition, as well as to fractionate the DNA. The % G+C community profiles for a site are reproducible as judged by the profiles from three replicate samples from one site (Fig. 2). The community as observed by the level of G+C resolution appeared to be stable over two years since the 1992 and 1994 profiles were similar (the standard deviations (s) of the mean curve were s = 0.14 for two replicates from the same soil sampling bag collected in 1992, and $2s$ = 0.40 between the two 1992 samples and the 1994 sample). High-molecular weight DNA of a limited size range was observed by agarose gel electrophoresis (data not shown) showing that the DNA separation in the gradient was not subject to potential bias of the varying molecular weight of randomly sheared DNA.

The G+C profiles of total soil DNA for all sites along the chronosequence is depicted in Figure 3. The profiles were standardized against DNA mixtures of known composition (Fig. 3, insert). Each profile had a distinct maximum between 55% and 70% G+C with a maximum of 63% (±2) G+C. This structure is typical also for

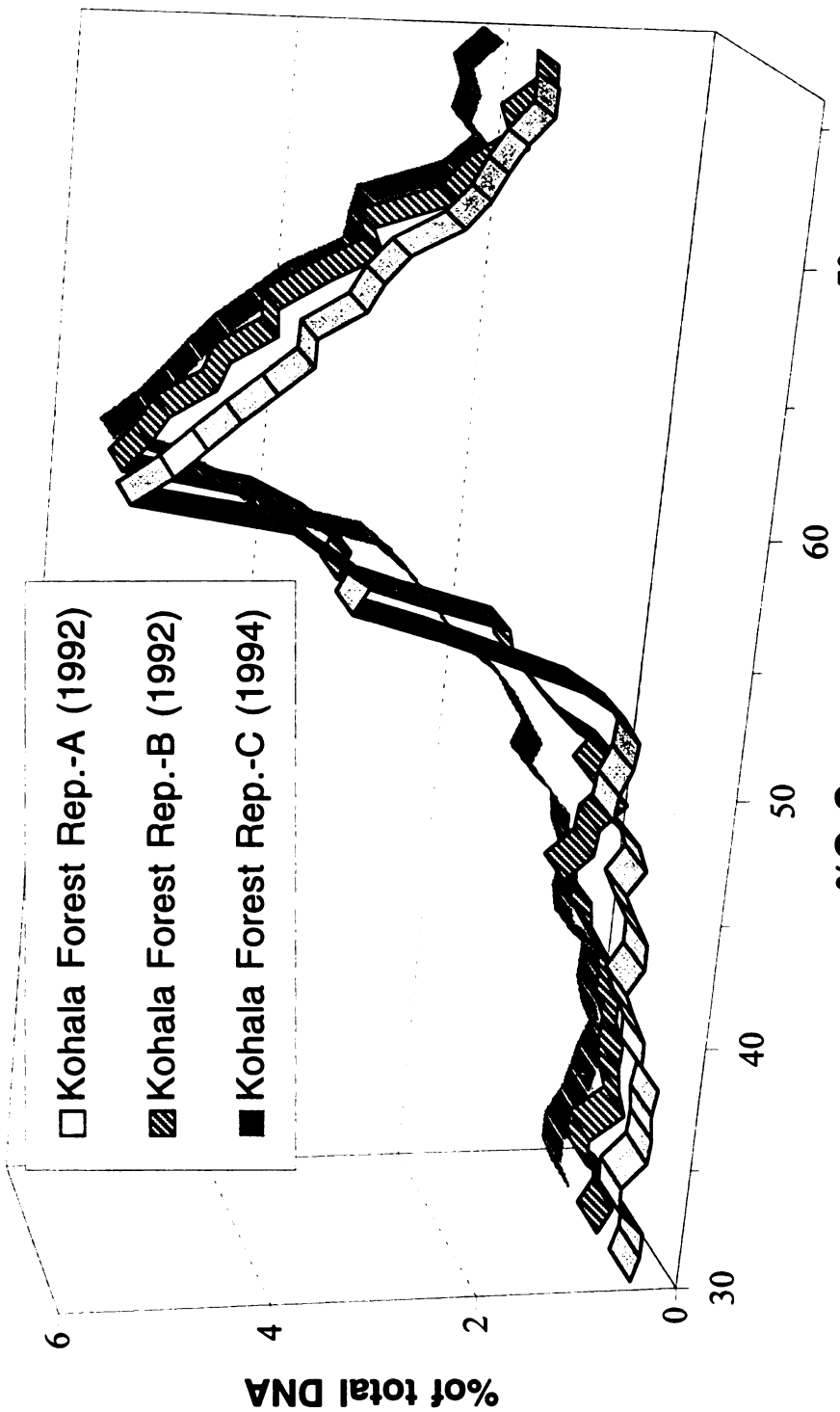


Figure 2. Replication of G+C profiles for three soil samples from the Kohala site (150,000 years). Replicates A and B were fragments of the same soil aggregate sampled in 1992. Replicate C was sampled at the same soil site 2 years later

Table 3. Comparison of extracellular and intracellular DNA extraction yields

Soil	Free DNA[a] (μg/g dw)	Total DNA[b] (μg/g dw)	Cell DNA[c] (μg/g dw)
Thurston	ND	6.4	6.5
Ola'a	0.1	14.1	14.2
Laupahoehoe	0.1	11.4	11.3
Kohala	ND	19.8	19.8

[a] Free DNA yield with extra cellular extraction protocol
[b] Total DNA yield with direct lysis protocol
[c] Cell DNA yield with direct lysis protocol after extra cellular extraction protocol
[d] Not detectable at a detection limit of <0.05 μg nucleic acid/g dw

midwestern agricultural soils (Chapter 2: Fig. 1). The ratio of DNA with high G+C content (55 to 70%) over DNA of low G+C content (30 to 55%) was approximately 2:1 for the two youngest soils, and increased to 4:1 and 3:1 for the two oldest sites, Laupahoehoe and Kohala, respectively. Over the chronosequence the profiles displayed increasingly smoother curves.

Community structure as revealed by analysis of phylotype abundance profiles of eubacterial rDNA libraries did not reveal an underlying community structure for the youngest or the oldest soil in the chronosequence (for youngest soil, see Chapter 2: Fig. 3A). Therefore, total soil DNA fractionation was used to select a simpler set of the community for analysis of structure. To lower soil bacterial diversity to a more manageable level I fractionated total community DNA by its G+C content (Fig. 3), and then analyzed bacterial rDNA amplified from five combined fractions over the range 62 to 65% G+C. Again, few repeating phylotypes were seen, hence the diversity was too high to see any structure (data not shown). Only when I studied the structure of single fractions of the %G+C gradient, was a distinct rank-abundance distribution of phylotypes seen within the clone library.

Phylotype abundance patterns. Two fractions of the %G+C based community profiles were chosen for comparative community structure analysis along the chronosequence. The 63% fraction was chosen because it was located within the major peak of DNA abundance typically found in previous soils studies (Chapter 2: Fig. 1). The 35% fraction was chosen to represent a community within the low biomass range of the soil extracted DNA (Fig. 3). After amplifying the SSU rDNA inserts

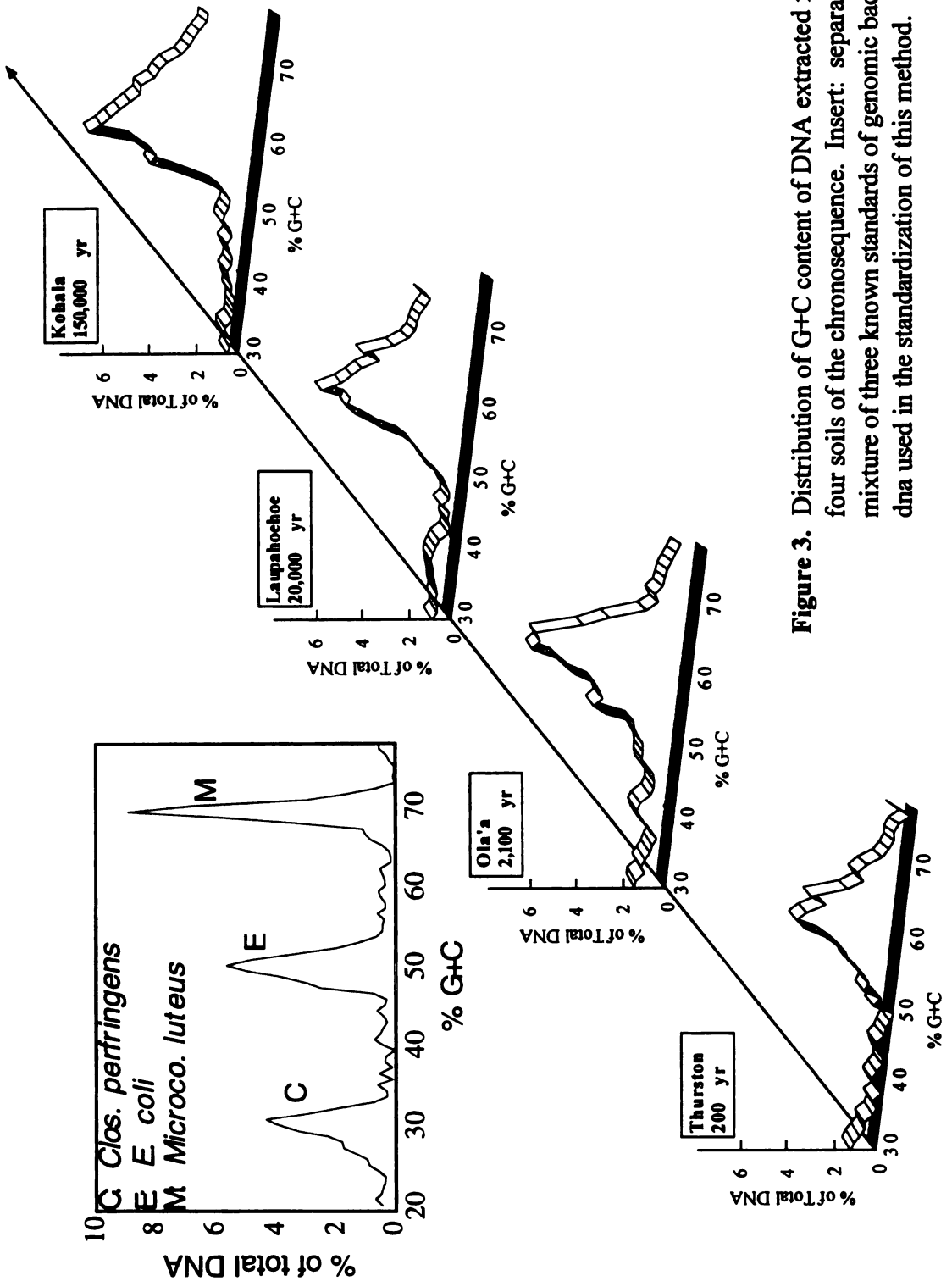


Figure 3. Distribution of G+C content of DNA extracted from the four soils of the chronosequence. Insert: separation of a mixture of three known standards of genomic bacterial dna used in the standardization of this method.

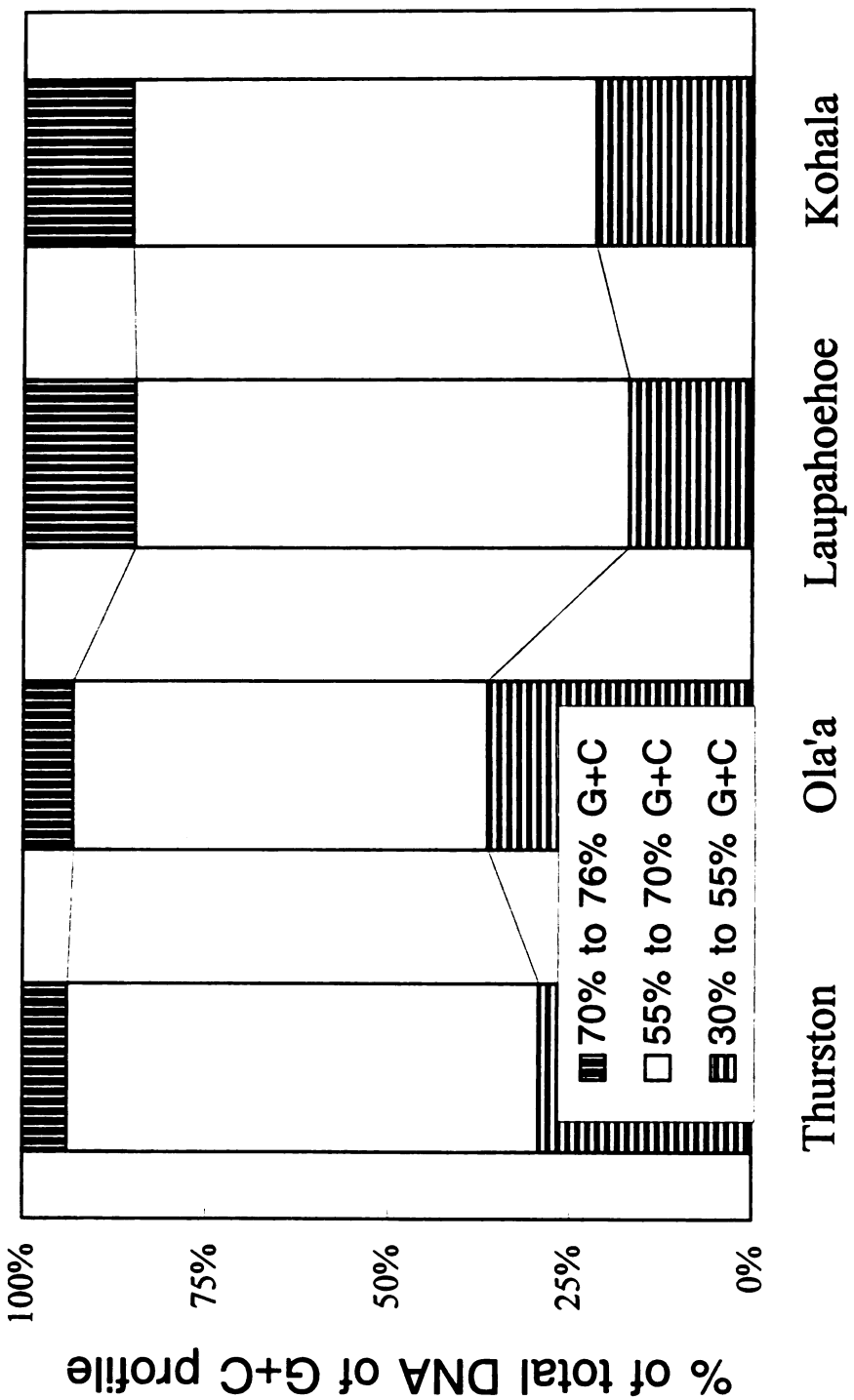


Figure 4. Contribution of different %G+C ranges to the community profiles for the four soils in the chronosequence

directly from clones they were digested with restriction enzymes and sorted by cluster analysis. Clones with similar restriction patterns were further differentiated when digested with a second combination of two restriction enzymes. The phylotype profile was repeatable in replicate samples as shown for three samples of the 63% fraction for the Laupahoehoe site (Fig. 5). Similar results were found with three samples from the oldest site, Kohala (data not shown). These data suggest not only that the findings from this method are repeatable since the two soil samples from the Kohala site show rather identical phylotype profiles. They also indicate that soil microbial communities that were sampled only three meters apart at the same site (Laupahoehoe) have very similar dominance structures (Table 4).

Figures 6 and 7 show the phylotype profiles for the 63% and the 35% G+C fractions of the chronosequence. Table 4 indicates the size of the clone libraries examined, and the relative abundance of the three best represented clones as a measure of dominance relative to the entire clone library. With increasing soil age the three most dominant SSU rDNA patterns represent 32%, 33%, 24(±5)%, and 67(±4)% of the SSU rDNA insert diversity within their respective 63% G+C fractions. The insert in Figure 6 illustrates the ratio between the three best represented clones and the remaining clones within each clone library along the chronosequence. In the 35% fractions the three dominant phylotypes constitute 20(±4.5)% of the total.

Rarefaction analysis. The quantitative effect of using a fraction of the community DNA on the diversity of phylotypes was evaluated using rarefaction curves (Simberloff, 1978; calculation see Appendix A). Rarefaction analysis is a technique for

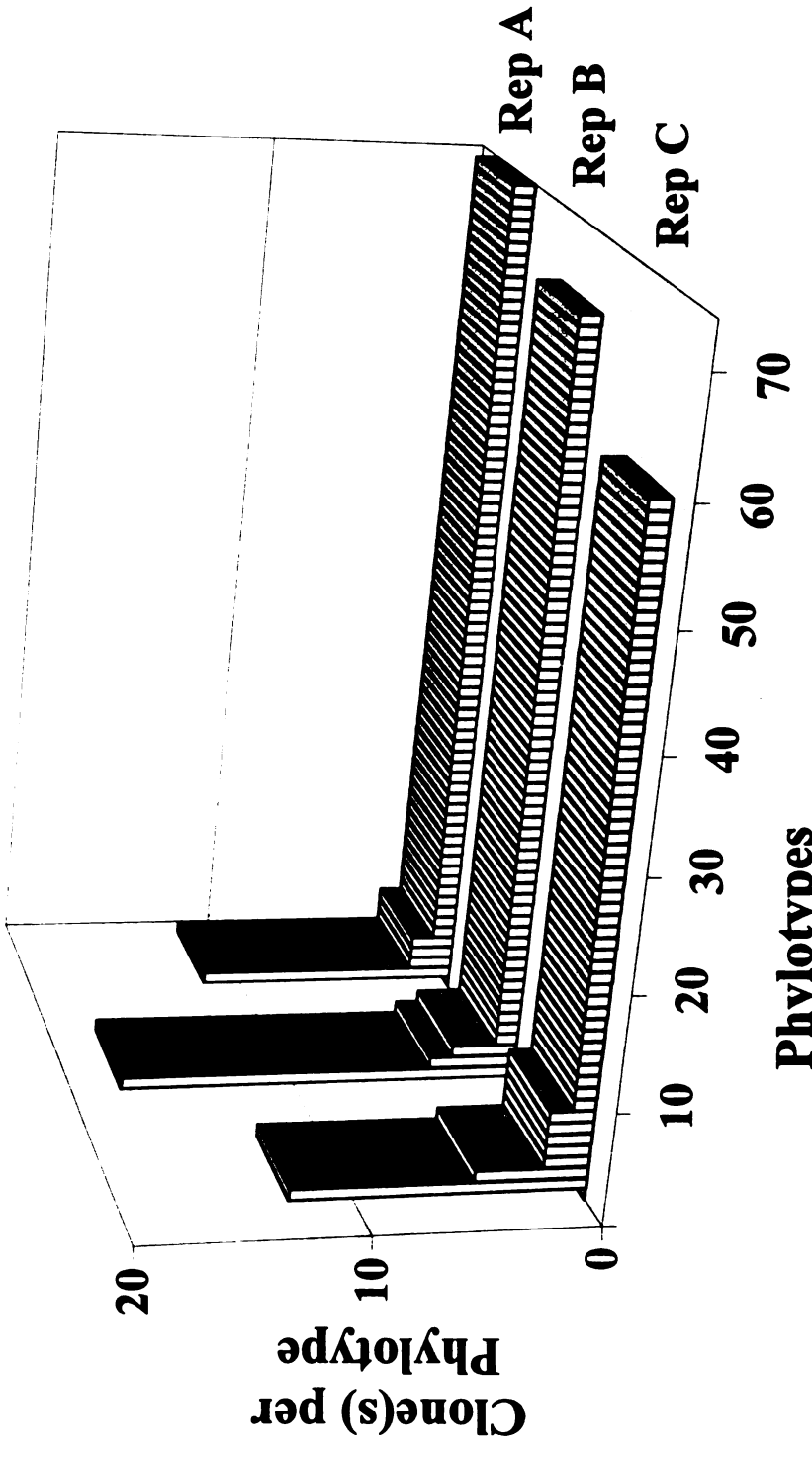


Figure 5. Replication of rank-abundance profiles for three soil samples of the mid-aged Laupahoehoe site (20,000 years)

Table 4. Number of clones and phylotypes in clone libraries and three most dominant phylotypes for each soil sample in the chronosequence. Similar phylotypes are indicated by bold numbers in identically shaded boxes. Underlined numbers indicate clones analyzed by nucleotide sequencing (Table 6).

Table 4. *Number of clones and phylotypes*

Table 4. Number of clones and phylotypes

63% G+C fractions

Clones or phylotypes		Thurston	Ola'a	Laupahoehoe Replicates[a,b]			Kohala Replicates[a,b]		
				A	B	C	A	B	C[c]
Total		76	77	79	89	87	79	80	78
Phylotypes		46	49	58	67	73	17	15	20
1st	(%)	13	17	16	20	14	43	48	47
2nd	(%)	11	9	6	4	2	15	18	9
3rd	(%)	8	6	3	3	2	9	6	6

35% G+C fractions

Total		59	76	ND	85	ND	82	75	81
Phylotypes		47	55	ND	67	ND	59	56	67
1st	(%)	7	11	-	11	-	9	4	6
2nd	(%)	7	8	-	6	-	5	4	4
3rd	(%)	5	7	-	6	-	5	4	4

[a] The similarities between the electrophoretic patterns of restriction fragments were analyzed by cluster analysis. Individual clones were grouped using a cut off of 97% similarity and a 5% error rate for the band position. The first, second and third most dominant phylotypes in each community were different from each other.

[b] The second and third most dominant phylotypes in the Laupahoehoe sample were identical to the first and second most dominant phylotypes in the Kohala sample, respectively.

[c] Sampling dates for A, B: 4/1992; for C: 5/1994

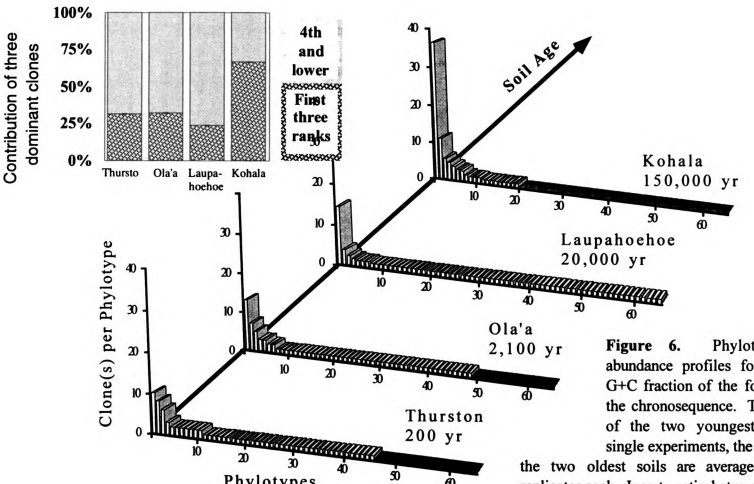


Figure 6. Phylotype rank abundance profiles for the 63% G+C fraction of the four soils of the chronosequence. The profiles of the two youngest soils are single experiments, the profiles of the two oldest soils are averages of three replicates each. Insert: ratio between the three best represented clones of the 63% G+C fractions and the remaining clones within each clone library along the chronosequence.

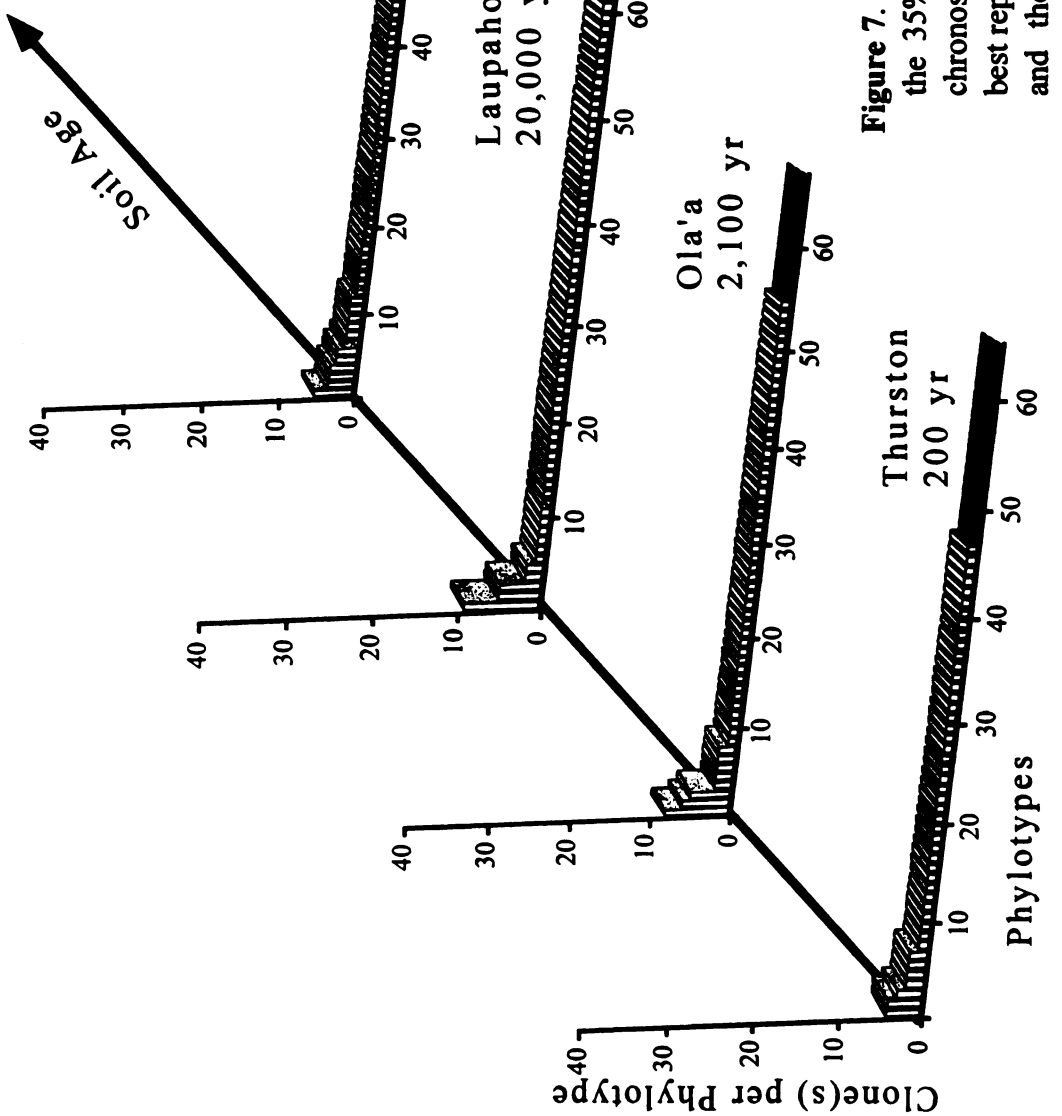


Figure 7. Phylotype rank abundance profiles for the 35% G+C fraction of the four soils of the chronosequence. Insert: ratio between the three best represented clones of the 35% G+C fraction and the remaining clones within each clone library along the chronosequence

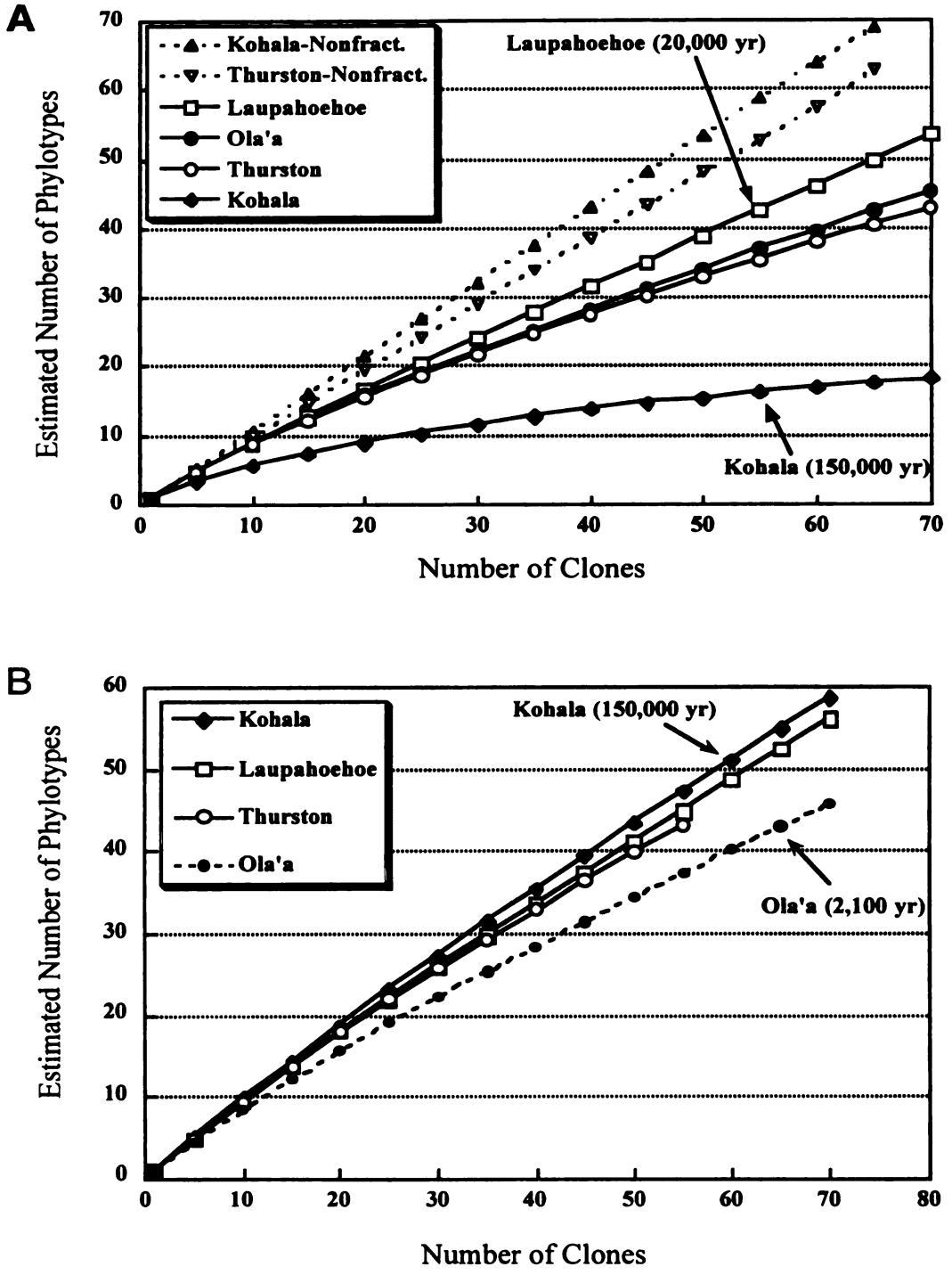


Figure 8. Rarefaction curves for the phylotypes found in the 63% G+C (A), and in the 35% (B) G+C fractions along the chronosequence. The X-axis depicts the expected number of phylotypes calculated from a random sample of individuals taken from the total population of phylotypes. A comparison with the rarefaction curve for phylotypes from nonfractionated whole community DNA is shown

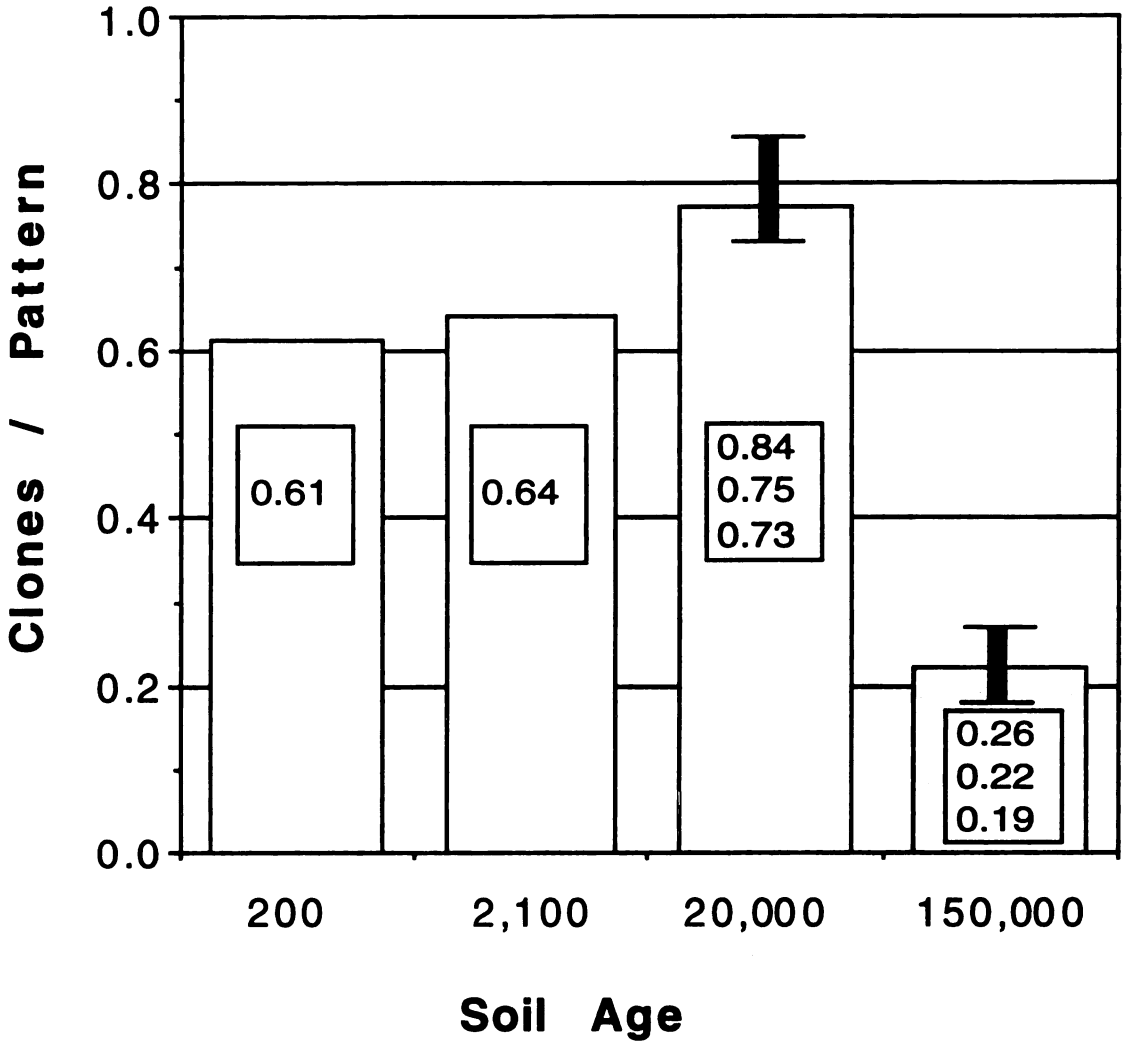


Figure 9. A simplified diversity index, defined as the ratio of all phylotypes of contributing clones in a particular clone library versus the total number of clones in that clone library, is shown for the four soils. The value of each sample is shown in each bar as well as the mean and the standard deviation

comparing the diversity of communities. Given the species-abundance distribution of a soil SSU rDNA clone library, rarefaction provides estimates of the species richness of sub-samples taken from it. When plotted against the subsample these estimates define the rarefaction curve. In Figure 8A I compare the cumulative diversity of total soil DNA and 63% G+C fractions thereof for the chronosequence. The high diversity throughout all 35% G+C fractions is summarized in Figure 8B. This analysis verified that the unfractionated DNA contained too many phylotypes to reveal any community structure but that fractionation of DNA by its G+C content did reduce the diversity to a level at which structures of dominance could be detected. However, this analysis does suggest that due to enormous diversity the phylotype sampling in these fractions is far from complete.

Community structure. Soil community structure changes can be seen across the chronosequence. The relative abundance of dominant community members in the high %G+C range is significantly expanded in the oldest soil (Fig. 6). The tendency to higher dominance over time appeared simultaneously in two ways. While the number of representative SSU rDNA clones per phylotype increased for the most dominant organism, the total portion of phylotypes in each soil library decreased. The most dominant clones appear in percent of total clones over time with 13% (200 years), 17% (2,100 years), 20% (20,000 years), and 35% (150,000 years), while the number of phylotypes for the respective soils decreased after an initial increase from 46 phylotypes (200 years) over 49 and 66 (±8, n=3) (20,000 years), to finally 17 (±2, n=3) phylotypes for the oldest soil (Table 4). Microbial species richness is reduced with

increasing soil age. While there is no statistically significant change in microbial species richness from the 200 year old Thurston to the 20,000 year old Laupahoehoe soil, species richness fell by 67% for the oldest soil, the 150,000 year old Kohala site (Fig. 9).

Comparative statistical analysis of entire rank-abundance profiles revealed a statistically significant difference only between the oldest Kohala site and the three younger sites ($p<0.05$) (Table 5; calculation see Appendix B: Table 3). Similar statistical analysis showed no significant difference between any of the four profiles of the 35% G+C fractions.

The increase of dominant structures in the oldest soil, and the decrease of phylotype richness are illustrated by a simplified diversity index, the ratio of phylotypes over the contributing clones in Fig. 9. This simplified diversity index is used to compare community dynamics graphically with other soil physical and chemical parameters in figures below.

A detailed statistical analysis of the ARDRA derived rank-abundance patterns is discussed in Appendix B. Test statistics for the Chi-Square and the Kruskal-Wallis tests were calculated for all phylotypes of a clone library sorted into three ranking groups (Appendix B: Fig.4). The first group is represented by the dominant phylotype. The second group consists of the second to fifth best represented phylotypes, and the third group covers the remaining phylotypes. The Chi-Square test indicated a statistically significant lack of homogeneity across the chronosequence for the three ranking groups ($\chi^2=139.2$; $p<0.05$). Similar calculations lead to a rejection of the null

Table 5. Statistical comparison of community patterns over time. After the test results for the Kruskal-Wallis statistic indicated nonhomogeneity between communities of the chronosequence the Multiple Comparisons Procedure was used to determine if the community structures are significantly different from each other ($p<0.05$). All clone libraries were grouped by ranks prior to the test. Details see Appendix B.

Pairwise comparison of soil sites[a]	Significant difference	
	35% G+C	63% G+C
Thurston and Ola'a	NO	NO
Thurston and Laupahoehoe	NO	NO
Thurston and Kohala	NO	**YES**
Ola'a and Laupahoehoe	NO	NO
Ola'a and Kohala	NO	**YES**
Laupahoehoe and Kohala	NO	**YES**

[a] Sites are grouped by soil age starting with the youngest soil

hypothesis of homogeneity for the 35% G+C fraction, although a Chi-Square test result of 19.4 for these fractions indicated much less difference between the four soil ages. A graphical comparison of determined and expected values for this test is shown in Appendix B: Figures 4 and 5 for the 63% G+C fractions of the chronosequence. The high diversity contributed by rare members of each clone library (Appendix B: Fig. 4: ranking group 6th-...) has a strong effect on the Chi-Square value when the dominance pattern becomes more accentuated with increasing soil age (Appendix B: Fig. 5) and the two oldest soils contribute the most with their extreme ratios of phylotype diversity over clone number. In addition to the Chi-Square test I also calculated the Kruskal-Wallis test for this chronosequence to determine which soil in the chronosequence has a community structure different from the others. In the Multiple Comparisons Procedure, a part of the Kruskal-Wallis test, I used all factorial combinations of the four soil ages to determine if there is a statistically significant difference between the structures of any two of the soil bacterial communites (Table 5). The Multiple Comparisons Procedure shows that the rank-abundance distribution of the Kohala clone library at 63% G+C is significantly different from the rank-abundance distributions of all three younger soils ($p<0.05$). There is, however, no significant difference amongst the three younger soils. For the rank-abundance distribution of the 35% G+C fractions the Multiple Comparisons Procedure shows no significant difference amongst the four soils of the chronosequence ($p<0.05$). However, at a coarser level of significance of $p<0.2$ the rank-abundance distribution of the Ola'a and the Laupahoehoe clone library are significantly different from the rank-abundance distributions of the oldest, the

Kohala soil. There is, however, no significant difference amongst the three younger soils and between the youngest and the oldest soil.

Phylotype overlap between sites. Most phylotypes are unique for each soil age. When I studied three samples from the same soil in the chronosequence for their overlapping restriction patterns I found that diversity within the same soil is much smaller than between soils of the chronosequence. When focusing on the three most dominant members of each soil an overlap of the same dominant restriction patterns could only be found once in the 63% fractions, between the two oldest soils at Laupahoehoe and Kohala (Table 4). However, the overlap of dominant restriction patterns within repeated samples drawn from the same chronosequence soil showed for Kohala soil samples that two out of three dominant members in the 63% fractions were identical. A similar result was found for three replicate samples drawn from the 20,000 year old Laupahoehoe soil (Table 4). As expected, the overlap of rare restriction patterns (*i.e.*, those found only once within a clone library) was low, in the range of 5 to 12% per site. These results indicate that variability within a soil site of a particular age is less than between different soils. Sampling distance had no effect on detecting difference in dominance structures for an order of magnitude. Two of the three repeated samples drawn from Kohala soil were taken from the same soil aggregate in one soil sampling bag. The three samples taken for comparison from the 20,000 year old Laupahoehoe soil were drawn at a distance of three meters apart from each other. This result corroborates the aptness of the method for comparative studies of soil microbial communities since differences between sites can be distinguished from those within

Table 6. Phylogenetic affiliations of SSU rDNA sequences from a chronosequence of Hawaiian rainforest soil (HRS). The relative abundance of each phylotype is indicated, as well as the similarity to the closest known species. Only unambiguously aligned regions were used for the analysis. All clones are from the 63% G+C fractions for each soil.

Soil site	Phylotype[a]	Relative abund. (%)[b]	Phylogenetic affiliation	Closest relative in RDP database	Similarity (%)
Thurston	HRS-12	13	Alpha Purple Bacteria	*Rhodopseudomonas viridis*[c]	83
	HRS-13	11	Gamma Purple Bacteria	*Pseudomonas syringae*[d]	98
	HRS-16	8	Alpha Purple Bacteria	*Azospirillum lipoferum*[e]	88
Laupahoehoe	HRS-53	20	Delta Purple Bacteria	*Polyangium* sp. str. Pl 4943[f]	45
(Replicate B)	HRS-62	4	*Fibrobacter* Phylum	env. str. MC 13[g]	72
	HRS-57	3	*Fibrobacter* Phylum	*Acidobacterium capsulatum* str. 161[g]	54
Kohala	HRS-46	43	*Fibrobacter* Phylum	env. str. MC 9[g]	70
(Replicate A)	HRS-55	15	*Fibrobacter* Phylum	*Acidobacterium capsulatum* str. 161[g]	65
(1992)	HRS-60	9	*Synthrophomonas* Group	*Desulfotomaculum geothermicum*[h]	42
Kohala	HRS-47	48	*Fibrobacter* Phylum	env. str. MC 26[g]	73
(Replicate B)	HRS-56	18	*Fibrobacter* Phylum	*Acidobacterium capsulatum* str. 161[g]	67
(1992)	HRS-50	6	*Fibrobacter* Phylum	env. str. MC 101[g]	73
Kohala	HRS-48	47	*Fibrobacter* Phylum	env. str. MC 27[g]	68
(Replicate C)	HRS-63	9	*Fibrobacter* Phylum	*Acidobacterium capsulatum* str. 161[g]	69
(1994)	HRS-51	6	*Synthrophomonas* Group	*Desulfotomaculum geothermicum*[h]	41

[a] HRS, environmental clone from Hawaiian Rainforest Soil

[b] Relative abundance of the clones belonging to a phylotype, calculated by dividing the number of clones belonging to that phylotype by the total number of clones analyzed

[c] RDP subdivision: *R.rubrum* Assemblage

[d] RDP subdivision: *Rhizobium-Agrobacterium* Group

[e] RDP subdivision: *Pseudomonas* Subgroup

[f] RDP subdivision: *Myxobacteria* Subdivision

[g] RDP subdivision: *Acidobacterium* Subdivision

[h] RDP subdivision: *Synthrophomonas* Assemblage

sites, at least in this young ecosystem.

Phylogenetic analyses. All of the soil clones sequenced affiliate with the *Eubacteria* as expected due to the specificity of the primer pair used (Table 6). None of the nucleotide sequences analyzed in this study matched known sequences from any cultured species, or were identical to sequences in the databases. Sequence similarity of partial SSU rDNA sequences to SSU rDNA sequences in the RDP database (Maidak *et al.* 1996) varied from 41% to 98%.

I subdivided the chronosequence nucleotide sequences into two general categories, those related to previously recognized bacterial groups (Olsen *et al.*, 1994) such as *Proteobacteria, Clostridia, Bacillus,* and *Syntrophmonas*, and those related to groups of novel descent such as *Fibrobacter*. All of the sequences of the only 35% G+C fraction analyzed (the youngest soil) were affiliated with the genus *Clostridium* (Chapter 2: Table 1). Of the three most dominant clones from the 63% G+C fractions of the chronosequence, only the youngest site did not have clones related to members of the group *Fibrobacter* (Table 6). Most members of the *Acidobacterium* subdivision in the *Fibrobacter* Phylum (Maidak *et al.*, 1996) are known only from rDNA sequence analyses. None of the nucleotide sequences analyzed in this study that belongs to the *Fibrobacter* Phylum is closely related to (< 27%) any previously described cultured species.

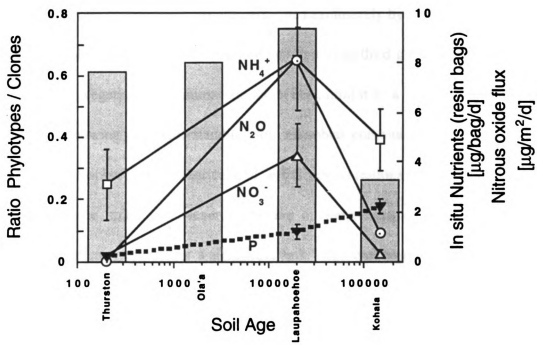


Figure 10. Plant available nitrogen and phosphorus determined in situ with anion-exchange resin bags that were buried from 5 to 8 month (data from Crews *et al.*, 1995). The data are overlaid with the diversity index from Figure 9

DISCUSSION

Soil microorganisms play an integral role in the maintenance of soil fertility and soil structure, and certain functions are performed exclusively by bacteria. Nevertheless, many soil ecological studies tend to ignore that important third group of organisms with a bias towards vegetative and animal life, or, at best, limit it to a black box of "microbial activity". Knowing and understanding soil microbial communities is essential for a sustainable biosphere, from agricultural efficiency questions to biodegradation of hazardous waste spills. However, achieving a comprehensive picture of a native bacterial community represents a significant task. The challenge becomes even greater when one moves this investigation into one of the most complex environments, soil, with its high species richness and evenness. Several recent studies have attempted to describe a few of the most common members of a soil community. However, I believe it is no longer enough to only inventory soil organisms, but we must also link that diversity to ecological processes.

In macroecological systems the following relationship between productivity and biodiversity is typical: In low-nutrient environments an increase in biodiversity seems to enhance productivity (Odum, 1998). In high-nutrient environments an increase in productivity increases dominance, and reduces diversity. This macroecological principle seems to be consistent with the microbial community dynamics observed at this

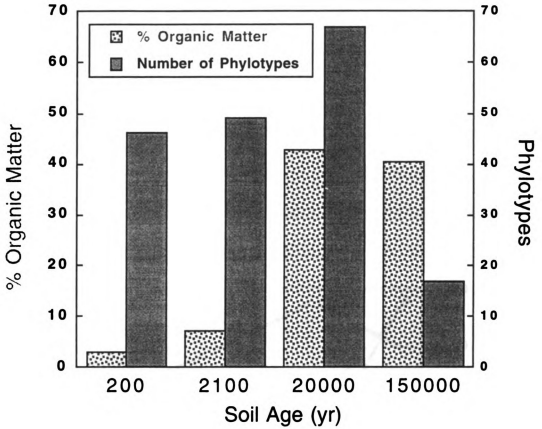


Figure 11. Comparison of soil organic matter content and phylotype diversity for soils of the chronosequence.

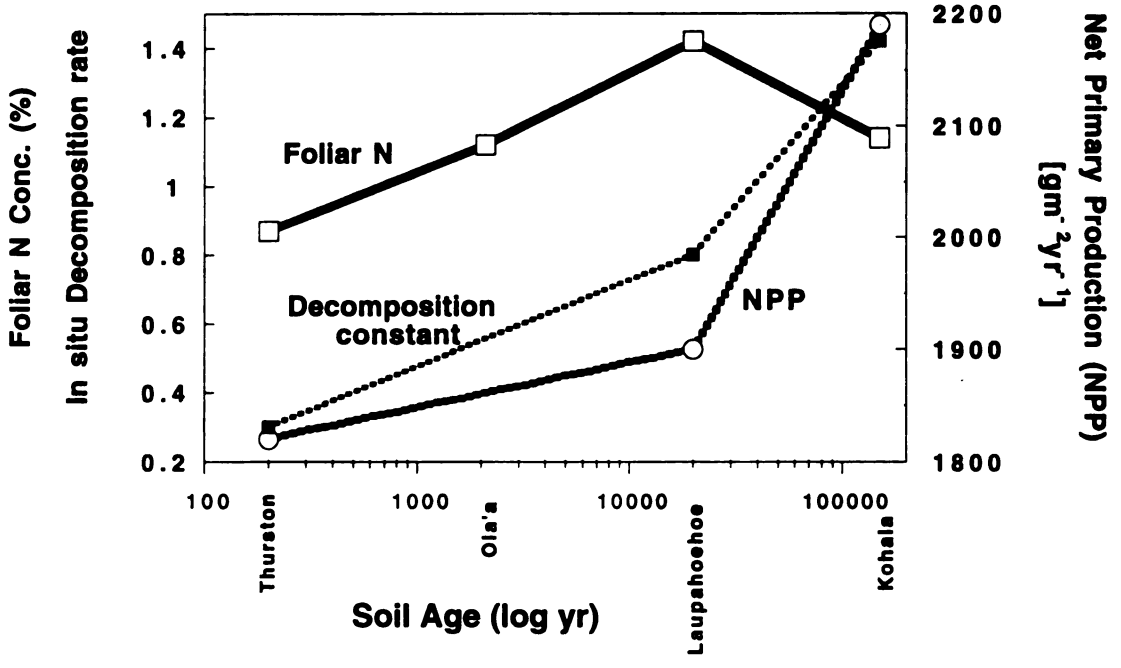


Figure 12. Exponential decomposition constants and foliar nitrogen concentration for leaf litter of the dominant plant *Metrosideros polymorpha*. Decomposition constants were determined over 2 years by in situ decomposition (data from Crews *et al.*, 1995). Foliar nitrogen concentration is calculated in % dry mass (data from Vitousek *et al.*, 1995). Net primary production is shown (data from Herbert, 1995).

chronosequence. A relationship between soil productivity and microbial diversity can be found when the three younger, less developed soils are compared with the oldest soil at the Kohala site. In the younger soils, the nutrient supply has not yet reached its maximum which is reflected by high phylotype diversity (Fig. 9). From the youngest soil to the mid-aged Laupahoehoe site, plant available nitrogen and phosphorus are increasing, with a nitrogen maximum at this age (Fig. 10). Between 2,100 years and 20,000 years of soil age the organic matter reaches a maximum (Fig. 11; Table 1). This coincides with a slightly higher phylotype diversity for the dominant (63% G+C) microbial fraction at the 20,000 year old Laupahoehoe site. The highest turnover rates for NH_4^+ -, NO_3^- -, and N_2O - nitrogen also occur at this site, which indicates maximal microbial activity, and soil productivity (Fig. 10; Crews *et al.*, 1995). After reaching this high level of nutrient availability at the mid-age site the resulting high productivity is followed by an increase in microbial dominance (63% fraction) at the expense of phylotype diversity (Fig. 6) at the next older soil.

Specialization is offered as the most reasonable underlying mechanism for this change in diversity at the oldest site. With time specialists for the decomposition of the abundant accumulated organic material may take over the available niches in the soil. This specialization would increase dominance at the expense of diversity, which is the pattern at the Kohala site. This assumed specialization is reflected in the decomposition dynamics along the chronosequence. Concentrations of elements in live tree leaves generally reflect available nutrient pools in the soil, for both the dominant tree *Metrosideros polymorpha* and for other tree and bush species across the chronosequence

(Vitousek *et al.*, 1995). While foliar nitrogen and phosphorus contents in *Metrosideros polymorpha* leaves do not change, the leaf litter decomposition rate increases significantly from the midage to the oldest soil, and climaxes in a three-fold faster decomposition rate compared to the youngest site (Fig. 12; Crews *et al.*, 1995). Control experiments showed that this increase in decomposition rate was site specific. Also, after two years of decomposition almost four times as much leaf litter mass remained at the mid-aged Laupahoehoe site relative to the oldest site, suggesting improved decomposition ability of the microbial community at that site. This difference in decomposition rate occurred despite the constancy of both precipitation and temperature across the chronosequence. With time the soil microbial community might develop and change to a K-selection strategy at Kohala, where efficiently competing decomposers operate at maximum decomposition rates.

This assumption is corroborated by the amount of nitrogen and phosphorus remaining within partially decomposed leaf litter after a period of two years (Crews *et al.*, 1995). Almost four times as much nitrogen and phosphorus remained in the leaf litter at the mid-aged Laupahoehoe site (34% of initial N and P) *versus* the oldest Kohala site (9% of initial N and P).

Over the course of soil development along the chronosequence the amount of plant available phosphorus is constantly increasing, while plant available nitrogen peaked at the midage Laupahoehoe site, and then fell to less than 50% that value at the oldest site (Fig. 10; Crews *et al.*, 1995). The C/N ratio stayed almost the same at the last two sites in the chronosequence suggesting that changes in plant available nutrients

or litter quality were not responsible for a change in microbial community structure, and hence specialization within the community is hypothesized (Table 1).

The dominant soil minerals might also contribute to the high microbial activity at the two oldest soils. Weathering processes are rapid both because of the wet tropical climate and the reactive nature of the lava parent material. The rapidly weatherable olivine, glass and plagioclase are completely consumed before 20,000 years (Vitousek *et al.*, 1997). Subsequently formed noncrystalline minerals (allophane, ferrihydrite, and imogolite) are characterized by large reactive surface areas that bind cations, phosphorus and soil organic matter very effectively (Wada, 1989), which aids productivity through accumulation of organic matter (Table 1; Fig. 11).

Specialization of the decomposing microbial population could have a mutualistic effect on the vegetative cover. The rapid decomposition of leaf litter with relatively high nutrient concentrations would constantly supply plants with high levels of available nutrients (Fig. 12). This rapid regeneration of available nutrients provides a positive feedback between the plants and the decomposing microbial community, in which high levels of available nutrients are maintained (Vitousek, 1982). Net primary production of the forest increases steadily between 200 and 20,000 years, but the average rate of net primary production exhibits a substantial increase after 20,000 years to peak at the oldest site at a level 20% over that of the youngest site (Fig. 12) (Herbert, 1995; Vitousek *et al.*, 1997), which is three times as high as found in midwestern forests.

Profiles of the fractionated community DNA. The G+C based fractionation gradients represented with excellent repeatability community profiles characteristic of

the relative composition of the soil community. The similarity of the replicate gradient profiles at Kohala (Fig. 2) demonstrated that these curves represent not a just "snapshot" in time but are repeatable at the same soil site over time. Hence, differences in profiles show the effects of major changes in the relative abundance of components of the soil microbial community. A population change of at least 10^5 to 10^6 cells is required to cause a change in G+C peak height in a total community DNA profile. The majority (60% (±3)) of the extracted soil DNA corresponds to the 55% to 68% G+C range which includes genera known to dominate soil bacterial communities, *e. g. Agrobacterium* (55-61%G+C), *Alcaligenes* (56-63%G+C), *Arthrobacter* (63-69%G+C), *and Pseudomonas* (58-66%G+C) (Alexander, 1977; Holben and Harris, 1995). Soil DNA found in the 29% to 47 % G+C range corresponds to genera like *Streptococcus* (35-40 %G+C), *Clostridium* (31-36 %G+C), and *Bacillus* (38-45 %G+C). Rarefaction analysis indicated reduced diversity in the individual % G+C fractions compared to the total DNA, especially for the 63% G+C series, but that rDNA diversity was far from exhausted in all clone libraries [83(±3) clones] of this fraction (Fig. 8A).

Bacterial diversity in soil environments was too high to be analyzed comparatively between different sites based on the direct sequencing approach. The studies of soil community structure found to date tend to concentrate on the relatively few and likely more abundant species, but do not recover the rich variety of organisms that occur in smaller numbers but encompass a much greater range of species. These microorganisms may be important in acting as a reservoir of different physiological

types, which may respond by population size increase following changes in the habitat. This would also imply a degree of built in redundancy in the microbial community, and that some organisms are remainders of past conditions. Knowledge of the composition of this reservoir will allow us to distinguish between a complete community shift and a change in only the dominant community members. Furthermore, evidence from ^{3}H-labeling experiments and G+C based fractionation of soil community DNA indicates that activity and growth are not confined to a small fraction of the bacterial biomass but are widespread among the genera represented in the profile (Harris, 1994). Hence, separation by G+C content uncovers new diversity and provides evidence that soil rDNA diversity is much higher than can be detected by direct rDNA analysis of total community DNA. This higher estimate of soil rDNA diversity is supportive of the high bacterial diversity estimate reported by Torsvik *et al.* (1990) based on rates of soil DNA reannealing.

DNA extraction. A significant concern in establishing a community picture from total community DNA extracted from an environment is whether cell lysis was partial or biased. The pre- and post-extraction direct microscopic counts indicate sufficient lysis efficiency was achieved. The high diversity found in the phylogenetic analysis of some of the rDNA clones together with finding a high number of sequences for sporeformers confirmed the reliability of my approach (Chapter 2: Fig.7). The efficiency of total community DNA extraction is crucial for a comprehensive evaluation of the soil microbial community. Initial difficulties with DNA extraction (not reported) might have been due to a parent-material effect. Weathered tephra, a Hawaiian volcanic

ash and the source of the parent material, is characterized by a high content of allophane at soils older than 2,100 years. Allophanes are amorphic hydrated aluminum silicates which are able to fix dissolved phosphate at high rates (Nanzyo *et al.*, 1993). For maximum yields of soil DNA, I saturated active sites on allophane by increasing the phosphate concentration in the extraction buffer one hundred times (to 100 mM phosphate). Extracting DNA from Andosols is also made difficult by their high content of organic matter. The allophanic clays in these soils help to stabilize soil organic matter such that soil organic matter contents two to three times higher that of nonallophanic materials are formed (Paul and Clark, 1996). The abundance of humic substances in soils, many of which coextract with nucleic acids in standard protocols, required post-extractive purification to allow for PCR amplification of the extracted community DNA.

The potential for bias is also present once comprehensive lysis is achieved. Biases can come from preferred amplification in the PCR reaction, the formation of chimeric SSU rDNA molecules, or the *rrn* copy number variation in different bacterial species (Chapter 1: Background). These biases have been extensively reported or studied elsewhere (Ward *et al.*, 1992; Stackebrandt *et al.*, 1993; Moré *et al.*, 1994; Ogram *et al.*, 1987; Tsai *et al.*, 1991; Liesack *et al.*, 1991; Suzuki *et al.*, 1996; Robinson-Cox *et al.*, 1995; Reysenbach *et al.*, 1992; Farrely *et al.*, 1995; Kopczynski *et al.*, 1994; Mead *et al.*, 1991; Finney, 1993). For a complete quantitative determination of a soil bacterial community, many of the techniques are still in need of refinement. I have used the at present best available methods as critically as possible to minimize the introduction of potential biases.

Phylogenetic analyses. In a recent comparative study of soil phylogenetic surveys based on rDNA sequence studies Kuske *et al.* (1997) show considerable commonality of sequences among the surveys that fall outside of the known bacterial domains (Olsen *et al.*, 1994). This remarkable abundance of members of this new phylogenetically diverse group in clone libraries across soils from vastly different geographic regions indicates that these organisms are dominant regardless of soil types, geographic regions, and overlying vegetation (Kuske *et al.*, 1997; Liesack *et al.*, 1992; Stackebrandt *et al.*, 1993; Ueda *et al.*, 1995; Borneman *et al.*, 1996; Borneman *et al.*, 1997). My findings support this thesis since members of the dominant biomass (63% G+C) of chronosequence soils analyzed also belong to the same novel branch (Table 6). Interestingly, they are all related to the clone cluster VI defined by Stackebrandt *et al.* (1993), which forms an individual line of descent currently under analysis (Ludwig *et al.*, 1997; N. Pace, personal communication). Kuske *et al.* (1997) made the first attempt to compare SSU rDNA sequence data among various other soils including their own, and discovered substantial similarities of phylogenetic types across different soils. Most of those sequences fall into several diverse groups distinct from previously recognized bacterial divisions. Of all nucleotide sequences analyzed in this study, only 40% are related to previously recognized bacterial groups (Olsen *et al.*, 1994). Of these, their closest relatives are members of culturable genera found in many types of soils, such as *Proteobacteria, Clostridia,* and *Bacillus*.

Comprehensive inventories? Figure 8 clearly demonstrated that complete diversity has not been captured in any of the communities analyzed, since

none of the rarefaction curves level off with increasing clone numbers. Zhou et al. (ESA abstract No. 226, 1997) attempted to analyze a continental soil clone library from 5 g pf soil with over 600 clones, but found this was still not large enough to capture full diversity, i.e. the respective rarefaction curve was still linear. The reason to avoid bigger clone libraries is not only the prohibitive cost and labor, but also the ecological value of such an attempt. No macroecological study attempts to gather all species of an environment. Instead, studies are restricted to a certain group, usually a functional group, which makes the analysis feasible. My attempt is analogous, namely to select and evaluate a fraction of the enormous soil bacterial diversity. The field needs to get away from the notion that one can be comprehensive. It is simply not yet possible.

Limitations. The interpretation of these data is limited by the fact that there are no comparable studies. Even though some of these data are based on triplicate experiments, generalizations are difficult to make; for example, only two fractions of total community DNA were investigated. All major changes described occurred in the fraction of high biomass, whereas no significant changes could be found in the G+C fraction representing more minor members of the community. This does not exclude the possibility that dominance within other high G+C fractions, *e.g.* 61% or 67% G+C fraction, might have very different responses to soil age. Further investigations on this soil series would be necessary to clarify this.

This chronosequence on the Island of Hawaii supplied a relatively simple system of developing soils to study and evaluate microbial community structure and its interaction with soil developmental processes over time. However, even though I know

that the Hawaiian islands are not recovering from the history of disturbances like the Pleistocene glaciations in temperate regions, the history of disturbances in the sampled area is not entirely known. Furthermore, I cannot claim long term stability or equilibrium stability in these tropical soils, since there is a constant input of nutrients and possibly microorganisms via the trade winds. Exogenous inputs of Asian dust deposition (Jackson *et al.*, 1971) might also be a constant source of nutrients and nonindigenous microorganisms. The possible influence of these inputs may prevent equilibrium in these communities. In addition, fungal activitiy might be the initial step of foliar decomposition. Also, the encountered community transition rates cannot be applied globally, since they probably occur more rapidly in Hawaii than in other soil sites due to abundant rainfall, warm year-round temperatures, and easily weathered primary minerals. Nevertheless, the understanding of community development gained on this Hawaiian chronosequence is an important first step to understanding developmental changes of microbial communities, and might be broadly applicable to continental or other island systems. The increase of microbial phylotype dominance along the chronosequence, which was offset by a decrease in phylotype diversity after the intermediate stage of soil development, might represent a typical development for the major fraction of microbial biomass in other mature soils.

APPENDIX A

APPENDIX A

STATISTICAL COMPARISON OF COMMUNITIES BY RAREFACTION CURVES

When analyzing community structure we need to know when we have sampled enough organisms to sufficiently characterize the community. When applied to SSU rDNA analyses of communities this equates to determining when a clone library is large enough to characterize the sampled bacterial community, *i.e.* when diversity in the 'number of species' versus 'number of individuals sampled' curve levels off (Gleason, 1922) (Fig. 1). With the rDNA method different taxa can be defined as clones with different restriction fragments derived form using several tetrameric endonucleases: these different groups have been termed phylotypes. More specifically, such analyses actually reflect the population distribution of amplifiable rDNA operons. While this does not represent species in a formal sense, it is a measure that groups like individuals, and hence "species" diversity measures can be used to characterize *rrn* populations.

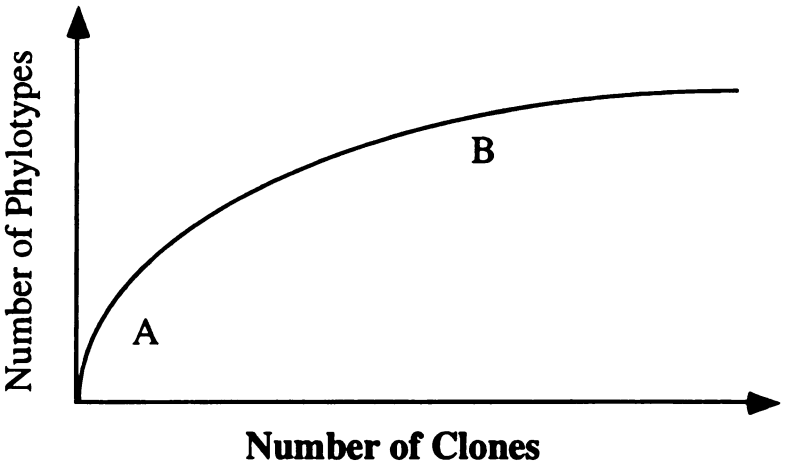


Figure 1. Sampling plot of the dependence of diversity on the number of individuals. Region A refers to the curve area of rapid accumulation of phylotypes associated with increasing number of clones sampled. As the number of clones added becomes larger, region B, the clone library has become large enough that new phylotypes are added at a much slower rate (adapted after Sanders 1968).

The clone libraries collected from an environment are finite samples of a much larger community. We can measure the phylotype abundance distribution within a clone library, but of ecological interest is the diversity of the bacterial community as a whole. To estimate the trend of species richness in a clone library with respect to the sampling curve (Fig. 1), rarefaction analysis can be used. Furthermore, the rarefaction curves of several communities plotted together render a graphical display of the differences in diversity of those communities.

Rarefaction should be used only for clone libraries that have been obtained by using standardized sampling and processing protocols. Rarefaction curves are not to be used to extrapolate beyond the number of clones collected (Tipper, 1979).

Rarefaction analysis. Rarefaction analysis is a statistical technique for comparing diversities of organismal communities. Given the species-abundance distribution of a collection, *i.e.* a clone library, rarefaction confers estimates of the species richness of sub-samples taken from it. When plotted against the subsample, these estimates define the rarefaction curve, which is a deterministic transform of the phylotype abundance distribution within the clone library. Rarefaction curves have been calculated to compare phylotype diversity of total soil community DNA versus fractions of this community DNA based on G+C content, and to compare these fractions among different soils. Rarefaction allows for unequal sample size and estimates the species richness (*i.e.* number of phylotypes), $E(S_m)$, that a total population of N individuals would have, if its size were randomly restricted (rarefied) to a sample of m individuals (*i.e.* clones taken from a clone library) (Tipper, 1979). Rarefaction is computed using the equation:

$$E(S_m) = \sum_{i=1}^{S}\left[1 - \binom{N - N_i}{m} \Big/ \binom{N}{m}\right], \qquad [1]$$

where S is the number of species in the parent collection, N_i is the number of individuals of the i-th species, N is the number of individuals in the parent collection (*i.e.* the entire clone library), and m is the number of individuals in the rarefied sample ($m<N$) (Simberloff, 1978). Since this calculation has no probabilistic basis, Simberloff (1978) created the software program SIM that calculates the variance of the expected species richness $E(S_m)$.

The SIM rarefaction program. The rarefaction program SIM is written in FORTRAN-IV for a maximum of 550 operational taxonomic units. This program computes the expected species richness ($E(S_N)$), the variance $Var(S_N)$ of this estimate, and the standard deviation (S.D.) of the estimate for rarefied, *i.e.* restricted, samples. These values are needed to compare the phylotype richness of clone libraries of unequal size. The data entry and output are shown in Figure 2; an example of rarefaction curves calculated with the software SIM can be seen in Figure 3.

Entry form:

N		NS		NUM												
89		67		15												
CP_1		CP_2		CP_3....CP_{ns}												
18	4	3	1	1	1	1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
1	1	1														
num_1		num_2		num_3... num_{NUM}												
1	5	10	15	20	25	30	35	40	45	50	55	60	65	70		

Output form:

```
        N      NS
       89      67
NUMBERS OF INDIVIDUALS IN THE DIFFERENT SPECIES
      18  4  3  1  1  1  1  1  1  1  1  1  1  1  1  1
       1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
       1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
       1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
       1  1  1

 N    E(S)    S.D.    VARIANCE
 1   1.0014   .3963    .1571
 5   4.6532   .7028    .4939
10   8.7892  1.1487   1.3195
15  12.7213  1.4472   2.0943
20  16.5618  1.6537   2.7348
25  20.3495  1.7871   3.1938
30  24.1009  1.9102   3.6490
35  27.8210  1.9898   3.9594
40  31.5156  2.0454   4.1836
45  35.1878  2.0714   4.2909
50  38.8413  2.0705   4.2869
55  42.4792  2.0391   4.1578
60  46.1044  1.9773   3.9097
65  49.7196  1.8793   3.5318
70  53.3271  1.7407   3.0300
Stop - Program terminated.
```

Figure 2. Example of data entry and output for the rarefaction program SIM using data from a fraction at the youngest Hawaii site, Thurston.

The data entry (top part of Fig. 2) requires the the following raw data:

- N, is the number of individual clones in the clone library (here: 89)
- NS, is the number of phylotypes

NUM, is the number of subsamples, num_i, of individual clones to be drawn from the clone library

- num_i, is a step in the array of subsamples in ascending order
- CP, is the number of clones in each phylotype
- NUMBERS OF INDIVIDUALS IN THE DIFFERENT SPECIES,

  is the number of clones in each phylotype sorted by descending abundance

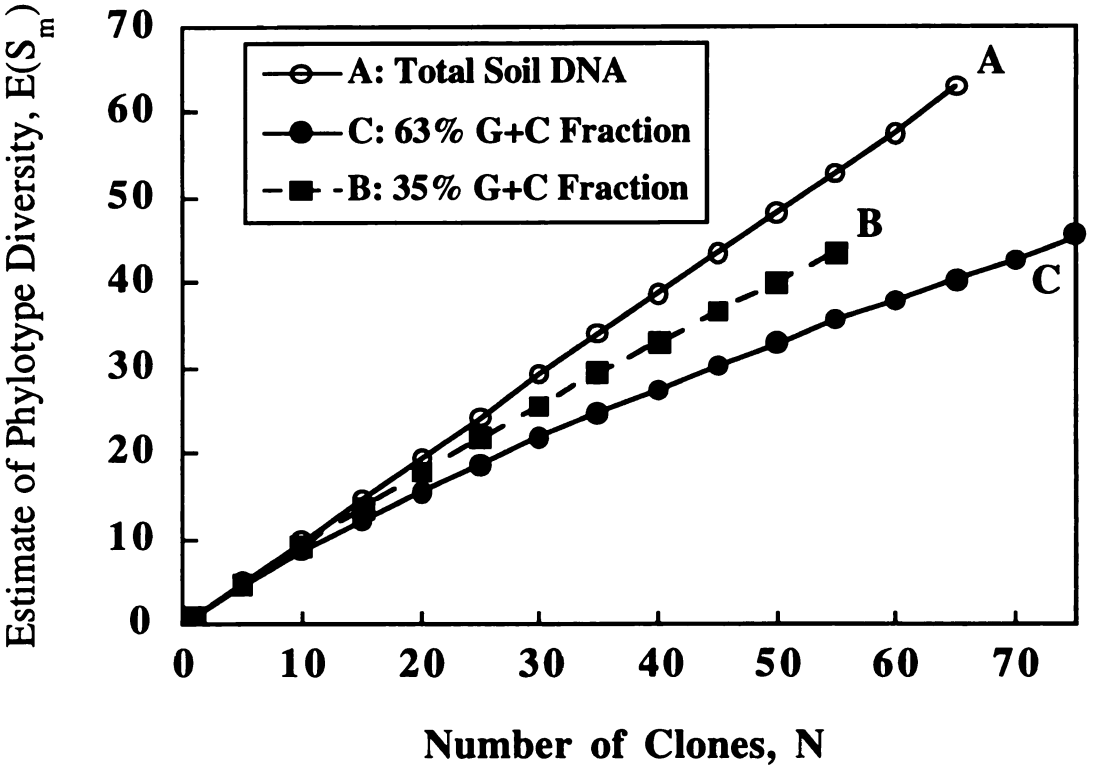


Figure 3. Rarefaction curves calculated by the rarefaction program SIM for comparison of three related clone libraries. Non-fractionated DNA (A) is compared with fractions of the same sample first separated by its G+C content for their phylotype diversity. Since rarefaction curves produce values of $E(S_m)$ only for integer values of m, the datapoints representing the rarefaction curve must be identified.

Rarefaction is a weak statistical technique, since identities of the species are not used.

Because the rarefaction curve is a mere transformation of the original phylotype-

Rarefaction is a weak statistical technique, since identities of the species are not used. Because the rarefaction curve is a mere transformation of the original phylotype-abundance distribution, I suggest here that comparison of rarefaction curves for other than graphical reasons should more profitably be replaced by direct comparison of those distributions, using the Kruskal-Wallis test statistic (Appendix B).

ACKNOWLEDGMENTS

I thank J. Dunbar for allowing me to use his updated version of Simberloff's (1978) FORTRAN program SIM.

REFERENCES

1. **Gleason, H. A.** 1922. On the relation between species and area. Ecology. **3**:158-162.

2. **Hurlbert, S.H.** 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology **52**:577-586.

3. **Sanders, H. L.** 1968. Marine benthic diversity: a comparative study. Am. Nat. **102**:243-282.

4. **Simberloff, D.** 1978. Use of rarefaction and related methods in ecology, p. 150-165. *In* K. L. Dickson, J. Cairns, Jr., and R. J. Livingston (ed.), Biological Data in Water Pollution Assessment: Quantitative and Statistical Analyses. Am. Soc. for Testing and Materials.

5. **Tipper, J. C.** 1979. Rarefaction and rarefiction - the use and abuse of a method in paleoecology. Paleobiology. **5**:423-434.

APPENDIX B

APPENDIX B

STATISTICAL ANALYSIS OF ARDRA DERIVED RANK-ABUNDANCE PATTERNS

The following represents an outline for the statistical analysis of ARDRA derived phylotype-abundance distributions in four steps, and its use is demonstrated for an analysis of a sequence of four soils of different age (Chapter 3). The tests and their use are:

1. **Chi-Square test**. Use the Chi-Square test to show that there is statistical evidence of a lack of homogeneity across time in the probabilities of SSU rDNA clones being in the various ranking groups.
2. **Kruskal-Wallis test.** Use the Kruskal-Wallis test to again provide statistical evidence of lack of homogeneity. When the null hypothesis is rejected, use Multiple Comparisons Procedure to pinpoint which of the possible 6 pairings of the 4 time periods are apparently different.
3. **Confidence intervals.** Give confidence intervals on the differences between the true portions in the most dominant phylotypes at different times. Then give similar

confidence intervals on the differences between the true portions of lower ranking phylotypes.

4. **Trendlines.** Create graphics and equations for trendlines fitted for the phylotype-abundance distribution curves.

To compare rank-abundance patterns from different soil samples we have to apply tests that are independent of population distributions and associated parameters, so-called nonparametric tests. This type of test becomes necessary since we cannot yet make any assumptions about the distribution of the population from which the samples are drawn. Both the Chi-Square and Kruskal-Wallis tests are non-parametric tests. They are different approaches for a parallel evaluation. I did not use ANOVA here since it compares averages and ignores the distribution aspects.

1. THE CHI-SQUARE TEST OF HOMOGENEITY

The Chi-Square test is designed to analyze the equality of proportions (Conover, 1980). The Chi-Square test statistic, χ^2, denotes the difference between observed and expected frequencies of an event. At $\chi^2=0$ the observed frequency of events is equal to the expected one. I applied the Chi-Square test to show that there is statistical evidence for a lack of homogeneity across different soil samples (*i.e.* across time) in the probabilities of being in the various ranking groups.

Contingency table

The observed data are sorted into a contingency table, a two way classification table where the rows represent ordered categories *c* and the columns symbolize the different populations *k* (Table 1) (Johnson et al., 1996).

Table 1: Contingency table to classify and rank different populations in categories

	c categories *J*				
k Populations *i*	Category 1:	Category 2:	...	Category c:	Row Total
Population 1	O_{11}	O_{12}	...	O_{1c}	n_1
Population 2	O_{21}	O_{22}	...	O_{2c}	n_I
Population 3	O_{31}	O_{32}	...	O_{3c}	n_3
...	...	...	...	...	...
Population *k*	O_{k1}	O_{k2}	...	O_{kc}	n_k
Column total t_j	t_1	t_2	...	t_c	Grand total *N*

k is the number of distinct populations (here: soil communities) from which samples have been drawn. *c* is the number of categories into which each sample has been classified. Note that the categories are ordered. The indices *i* and *j* stand for *i*, rows, and *j*, columns. n_i is the sample size of the sample from the i^{th} population. *Oij* is an observed count in population *j* that fall into the i^{th} category. The grand total *N* is defined as the total number of observations (sum of all n_i) of all cells in the table. t_j is the sum of observed counts across *k* populations within a category

Hypothesis

The objective is to determine if the observed proportions in each category are nearly the same for all populations (Conover, 1980).

Null hypothesis H_0: Homogeneity exists across the chronosequence between the four different clone populations of each clone library. The probabilities for all populations are the same in each category or ranking group. $H_o: p_1 = p_2 = p_3 = p_4$

Alternative hypothesis H_1: Nonhomogeneity across the time periods. Probabilities for all populations are different in each category or ranking group. $H_1: p_1 \neq p_2 \neq p_3 \neq p_4$

Test Statistic

The objective is to test whether the populations are homogeneous with respect to cell probabilities (Conover, 1980). In formula [1] the χ^2 statistic is calculated by comparing observed counts with expected counts.

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}; \qquad [1]$$

The indices *i* and *j* stand for *i*, rows, and *j*, columns. *Oij* is an observed count, and *Eij* is the expected cell frequency in any table cell *ij*. The expected cell frequency is defined as

$$E_{ij} = \frac{(\text{row total for row } i)(\text{column total for column } j)}{\text{grand total } N}; \qquad [2]$$

to determine the degrees of freedom for this test statistic we calculate

d.f. = (number of rows - 1)(number of columns -1). [3]

The region of rejection for the null hypothesis is defined by a cut off criterion α which can be found in tables for Chi-Square distribution: $\chi^2 \geq \chi_\alpha^2$ [4]

Example

As an example I calculated the test statistic for the chronosequence described in Chapter 3. The phylotypes were divided into three categories or ranking groups, the most dominant phylotype in each clone library, the second through the fifth most abundant phylotype, and finally all phylotypes in the sixth and higher position. The subdivision into the three described categories yielded the best comparative results. The clone libraries constructed for the four soil ages represent the populations in the contingency table (Table 2). The expected cell frequency E_{ij} was calculated using formula [2].

Table 2: Contingency table for four soil bacterial communities from soils of different age. Observed counts are classified into three ordered categories. The number in parentheses denote the expected cell frequencies. All data are derived from the 63% G+C fraction. The data for the two oldest soils (populations k_3 and k_4) are contributed by three replicates each.

Population:	Soil Age	Category 1: Phylotype 1	Category 2: Phylotype 2⇒5	Category 3: Phylotype 6 plus	Row Total
k_1	200 yrs	10 (20.6)	19 (16.4)	47 (39.0)	76
k_2	2,100 yrs	13 (20.9)	18 (16.6)	46 (39.5)	77
k_3	20,000 yrs	43 (69.2)	27 (55.0)	185 (130.9)	255
k_4	150,000 yrs	109 (64.3)	75 (51.1)	53 (121.6)	237
	Column total across time	175	139	331	645

Following equation [3], the test statistic will have six degrees of freedom. In a Chi-Square table we find $\chi_\alpha^2 = 12.59$ to be the limit at a typical level of significance of $p<0.05$ and six degrees of freedom (Johnson et al., 1996). We reject the null hypothesis H_0 if $\chi^2 \geq 12.59$. Following equation [1] the test statistic has a value of $\chi^2 = 139.2$. Since formula [4] is met we can reject the null hypothesis that there is homogeneity across soil age for the three ranking groups. Similar calculations lead to a rejection of the null hypothesis for the 35% G+C fraction, although a Chi-Square test result of 19.4 ($p<0.05$) in Chapter 3 indicated much less difference between the four soil ages.

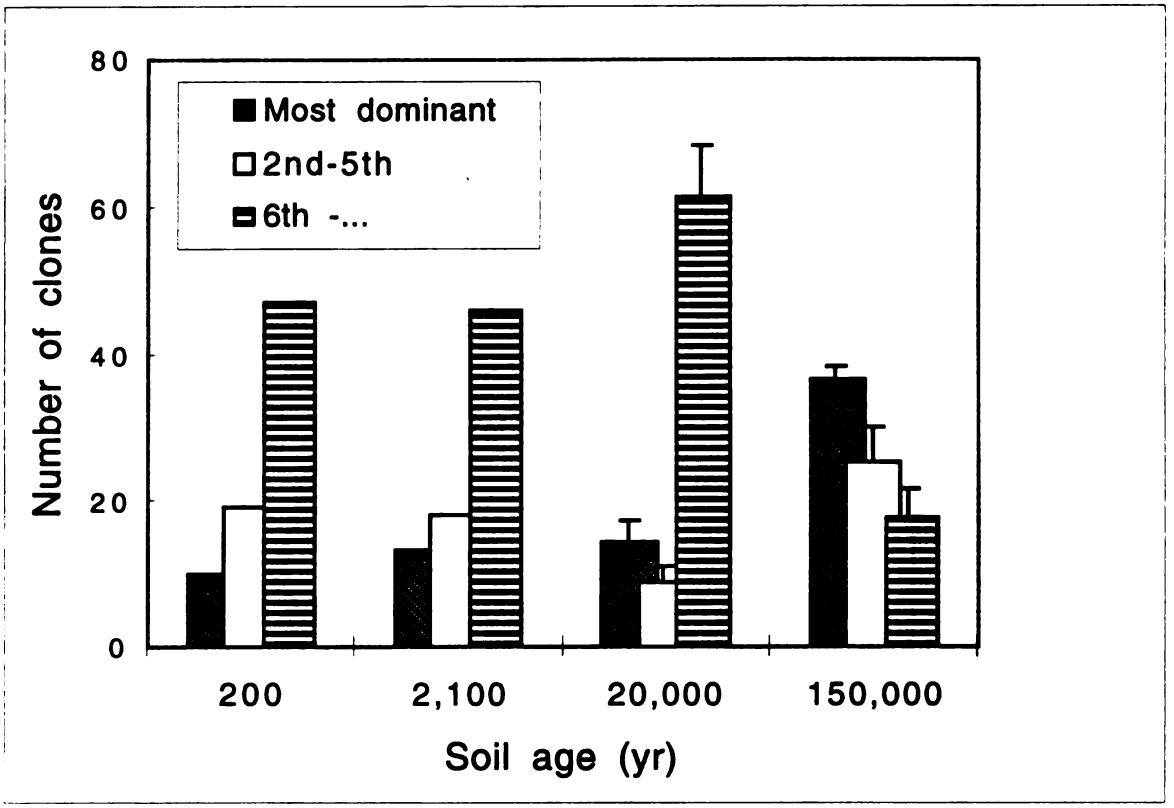


Figure 4. Clones per phylotype group for the 63% G+C fractions of a chronosequence of four soils. The two oldest soils are shown as the mean of three clone library replicates.

A graphical comparison of determined and expected cell values is shown in Figure 3 and Figure 4 for the 63% G+C fractions of the chronosequence. The high diversity contributed by rare members of each clone library (Fig. 3, ranking group 6th-...) has a strong effect on the Chi-Square value when the dominance pattern becomes more structured with increasing soil age (Fig. 4). The two oldest soils contribute the most to the value of Chi-Square, especially their extreme ratios of phylotype diversity over clone number (Fig. 5).

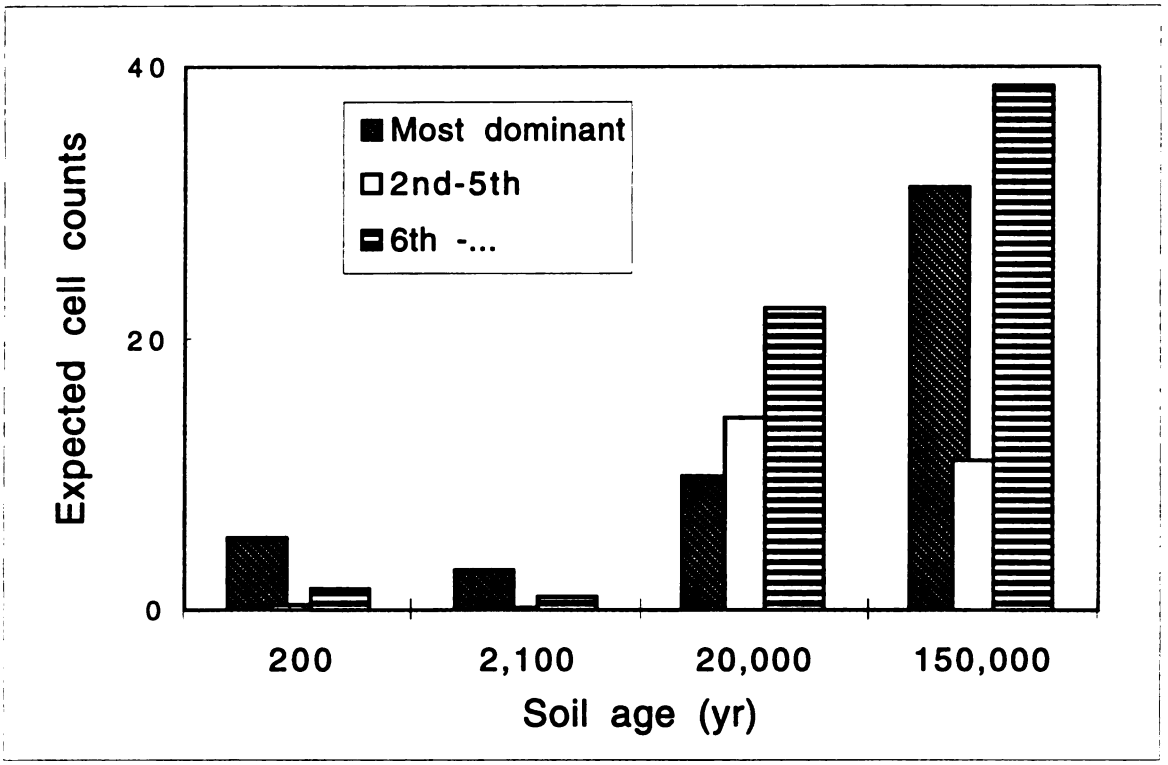


Figure 5. Contribution of expected cell frequencies in the contingency table (Table 2) to the value of Chi-Square for the 63% G+C fractions of a chronosequence of four soils (equation [2]).

Since the test statistic had a value of $\chi^2 = 139.2$ we rejected the null hypothesis H_0 of $\chi^2 \geq 12.59$.

The Chi-Square test successfully disproved the null hypothesis, and gave statistical evidence for a lack of homogeneity across soil samples of different age in the probabilities of being in the three ranking groups.

2. THE KRUSKAL-WALLIS TEST

The Kruskal-Wallis test is designed to analyze different, independent populations of more than two samples to make comparisons between them (Conover, 1980). It is a function of the ranks of all observations in the combined sample. After briefly explaining the test's theory I will demonstrate how to apply the Kruskal-Wallis test using actual data from the chronosequence of four soils in Chapter 3.

Assumptions. The following assumptions are made when the Kruskal-Wallis test is used (Conover, 1980):

1. All samples are random samples from their respective populations
2. There is independence within each sample, as well as mutual independence among the various samples
3. The measurement scale is ordinal or higher

Hypotheses:

Null hypothesis H_0: All of the population distribution functions are identical. *I. e.*, all four time periods of the chronosequence have rank-abundance distributions equal for dominance and diversity of the soil bacterial community.

Alternative hypothesis H_1: At least one of the populations tend to yield larger observations than at least one of the other populations. That is, at least one of the four time periods of the chronosequence is

characterized by a greater degree of dominance and diversity than at least one of the remaining time periods

Contingency table. Again a contingency table is made (Table 1), and ranks are computed for each cell. R_j is the average rank in the j[th] column, and t_j is again the column total,

$$\overline{R_1} = \frac{t_1 + 1}{2}; \quad \overline{R_2} = t_1 + \frac{t_2 + 1}{2}; \quad \overline{R_3} = t_1 + t_2 + \frac{t_3 + 1}{2}; \ldots\ldots \overline{R_c} = \sum_{j=1}^{c-1} t_i + \frac{(t_c + 1)}{2}; \qquad [1]$$

whereas the sum of ranks on the i^{th} row is

$$R_i = \sum_{j=1}^{c} O_{ij} \overline{R_j} \qquad [2]$$

where c is the number of categories.

Test Statistic T

Rejecting the null hypothesis. If the null hypothesis H_o is true, then the contributions of the respective ranks $\overline{R_1}$, $\overline{R_2}$, $\overline{R_3}$ in each sample should each be the same (proportional to sample size n_i). Differences between samples change contributions of $\overline{R_1}$, $\overline{R_2}$, $\overline{R_3}$, and increase the test statistic T considerably (Spiegel, 1994). After calculating the test statistic T the null hypothesis H_o is rejected if $T > \chi^2_{1-\alpha}$ (at the level of significance α). The critical region of a typical level of significance a = 0.05 corresponds to values of T greater than the 0.95 quantile of a Chi-Square random variable with k-1 degrees of freedom which can be found in tables for Chi-Square distribution (Johnson *et al.*, 1996).

$$T = \frac{1}{S^2}\left(\sum_{i=1}^{k}\frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right) \qquad [3]$$

The value of the test variability, S^2, is calculated from the equation

$$S^2 = \frac{1}{N-1}\left[\sum_{j=1}^{c} t_j\left(\overline{R_j}\right)^2 - \frac{N(N+1)^2}{4}\right] \qquad [4]$$

Multiple Comparisons Procedure. If H_o is rejected (and only then), the following multiple comparisons procedure can be employed to determine which pairs of populations tend to differ (Conover, 1980). The number of possible pairwise comparisons for k populations is

$$\binom{k}{2} = \left(\frac{k!}{2!2!}\right) \qquad [5]$$

Assume that the i^{th} and j^{th} populations are different, if the following inequality is satisfied:

$$\left|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right| > t_{1-(\alpha/2)}\sqrt{A} \qquad [6]$$

where

$$A = S^2\left(\frac{N-1-T}{N-k}\right)\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$$

$R_i$and R_j are the rank sums of the two samples, $t_{1-a/2}$ is the (1- a/2) quantile of the t-distribution with N-k degrees of freedom (see tables for t distribution (Johnson *et al.*, 1996)).

This procedure is repeated for all pairs of populations (see Table 4 below as an example for k=4 populations).

Example

As an example we calculate the test statistic for the chronosequence described in Chapter 3, four different soil communities (populations k=4), each representing a different soil age; n_1, n_2, n_3, n_4. are the number of clones investigated for each soil. All data are derived from the 63% G+C fraction. Again, as for the Chi-Square test above, all observations are pooled in a contingency table, and the phylotypes were divided into three categories or ranking groups. The most dominant phylotype in each clone library is treated as another rank, $\overline{R_1}$, the second to fifth most dominant clones are treated as the same rank $\overline{R_2}$, and finally all of the remaining sixth and higher clones as grouped as a third rank, $\overline{R_3}$.

The Kruskal-Wallis Test Step by Step

1. Compute average ranks for the different classification levels. Then compute row ranks using both: (a) average ranks from step 1,
   and (b) all observed counts
2. Compare the error variance S^2
3. Compute the Kruskal-Wallis test statistic T
4. Determine, if T lies within the 'tail' of χ^2
5. Reject homogeneity between populations (soil samples)
6. Use Multiple Comparisons Procedure to compare populations pairwise.

1. Computing the ranks using a contingency table (Table 1).

First, compute the average ranks $\overline{R_1}$, $\overline{R_2}$, $\overline{R_3}$ for each of the respective ranking groups using equation [1].

$$\overline{R_1} = \frac{175+1}{2} = 88; \quad \overline{R_2} = 88 + \frac{139+1}{2} = 158; \quad \overline{R_3} = 88 + 158 + \frac{331+1}{2} = 412;$$

Second, compute the rank totals $\overline{R_i}$'s for all rows using equation [2].

$R_1 = 10(88) + 19(158) + 47(412) = 23{,}246$; same calculation for R_2 through R_4

2. Determining the test variability S^2. The total number of clones in all samples is N=645. The number of categories is c=3. The value of S^2 is calculated using equation [4]: S^2=28988.93

3. Rejecting the null hypothesis. Using the result for test variability S^2 in equation [3] the value of the test statistic T is determined as T=125.4. For k-1=3 degrees of freedom, the quantile found in the Chi-Square distribution table for the common level of significance α=0.05 is 7.81. Since the test statistic T is of higher value than 7.81, the null hypothesis H_0 of identical populations and hence, identical population distributions, is rejected. This allows us to continue with the Multiple Comparisons Procedure.

4. Multiple Comparisons Procedure. Equation [6] denotes the inequality that has to be satisfied if two populations i and j are called significantly different. Following the

equation N-k the test statistic will have 641 degrees of freedom. From the tables for t distribution, the standard variable $t_{1-(\alpha/2)}$ quantile with N-k =641 degrees of freedom is $t_{1-(\alpha/2)}$ =1.96 for a typical α=0.05. According to equation [5] we can make $\binom{4}{2}$=6 different pairwise comparisons of soil samples of different age (populations) in the group of four. R_i and R_j are the rank sums of two samples each in a pairwise comparison (Table 3). The value of constant b is obtained using equations [3] and [4]

$$\sqrt{b} = \sqrt{\left(S^2 \frac{N-1-T}{N-k}\right)} = 153.15.$$

Table 3. Results for the Multiple Comparisons Procedure of four different soil bacterial communities.

Populations i and j	$\left\lvert \frac{R_i}{n_i} - \frac{R_j}{n_j} \right\rvert$	$t_{1-(\alpha/2)}\sqrt{b}\sqrt{\frac{1}{n_i} - \frac{1}{n_j}}$	Statement[a]
1 and 2	10.79	48.53	False
1 and 3	19.34	39.23	False
1 and 4	144.33	39.57	True
2 and 3	30.13	39.03	False
2 and 4	133.54	39.37	True
3 and 4	163.67	27.08	True

[a] That the compared populations have a different distribution

The Kruskal-Wallis test will indicate if there is a statistically significant difference between any two of the means. In half of the cases the second column

exceeds the third column. The Multiple Comparisons Procedure shows that the rank-abundance distribution of the Kohala clone library at 63% G+C is significantly different from the rank-abundance distributions of all three younger soils ($p<0.05$). There is, however, no significant difference amongst the three younger soils. The second column exceeds the third column in all of the cases for the 35% G+C fractions of the chronosequence. The Multiple Comparisons Procedure shows that the rank-abundance distribution of the four clone libraries from the chronosequence 35% G+C fraction shows no significant difference amongst the four soils ($p<0.05$). However, at a coarser level of significance of $p<0.2$, in one third of all cases the second column exceeds the third column. The Multiple Comparisons Procedure reveals that the rank-abundance distribution of the Ola'a and the Laupahoehoe clone library from the 35% G+C fraction are significantly different from the rank-abundance distributions of the oldest (Kohala) soil. There is, however, no significant difference amongst the three younger soils and between the youngest and the oldest soil.

3. CONFIDENCE INTERVALS

To statistically support differences in relative abundance of the most dominant phylotypes between soils in the chronosequence, I used confidence intervals. A confidence interval is an estimate. Here we estimate the true probability of a single *rrn* clone drawn from the soil and belonging to some dominance rank. The estimate is the fraction of bacteria in that rank for a specific sample. Suppose p_1 is the unknown probability of being in the most dominant group for time period 1, and further suppose p_2 is the unknown probability of being in the most dominant group for time period 2. If a 95% confidence interval estimate is (0.07, 0.19) you are "95% confident" that the most dominant group comprised at least 7% more of the population in Time period 1 than the most dominant group in Time period 2 comprised of the population in the second time period. If the confidence interval p_1-p_2 contains only positive values we can be confident that p_1 is bigger than p_2. The confidence interval shows a lack of homogeneity if it does not include 0, *i.e.* there is no statistically significant difference.

The Chi-Square- and the Kruskal Wallis-tests test for lack of homogeneity at a coarser level than confidence intervals. In these tests rank distributions between two communities are viewed as two separate entities and hence different from each other. The confidence interval estimates on p_1- p_2 are more specific. They allow one to pick a certain dominance rank, and to estimate how much the contribution of that specific rank to the whole population differs across two time groups.

Calculate the confidence interval for p_1- p_2, where p_1 is the true probability of being in the most dominant rank group in a certain clone library 1, while p_2 is the true

probability of being in the most dominant rank group in a certain clone library 2. An approximate $100(1-\alpha)\%$ confidence interval for p_1- p_2 is (Johnson *et al.*, 1996):

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \qquad [1]$$

for $0<p_i<1$, and large sample sizes n_1 and n_2 from the two populations that are compared.

In the example of the chronosequence I assume for the probability multiplier $z_{\alpha/2} = 1.96$ which corresponds to a level of significance of p <0.05. As an example I formulated the result for the 63% fractions of the chronosequence (Chapter 4): I am 95% confident that the most dominant group comprised at least 7% more of the clone library at the oldest site, Kohala, than the most dominant group at the youngest soil, Thurston. Also, I am 95% confident that the most dominant group comprised at least 5% more of the clone library at the 20,000 year old Laupahoehoe than the most dominant group at the youngest soil (Thurston). The confidence intervals of factorial combinations of other soils did not show significant differences.

4. TRENDLINES

Graphical trendlines are regression curves calculated to fit the phylotype-abundance distribution curves for the respective soil ages (Fig.6) (Spiegel, 1994). These trendlines are based on the formula of a geometric curve $y = \alpha x^{\beta}$, where, in a rank-abundance distribution, y-values denote phylotype abundance while x-values represent the rank of the respective phylotype. In comparing curves, the factor α is influenced by the number of dominant phylotypes, while factor β is dependent on the frequency of rare phylotypes, and hence a measure of phylotype diversity. Beta gains influence with increasing values of x. Trendlines can easily be calculated using the software program Excel 5.0 (Microsoft Coorporation) (Fig. 6).

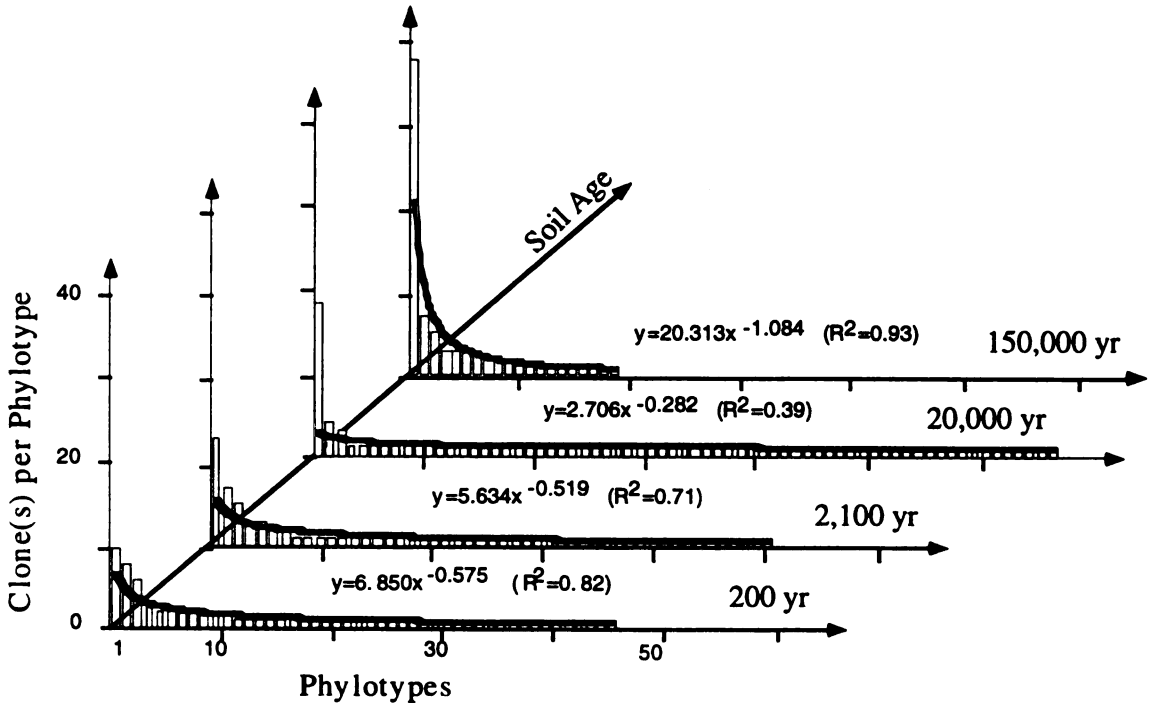


Figure 6. Regression trendlines and their equations calculated to fit the rank-abundance profiles for the four soils of the chronosequence. All graphics describe the 63% G+C fractions for each respective soil site.

R^2 is the percentage of all variability in the frequencies which is accounted for by the fitted trend-line in each case (Johnson *et al.*, 1996). Here, R^2 is a descriptive statistic, and I am not assuming any underlying probability distribution. As an example I sorted the different factors for the four soils of the chronosequence described in Chapter 4 (Table 4).

Table 4. Comparison of factors α and β in the regression equation $y = \alpha x^{\beta}$ for rank-abundance distributions of four soils of different age (63% G+C fraction)

Soil site	Soil age (y)	Alpha	Beta	R^2
Thurston	200	6.850	-0.575	0.816
Ola'a	2,100	5.634	-0.519	0.707
Laupahoehoe	20,000	2.706	-0.282	0.390
Kohala	150,000	20.313	-1.084	0.928

In this study trendlines are merely used as a graphical approach to show the trend of diversity *versus* dominance across age differences, and to express this trend numercially. There is, however, no statistical merit to trendlines in this study, since we cannot actually compare them between different soils. Regression analysis treats the series of phylotypes as a continuum rather than as a series of independent single events that it is. The strong dependence of residuals (individual deviations from the regression

line) is thus not taken into account. Therefore, I did not calculate confidence intervals or tests of hypotheses based on trendlines. However, trendlines are useful in determining the similarity of replictates (*e.g.* 20,000 year series in Table 5). Table 5 summarizes the results for the trendline equations for all soils investigated.

Table 5. Comparison of factors α and β in the geometric regression equation $y = \alpha x^{\beta}$ for rank-abundance distributions of all soils investigated (63% G+C fraction)

Soil Age (y)	Vegetation[a]	Alpha	Beta	R^2
200	F	6.850	-0.5746	0.8160
2,100	F	5.634	-0.5189	0.7069
20,000	F- Rep A	2.195	-0.2171	0.3853
20,000	F- Rep B	2.706	-0.2816	0.3899
20,000	F- Rep C	3.654	-0.3742	0.6101
150,000	P- Rep A	9.564	-0.6261	0.7163
150,000	P- Rep B	1.831	-0.7052	0.7053
150,000	F	20.310	-1.0840	0.9278
150,000	F- Rep A	30.505	-1.3242	0. 9548
150,000	F- Rep B	9.564	-1.4311	0. 9443

[a] F stands for forest, P for pasture soils. Rep stands for replicate.

SUMMARY AND RECOMMENDATIONS

For this summary we use again the example of the Hawaiian chronosequence of four soils (Chapter 4). The variable in this soil series is soil age. For each soil age we created rank-abundance profiles of phylotypes derived from SSU rDNA clones. The statistical analysis should determine

(i) if the rank-abundance profiles are different from each other [Chi-Square test], and if so,

(ii) which of all possible pairings of the four soil ages are different [Kruskal-Wallis test], and finally,

(iii) amongst those different pairings which rank group of SSU rDNA clones contributed significantly to that difference [Confidence intervals].

This sequence of statistical analyses is appropriate if one wants to compare independent profiles for their significant differences. The observed data have to be arranged in increasing order according to some property such as quality or value (*e.g.* rank); samples can have different sizes (*e.g.* different clone libraries) (Conover, 1980). The Chi-Square test is less powerful than the Kruskal-Wallis test. In addition to showing that data collections (*e.g.* rank-abundance profiles) are different the Kruskal - Wallis test statistic is also a function of the ranks of the observations. However, it involves more effort than the Chi-Square test. Both, the Chi-Square test as well as the Kruskal-Wallis test are valid only for large samples (Johnson et al., 1996). Cell frequencies no smaller than ≥ 5 are normally required. Replicate analyses of soil from

the same site are necessary to determine if within site variation between rank-abundance profiles is significantly smaller than between site variation.

For a quick descriptive comparison of the four rank-abundance profiles without any statistical assertion trendlines can be made. Trendlines can however be useful to compare two replicates. The calculations necessary for all tests mentioned are quickly done with any advanced spreadsheet software (*e.g.* Excel by Microsoft) or with common statistical software packages (*e.g.* Jump by SSPS).

The statistical methods outlined in this Appendix go partly beyond what was used for this study. Their comprehensiveness should help colleagues who work on similar questions in finding the appropriate approach to particular problems with rank-abundance based community studies.

ACKNOWLEDGMENTS

I thank James Demopolos (Department of Statistics, MSU) for his help in preparing Appendix B.

REFERENCES

1. **Conover, W. J.** 1980. Practical Nonparametric Statistics (2nd ed.), pp. 229-234. John Wiley & Sons, New York, New York, U.S.A.

2. **Johnson, R. A. and G. K. Bhattacharyya.** 1996. Statistics: Principles and Methods (3rd ed.). pp. 554-558. John Wiley & Sons, New York, New York, U.S.A.

3. **Spiegel, M. R.** 1994. Schaum's Outline Series of Theory and Problems of Statistics, 2nd ed. McGraw-Hill, Inc., New York, New York, U.S.A.

Recommended reading to expand on the statistical analysis:

Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren (eds.). 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, England.

BIBLIOGRAPHY

LIST OF REFERENCES

1. **Alexander, M.** 1977. Introduction to Soil Microbiology. John Wiley & Sons, New York, New York, USA.

2. **Amann, R., N. Springer, W. Ludwig, H.-D. Görtz, and K. H. Schleifer.** 1991. Identification *in situ* and phylogeny of uncultured bacterial endosymbionts. Nature **351:**161-164.

3. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143-169.

4. **Arah, J. R. M., and K. A. Smith.** 1985. Anaerobic micro-environments in soil and the occurrence of anaerobic bacteria, p. 247-261. *In* V. Jensen, A. Kjøller, and L. H. Sørensen. (ed.), Microbial Communities in Soil Microbiology. Elsevier Applied Science Publishers, London, England.

5. **Atlas of Hawaii**. 1983. . University of Hawaii Press, Honolulu, Hawaii, USA.

6. **Betzer, P. R., K. L. Carder, R. A. Duce, J. T. Merill, N. W. Tindale, M. Uematsu, D. K. Costello, R. W. Young, R. A., Feely, J. A. Breland, R. E. Bernstein, and A. M. Greco.** 1988. Long-range transport of giant mineral aerosol particles. Nature **336:**568-571.

7. **Boivin-Jahns, V., A. Bianchi, R. Ruimy, J. Garcin, S. Daumas, and R. Christen.** 1995. Comparison of phenotypical and molecular methods for the identification of bacterial strains isolated form a deep subsurface environment. Appl. Environ. Microbiol. **61:**3400-3406.

8. **Borneman, J., P. W. Skroch, K. M. O'Sullivan, J. A. Paulus, N. G. Rumjanek, J. L. Jansen, J. Nienhus, and E. W. Triplett.** 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. Appl. Environ. Ecol. **62:**1935-1943.

9. **Borneman, J., and E. W. Triplett.** 1997. Molecular microbial diversity in soils from Eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl. Environ. Microbiol. **63:**2647-2653.

10. **Carlquist, S. C.** 1980. Hawaii, A Natural History. Pacific Tropical Botanical Garden, Honolulu, Hawaii, USA.

11. **Clague, D. A., and G. B. Dalrymple.** 1981. The Hawaiian-Emperor volcanic chain: geologic evolution, p.5-54. *In* R. W. Decker, T. C. Wright, and P. H. Stauffer (ed.), Volcanism in Hawaii. United States Geological Survey, Washington, D.C., USA.

12. **Crews, T. E., K. Kitayama, J. H. Fownes, R. H. Riley, D. A. Herbert, D. Mueller-Dombois, and P. M. Vitousek.** 1995. Changes in soil phosphorous fractions and ecosystem dynamics across a long chronosequence in Hawaii. Ecology **76:**1407-1424.

13. **Cuddihy, L. W., and C. P. Stone.** 1990. Alteration of native Hawaiian Vegetation: effects of humans, thier activities and introductions. Unversity of Hawaii Press, Honolulu, Hawaii, USA.

14. **D'Aquila, R. T., L. J. Bechtel, J. A. Videler, J. J. Eron, P. Gorczca, and J. C. Kaplan.** 1991. Maximizing sensitivity and specificity of PCR by preamplification heating. Nucleicl. Acids Res. **19:**3749.

15. **Deharveng, L., and A. Bedos.** 1993. Factors influencing diversity of soil Collembola in a tropical mountain forest (Doi Inthanon, Northern Thailand). *In* M. G. Paoletti, W. Foissner, D. Coleman (ed.), Soil Biota, Nutrient Cycling, and Farming Systems. Lewis Publishers, Boca Raton, Florida, USA.

16. **Department of Geography,** University of Hawaii. 1983. Atlas of Hawaii, 2nd ed. University of Hawaii Press, Honolulu, Hawaii, USA.

17. **Ellenbroek, F. M., and T. E. Cappenberg.** 1991. DNA synthesis and tritiaited thymidine incorporation by heterotrophic freshwater bacteria in continuous culture. Appl. Environ. Microbiol. **57:**1675-1682.

18. **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. **61:**2798-2801.

19. **Ferris, M. J., A. L. Ruff-Roberts, E. D. Kopczynski, M. M. Bateson, and D. M. Ward.** 1996. Enrichment culture and microscopy conceal diverse thermophilic Synechococcus populations in a single hot spring microbial mat habitat. Appl. Environ. Microbiol. **62:**1045-1050.

20. **Finney, M.** 1993. Molecular cloning of PCR products, p. 15.7.1-15.7.6. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current Protocols in Molecular Biology, vol. 2. Greene Publishing Associates and Wiley-Interscience, New York, New York, USA.

21. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209:**457-463.

22. **Fulthorpe, R. R., A. N. Rhodes, and J. M. Tiedje.** 1996. Pristine soils mineralize 3-chlorobenzoate and 2,4-dichlorophenoxyacetate via different microbial populations. Appl. Envrion. Microbiol. **62:**1159-1166.

23. **Giambelluca, T. W., M. A. Nullet, and T. A. Schroeder.** 1986. Rainfall Atlas of Hawaii. Department of Land and Natural Resources, Honolulu, Hawaii, USA.

24. **Giovannoni, S., T. Britschgi, C. L. Moyer, and K. G. Field.** 1990. Genetic diversity in Sargasso sea bacterioplankton. Nature **345:**60-63.

25. **Glasel, J.** 1995. Validity of nucleic purities monitored by 260nm/280nm absorbance ratios. BioTechniques **18:**62-63.

26. **Gray, M. W., D. Sankoff, and R. J. Cedergren.** 1984. On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res. **12:**5837-5852.

27. **Harris, D.** 1994. Analyses of DNA extracted from microbial communities, p. 111-118. *In* K. Ritz, J. Dighton, and K. E. Giller (ed.), Beyond the Biomass. Compositional and Functional Analysis of Soil Microbial Communities. John Wiley & Sons, Chichester, England.

28. **Hawaii Scientific Drilling Project Team.** 1996. Hawaii scientific drilling project: summary of preliminary results. GSA Today **6:**1-8

29. **Herbert, D. A., and J. H. Fownes.** 1995. Phosphorus limitation of forest leaf area and net primary productivity on a weathered tropical soil. Biochemistry **29:**223-235.

30. **Hermjakob, H.** The rapid identification by fragment length evaluation. [Online] Available http://bibiserv.techfak.uni-bielefeld.de/RIFLE/E.welcome.html, 02/26/1997.

31. **Holben, W. E.** 1994. Isolation and purification of bacterial DNA from soil, p. 727-751. *In* R. Weaver *et al.* (ed.), Methods of Soil Analysis, Part 2. Microbiological and Biochemical Properties. Soil Science Society of America Book Series no. 5. Soil Science Society of America, Madison, Wisconsin, USA.

32. **Holben, W. E., and D. Harris.** 1995. DNA-based monitoring of total bacterial community structure in environmental samples. Mol. Ecol. **4:**627-631.

33. **Hotchkiss, S. C., and J. O. Juvik.** 1993. Pollen record from Kaau Crater, Oahu, Hawaii: evidence for a dry glacial maximum. Bull. Ecol. Soc. Amer. **74:**282.

34. **Hurlbert, S. H.** 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology **52:**577-586.

35. **Huston, M. A.** 1994. Biological Diversity. Cambridge University Press, Cambridge, England.

36. **Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White** (ed.). 1990. PCR Protocols. Academic Press, San Diego, California, USA.

37. **Jackson, M. L., T. W. M. Levelt, J. K. Syers, R. W. Rex, R. N. Clayton, G. D. Sherman, and G. Uehara.** 1971. Geomorphological relationships of troposherically derived quartz in the soils of the Hawaiian Islands. Soil Sci.Soc.Am. Proc. **35:**515-525.

38. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21-132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, New York, New York, USA.

39. **Kim, J., J. R. Cole, E. Torng, and S. Pramanik.** 1996. Presented at the Presented at the Intelligent Systems for Molecular Biology '96. Proceedings of the Fourth International Conference on Computational Biology, USA.

40. **Kitayama, K., and D. Mueller-Dombois.** 1995. Vegetation changes along gradients of long-term soil development in the Hawaiian montane rainforest zone. Vegetatio **120:**1-20.

41. **Kitayama, K.** 1996. Soil nitrogen dynamics along a gradient of long-term soil development in a Hawaiian wet montane rainforest. Plant and Soil **183:**253-262.

42. **Komatsoulis, G. A. and M. S. Waterman.** 1997. A new computational method for detection of chimeric 16S rRNA artifacts generated by PCR amplification from mixed bacterial populations. Applied and Environmental Microbiol. **63:**2338-2346.

43. **Kopczynski, E. D., M. M. Bateson, and D. M. Ward**. 1994. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. **60:**746-748.

44. **Kuske, C. R., S. M. Barns, and J. D. Busch**. 1997. Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. Appl. Environ. Microbiol. **63:**3614-3621.

45. **Labarca, C., and K. Paigen**. 1980. A simple, rapid and sensitive DNA assay procedure. Anal. Biochem. **102:**344-352.

46. **Lee, S.-Y., J. Bollinger, D. Bezdicek, and A. Ogram.** 1996. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitiative PCR method. Appl. Environ. Microbiol. **62:**3787-3793.

47. **Levine, S. H**. 1976. Competitive interactions in ecosystems. Am. Nat. **110:**903-910.

48. **Liesack, W., H. Weyland, and E. Stackebrandt.** 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed culture of strictly barophylic bacteria. Microb. Ecol. **21:**191-198.

49. **Liesack, W. and E. Stackebrandt**. 1992. Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. **174:**5072-5078.

50. **Liesack, W., P. H. Janssen, F. A. Rainey, N. L. Ward-Rainey, and E. Stackebrandt.** 1997. Microbial diversity in soil: the need for a combined approach using molecular and cultivation techniques, p.375-439. *In* J. D. van Elsass, J. T. Trevors and E. M. H. Wellington (ed.), Modern Soil Microbiology. Marcel Dekker, Inc., New York, New York, USA.

51. **Ludwig, W., S. H. Bauer, M. Bauer, I. Held, G. Kirchhof, R. Schulze, I. Huber, S. Spring, A. Hartmann, and K. H. Schleifer.** 1997. Detection and *in situ* identification of representatives of a widely distributed new bacterial phylum. FEMS Microbiol. Let. **153:**181-190.

52. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and and C. R. Woese.** 1996. The RDP (Ribosomal Database Project). Nucleic Acids Res. **24:**82-85.

53. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a labratory manual. Cold Spring Harbor Labratory, Cold Spring Harbor, New York, USA.

54. **Mead, D. A., N. K. Pey, C. Herrnstadt, R. A. Marcil, and L. M. Smith.** 1991. A universal method for the direct cloning of PCR amplified nucleic acid. Bio/Technology **9:**657-663.

55. **Merrill, J. T.** 1989. Atmospheric long-range transport to the Pacific Ocean, p. 15-50. *In* J. P. Riley and R. Chester (ed.), Chemical oceanography, vol. 10. Academic Press, New York, New York, USA.

56. **Moré, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse, and E. L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl. Microbial. Ecol. **60:**1572-1580.

57. **Moyer, C. L., F. C. Dobbs, and D. M. Karl.** 1994. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. **60:**871-879.

58. **Muyzer, G.** 1995. Protocols for the DGGE, p. 8-14. *In* K. Smalla and G. Muyzer (ed.) EU-Workshop on the Application of DGGE and TGGE in Microbial Ecology, Braunschweig, Germany. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig, Germany.

59. **Nannipieri, P., S. Greco, and B. Ceccanti.** 1990. Ecological significance of the biological activity in soil, p. 293-356. *In* J.-M. Bollag and G. Stotzky (ed.), Soil Biochemistry, vol. 6. Marcel Dekker, Inc., New York, USA.

60. **Nanzyo, M., R. Dahlgren, and S. Shoji.** 1993. Chemical characteristics of volcanic ash soils, p. 145-187. *In* M. Nanzyo, S. Shoji, and R. Dahlgren (ed.), Volcanic Ash Soils. Elsevier, Amsterdam, The Netherlands.

61. **Nübel, U., B. Engelen, A. Felske, J. Snaidr, A. Wieshuber, R. I. Amann, W. Ludwig, and H. Backhaus.** 1996. Sequence heterogeneities of genes encoding 16S rRNAs in Paenibacillus polymyxa detected by temperature gradient gel electrophoresis. J. Bacteriol. **178:**5636-5643.

62. **Nüsslein, K., and J. M. Tiedje.** 1998. Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from DNA fractionated on the basis of its guanine and cytosine composition. Appl. Environ. Microbiol. **64:**1283-1289.

63. **Odum, E. P.** 1998. Productivity and biodiversity: a two-way relationship. Bull. Ecol. Soc. Am. **79:**125.

64. **Ogram, A., G. S. Sayler, and T. Barkay.** 1987. The extraction and purification of microbial DNA from sediments. J. Microbiol. Methods **7:**57-66.

65. **Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl.** 1986. Microbial ecology and evolution: a ribosomal RNA approach. Ann. Rev. Microbiol. **40:**337-365.

66. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:**41-48.

67. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176:**1-6.

68. **Paul, E. A. and F. E. Clark.** 1996. Soil Microbiology and Biochemistry, 2nd ed. Academic Press, San Diego, California, USA.

69. **Peck, T. R., D. Megel, K. Eik, , D. A. Whitney, D. Warncke, R. C. Munter, J. R. Brown, D. Knudsen, W. C. Dahnke, D. J. Eckert, D. Beegle, P. E. Fixen, and E. E. Schulte.** 1988. Recommended Chemical Soil Test Procedure for the North Central Region. North Dakota Agricultural Experiment Station, North Dakota State University, Fargo, North Dakota.

70. **Reiners, W. A., A. F. Bouwman, W. F. J. Parsons, and M. Keller.** 1994. Tropical rainforest conversion to pasture: changes in vegetation and soil properties. Ecol. Appl. **4:**363-377.

71. **Reysenbach, A.-L., L. J. Giver, G. S. Wickham, and N. R. Pace.** 1992. Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. **58:**3417-3418.

72. **Rheims, H., C. Sproer, F. A. Rainey, and E. Stackebrandt.** 1996. Molecular biological evidence for the occurrence of uncultured members if the actinomycete line of descent in different environments and geographical locations. Microbiology **142:**2863-2870.

73. **Riley, R. H., and P. M. Vitousek.** 1995. Nutrient dynamics and nitrogen trace gas flux during ecosystem development in montane rain forest. Ecology **76:**292-304.

74. **Robinson-Cox, J. F., M. M. Bateson, and D. M. Ward.** 1995. Evaluation of the nearest-neighbor methods for detection of chimeric small-subunit rRNA sequences. Appl. Environ. Microbiol. **61:**1240-1245.

75. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. G. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Ehrlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487-491.

76. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406-425.

77. **Schmidt, T. M., E. F. DeLong, and N.R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. **173:**4371-4378.

78. **Schmidt, T. M., B. Pace, and N.R. Pace.** 1991. Detection of DNA contamination in Taq polymerase. BioTechniques **11:**176-177.

79. **Simberloff, D.** 1978. Use of rarefaction and related methods in ecology, p. 150-165. *In* K. L. Dickson, J. Cairns, Jr., and R. J. Livingston (ed.), Biological Data in Water Pollution Assessment: Quantitative and Statistical Analyses. American Society for Testing and Materials, West Conshohocken, Pennsylvania, USA.

80. **Sørheim, R., V. L. Torsvik, and J. Goksøyr.** 1989. Phenotypical divergences between populations of soil bacteria isolated on different media. Microb. Ecol. **17:**181-192.

81. **Soil Survey Staff.** 1973. Soil survey of the islands of Kauai, Oahu, Maui, Molokai, and Lanai, state of Hawaii. U.S. Government Printing Office, Washington, D.C., USA.

82. **Spaans, E. J. A., G. A. M. Baltissen, J. Bouma, R, Miedema, A. L. E. Lansu, D. Schoonderbeek, and W. G. Wielemaker.** 1989. Changes in physical properties of young and old volcanic surface soils in Costa Rica after clearing of tropical rain forest. Hydrogeol. Processes **3:**383-392.

83. **Stackebrandt, E., W. Liesack, and B. M. Goebel.** 1993. Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J. **7:**232-236.

84. **Strunk, O., O. Gross, B. Reichel, M. May, S. Hermann, N. Stuckmann, B. Nonhoff, M. Lenke, A. Ginhart, T. Ludwig, A. Bode, K.-H. Schleifer, and W. Ludwig.** 1998. Arb: A software environment for sequence data. Nucleic Acids Res., in press. Also, http://www.mikro.biologie.tu-muenchen.de/ARB.

85. **Suzuki, M. T., E. B. Sherr, and B. F. Sherr.** 1993. DAPI direct counting underestimates bacterial abundances and average cell-size compared to AO direct counting. Limnol. Oceanogr. **38:**1566-1570.

86. **Suzuki, M. T., and S. J. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Micrbiol. **62:**625-630.

87. **Swofford, D. L.** 1990. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0. Illinois Natural History Survey, Champaign, Illinois, USA.

88. **Tiedje, J. M., J.-Z. Zhou, K. Nüsslein, C. L. Moyer, and R. R. Fulthorpe.** 1997. Extent and patterns of soil microbial diversity, p. 35-41. *In* M. T. Martins *et al.* (ed.) Progress in Microbial Ecology. Brazilian Society for Microbiology, Sao Paulo, Brazil.

89. **Torsvik, V., J. Goksoyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. **56:**782-787.

90. **Torsvik, V., J. Goksøyr, F. L. Daae, R. Sørheim, J. Michalsen, and K. Salte.** 1994. Use of DNA analysis to determine the diverstiy of microbial communities, p.39-48. *In* K. Ritz, J. Dighton and K. E. Giller (ed.), Beyond the Biomass. John Wiley & Sons, Chichester, England.

91. **Tsai, Y.-L., and B. H. Olsen**. 1991. Rapid method for direct extraction of DNA from soils and sediments. Appl. Environ. Microbiol. **57:**1070-1074.

92. **Ueda, T., Y. Suga, and T. Matsuguchi.** 1995. Molecular phylogenetic analysis of a soil microbial community in a soybean field. Eur. J. Soil Sci. **46:**415-421.

93. **Vitousek, P. M**. 1982. Nutrient cycling and nutrient use efficiency. Am.Nat. **119:**553-572.

94. **Vitousek, P. M., and L. R. Walker**. 1989. Biological invasion by Myrica faya in Hawaii: plant demography, nitrogen fixation, and ecosystem effects. Ecol. Monographs. **59:**247-265.

95. **Vitousek, P. M., D. R. Turner, and K. Kitayama.** 1995. Foliar nutrients during long-term soil development in Hawaiian montane rain forest. Ecology **76:**712-720.

96. **Vitousek, P. M., O. A. Chadwick, T. E. Crews, J. H. Fownes, D. M. Hendricks, and D. Herbert.** 1997. Soil and ecosystem develoomment across the Hawaiian Islands. GSA Today, in press.

97. **Wada, K.** (ed.). 1989. Allophane and Imogolite, vol. No.1, Madison, Wisconsin, USA.

98. **Wang, G. C.-Y., and Y. Wang.** 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Appl. Environ. Microbiol. **63:**4645-4650.

99. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature **345:**63-65.

100. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature, p. 219-286. *In* K. C. Marshall (ed.), Advances in Microbial Ecology, vol. 12. Plenum Press, New York, New York, U.S.A.

101. **Weisblum, B., and E. Haenssler.** 1974. Fluorometric properties of the bisbenzimidazole derivative Hoechst 33258, a fluorescent probe specific for AT concentration in chromosomal DNA. Chromosoma **46:**255-260.

102. **Weisburg, W. G., J. G. Tully, D. L. Rose, J. P. Petzel, H. Oyaizu, D. Yang, L. Mandelco, J. Sechrest, T. G. Lawerence, J. Van Etten, J. Maniloff, and C. R. Woese.** 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. **171:**6455-6467.

103. **Weisburg, W. W., S. M. Barns, D. A. Pelletier, and D. J. Lane.** 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. **173:**697-703.

104. **Weller, R., J. W. Weller, and D. M. Ward.** 1991. 16S rRNA sequences of uncultivated hot spring cyanobacterial mat inhabitiants retrieved as randomly primed cDNA. Appl. Environ. Microbiol. **57:**1146-1151.

105. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221-271.

106. **Woese, C. R., O. Kandler, and M. L. Wheelis**. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. **87:**4576-4579.

107. **Wright, T. C., and R. T. Helz.** 1987. Recent advances in Hawaiian petrology and geochemistry, p.625-640. *In* R. W. Decker, T. C. Wright, and P. H. Stauffer (ed.), Volcanism in Hawaii. U.S. Geological Survey Professional Paper 1350, United States Government Printing Office, Washington, D.C., USA.

108. **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. Rapid method for the recovery of DNA from soils of diverse composition. Appl. Environ. Microbiol. **62:**316-322.

109. **Zhou, J., M. E. Davey, J. B. Figueras, E. Rivkina, D. Gilichinsky, and J. M. Tiedje.** 1997. Phylogenetic diversity of a bacterial community determined from Siberian Tundra soil DNA. Microbiology **143:**3913-3919.