THE TERMINAL-RESTRICTION FRAGMENT LENGTH POLYMORPHISM ASSAY AND ITS USE IN DETERMINING BACTERIAL COMMUNITY SUCCESSION AT HYDROTHERMAL VENTS

by

Jeff Engebretson

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

Moheb A. Ghali, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Craig L. Moyer

Emily L. Peele
Dr. Emily R. Peele

Robin A Matthews
Dr. Robin A. Matthews

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I agree that the Library shall make its copies freely available for inspection. I further agree that copying of this thesis in whole or in part is allowable only for scholarly purposes. It is understood, however, that any copying or publication of this thesis for commercial purposes, or for financial gain, shall not be allowed without my written permission.

Signature____________________

Date________________________

THE TERMINAL-RESTRICTION FRAGMENT LENGTH POLYMORPHISM ASSAY AND ITS USE IN DETERMINING BACTERIAL COMMUNITY SUCCESSION AT HYDROTHERMAL VENTS

A Thesis
Presented to
the Faculty of
Western Washington University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

By

Jeff Engebretson

November, 2002

THE TERMINAL-RESTRICTION FRAGMENT LENGTH POLYMORPHISM ASSAY AND ITS USE IN DETERMINING BACTERIAL COMMUNITY SUCCESSION AT HYDROTHERMAL VENTS

by

Jeff Engebretson

ABSTRACT

The focus of this study was to analyze bacterial community succession at hydrothermal vents using terminal-restriction fragment length polymorphism (T-RFLP). In chapter one, a set of small subunit ribosomal RNA gene (SSU rDNA) sequences ($n = 14,870$) was probed *in silico* with bacterial-specific oligonucleotides resulting in 4603 (31%) sequence variants that were successfully detected. A suite of 18 restriction endonucleases (REs) were classified based upon their ability to resolve uniquely-sized terminal-restriction fragments (T-RFs). Further analysis of T-RFLP was performed by iterative, random sampling from the derived database of T-RFs. All REs used in this study detected $\leq$70% of OTUs at richness values greater than 50 T-RFs per community. Based on these results, the most efficacious use of T-RFLP is in studies that interrogate communities with low to intermediate species richness. In chapter two, T-RFLP patterns of hydrothermal vent bacterial communities within microbial growth chambers (MGCs) were analyzed. Over a five-year sampling period, 52 MGCs were recovered after either a short-term (days) or long-term (annual) deployment. The average number of populations (a measure of species richness) within communities that developed in short-term deployed MGCs was significantly lower than those detected in long-term deployed MGCs. All short-term MGC communities were dominated by pioneer populations indicative of *epsilon-Proteobacteria*, of which, specific T-RF groups were recognized at vent sites throughout the five-year sampling period. The dominant colonizing populations at the vents studied were spatially constricted throughout the five-year sampling period regardless of changes in vent fluid characteristics. The long-term MGCs showed evidence of succession by an increased occurrence of numerous, unidentified populations.

ACKNOWLEDGMENTS

This thesis was made possible due to the financial, technical, and moral support of numerous individuals and institutions. I would like to thank my primary research advisor, Craig Moyer, for providing the ideas, tools, and knowledge that were necessary to accomplish the goals of this study. I would also like to extend my appreciation to Emily Peele and Robin Matthews for their insight and editing abilities. All three committee members supplemented their support by teaching classes that helped inspire the direction of my research. I would like to thank my family and friends, especially Susan O'Donnell, for their unwavering support, encouragment, and advice throughout this endeavour. This project was funded by the Washington State Sea Grant Office, The Bureau of Faculty Research, and the Biology Department at Western Washington University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER 1

Fidelity of Select Restriction Endonucleases in Determining Microbial Diversity by Terminal-Restriction Fragment Length Polymorphism

ABSTRACT

An evaluation of 18 restriction endonucleases (REs) for use in terminal-restriction fragment length polymorphism (T-RFLP) was performed using species richness, species density indices, rarefaction and computer simulations on 4603 bacterial small subunit ribosomal RNA gene (SSU rDNA) sequences. T-RFLP has become a commonly used method to screen environmental samples for precursory identification and community comparison studies due to its precision and high-throughput capability. The accuracy of T-RFLP in describing a community has not yet been thoroughly evaluated. In this paper, I attempt to classify REs based upon their ability to resolve unique terminal-restriction fragments or operational taxonomic units (OTUs) from a database of gene sequences. Furthermore, I assess the predictive accuracy of T-RFLP at values of varying community richness (n = 1, 5, 10, 50, and 100). The classification of RE fidelity was performed by measuring richness and density on the entire database of terminal-restriction fragments. Further analysis of T-RFLP accuracy at determining richness was performed by iterative, random sampling from the derived database of T-RFs. It became apparent that two constraints were influential for measuring the fidelity of a given RE: (i) the ability to resolve unique sequence variants and (ii) the number of T-RFs that fell within a measurable size range. The latter constraint was found to be more significant in estimating RE fidelity. Of the 18 REs examined, *Bst*UI, *Dde*I, *Sau*96I, and *Msp*I had the highest frequency of resolving single populations in model communities. All REs used in this study detected $\leq$70% of OTUs at richness values greater than 50 T-RFs per community. Based on the results of these *in silico* experiments, the most efficacious use of T-RFLP for microbial diversity studies are ones that address situations of low to intermediate species richness (e.g., colonization, early successional stages, biofilm formation).

INTRODUCTION

Studies in microbial ecology have come to rely heavily on molecular-based analyses due to the difficulty arising from exclusively characterizing a natural community by cultivation and microscopy (Amann *et al*., 1995; Giovannoni *et al*., 1990; Jeanthon, 2000; Pace, 1997). The molecular tool commonly used for examining microbial communities is the small subunit ribosomal RNA gene (SSU rDNA). The SSU rDNA is valuable in that it is present in all known extant organisms; it has a relatively slow rate of evolution with apparent conserved and variable regions; and, it has been characterized for a broad array of organisms (Moyer, 2001; Woese, 1987). After polymerase chain reaction (PCR)-based amplification of the gene, the goal is to determine SSU rDNA sequence information and to infer characteristics (e.g., phylogeny and perhaps metabolic capacity) based on the description of closely-related taxa (Head *et al*., 1998; Horz *et al*., 2001; Lynch, 2000). The process of characterizing a microbial community by this laborious, sequence-based methodology is often precluded or replaced by a screening technique that will elucidate the complexity of a community. There are numerous procedures available to characterize a community of gene amplicons. Terminal-Restriction Fragment Length Polymorphism (T-RFLP) is one of the procedures (Avannis-Aghajani *et al*., 1994; Clement *et al*., 1998; Liu *et al*., 1997) that can be used to track spatial and temporal changes in SSU rDNAs from microbial communities (Dunbar *et al*., 2001; Sessitsch *et al*., 2001; see Osborn *et al*., 2000 for a review).

The T-RFLP technique has become a common diagnostic and screening method due to its high sensitivity and ability to rapidly acquire precise data compared to more laborious or imprecise forms of analysis such as DGGE (Ferris *et al.*, 1997; Horz *et al.*, 2001; Moeseneder *et al.*, 1999; Muyzer, 1999) and RFLP (Acinas *et al.*, 1997; Delong *et al.*, 1993; Heyndrickx *et al.*, 1996; Moyer *et al.*, 1994; Rath *et al.*, 1998; Vaneechoutte and Rossau, 1992; Weidner *et al.*, 1996). The T-RFLP technique utilizes one or more fluorescently-labeled primers during PCR. After enzymatic digestion of PCR amplicons, each unique terminal-restriction fragment (T-RF) can be defined as an operational taxonomic unit (OTU) and may often be inferred as a single population within a community. The validity of this inference has not yet been fully explored. The choice and number of restriction endonucleases (REs) should be screened for maximum fidelity of the T-RFs for a given sample. I define fidelity as the ability of a restriction endonuclease to identify the number of SSU rDNA sequence variants in a community via T-RF size distributions. By choosing the appropriate number and types of REs, the investigator will increase the probability that the T-RF community reflects the natural diversity of microbial populations.

The focus of this study was to explore (i) whether sequence variants were more clearly resolved by the use of selected REs and (ii) to measure the success of REs at detecting sequence variants of communities of varying richness. I used traditional diversity measurements and computer simulations to explore the resolving power of select tetrameric REs given a specific set of PCR primers used to amplify an ~1460 bp region of the SSU rDNA from organisms of the domain *Bacteria*. The derived database of T-RFs for each RE was considered a community and the calculation of richness and density indices provided a measure for the resolving power of a given RE. The fidelity of both REs and the T-RFLP assay under varying SSU rDNA richness conditions was measured by utilizing computer-simulation experiments.

MATERIALS AND METHODS

Algorithm to determine PCR amplicon and T-RF length

A database of available unaligned SSU rDNA sequences (n = 14,870) was acquired from the Ribosomal Database Project Release 8.1 (Maidak *et al.*, 2001) on July 30, 2001. A computer program was written in Qbasic programming language to perform both *in silico* PCR probing and RE digestion. The degenerate PCR primers used for this computer analysis were located at *Escherichia coli* positions 49 - 68F (5' - TNA NAC ATG CAA GTC GNN CG - 3') and 1492 - 1510R (3' - AAG TCG TAA CAA GGT ANC N - 5') (e.g., Emerson and Moyer, 2002; Moyer *et al.*, 1994). Sequences containing greater than or equal to 80% similarity to the primers were selected for continued analysis. Exterior ends were trimmed and PCR amplicon lengths were checked to verify DNA sequence sizes of ~1460 bp. The selection of REs (summarized in Table 1.1) was an attempt to screen all discovered tetrameric (4-base cutter) restriction endonucleases. T-RF lengths were calculated by counting the number of string characters from the terminal end to the first cutting site for a given RE. The compiled T-RF lengths were stored in a new database.

Calculation of OTU richness and density

The database of T-RFs for each RE was considered a community to evaluate individual REs. Richness is defined as the number of unique OTUs for a given RE. Density indices, which are traditionally misidentified as richness indices, were calculated on each RE-community within a desired size range (50-500 bp) to provide an initial

Table 1.1. A summary of the number of OTUs and density indices associated with RE digests from 4603 SSU rDNA sequences. The rank for each RE is shown in parentheses.

Restriction endonuclease	Recognition site	Total no. of OTUs	No. of OTUs within 50-500 bp	No. of T-RFs within 50-500 bp	D_1[a, b]	D_2[a, c]
*Hpy*188III	TC^NNGA	571 (1)	320 (2)	3129 (11)	38.5 (1)	6.29 (1)
*Hha*I[d]	GCG^C	563 (2)	271 (5)	2817 (15)	32.8 (5)	5.36 (5)
*Dpn*II[d]	^GATC	542 (3)	255 (9)	3430 (9)	29.8 (10)	4.87 (10)
*Scr*FI	CC^NGG	532 (4)	263 (7)	2697 (16)	31.6 (6)	5.17 (6)
*Bfa*I[d]	C^TAG	524 (5)	244 (12)	2876 (14)	29.3 (11)	4.79 (11)
*Rsa*I	GT^AC	513 (6)	253 (10)	3011 (12)	31.0 (7)	5.06 (7)
*Tsp*509I	^AATT	512 (7)	295 (3)	2682 (17)	36.7 (2b)	5.99 (2b)
*Mse*I	T^TAA	502 (8)	235 (14)	2247 (18)	30.3 (9)	4.96 (9)
*Hpy*CH4IV[d]	A^CGT	456 (9)	230 (15)	3973 (6)	24.1 (18)	3.94 (18)
*Hinf*I	G^ANTC	422 (10)	248 (11)	4103 (3)	28.4 (12)	4.64 (12)
*Msp*I[d]	C^CGG	406 (11)	322 (1)	4027 (5)	36.7 (2a)	5.99 (2a)
*Hae*III	GG^CC	346 (12)	241 (13)	3956 (7)	27.6 (14)	4.51 (14)
*Hpy*188I	TCN^GA	345 (13)	283 (4)	3258 (10)	33.9 (4)	5.55 (4)
*Bsl*I	CCN_5^NNGG	315 (14)	265 (6)	3708 (8)	30.8 (8)	5.04 (8)
*Alu*I	AG^CT	294 (15)	210 (18)	2993 (13)	25.5 (15)	4.18 (15)
*Bst*UI[d]	CG^CG	292 (16)	262 (8)	4353 (1)	27.9 (13)	4.56 (13)
*Dde*I	C^TNAG	280 (17)	228 (16)	4130 (2)	24.6 (17)	4.03 (17)
*Sau*96I	G^GNCC	273 (18)	221 (17)	4088 (4)	25.4 (16)	4.16 (16)

[a] calculated for T-RFs within 50-500 bp. Data were randomly removed to 2247 T-RFs per RE to remove the effects of sample size.

[b] for a community of 2247 individuals the range of possible values are 0-291.

[c] for a community of 2247 individuals the range of possible values are 0-47.

[d] RE has one or more known isoschizomers.

evaluation of the resolving power. D_1 is the Margalef Index (Margalef, 1958) and is calculated as follows:

$$D_1 = (S - 1) / \ln n$$

where S is the number of OTUs and n is the number of T-RFs. The minimum value of Margalef's index is zero (i.e., when the number of OTUs is 1) and the maximum value is $(n-1)/(\ln n)$ (i.e., when each OTU is represented by one T-RF). D_2 is the Menhinick index (Menhinick, 1964) and is calculated as follows:

$$D_2 = S / \mathrm{sqrt}\,(n)$$

Menhinick's index approaches zero when there are a high number of individuals but few OTUs and, like Margalef's index, approaches a maximum when the number of species is equal to the number of individuals. The results of D_1 and D_2 are sensitive to variations in sample size (Ludwig and Reynolds, 1988); therefore, data were normalized by randomly removing T-RFs from each RE profile until all sample sizes were equal to 2247 T-RFs per RE profile (2247 corresponds to the lowest common denominator of T-RFs within 50-500 bp found by *Mse*I). The rarefaction algorithm (Hurlbert, 1971; Heck *et al.*, 1975; Simberloff, 1978; Tipper, 1979; Siegel, 1990; Brewer and Williamson, 1994; Hughes *et al.*, 2001) is a deterministic transform of a community's species-abundance distribution independent of sample size and is based on the following equation:

$$E(S_n) = \sum_{i=1}^{S} \left[1 - \binom{N - N_i}{n} \Big/ \binom{N}{n} \right]$$

where $E(S_n)$ is the expected number of unique T-RFs in a sample of n T-RFs selected at random from a collection containing N T-RFs, S unique T-RFs, and N_i T-RFs in the i^{th} unique T-RF. The rarefaction curves were calculated using Analytic Rarefaction (version 3.1; S.M. Holland [http://www.uga.edu/~strata/software]).

Iteration detection of restriction endonuclease fidelity

Computer simulations were performed by repeated (n = 100) random sampling without replacement from the T-RF database for each RE profile. The simulation was designed to obtain communities with SSU rDNA richness values of 1, 5, 10, 50, and 100 members. Community detection values were expressed as the probability of detecting a uniquely-sized T-RF or OTU within the range of 50-500 bp. To determine overall T-RFLP efficacy, multiple RE profiles were combined for a single community. The community detection values were randomly sorted for each of the 100 independent samplings at each richness value. The average maximum community detection values were chosen for successive random RE selections.

RESULTS

The 4603 (31% of RDP vers. 8.1) sequences used in this analysis were uniformly distributed throughout the domain *Bacteria* (Table 1.2). The richness and density indices evaluated for those sequences showed that the REs used in this study have a high degree of variability in their ability to resolve OTUs from T-RF data (Table 1.1). *Hpy*188III (rank = 1st) and *Hha*I (rank = 2nd) were able to distinguish the greatest number of OTUs (571 and 563, respectively) while *Dde*I (rank = 17th) and *Sau*96I (rank = 18th) distinguished less than 300 OTUs (Table 1.1).

When T-RFs outside the range of 50-500 bp were removed (in an effort to represent the realistic resolving power of capillary and gel electrophoresis technology), the number of T-RFs associated with each RE was no longer equal at 4603 and a different hierarchy of REs in detecting OTUs was found. *Bst*UI ranked highest among the REs containing T-RFs within 50-500 bp (Table 1.1). The disparity in rank after removal of T-RFs outside the desired range of 50-500 bp is mainly due to the existence of highly conserved cut sites. For example, *Bfa*I has two highly conserved terminal cut sites within the range of 570-601 bp and 745-760 bp. The resulting exclusion of 829 (18%) SSU rDNA T-RFs (data not shown) thereby decreased the rank of *Bfa*I from 5th to 12th. Conversely, increases in rank (e.g., *Msp*I changed from 11th to 1st) occurred when the majority of T-RFs for a RE were within the specified range of 50-500 bp (see Marsh *et al.*, 2000 for a web-based tool to sort T-RFs by size).

Both T-RF richness and density indices showed that *Hpy*188III, *Msp*I, *Tsp*509I, and *Hpy*188I resolved the greatest number of OTUs in the range of 50-500 bp (Table 1.1). To remove the effect of unequal sample sizes on evaluating density indices,

Table 1.2. Phylogenetic distribution of the 4603 RDP (Vers. 8.1) sequences having primer sites matching the primers included in this study.

Major taxon	# of sequences (% coverage)
Bacteria	
Proteobacteria	
α subdivision	631 (32%)
β subdivision	290 (27%)
γ subdivision	824 (28%)
δ subdivision	168 (31%)
ε subdivision	95 (30%)
High GC gram positive	926 (40%)
Bacillus subdivision	476 (38%)
Other gram positive	429 (31%)
Flexibacter-Cytophaga-Bacteroides	132 (17%)
Cyanobacteria and Chloroplasts	150 (29%)
Spirochetes	144 (22%)
Others	338 (26%)
Total	4603 (31%)

amplicons were randomly removed until the sample size for all communities equaled 2247 (the lowest common denominator of T-RFs cut within 50-500 bp by *Mse*I). Because effect of randomly removing data is nonmeasurable, the rarefaction procedure was used to compare communities of unequal sample size (Hurlbert, 1971). The rarefaction curves agreed with the Margalef and Menhinick indices showing the same REs as the top performers in terms of their ability to identify the greatest number of OTUs (Figure 1.1).

When 100 random communities were assembled at discrete values of richness (i.e., 1, 5, 10, 50, and 100), *Bst*UI outperformed other REs at low values of richness (Figure 1.2). When community richness was set to 1, *Bst*UI was able to detect the single population 98% of the time, while *Mse*I performed poorest by detecting the population in only 50% of the communities. When richness values were increased to 5 and 10, a similar, yet declining resolution continued with *Bst*UI still outperforming other enzymes. Once the T-RF richness was increased to 50, all REs detected ≤71% of the OTUs but there were 7 REs (*Bst*UI, *Dde*I, *Sau*96I, *Msp*I, *Hin*fI, *Hae*III, and *Bsl*I) that still outperformed the others. When the T-RF richness was set to 100, all REs performed uniformly poor by detecting within the range of 58% (*Msp*I) to 35% (*Mse*I) of the sequence variants. This decrease in fidelity is explained by an increasing frequency of identical T-RF sizes within a single community.

When REs were chosen in a random, sequential series, a plot was generated to characterize the number of enzymes needed to increase the detection of OTUs (Figure 1.3). At richness equal to 1, the probability of detecting the T-RF reached 100% after using 5 REs. At richness values of 5 and 10, similar curves were generated where the probability of detecting OTUs increased rapidly until after the use of 3 to 5 enzymes, at

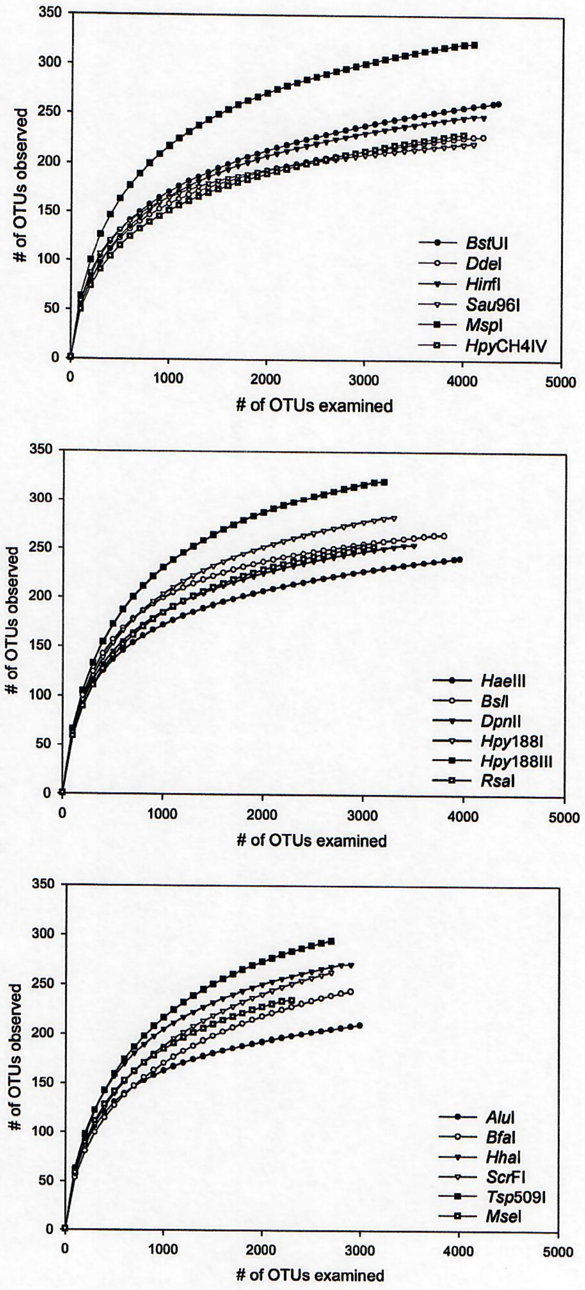

Figure 1.1. Rarefaction analysis curves for the 18 restriction endonucleases used in this study. Restriction endonucleases are divided into 3 graphs for illustrative purposes and are ordered sequentially by the number of terminal-restriction fragments that fell within the measurable size range of 50-500 bp.

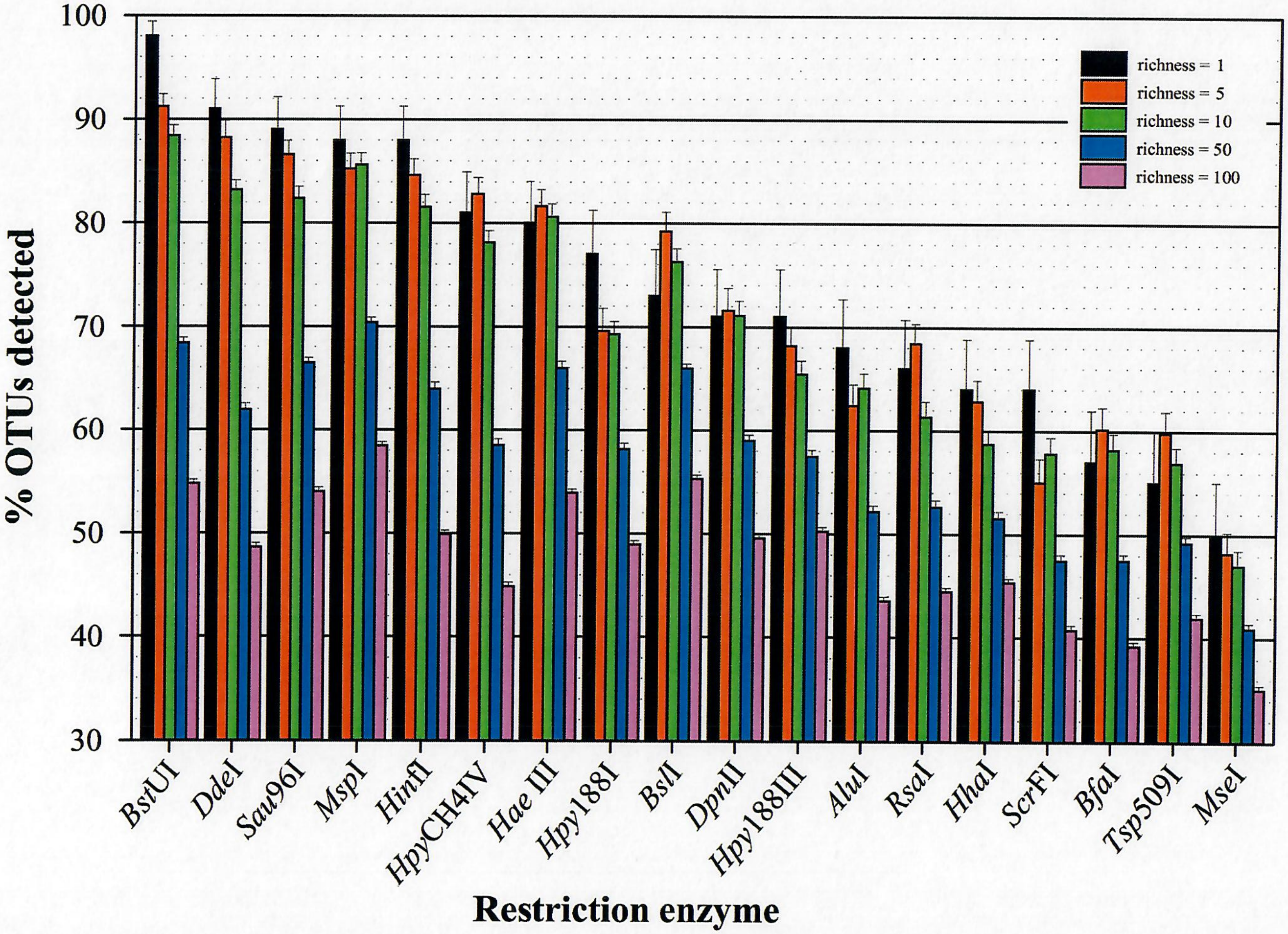

Figure 1.2. The percent OTUs detected for each restriction endonuclease at five discrete values of richness (1, 5, 10, 50, and 100). The enzymes are ordered from left to right in decreasing order of % OTUs detected with respect to a community richness of one. Error bars indicate standard error about the mean.

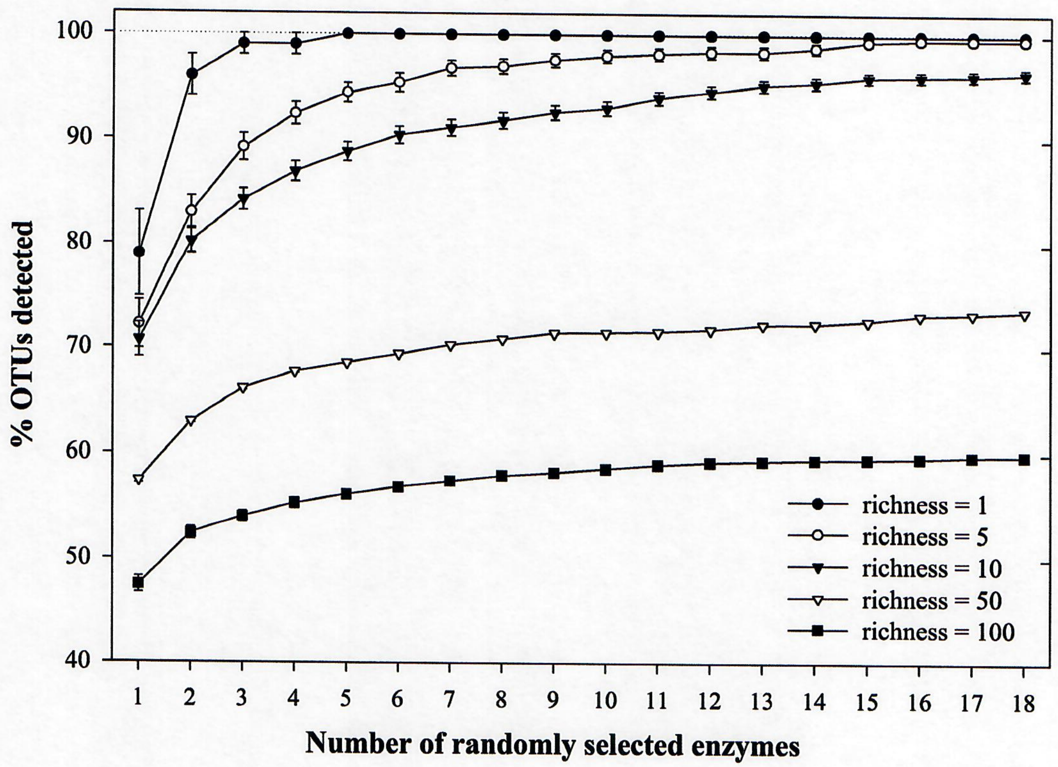

Figure 1.3. The percent OTUs detected by randomly selecting from the pool of 18 restriction endonucleases at five discrete values of richness (1, 5, 10, 50, and 100). Error bars indicate standard error about the mean. Lines are drawn between points to illustrate trends.

which point the probability of increasing the detection was gradual. At richness values of 50 and 100, the detection ability only gradually increased to a maximum of 74% and 60%, respectively. The fraction of OTUs detected for each richness profile only gradually increased after four successive REs were used (Figure 1.3).

A second simulation was generated by choosing the subset of four REs (*Bst*UI, *Dde*I, *Sau*96I, and *Msp*I) that had the highest fidelity for detecting single-population communities (Figure 1.2). The second simulation was analyzed at richness values of 1, 5, and 10 because all REs had performed poorly at richness values of 50 and 100. At richness equal to 1, the probability of detecting the single population reached 100% after using only 2 REs. At richness values of 5 and 10, the top 4-performing REs were able to detect 98.8 ± 0.48 % and 94 ± 0.72 %, respectively (Figure 1.4).

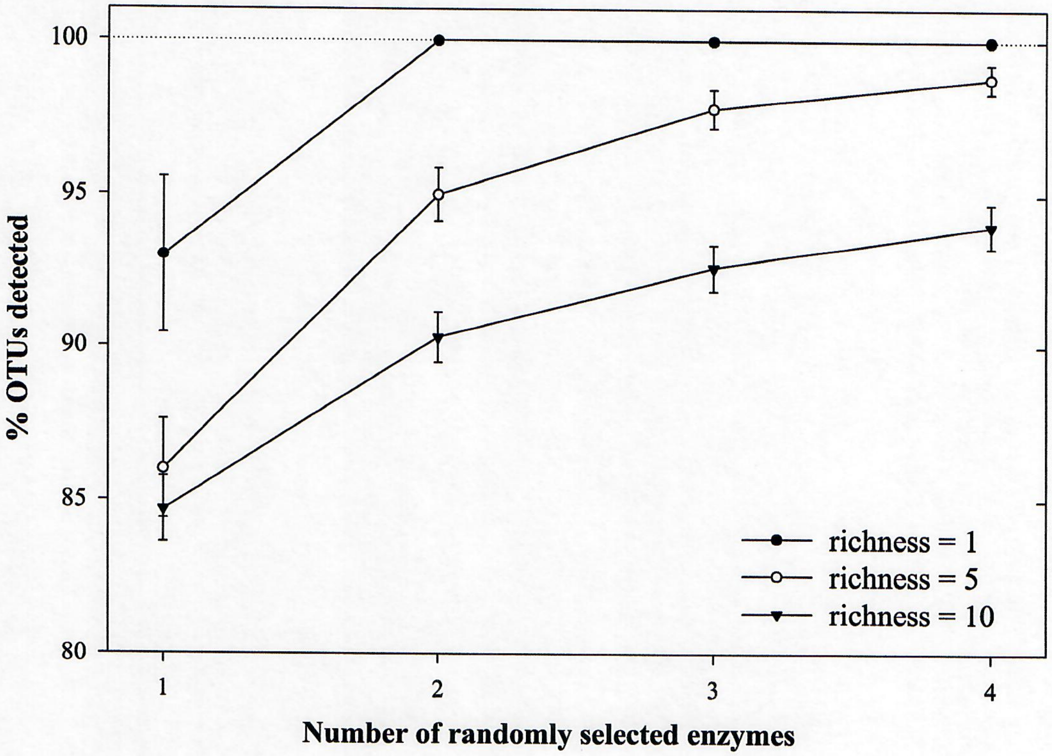

Figure 1.4. The percent OTUs detected by randomly selecting from a pool of four restriction endonucleases of highest fidelity (*Bst*U I, *Dde* I, *Sau*96 I, and *Msp* I) at three discrete values of richness (1, 5, and 10). Error bars indicate standard error about the mean. Lines are drawn between points to illustrate trends.

DISCUSSION

The T-RFLP assay provides microbial ecologists with a rapid method to estimate community diversity and to screen samples for facilitating the prioritization of sequencing effort. Community comparisons and other downstream analyses of T-RFLP data have been adopted via an assortment of statistical methods such as similarity indices (Dollhopf *et al.*, 2001, Watts *et al.*, 2001), hierarchical clustering algorithms (Bruce and Hughes, 2000; Dunbar *et al.*, 2000; Dunbar *et al.*, 2001; Liu *et al.*, 1997; Moeseneder *et al.*, 1999; Sessitsch *et al.*, 2001), principal components analyses (Clement *et al.*, 1998; Dollhopf *et al.*, 2001; Franklin *et al.*, 2001), and self-organizing maps (Dollhopf *et al.*, 2001). This study has focused on evaluating the ability of REs to resolve true community diversity by estimating confidence with respect to specific REs. Furthermore, I have addressed whether T-RFLP could be used as an accurate assay in characterizing microbial diversity by combining data from multiple RE analyses for a single community. A single RE digest on a community with richness greater than 100 may show less than 50% of the true diversity (Figure 1.2). For example, Dunbar *et al.* (2000) found only 20 T-RFs from a soil community that contained 154 RFLP-derived clones. Consequently, the use of clustering, ordination, and neural networks may not give desirable results on high-diversity community comparisons. In our lab, we concatenate the results of 4-7 RE digests for a single community yet rarely achieve meaningful similarities when performing cluster analysis on communities of high diversity. However, identifying dominant populations or doing statistical analyses on communities of low diversity still remains robust when utilizing T-RFLP data.

The limitations of electrophoresis technology to accurately and precisely size fragments within specific size ranges are well documented (Grossman *et al.*, 1992; Lerman and Frisch, 1982; Luckey *et al.*, 1990; Luckey *et al.*, 1993; Lumpkin *et al.*, 1985; Manabe *et al.*, 1994; Swerdlow *et al.*, 1991). The velocity at which a DNA fragment moves through a sieving matrix, such as agarose or polyacrylamide, is not linearly correlated with size. Small DNA fragments obtain a high degree of separation as they travel rapidly through the matrix, thereby allowing high precision in sizing. Unfortunately, there are a number of problems with the inclusion of small fragments (<50 bp) in T-RFLP analysis, including the loss of small DNA fragments associated with the purification of samples; the unknown effects of Brownian motion; and, the existence of residual PCR artifacts that may result in anomalous data (e.g., primer dimers). Because the migration time interval, or mobility, between fragments decreases as DNA fragment size increases, a maximum size (in bp) exists where resolution of uniquely-sized DNA fragments is no longer possible. Consequently, the inclusion of large T-RFs (>500 bp) for data analysis is not recommended for fragment-analyzing technology. The advent of combined technologies with pulsed field gel electrophoresis (Schwartz and Cantor, 1984) may eventually lead to higher precision base calling of large DNA fragments for T-RFLP analysis.

The resolvability of a RE is its potential to detect OTUs from a set of sequence variants based on T-RF size distributions. The resolution can be directly analyzed by the use of diversity measurements (e.g., density indices, rarefaction curves). The fidelity of a RE is a more accurate measure and includes constraints involved in the T-RFLP assay. Fidelity can only be measured by simulatory modeling because specific values of constraints are unknown. In this study, the ranking of REs based on the number of T-RFs within 50-500 bp was more important than resolving ability in determining RE fidelity (compare rank values in Table 1.1 with Figure 1.2). If fragment-analyzing technology included a larger range of T-RFs (e.g., 50-1500 bp), resolvability complemented with a suitable PCR primer set would likely be the best proxy for determining RE fidelity.

The random selection of sequences from the RDP database is not likely representative of a random community found in nature due to the bias of clinical and repetitive entries found in the database (e.g., there are ~45 SSU rRNA sequences from *Escherichia coli* in RDP Release 8.1). Currently, there are no available methods in molecular microbial ecology that accurately reflect a complete community found in nature. Consequently, we have no measure for what an accurate natural microbial community is in its entirety. The results of this study therefore, show a lower threshold of T-RFLP range of detection.

Selection of appropriate primers for PCR-based microbial ecological studies is an ongoing task. The appropriate weighting of parameters that describe a suitable primer (e.g., melting temperature, % similarity) is not yet agreed upon and must be determined empirically. I chose a previously tested degenerate primer set (Emerson and Moyer, 2002; Moyer *et al*., 1994) based on its ability to detect the greatest percentage of bacterial sequences from the RDP database when compared with other primer sets (data not shown). A further prerequisite to selection of primers was to use bacterial primers spanning the majority of the SSU rDNA gene. The results of this investigation would not significantly vary if alternative terminal primers close to the 5'-end of the gene were selected.

This study has demonstrated that the selection and number of REs is an important parameter when desiring an accurate representation of the diversity of a microbial community. The T-RFLP technique is likely a valuable screening tool when studying spatiotemporal changes of natural communities with relatively low to intermediate species richness. This technique may not be an adequate tool to characterize complex microbial populations.

LITERATURE CITED

Acinas, S., F. Rodriguez-Valera, and C. Pedros-Alio. 1997. Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. FEMS Microbiol. Ecol. 24: 27-40.

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

Avannis-Aghajani, E., K. Jones, D. Chapman and C. Brunk. 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. Biotechniques 17: 144-149.

Brewer, A., and M. Williamson. 1994. A new relationship for rarefaction. Biodiv. Conserv. 3: 373-379.

Bruce, K.D., and M.R. Hughes. 2000. Terminal restriction fragment length polymorphism monitoring of genes amplified directly from bacterial communities in soils and sediments. Mol. Biotechnol. 16: 261-269.

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

Delong, E.F., D.G. Franks, and A.L. Alldredge. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. 38: 924-934.

Dollhopf, S.L., S.A. Hashsham, and J.M. Tiedje. 2001. Interpreting 16S rDNA T-RFLP data: application of self-organizing maps and principal components analysis to describe community dynamics and convergence. Microb. Ecol. 42: 495-505.

Dunbar, J., L.O. Ticknor, and C.R. Kuske. 2000. Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment length analysis. Appl. Environ. Microbiol. 66: 2943-2950.

Dunbar, J., L.O. Ticknor, and C.R. Kuske. 2001. Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. Appl. Environ. Microbiol. 67: 190-197.

Emerson, D., and C.L. Moyer. 2002. Neutrophilic Fe-oxidizing bacteria are abundant at the Loihi Seamount hydrothermal vents and play a major role in Fe oxide deposition. Appl. Environ. Microbiol. 68: 3085-3093.

Ferris, M.J., and D.M. Ward. 1997. Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 63: 1375-1381.

Franklin, R.B., J.L. Garland, C.H. Bolster, and A.L. Mills. 2001. Impact of dilution on microbial community structure and functional potential: comparison of numerical simulations and batch culture experiments. Appl. Environ. Microbiol. 67: 702-712.

Giovannoni, S.J., T.B. Britschgi, C.L. Moyer, and K.G. Field. 1990. Genetic diversity of Sargasso Sea bacterioplankton. Nature 345: 60-63.

Grossman, P.D., S. Menchen, and D. Hershey. 1992. Quantitative analysis of DNA-sequencing electrophoresis. Genet. Anal. Tech. Appl. 9: 9-16.

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

Heck, K.L., G.V. Belle, and D. Simberloff. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecology 56: 1459-1461.

Heyndrickx, M., L. Vauterin, P. Vandamme, K. Kersters, and P. De Vos. 1996. Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phylogeny and taxonomy. J. Microbiol. Meth. 26: 247-259.

Horz, H., M.T. Yimga, and W. Liesack. 2001. Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of *pmoA*, *mmoX*, *mxaF*, and 16S rRNA and ribosomal DNA, including *pmoA*-based terminal restriction fragment length polymorphism profiling. Appl. Environ. Microbiol. 67: 4177-4185.

Hughes, J.B., J.J. Hellmann, T.H. Ricketts, and B.J.M. Bohannan. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl. Environ. Microbiol. 67: 4399-4406.

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

Jeanthon, C. 2000. Molecular ecology of hydrothermal vent microbial communities. Antonie Van Leeuwenhoek 77: 117-133.

Lerman, L.S. and H.L. Frisch. 1982. Why does the electrophoretic mobility of DNA in gels vary with the length of the molecule? Biopolymers 21: 995-997.

Liu, W., T.L. Marsh, H. Cheng, and L.J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63: 4516-4522.

Luckey, J.A., H. Drossman, A.J. Kostichka, D.A. Mead, J. D'Cunha, T.B. Norris, and L.M. Smith. 1990. High speed DNA sequencing by capillary electrophoresis. Nucleic Acids Res. 18: 4417-4421.

Luckey, J.A., T.B. Norris, and L.M. Smith. 1993. Analysis of resolution in DNA sequencing by capillary gel electrophoresis. J. Phys. Chem. 97: 3067-3075.

Ludwig, J.A., and J.F. Reynolds. 1988. Statistical Ecology. John Wiley & Sons, New York, New York.

Lumpkin, O.J., P. Déjardin, and B.H. Zimm. 1985. Theory of gel electrophoresis of DNA. Biopolymers 24: 1573-1593.

Lynch, K.S. 2000. Bacterial community structure and phylogenetic diversity of hydrothermal vents at Axial Volcano, Juan de Fuca Ridge. Master's Thesis, Western Washington University.

Maidak, B.L., J.R. Cole, T.G. Lilburn, C.T. Parker Jr., P.R. Saxman, R.J. Farris, G.M. Garrity, G.J. Olsen, T.M. Schmidt, and J.M. Tiedje. 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. 29: 173-4.

Manabe, T., N. Chen, and S. Terabe. 1994. Effects of linear polyacrylamide concentrations and applied voltages on the separation of oligonucleotides and DNA sequencing fragments by capillary electrophoresis. Anal. Chem. 66: 4243-4252.

Margalef, R. 1958. Information theory in ecology. Gen. Syst. 3: 36-71.

Marsh, T.L., P. Saxman, J. Cole, and J. Tiedje. 2000. Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. Appl. Environ. Microbiol. 66: 3616-3620.

Menhinick, E.P. 1964. A comparison of some species-individual diversity indices applies to samples of field insects. Ecol. 45: 859-861.

Moeseneder, M.M., J.M. Arrieta, G. Muyzer, C. Winter, and G.J. Herndl. 1999. Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 65: 3518-3525.

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 community at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. 60: 871-879.

Moyer, C.L. 2001. Molecular phylogeny: applications and implications for marine microbiology. Methods Microbiol. 30: 375-394.

Muyzer, G. 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. Curr. Opin. Microbiol. 2:317-322.

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

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. Science 276: 734-740.

Rath, J., K.Y. Wu, G.J. Herndl, and E.F. Delong. 1998. High phylogenetic diversity in a marine-snow-associated bacterial assemblage. Aquat. Microb. Ecol. 14: 261-269.

Schwartz, D.C., and C.R. Cantor. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37: 67-75.

Sessitsch, A., A. Weilharter, M.H. Gerzabek, H. Kirchmann, and E. Kandeler. 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl. Environ. Microbiol. 67: 4215-4224.

Siegel, A.F. 1990. Rarefaction curves, p. 623-626. *In* S. Kotz and N.L. Johnson (eds.), Encyclopedia of statistical sciences, Vol. 7. New York: John Wiley and Sons.

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 (eds.), Biological data in water pollution assessment: quantitative and statistical analyses, American Society for Testing and Materials.

Swerdlow, H., J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S. Wu, and N.J. Dovichi. 1991. Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. Anal. Chem. 63: 2835-2841.

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

Vaneechoutte, M., and R. Rossau. 1992. Rapid identification of bacteria of the *Comamonadaceae* with amplified ribosomal DNA-restriction analysis (ARDRA). FEMS Microbiol. Lett. 93: 227-234.

Watts, J.E.M., Q. Wu, S.B. Schreier, H.D. May, and K.R. Sowers. 2001. Comparative analysis of polychlorinated biphenyl-dechlorinating communities in enrichment cultures using three different molecular screening techniques. Environ. Microbiol. 3: 710-719.

Weidner, S., W. Arnold, and A. Puhler. 1996. Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. Appl. Environ. Microbiol. 62: 766-771.

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

CHAPTER 2

Bacterial Pioneer Populations, Succession, and Community Comparisons at Northeast Pacific Hydrothermal Vents

ABSTRACT

Terminal-restriction fragment length polymorphism (T-RFLP) patterns were used to track populations of bacteria occurring within microbial growth chambers (MGCs) deployed at seven diffuse-flow (T_{max} = 100°C) hydrothermal vent orifices located within the caldera of Axial Volcano, Juan de Fuca Ridge. For comparison, two diffuse vents located at the Magic Mountain area on the Explorer Ridge were also examined. Over a five-year sampling period, 52 MGCs were recovered after either a short-term (days) or long-term (annual) deployment. Upon recovery, genomic DNA was extracted and amplified using bacterial-specific PCR primers to generate 5' fluorescently labeled amplicons of small subunit rRNA genes (i.e., SSU rDNAs). These PCR amplicons were digested with multiple tetrameric restriction endonucleases and the respective community diversity and succession patterns were characterized. The average number of populations (a measure of species richness) within the community that developed in short-term deployed MGCs was significantly lower than those detected in long-term deployed MGCs. All short-term MGC communities were dominated by pioneer populations indicative of *epsilon-Proteobacteria*, of which, specific T-RF groups were recognized at vent sites throughout the five-year sampling period. The colonizing populations at the vents studied showed low temporal variability with respect to their community structure and no apparent correlation with vent fluid characteristics. The dominant colonizing populations at Magic Mountain vent sites were similar to each other, though different from those found at Axial Volcano. The long-term MGCs showed evidence of succession by an increased occurrence of numerous other unidentified populations. The discovery that all primary colonizing populations were most similar to known lineages of *epsilon-Proteobacteria* detected from hydrothermal vents located worldwide provides further evidence that a few cosmopolitan populations are capable of acting as the primary microbial successors of newly-formed hydrothermal vent systems.

INTRODUCTION

A mid-ocean ridge system occurs in areas of enhanced seafloor spreading where new boundaries of lithospheric plates are being produced by tectonic forces derived from convective cells deep within the earth's mantle (Atwater and Molnar, 1973). The rapid cooling of and diastrophism associated with extruding magma (>1200°C) as it interacts with low-temperature oceanic bottom water (~2°C) creates an elaborate network of subseafloor lava tubes and interstices. Oceanic water is cycled through this network by the replacement of buoyant, superheated water (~350°C) beneath the crust. The interaction of heat, seawater, and fresh basalt results in a continuous flow of element-rich, reducing, acidic fluid emitting from the seafloor (Seyfried and Mottl, 1995). In many cases, the hydrothermal fluid is the main source of sustenance for a complex ecosystem of vent-associated fauna. The physical and chemical characteristics of vents might resemble environments of early Earth (>3 Gya) and may be homologs of where life originated and evolved (Baross and Hoffman, 1985). This type of submarine volcanism and its associated hydrothermal pumping accounts for greater than 75% of all volcanic activity on the planet (Van Dover, 2000).

When hydrothermal vents at the Galapagos rift were first observed in 1977 (Corliss *et al.*, 1979; Grassle *et al.*, 1979; Lonsdale, 1977), the researchers were surprised to discover a plethora of fauna thriving in the sulfide-rich fluids that exude from deep within the basaltic seafloor. Until that discovery, the deep sea was considered a nutrient-limited environment that was entirely dependent on sinking materials derived from primary production in the photic zone. It was also thought that macrofauna could not exist in hydrothermal fluids because sulfide binds to iron in biological systems which would inhibit cytochrome-*c* oxidase activity of the electron transport system and disrupt the ability for hemoglobin to transport oxygen (Van Dover, 2000). These "poisonous" factors initially puzzled researchers as to how a community could be so robust at hydrothermal vents. It has since been hypothesized that chemosynthetic microbes and their interactions with hydrothermal fluids are the primary forms of nutrient supply to a vent ecosystem.

At the interface where seawater and the hydrothermal fluids mix, there are sharp gradients of pH and temperature allowing for a variety of microbial metabolisms. These gradients combined with high concentrations of solutes and dissolved gases (e.g., H_2, H_2S, $S_2O_3^{2-}$, S_0, NH_4^+, Fe^{2+}) result in unique populations within a microbial mat community with high rates of aerobic and anaerobic metabolic activity (Moyer *et al.*, 1995; Polz and Cavanaugh, 1995; Muyzer *et al.*, 1995; Kleinkauf, 2000; Lynch, 2000; Reysenbach *et al.*, 2000; Corre *et al.*, 2001; Teske *et al.*, 2002; for a review see Jeanthon, 2000). The rates of carbon uptake associated with specific metabolic pathways relative to total community metabolism have not yet been measured (Karl, 1995). However, because of the high redox potential and biochemical availability of reduced sulfur compounds, chemolithotrophy in the form of sulfur oxidation is likely the most common form of metabolism supporting organic carbon production at many hydrothermal vent systems (Ruby *et al.*, 1981; Jannasch and Mottl, 1985; Jannasch and Wirsen, 1985). This hypothesis is further supported by the selective isolation of sulfur-oxidizing bacteria (Ruby *et al.*, 1981; Jannasch, 1983; Tuttle *et al.*, 1983; Wirsen *et al.*, 1998; Taylor *et al.*, 1999; Brinkhoff *et al.*, 1999) and dominant phylogenetic members of a community being most similar to known sulfur-oxidizing isolates (Muyzer *et al.*, 1995; Moyer *et al.*, 1995;

Polz and Cavanaugh, 1995; Sievert *et al.*, 2000; Kleinkauf, 2000; Lynch, 2000). Based on a coupled thermodynamic and geochemical constraint model, McCollum and Shock (1997) estimate that sulfur oxidation accounts for >95% of energy assimilation within a hydrothermal vent system.

The populations of microbes living within vent fluid are primarily chemolithotrophic and thus dependent on the chemical signature of the vent effluent. It may be possible to detect this metabolic dependence by tracking microbial populations when the vent-fluid chemistry changes. Butterfield *et al.* (1997) synthesized a generic model to show how vent-fluid chemistry changes after a volcanic, or plume event. Immediately following a volcanic event the dominant fluids are reduced, H_2S-rich and the dominant microbial populations are potentially ones that oxidize sulfide. As heat flux through the system decreases, Fe-rich fluids play a larger role and iron oxidizers may dominate the microbial population. The dominance of an iron-oxidizing bacterium at various hydrothermal vents of the Loihi Seamount has since been reported (Emerson and Moyer, 2002).

The bacterial diversity at hydrothermal vents is relatively low compared to diversity found in other environments. Using DNA reassociation kinetic values, Torsvik *et al.* (1990) found that there are approximately 4000 unique bacterial genomes in 1 gram of forest soil. A phylogenetic study of four Arizona soils (Dunbar *et al.*, 2000) found over 130 unique phylotypes that spanned 21 bacterial divisions (Dunbar *et al.*, 2002). A phylogenetic survey of a marine snow-associated bacterial assemblage detected 90 unique restriction fragment length polymorphism patterns from 95 bacterial clones, representing a sampling coverage of only 5% (DeLong *et al.*, 1993; Rath *et al.*, 1998). By comparison, Polz and Cavanaugh (1995) used slot-blot hybridization values to show that a single dominant phylotype contributed to over half of a sulfide-associated vent community on the Mid-Atlantic Ridge. Similarly, when bacterial mat communities were examined at Loihi Seamount, the 12 detected operational taxonomic units (OTUs) were dominated by only two populations (Moyer *et al.*, 1994).

A preliminary census of microbial mat communities at Axial Volcano found that diversity is higher than other sampled vent communities. In examining the amplified ribosomal DNA restriction analysis or ARDRA patterns of four bacterial mat communities, Lynch (2000) discovered communities that contained up to 11 multi-occurrence OTUs and contained a high number (26 to 48) of single occurrence OTUs at any given sample site. The study also determined that seven of the multi-occurrence OTUs clustered within the *epsilon-Proteobacteria,* while those remaining were distributed among the *delta-* and *gamma-Proteobacteria* subdivisions, and within the Flexibacter-Cytophaga-Bacteroides and Green Non-Sulfur Bacteria divisions. Rarefaction analyses and coverage estimates showed that bacterial community diversity was higher than expected.

The initial colonization of inert surfaces by microorganisms occurs because relatively dilute nutrients become concentrated at surfaces through adsorption (Atlas and Bartha, 1998). Characterization of the initial colonizing bacterial populations at hydrothermal vents is poorly understood. When a growth chamber was deployed for five days atop a Mid-Atlantic Ridge vent orifice, the colonizing bacterial populations were dominated (28%) by a distinct lineage of *epsilon-Proteobacteria* (Reysenbach *et al.*, 2000). A similar *in situ* experiment found the rapid production of filamentous sulfur mat within the flow of vent fluid and at least two distinct microbial morphotypes (Taylor *et al.*, 1999). Studies of other environments have found a predictable colonization of specific genera in oral microbial communities (Kolenbrander, 2000) and salt marshes (Dang and Lovell, 2000; Dang and Lovell, 2002). However, no apparent order to bacterial colonization and succession were found in biofouling marine habitats (Cooksey and Wigglesworth-Cooksey, 1995).

The use of selective enrichment culture methodologies to characterize microbial populations consistently underestimates the numbers and types of microbes present within a community (Amann *et al.*, 1995; Brock, 1987; Hugenholz *et al.*, 1998; Ward *et al.*, 1992). Culturable bacteria likely represent <5% of bacteria present in a community (Dunbar *et al.*, 1997; Head *et al.*, 1998) and metabolic inferences may indicate nothing about dominant metabolisms within the environment. The use of microscopic analysis to determine community structure is also limited by morphological similarities for phylogenetically distinct taxa. Therefore, this study has focused on using molecular component analyses to study the microbial populations at hydrothermal vents.

The small subunit ribosomal RNA gene (SSU rDNA) has proven an effective "historical document" in analyzing microbial diversity (Woese, 1987). The gene is present in all known extant organisms, is a vital component in protein synthesis machinery, and has no recorded evidence of lateral gene transfer which may skew the accuracy of observed evolutionary relationships. Furthermore, SSU rDNA carries great utility as a phylogenetic descriptor for discrete taxa due to predictable conserved and variable regions in primary structure and a conserved secondary structure (Ward *et al.*, 1992). An *a priori* knowledge of conserved sequence regions can be used to develop probes for detection of multiple taxa while differences in variable regions allow for diagnosis, classification, and evolutionary analyses. Once SSU rDNA content for a community is measured, the contextual placement of taxa is only as meaningful as the size and quality of the database in which one is making comparisons (Moyer, 2001). The Ribosomal Database Project (RDP Release 8.1) now contains over 50,000 aligned SSU rDNA sequences in a phylogenetically organized manner (Maidak *et al.*, 2001).

The main goal of this study was to use terminal-restriction fragment length polymorphism (T-RFLP; Avannis-Aghajani *et al.*, 1994; Liu *et al.*, 1997; Clement *et al.*, 1998) to better understand the colonization, succession, and community comparisons of bacterial populations at hydrothermal vents. T-RFLP is a rapid profiling technique taking the molecular-components approach to track spatial and temporal changes in SSU rDNA communities (Dunbar *et al.*, 2000; Sessitsch *et al.*, 2001; for a review see Osborn *et al.*, 2000) and has been successfully used to detect populations at hydrothermal vents (Emerson and Moyer, 2002). The method is especially advantageous for characterizing numerous samples with low to intermediate diversity (e.g., Chap. 1), as may be the case with colonization experiments. Microbial growth chambers (MGCs) were placed within the flow of seven actively venting orifices for both short- and long-term periods (days and years, respectively) from 1998-2001. Additionally, during 2002, short-term experiments were deployed at the Magic Mountain area of Explorer Ridge. The MGCs were designed to emulate newly-formed basaltic lava with a large surface area for microbial colonization. The potential effects of meiofauna and macrofauna grazing were removed by enclosing the colonization surfaces within nylon mesh. The community genomic DNA (gDNA) was extracted, and SSU rDNA genes from the domain *Bacteria* were amplified with an end-labeled fluorescent oligonucleotide. The community of SSU rDNA amplicons were subjected to a set of restriction endonucleases and the resulting end-labeled fragments were analyzed with respect to colonizing populations, community structure and diversity.

MATERIALS AND METHODS

Microbial Growth Chambers

Microbial growth chambers (MGCs) were constructed of a cluster of three 3" sections of 4" o.d. plexiglass cylinders, surrounded on top and bottom by 202 µm Nytex (Nytex, Tetko, Inc., Kansas City, MO) to exclude macrofauna and meiofauna from grazing (Figure 2.1). The Nytex was sealed to the edges of the plexiglass using all-purpose clear glue (Arrow Fastener Co., Saddlebrook, NJ). The interior of the chambers contained ~300 g of hand-woven, 8 µm diameter glass wool (Corning, Inc., Corning, NY) as a substrate for microbial colonization. The structural integrity was tested and maintained in 100°C water at 1 atm pressure. A stainless steel eye-bolt was secured between the cylinders for negative buoyancy and a polyurethane line was attached to facilitate deployment and recovery operations.

Site Characteristics, Deployment and Recovery of Samples

MGCs were deployed and recovered with the remotely-operated vehicle (ROV) *ROPOS* during annual oceanographic cruises (September 1998, July 1999, July 2000, July 2001, July 2002) aboard NOAA Ship Ronald H. Brown and R/V Thomas G. Thompson in association with the New Millenium Observatory (NeMO; Embley and Baker, 1999). MGCs were placed within the diffuse flow of low-temperature (≤100°C) hydrothermal vent sites at the caldera of Axial Seamount (45°55'N; 130°00'W) on the Juan de Fuca Ridge and at the Magic Mountain vent area (49°45'N; 130°15'W) on the Explorer Ridge (Figure 2.2). Axial Seamount is an active volcano located approximately 400 km from the coast of Oregon and is centered at the intersection of the Juan de Fuca

Figure 2.1. Photo of microbial growth chambers before and after being placed within the flow of a hydrothermal vent.

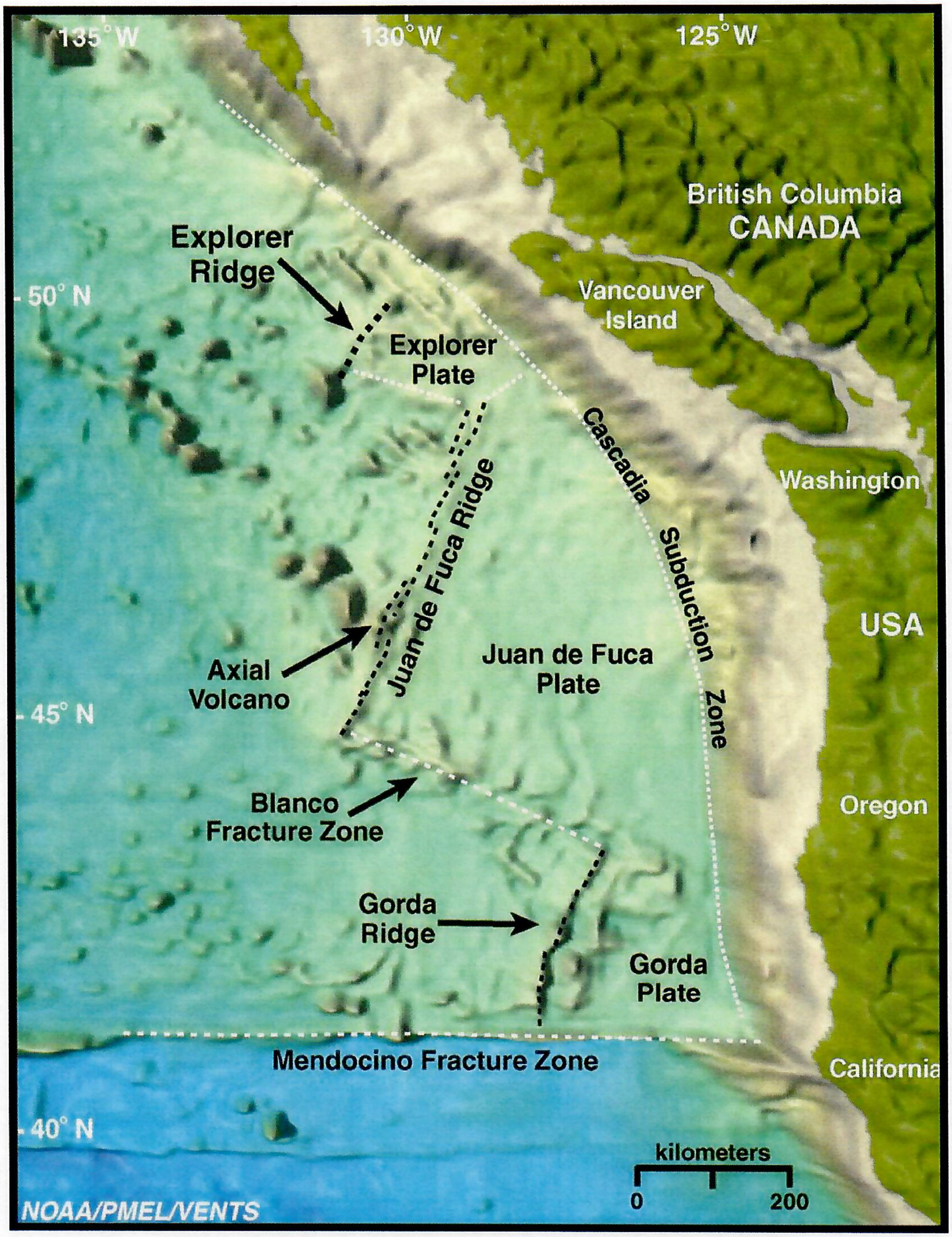

Figure 2.2. Map of the Northeast Pacific seafloor and its associated geophysical features. Samples for this study were taken from Axial Volcano and Explorer Ridge. Figure courtesy of S. Merle, R. Embley, and W. Chadwick, unpublished data; Embley *et al.*, (1999).

Ridge and the Cobb Seamount chain. Axial is characterized by intense volcanic activity from the Cobb-Eikelberg hotspot juxtaposed on the extensional field of the Juan de Fuca spreading ridge (Chase *et al.*, 1985; Davis, 2000; Johnson and Embley, 1990). The volcano rises to a summit depth of 1500 m, approximately 1100 m above the surrounding Eastern Pacific seafloor. The caldera of the volcano (8 km x 3 km) holds a diverse network of hydrothermal systems including three discovered vent fields ('98 Flow, ASHES, CASM) and 25-30 actively venting orifices (Figure 2.3). Typical communities within the vicinity of active vent fields include microbial mats, vestimentiferan tubeworms, molluscs, annelids, and arthropods (Tunnicliffe *et al.*, 1985). Intense volcanic activity was observed at Axial in January 1998 when seismic activity persisted for 12 days and included over 8000 earthquakes as recorded by the SOSUS network (Dziak and Fox, 1999). The volcanic activity resulted in newly encrusted seafloor (Baker *et al.*, 1999; Embley *et al.*, 1999) and evidence of iron and sulfur enrichments in the water column (Feely *et al.*, 1999). The event response team found bacterial numbers in the neutrally-buoyant hydrothermal plume waters to be twice those found in the background values (Cowen *et al.*, 1999). The unique setting and proximity to major oceanographic research institutions has enabled Axial Volcano to be one of the most intensely studied areas of seafloor in the world. The Magic Mountain vent area is located on the Explorer Ridge approximately 425 km from Axial Volcano (Figure 2.2).

Over the 5 year sampling period, 52 MGCs were recovered at seven hydrothermal vent sites at Axial Seamount, and 2 MGCs were recovered at the Magic Mountain area. The samples were either collected on the same cruise as deployment or during following cruises in an effort to achieve both short-term (days) and long-term (annual) time series.

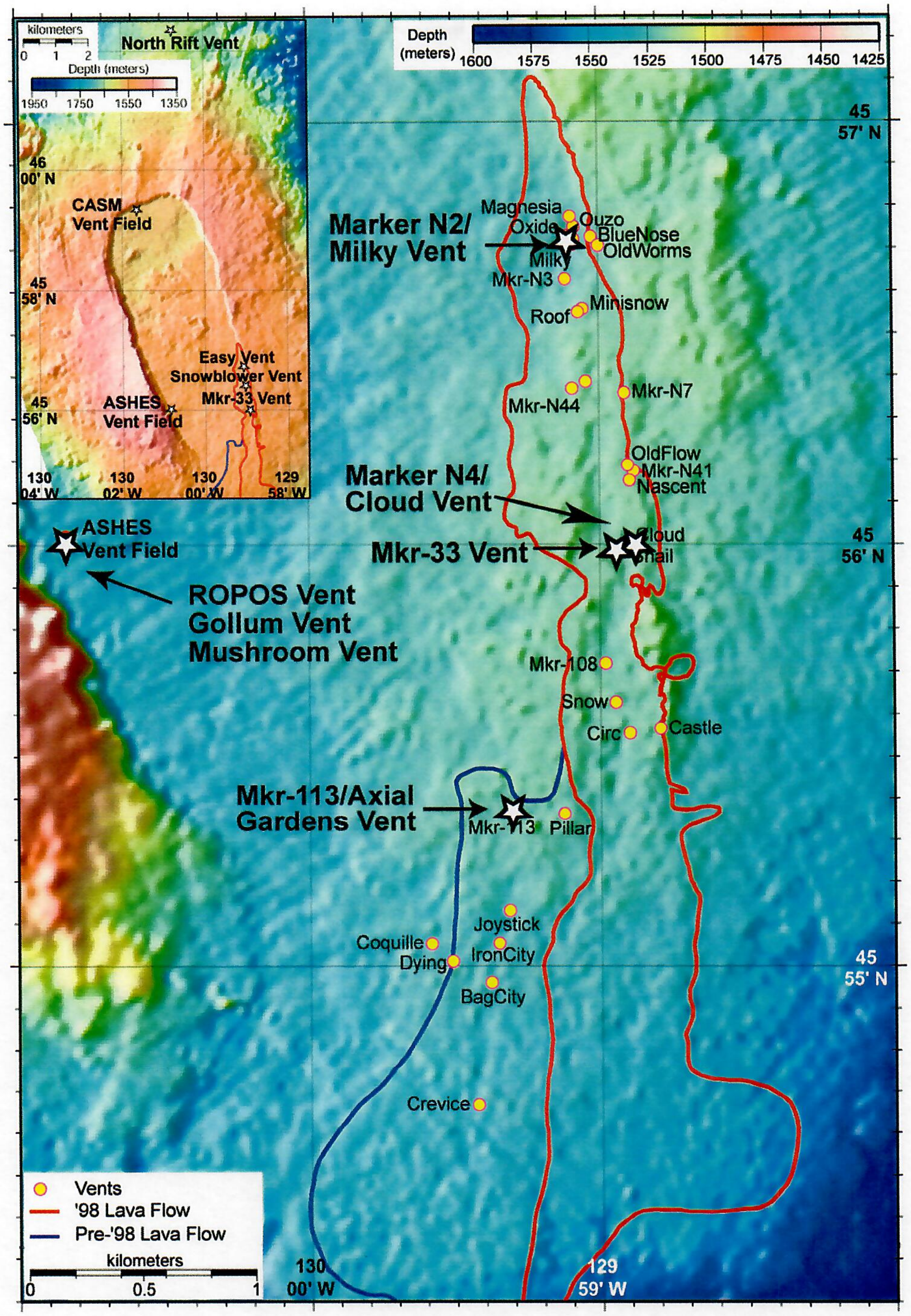

Figure 2.3. Topographic map of Axial Volcano. Inset map shows the caldera structure and exploded view indicates discovered hydrothermal vents. Vents labeled with a star indicate where samples were collected for this study. Figure courtesy of S. Merle, R. Embley, and W. Chadwick, unpublished data; Embley *et al.*, (1999).

Upon retrieval, glass wool was aseptically removed from each chamber and placed either on dry ice or in a –80°C ultrafreezer until transfer to shore, where the samples were then stored at -80°C until processing.

Cellular Extraction

Glass wool from single MGC cylinders was thawed on ice and suspended in a rotating flask containing 500 ml 1X phosphate-buffered saline for 30 min. The detached cellular biomass was decanted and spun for 5 min (10,000 x g) to form a cell pellet. This procedure was repeated (2-3 times) until the glass wool appeared visibly free of organic matter. The concentrated organic material was aliquoted into ~0.5 g (wet weight) samples and was either stored at -80°C or immediately processed.

Genomic DNA Extraction and Purification

DNA was extracted according to the protocol of the FastDNA SPIN Kit (Q-BIOgene, Carlsbad, CA) with the following modifications. A single aliquot of organic material (~0.5 g) was added to the bead solution. The cell lysis procedure was performed using a Mini-BeadBeater (BioSpec Products, Inc., Bartlesville, OK) at 3800 RPM for 1 min. After DNA isolation, the DNA was further purified by size exclusion of low molecular weight solubilized material (<50,000 daltons) using Microcon 50 centrifugal filters (Millipore, Bedford, MA). Each sample was quantified for DNA yield and purity by measuring light absorbance curves from 220 to 320 nm with a spectrophotometer (HP model 8452A, Agilent Technologies, Palo Alto, CA). An aliquant of each sample was examined for size and purity by electrophoresis in a 1% SeaKem GTG agarose (BMA

BioProducts, Rockland, ME) gel suspended in 1X TAE buffer. The gel was stained in a 1X TAE solution containing 1 µg/mL ethidium bromide for 30 min and destained in distilled H_2O for 30 min. The gel was visualized by UV excitation and the images were captured using an EDAS 290 gel documentation system (Eastman Kodak Company, Rochester, NY).

PCR Amplification of SSU rDNA

Amplifications were performed on genomic DNAs from environmental samples using the forward primer (5' – TNA NAC ATG CAA GTC GRR CG) corresponding to positions 49 to 68 of the *Escherichia coli* rRNA numbering system (Brosius *et al.*, 1978), and the reverse primer (5' – RGY TAC CTT GTT ACG ACT T) corresponding to positions 1510 to 1492, where R is purine analog dK, Y is pyrimidine analog dP, and N is an equal mixture of both analogs at a single position (Glen Research, Sterling, VA). For T-RFLP experiments, the forward primer was gel purified and labeled at the 5' end with the phosphoramidite dye 6-FAM (Applied Biosystems, Foster City, CA). PCRs were performed using 25-100 ng community DNA, 5 U *Taq* DNA polymerase (Promega, Madison, WI), 1X PCR buffer (Promega), 2.5 mM $MgCl_2$, 200 µM each dNTP, 10 µg acetylated bovine serum albumin (BSA; Promega), 1 µM each forward and reverse primers, and Optima water (Fisher Scientific, Fair Lawn, NJ) to a total volume of 50 µl. Before BSA and *Taq* polymerase were added, reaction mixtures were heated to 95°C for 2 min and then chilled on ice. The reaction tubes were placed in a GeneAmp 9700 thermocycler (Applied Biosystems) under the following conditions: 30 cycles of denaturation (94°C for 1 min), hybridization (56°C for 90 sec), and elongation (72°C for

3 min), with a final 7 min elongation step. Amplified PCR products were examined by gel electrophoresis to estimate PCR efficiency. Every reaction series included a negative control, which did not contain template DNA. Only reaction series with negative controls not yielding any amplification products were carried through subsequent analyses.

T-RFLP Conditions and Analyses

Amplicons were digested with some or all of the following restriction endonucleases: *Alu*I (AG^CU), *Bst*UI (CG^CG), *Hae*III (GG^CC), *Hha*I (GCG^C), *Hin*fI (G^ANTC), *Msp*I (C^CGG), and *Rsa*I (GT^AC). The treatments used 5 U enzyme (New England Biolabs, Beverly, MA), 1X NEBuffer 2 (New England Biolabs), Optima water, and 15 µl PCR product for a 30 µl enzymatic reaction. Each reaction was incubated at 37°C for 3 hrs, except *Bst*UI which was incubated at 60°C. An aliquant of each sample was examined by gel electrophoresis in a 3% NuSeive 3:1 gel (BMA BioProducts) suspended in 1X TBE buffer. The gel was stained in a 1X TBE solution containing 1 µg/mL ethidium bromide for 30 min and destained in distilled H_2O for 30 min. The remaining samples prepared for T-RFLP were purified of salts according to the protocol of the Qiagen Nucleotide Removal Kit (Qiagen, Valencia, CA) in an effort to remove small anions that may reduce the signal during electrokinetic injection by capillary electrophoresis. The purified samples were eluted in 50 µl of 50 mM Tris, pH 8.5. The reactions were run by adding 2 µl sample to 10 µl deionized formamide and 0.5 µl of ROX-labeled GS500 internal size standard (Applied Biosystems). Each sample was denatured by heating at 95°C for 5 min and transferred to ice. The size and magnitude of the fluorescently labeled terminal restriction fragments were determined by capillary electrophoresis using an ABI 3100 Genetic Analyzer with the Collection Analysis and GeneScan software (Applied Biosystems).

T-RFLP Data Management and Normalization

The chromatograms formulated by GeneScan were imported into Genotyper (Applied Biosystems). Electrophoretic data were converted into tabular format based on peak size and peak height with the exclusion of peak heights less than 5% relative fluorescence units of the highest peak in the chromatogram. Peak sizes less than 50 bp and greater than 500 bp were also excluded. The small peak heights were considered noise and excluded from further analysis. Peak sizes greater than 500 bp were excluded from the analysis due to limitations of the size standard and potential limitations of electrophoresis. Peak sizes less than 50 bp were excluded due to primer dimers and other small charged molecules that may skew the analysis. It is worthy to note that the exclusion of data based on peak size, in this case, does not likely result in the ultimate exclusion of populations within the community because multiple restriction endonucleases were used on each community. The relative fluorescence units for each T-RF were standardized as percentiles of the whole community.

Derivation of Short-Term T-RF Groups

The database of derived T-RF lengths used in chapter one was used to compare with experimentally-derived T-RF lengths from this study. Included in the database were the following additions: representative hydrothermal vent clone isolates from Axial Volcano (AXB OTUs; *n* = 15; Lynch, 2000), and representative hydrothermal vent clone isolates from an *in situ* growth chamber at a Mid-Atlantic Ridge vent site ($n = 22$; Corre *et al.*, 2001). A scoring algorithm was designed to resolve the most likely database candidates with experimental data (Nilsson and Strom, 2002). If a database entry contained a T-RF size within ±4 bp of experimental data, then the entry would receive a score of 1. Database entries received an additional "point" for each successive matching T-RF. For example, if a database entry had three matching T-RF sizes from the seven restriction endonucleases used then that database entry would receive a score of 3. Each database entry that matched at least one of the experimentally-derived T-RF sizes from short-term deployed MGCs or contained a T-RF outside the readable range of 50-500 bp received a score of 7 and thus remained in the database. The remaining entries were subdivided into T-RF groups based on a heuristic algorithm with the capacity to distinguish among groups. For example, the seven restriction endonucleases used in this study were unable to distinguish among five *Campylobacter* spp. sequence entries; therefore, a T-RF group called CAMP was created. Each T-RF group was checked for presence or absence within a community at increasing levels of fluorescence intensity (based on the standardized height of an electrophoretic peak).

Statistical Analyses of Microbial Growth Chambers

MGCs placed for short- and long-term experiments were compared based on their respective terminal-restriction fragment patterns using agglomerative, hierarchical clustering. Euclidean distances between MGC communities were calculated and Ward's clustering algorithm was used to agglomerate clusters (Ward, 1963). Boxplot analysis was performed to provide a graphical depiction of the difference between T-RFs of short- and long-term MGC communities. The statistical analyses were performed using R, a freely available statistical software package (Ihaka and Gentleman, 1996).

RESULTS

Microbial Growth Chamber T-RFLP Analyses

DNA was extracted from 52 microbial growth chamber (MGC) -derived communities. Thirty representative cylinders from short- and long-term MGCs were processed and analyzed via T-RFLP to reduce sampling redundancy. Of the 30, 16 consisted of short-term (days) incubation periods and 14 were recovered after long-term (years) periods of deployment. Each MGC was analyzed by some combination of seven individual restriction endonuclease digests (Table 2.1). Of the seven vent sites sampled at Axial, four were in the 1998 eruption lava field (Marker 33, Marker N4, Marker 113, and Marker N2) and three were in the ASHES vent field (ROPOS vent, Mushroom vent, and Gollum vent). Short-term data were achieved at Marker 33 in the sampling years 1998-2001, at Marker N4 for 1998-2000, and at Marker 113 and Marker N2 for the 1998 sampling year only. Long-term data were achieved at Marker 33 in 1999-2001, at Marker N4 for 1999 and 2001, at Marker 113 for 1999 and 2001, at Gollum vent in 1999-2001, at Mushroom vent in 2000, and at ROPOS vent in 2000-2001. Two vent sites were also sampled from Explorer Ridge at the Magic Mountain area. Both of these vents, Marker 81 and Marker 73, consisted of short-term deployments in 2002 (Table 2.1).

Colonization and Succession

The number of T-RF peaks between short- and long-term deployed MGC communities contrasted in that diversity in short-term experiments was lower than in long-term experiments (Figure 2.4). To compare contrasting diversities in terms of species richness among populations detectable by T-RFLP, the electropherogram with

Table 2.1. Summary of location, duration, and collection year for MGC communities used in this study. Also listed are the restriction endonuclease treatments for which electropherograms were produced. T-RFLP analysis was performed on glass wool obtained from single cylinders of the microbial growth chambers.

Region	Area	Vent	Incub. Time (days)	Collect. Year	Restriction endonucleases						
					*Alu*I	*Bst*UI	*Hae*III	*Hha*I	*Hinf*I	*Msp*I	*Rsa*I
Axial	Mk33/Cloud	Mk33	2	1998			X			X	
	Mk33/Cloud	Mk33	2	1998			X			X	
	Mk33/Cloud	Mk33	14	1998			X			X	
	Mk33/Cloud	Mk33	16	1998	X	X	X	X	X	X	X
	Mk33/Cloud	Mk33	16	1998			X			X	
	Mk33/Cloud	Mk33	5	1999	X	X		X	X		X
	Mk33/Cloud	Mk33	7	2000	X	X		X	X		X
	Mk33/Cloud	Mk33	9	2001	X	X		X	X		X
	Mk33/Cloud	N4	2	1998			X			X	
	Mk33/Cloud	N4	2	1998	X	X	X	X	X	X	X
	Mk33/Cloud	N4	5	1999	X	X		X	X		X
	Mk33/Cloud	N4	7	2000	X	X		X	X		X
	Axial Garden	Mk113	4	1998	X	X	X	X	X	X	X
	Milky	N2	12	1998	X	X		X	X		X
	Mk33/Cloud	Mk33	296	1999	X	X	X	X	X	X	X
	Mk33/Cloud	Mk33	375	2000	X	X	X	X	X	X	X
	Mk33/Cloud	Mk33	375	2001	X	X	X	X	X	X	X
	Mk33/Cloud	N4	297	1999	X	X	X	X	X	X	X
	Mk33/Cloud	N4	378	2001	X	X	X	X	X	X	X
	Axial Garden	Mk113	308	1999	X	X	X	X	X	X	X
	Axial Garden	Mk113	1063	2001	X	X	X	X	X	X	X
	ASHES	Gollum	307	1999	X	X	X	X	X	X	X
	ASHES	Gollum	360	2000	X	X	X	X	X	X	X
	ASHES	Gollum	339	2001	X	X	X	X	X	X	X
	ASHES	Gollum	1056	2001	X	X	X	X	X	X	X
	ASHES	ROPOS	672	2000	X	X	X	X	X	X	X
	ASHES	ROPOS	340	2001	X	X	X	X	X	X	X
	ASHES	Mushroom	669	2000	X	X	X	X	X	X	X
Explorer	Magic Mtn	Mk81	5	2002	X			X		X	X
	Magic Mtn	Mk73	5	2002	X			X		X	X

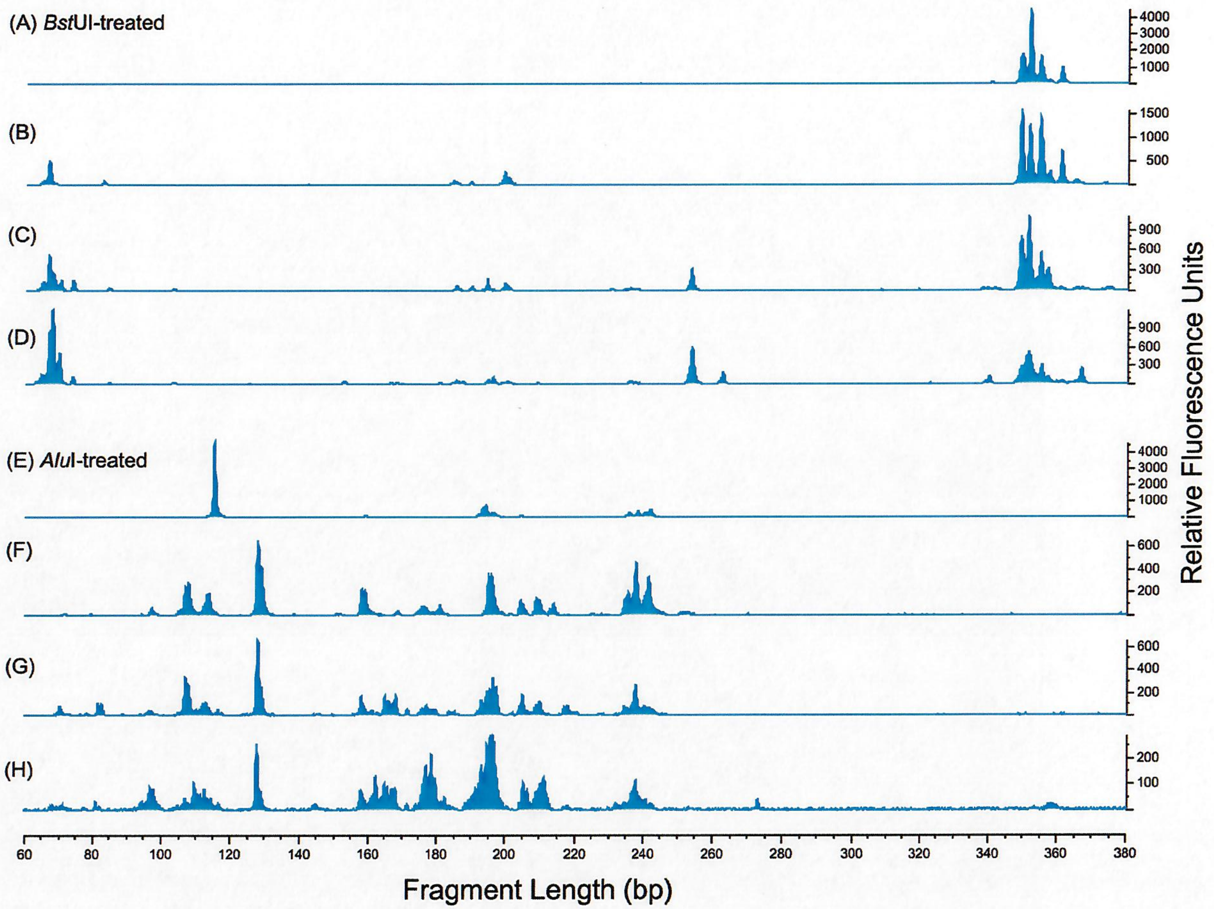

Figure 2.4. T-RFLP profiles from four MGCs placed at the Marker 33 hydrothermal vent. Terminal fragments were generated from SSU rDNA amplicons digested with (A-D) *Bst*UI and (E-H) *Alu*I. (A, E) Collected in 1998 after a 16-day deployment; (B, F) collected in 1999 after a 296-day deployment; (C, G) collected in 2000 after a 375-day deployment; (D, H) collected in 2001 after a 375-day deployment.

the greatest number of T-RFs (i.e., the best indicator of species richness) was chosen for each MGC. The resulting data were binned and examined for homoscedasticity and a normal distribution (prerequisites for parametric statistical testing). The Shapiro-Wilks test is specifically designed to detect departures from normality, where a $p > 0.05$ implies normality, and suggested that both short- and long-term data were distributed normally ($p = 0.13$ and 0.44, respectively). A notched boxplot analysis indicated T-RF frequencies for long-term deployed MGCs were distributed with relatively higher variability while T-RF frequencies of short-term deployed MGCs were relatively homoscedastic. The notched boxplots supported the evidence that species richness increased over long-term incubation periods (Figure 2.5). The difference between time zero and the average number of T-RFs for short-term deployed MGCs ($x_{avg} = 8$) implies a first order approximation of bacterial colonization rate at hydrothermal vents. The difference between short- and long-term deployed MGCs implies a reduced level of colonization with time.

Community Similarity Patterns

The analysis of community similarity based on T-RF patterns was done by performing agglomerative, hierarchical clustering on short- and long-term deployed MGCs. The MGCs placed for short-term durations were clustered based on the cumulative results of the three restriction endonuclease digestions (*Alu*I, *Hha*I, and *Rsa*I) that were empirically determined to provide sufficient taxonomic information necessary for categorizing T-RF groups. The MGC communities placed for short-term durations clustered into two relatively distinct groups, each of which was characterized by

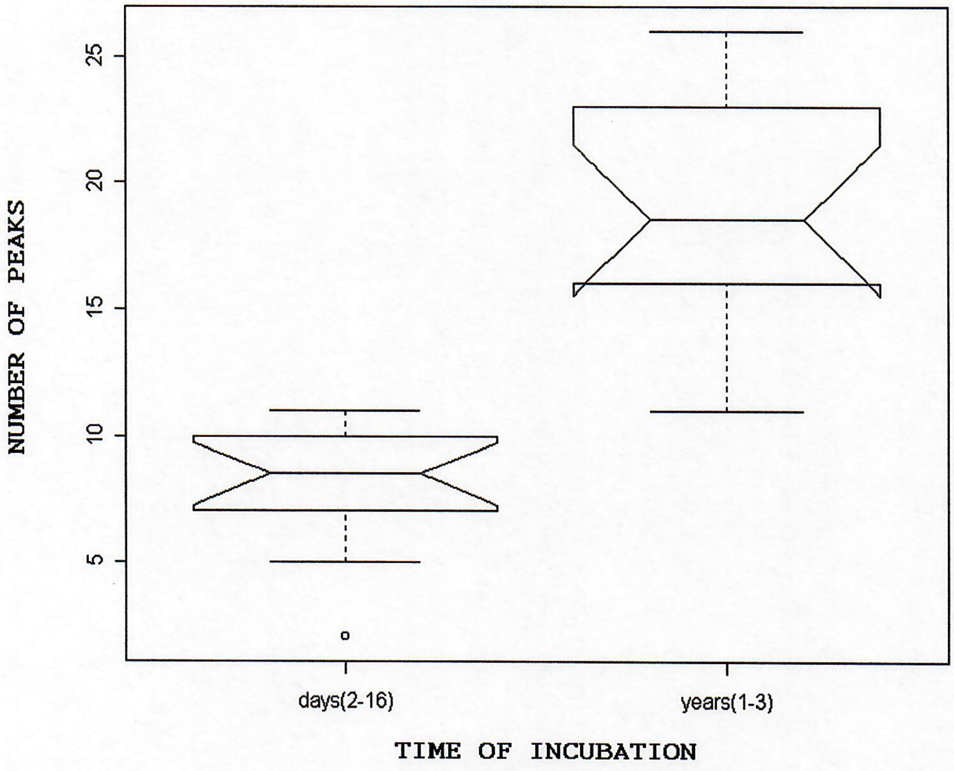

Figure 2.5. Boxplot of temporally-categorized data versus the number of T-RFs as determined by the analysis software. The number of T-RFs for each microbial growth chamber was chosen based on the restriction endonuclease that identified the greatest number of T-RFs for each community. The centered line is the median and the borders are the interquartile values (the medians of the upper and lower values). The whiskers indicate minimum and maximum values while the circle indicates an outlier defined as a distance that is greater than 1.5X the value of the interquartile range from that end of the box. The notch width = median +/- 1.58 IQR / sqrt (n). If the data are normal then the above equation is an estimate of 1.96 standard errors. Since the notches of the two plots do not overlap then the medians are significantly different at the 5 percent level.

hydrothermal vents from which multiple time series were taken (Marker 33 and Marker N4). The Marker 113 community was inclusive to the Marker 33 group. The two Explorer Ridge vents (Marker 73 and Marker 81) were also loosely clustered into the Marker 33 group, but were by far most similar to each other. The Marker N2 community grouped with the Marker N4 communities, but was the most dissimilar within its respective cluster (Figure 2.6).

The MGCs placed for long-term durations were clustered using T-RF patterns from the seven restriction endonucleases. One of the two long-term deployed MGCs collected in 2001 from Gollum Vent clustered autonomously from the other long-term MGC communities. This uniqueness is likely a reflection of the apparent low diversity compared to other long-term placed MGC communities (Figure 2.7). The MGCs placed for long-term durations did not cluster into recognizable spatial or temporal patterns based on the results of agglomerative, hierarchical cluster analysis (Figure 2.7).

Pioneer Populations

Experimentally-derived T-RFs for MGCs incubated for short-term durations were compared to binned theoretical T-RFs from the Ribosomal Database Project. For example, a single T-RF peak size of 351 bp was found in the *Bst*UI-digested community for a short-term MGC deployment. There are 687 sequence variants in the database with a *Bst*UI terminal cut site of 351 ± 4 bp. The same MGC community contained a single 498 bp peak size for the *Hha*I digest. There are 11 sequence variants in the database containing both *Bst*UI and *Hha*I cut sites. A diagnostic flowchart was created to classify

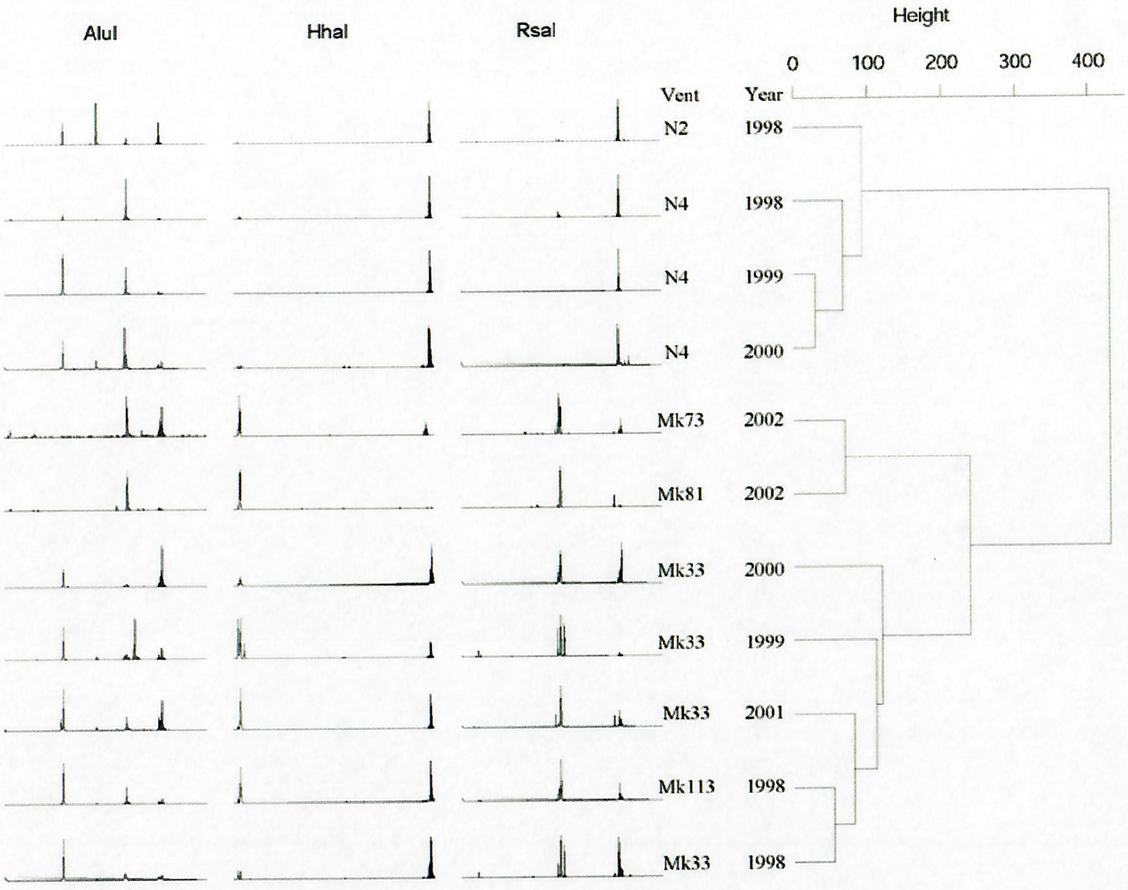

Figure 2.6. Agglomerative, hierarchical clustering dendrograms of short-term MGC communities based on similarities of terminal-restriction fragment size and height patterns. Samples from multiple restriction endonuclease digestions (*Alu*I, *Hha*I, and *Rsa*I) were concatenated for a single MGC. Height units are a result of the minimum variance clustering algorithm (Ward, 1963). Similar dendrogram topologies were observed when other clustering methods were used.

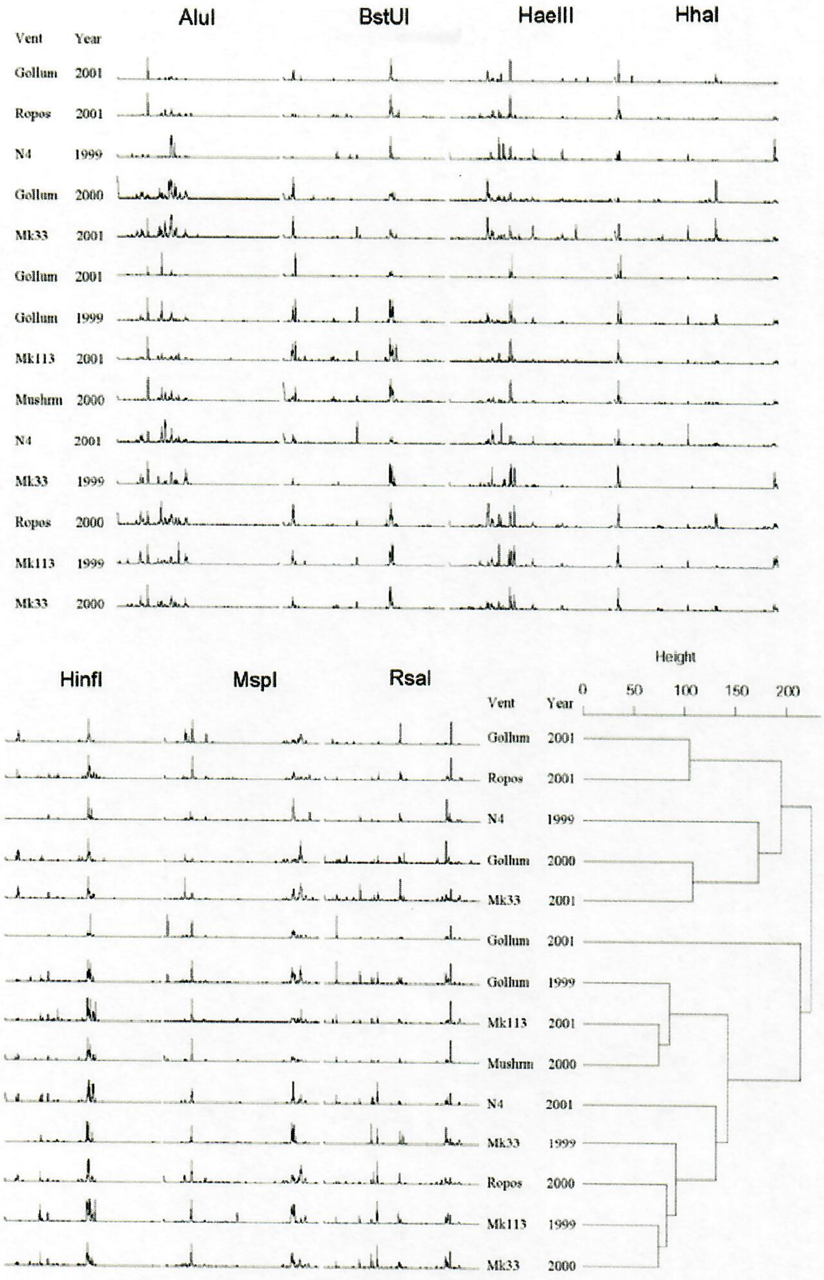

Figure 2.7. Agglomerative, hierarchical clustering dendrograms of long-term MGC communities based on similarities of terminal-restriction fragment size and height patterns. Samples from multiple restriction endonuclease digestions (*Alu*I, *Bst*UI, *Hae*III, *Hha*I, *Hin*fI, *Msp*I, and *Rsa*I) were concatenated for a single MGC. Height units are a result of the minimum variance clustering algorithm (Ward, 1963). Similar dendrogram topologies were observed when other clustering methods were used.

sequence variants into T-RF groups to facilitate the observation of population distributions (Figure 2.8).

All MGCs that were deployed for short-term periods showed collective *Hae*III-, *Msp*I-, and *Bst*UI-cut T-RF sizes that were inclusively correlated to 18 taxa within the subdivision *epsilon-Proteobacteria* (Table 2.2). The 18 sequence entries whose T-RF sizes were inclusive to short-term electropherograms were subcategorized into 5 T-RF groups based on their specific associated *Hha*I-, *Rsa*I-, and *Alu*I-cut T-RF patterns within the diagnostic flowchart (Figure 2.6; Figure 2.8).

A graduated presence/absence table was generated by corresponding the relative peak heights of diagnostic T-RF sizes with their respective T-RF groups (Table 2.3). Marker 33 consisted primarily of AXB9 group and PVBS group throughout sampling years 1998-2001. Marker N4 was dominated by the AXB9 and EPIS groups throughout sampling years 1998-2000. Marker 113 was similar to Marker 33 in that it was dominated by AXB9 group and PVBS group during the 1998 sampling year. Marker N2 consisted primarily of AXB2 during the 1998 sampling year. Both MGCs recovered from the Explorer Ridge vents were dominated by the PVBS group (Table 2.3). The categorized T-RF group patterns distinguished by a graduated presence/absence analysis were similar to the dendrogram clusters for short-term incubated MGC communities (Figure 2.6; Table 2.3).

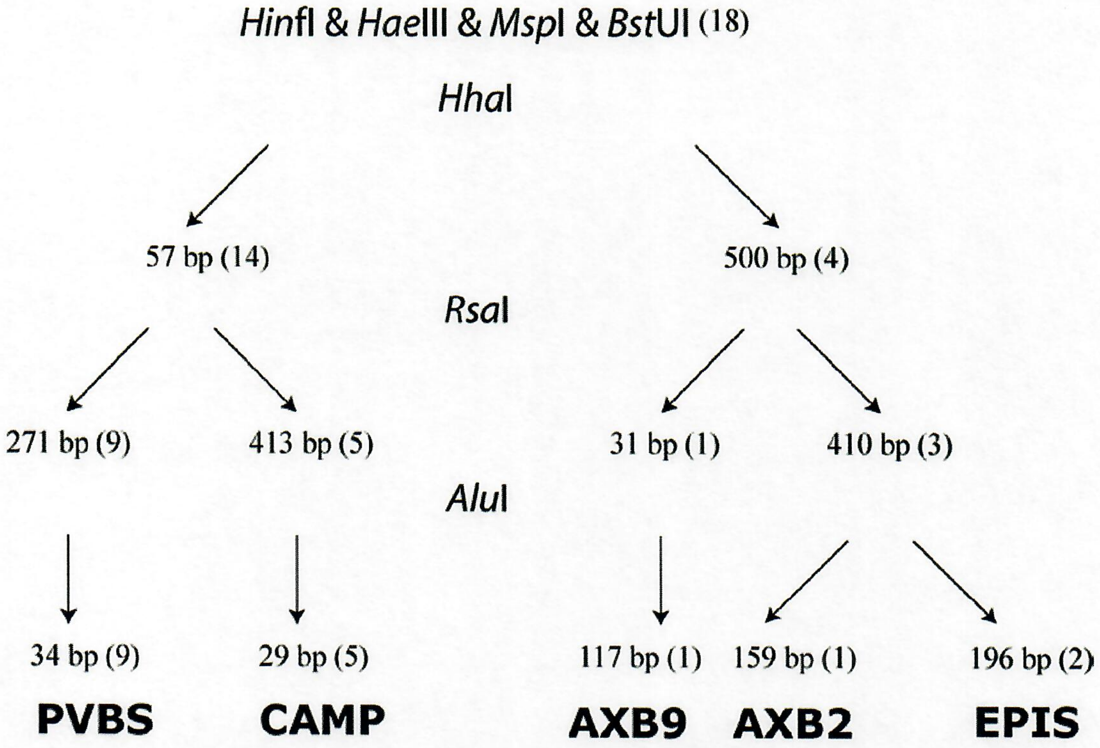

Figure 2.8. A flowchart of the diagnostics that were performed to distinguish the five T-RF groups associated with pioneer bacterial populations. The numbers in parentheses indicate the number of sequence variants that match the T-RF size.

Table 2.2 Names of taxa whose SSU rDNA sequences have corresponding T-RF sizes with pioneer populations at hydrothermal vents in this study.

Species or Clone (Genbank Accession No.)	T-RF Group	Reference
AXB OTU9	AXB9	Lynch, 2000
AXB OTU2	AXB2	Lynch, 2000
symbiont of *Rimicaris exoculata* (U29081)	EPIS	Polz and Cavanaugh, 1995
symbiont of *Alvinella pompejana* (L35520)	EPIS	Haddad *et al.*, 1995
Campylobacter cf. lanienae (AF043422-24)	CAMP	Logan *et al.*, 2000
Campylobacter jejuni (Z29326)	CAMP	Kim *et al.*, 1993
Campylobacter jejuni (Y19244)	CAMP	Krogfelt *et al.*, unpublished
PVB OTU2 (U15100-U15102)	PVBS	Moyer *et al.*, 1995
PVB OTU3 (U15103-U15105)	PVBS	Moyer *et al.*, 1995
PVB OTU8 (U15107)	PVBS	Moyer *et al.*, 1995
VC1.2Cl32 (AF367492)	PVBS	Corre *et al.*, 2001
strn. JTB146 (AB015257)	PVBS	Li *et al.*, 1999

Table 2.3. Graduated presence/absence table of pioneer T-RF groups as they occur at hydrothermal vents in this study. The numbers reflect a scale associated with percentiles of the whole community as seen using three diagnostic restriction endonucleases (*Hha*I, *Rsa*I, and *Alu*I). Some microbially significant physical and chemical characteristics of the sample sites are included for comparison.

	Vent	**Year**	**T-RF groups**					**Vent conditions**[a]		
			PVBS	**AXB9**	**EPIS**	**AXB2**	**CAMP**	**Temp (°C)**	**pH @ 22°C**	**H_2S (µmol/kg)**
Axial Volcano	Mk33	1998	2	5	+	+	1	39.0	4.99	2100
		1999	3	2	+	+	+	67.6	4.92	2300
		2000	2	1	+	--	2	24.8	6.15	17.2
		2001	4	2	+	+	1	12.4	6.05	73.5
	MkN4	1998	+	1	7	+	+	20.6	5.34	719
		1999	+	4	4	+	--	17.0	5.99	10.5
		2000	--	2	3	+	+	15.9	n.d.	79.0
	Mk113	1998	4	5	--	--	1	20.0	5.62	740
	MkN2	1998	+	1	1	3	+	8.0	6.22	31.0
Explorer Ridge	Mk73	2002	6	--	1	--	1	n.d.	n.d.	n.d.
	Mk81	2002	7	--	+	--	+	n.d.	n.d.	n.d.

[a] courtesy of David Butterfield, NOAA PMEL.

--	not present
+	1-9%
1	10-19%
2	20-29%
3	30-39%
4	40-49%
5	50-59%
6	60-69%
7	70-100%

DISCUSSION

This study employed a novel sampling technique through the deployment and recovery of MGCs to observe representative bacterial communities growing on surfaces exposed to hydrothermal vent effluent. The goal was to observe bacterial colonization and community succession at various hydrothermal vents within and distal from Axial Volcano by utilizing the high-throughput T-RFLP technique.

The use of molecular biological techniques to detect and study naturally occurring microbial populations has some distinct advantages, encountered herein, when compared to other methods. Estimates of biomass, average growth rate, and carbon production can be effectively obtained through the use of microscopy, culturing, and radiolabeled metabolic precursor molecules. Unfortunately for microbial ecologists, this suite of methods does not adequately address questions of diversity because they are limited to identifying a small fraction of a bacterial community (Pedros-Alio, 1993; Amann *et al.*, 1995; Dykhuizen, 1998). The substituted probing, via PCR, of a community for SSU rDNA has proved an effective tool in answering questions about microbial diversity, ecology, and evolution (Head *et al.*, 1998).

There are also specific intrinsic limitations of molecular biological techniques when used to assess bacterial diversity and experimental outcomes must be scrutinized within this context. Genomic DNA extraction techniques often utilize a physical means (i.e., rigorous shaking or "bead-beating") to free nucleic acids from within cells. The resulting extracted DNA may be sheared or preferentially broken within weak (e.g., TA-rich) regions. This could cause a potential bias in the analyzed bacterial population due to fewer contiguous template DNAs available for amplification. To minimize DNA breakage, optimization of lysis times can be achieved by observing for high molecular weight genomic DNA (> 12 kilobases) on an agarose gel. When probing a natural community for a large subset of populations (e.g., the domain *Bacteria*), the design of degenerate PCR primers (primers with the ability to hybridize more than one target sequence) in relatively conserved regions of the SSU rDNA is essential to reduce the effects of PCR selection or preferential binding and an ultimate distortion of the observed representative populations (Lane, 1991). The potential for artifact sequences (i.e., chimeras, mutations, and heteroduplexes) generated during PCR were minimized by reducing DNA shearing, and minimizing the number of PCR cycles (n = 30) needed to obtain a PCR product (Qiu *et al.*, 2001). It is impossible to evaluate the effects of PCR-generated artifacts during T-RFLP analysis, but PCR conditions during this study were similar to a sequence-based study from this laboratory where PCR artifacts were found to be minimal (Lynch, 2000).

The T-RFLP technique is sensitive and precise; however, it also yields low resolution when analyzing complex microbial communities and is most aptly suited to replicated, field-scale studies and low to intermediate diversity situations (Chap. 1; Kuske *et al.*, 2002; Dunbar *et al.*, 2000). When a single population dominates a community, it is possible to infer evolutionary lineage or a contextual phylogenetic placement by comparing derived T-RF sizes for multiple restriction endonuclease digestions with theoretical T-RF sizes of published sequence data. When diagnosing T-RFs, it is necessary to include a range of theoretical T-RF sizes (e.g., ±4 bp) as an accuracy within 1 bp is rarely achieved. Some investigators have reported the discrepancy (Clement *et al.*, 1998; Nilsson and Strom, 2002) and, curiously, when the analysis is performed on a capillary electrophoresis unit, the observed T-RF sizes are consistently less than theoretically predicted (Nilsson and Strom, 2002).

The similarity in the number of T-RF peaks associated with each of the incubation times, regardless of vent location or sampling year, provides evidence of bacterial community succession (Figure 2.5). The introduction of an inert colonization substrate exposed to vent effluent allowed for the proliferation of pioneer bacteria. The initial colonizers were likely chemolithotrophic bacterial populations with efficient dispersal mechanisms, and high growth rates (Atlas and Bartha, 1998) when exposed to a variety of reduced solutes (Karl, 1995). Multiple metabolic strategies and ecological niches became available through time, allowing for succession as realized by an increase in the number of bacterial populations (Atlas and Bartha, 1998). Additional sampling on a refined time scale would likely enhance this characterization of succession and yield a broader context to bacterial colonization rate at hydrothermal vents.

The *epsilon-Proteobacteria* subdivision consists primarily of taxa adapted to utilize multiple metabolic pathways in the oxidation and/or reduction of sulfur compounds in microaerophilic and anaerobic environments (Vandamme *et al.*, 1991). The finding that T-RF patterns from pioneer populations in this study are exclusive to *epsilon-Proteobacteria* generally supports the accepted hypothesis that sulfur-oxidation is the predominant source of energy for the reduction of CO_2 to organic carbon (Jannasch, 1995; McCollum and Shock, 1997). Results from this study also suggest that certain *epsilon-Proteobacteria* have exploitive dispersal and settling mechanisms providing an advantage in early colonization which may explain their cosmopolitan nature within free-living hydrothermal vent bacterial communities (Moyer *et al.*, 1995; Polz and

Cavanaugh, 1995; Kleinkauf, 2000; Lynch, 2000; Corre *et al.*, 2001; Longnecker and Reysenbach, 2001; Reysenbach *et al.*, 2000).

By utilizing the T-RFLP approach, this study distinguished the relative dominance of five specific *epsilon-Proteobacteria* pioneer T-RF groupings at hydrothermal vents through space and time (Figure 2.8; Table 2.3). The PVBS group was most similar to closely-related clones isolated from Pele's vents of the Loihi Seamount which were inclusive to the Thiovulum lineage (Moyer *et al.*, 1995), a clone from the Snake Pit vent on the Mid-Atlantic Ridge (Corre *et al.*, 2001), and a clone from a cold-seep community near *Calyptogena* clams in the Japan trench (Li *et al.*, 1999). The EPIS group was most similar to episymbionts associated with the hydrothermal vent shrimp *Rimicaris exoculata* (Polz and Cavanaugh, 1995), and the hydrothermal vent polychaetous annelid *Alvinella pompejana* (Haddad *et al.*, 1995). The shrimp and worm epibionts were isolated from Mid-Atlantic Ridge and East Pacific Rise habitats, respectively. The AXB2 and AXB9 T-RF groups are each most similar to phylotypes isolated from microbial mat communities at Axial Volcano (Lynch, 2000). The CAMP group is most similar to variants of *Campylobacter* spp. which were isolated as human pathogens (Logan *et al.*, 2000; Kim *et al.*, 1993; Krogfelt *et al.*, unpublished). The taxa associated with the five groupings distinguished in this study (Table 2.2) serve as taxonomic markers to track experimentally-defined populations, and yield only an inferred evolutionary lineage.

Due to the metabolic reliance of a chemosynthetic community on the physico-chemical conditions of hydrothermal vent fluids, the *a priori* hypothesis was that geochemistry, specifically [H_2S], would correlate with the chemolithotrophic populations of the community (Butterfield *et al.*, 1997). The patterns associated with pioneer populations do not appear to be tightly-regulated by changes in the physico-chemical conditions of vent fluid (Table 2.3; Figure 2.6). Furthermore, the temporal population constancy at Marker 33 and Marker N4 indicates that the dispersal of bacteria to seafloor microbial mat populations may be dominated by a localized subseafloor community. Alternatively, the conserved populations observed in this study may be a result of differences between fluid-flow regimes present at Marker 33 and Marker N4. The MGCs placed at Marker 33 were likely exposed to more interactions with seawater because the flow of venting fluid was focused. MGCs deployed at Marker N4 were placed in a hole that was ~0.5 m diameter and ~1 m depth. The MGCs at Marker N4 were saturated by venting fluids and not exposed directly to seawater during incubation. The EPIS group may be more sensitive to direct exposure with seawater and/or the PVBS group may require more mixing with seawater (Table 2.3).

The increased number of T-RF peaks associated with long-term deployed MGCs made it challenging to correlate peaks with taxa. Peak sizes associated with pioneer groups were observed in long-term deployments (Figure 2.4) but their relative abundances were indeterminate. The T-RF patterns of long-term deployed MGCs did not cluster into recognizable patterns associated with spatial or temporal variability (Figure 2.7). The apparent random distribution of individual populations may indicate that the array of populations and consequently the genetic diversity is malleable once the number of new ecological niches is reduced or eliminated. The stability of a community is conventionally attributed by the persistence of individual populations because persistence is related to the fitness of individuals (Grimm et al., 1992). A more rigorous proxy of community stability is based on the relative contribution of metabolic rates to the ecosystem irrespective of population persistence (Fernandez et al., 1999; Tilman, 1996). The complexity of physico-chemical gradients and heterogeneous microhabitats within long-term deployed MGCs is unknown but assumed to be greater than that of the more homogeneous short-term deployed MGCs where pioneer populations arise. The functional or metabolic diversity may be a better descriptor of relationships between vent chemistry and microbial community structure. For example, rates of sulfur- and iron-oxidation, anaerobic metabolisms, and heterotrophy could all be potential constraints of functional diversity once the ecological niches are established. In addition, the community synopsis that was achieved in this study would not delineate fluctuations in population levels and types that are subjected to pressures such as predator-prey mechanisms, interspecific competition, disease, starvation-survival, and allochthonous inputs.

In summary, this study demonstrates a novel approach for estimating bacterial community structure, succession, and diversity from hydrothermal vents using *in situ* microbial growth chambers combined with molecular biological techniques. The high degree of sampling in this study shows evidence for predictable early colonization by closely-related bacterial populations. Community succession was observed by an increase in the number of populations, but the inferred identity of successional populations could not be established beyond the pioneer populations. It was taxonomically determined that the pioneer populations are exclusively composed of members of the subdivision *epsilon-Proteobacteria* similar to those determined from previous studies at hydrothermal vents occurring worldwide. The T-RFLP assay was also successfully used to distinguish among groups of *epsilon-Proteobacteria* and to detect a previously unobserved temporal population constancy at individual hydrothermal vents. The pioneer populations comprising the bacterial communities at Marker 33 and Marker N4 were locally restricted throughout multiple sampling seasons regardless of fluctuations in geochemistry indicating a potential for dispersal from separate subsurface communities.

LITERATURE CITED

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.

Atlas, R.M., and R. Bartha. 1998. Microbial Ecology: Fundamentals and applications (4th ed.). Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Atwater, T., and P. Molnar. 1973. Relative motion of the Pacific and North American plates deduced from sea-floor spreading in the Atlantic, Indian, and South Pacific oceans. Stanford Univ. Publ. Geol. Sci. 13: 136-148.

Avannis-Aghajani, E., K. Jones, D. Chapman, and C. Brunk. 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. Biotechniques 17: 144-149.

Baker, E.T., C.G. Fox, and J.P. Cowen. 1999. *In situ* observations of the onset of hydrothermal discharge during the 1998 submarine eruption of Axial Volcano, Juan de Fuca Ridge. Geophys. Res. Lett. 26: 3445-3448.

Baross, J.A., and S.E. Hoffman. 1985. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. Origins of Life 15: 327-345.

Brinkhoff, T., S.M Sievert, J. Kuever, and G. Muyzer. 1999. Distribution and diversity of sulfur-oxidizing *Thiomicrospira* spp. at a shallow-water hydrothermal vent in the Aegean Sea (Milos, Greece). Appl. Environ. Microbiol. 65: 3843-3849.

Brosius, J., J.L. Palmer, H.P. Kennedy, and H.F. Noller. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. USA 75: 4801-4805.

Brock, T.D. 1987. The study of microorganisms *in situ*: progress and problems. Symp. Soc. Gen. Microbiol. 41: 1-17.

Butterfield, D.A., I.R. Jonasson, G.J. Massoth, R.A. Feely, K.K. Roe, R.E. Embley, J.F. Holden, R.E. McDuff, M.D. Lilley, and J.R. Delaney. 1997. Seafloor eruptions and evolution of hydrothermal fluid chemistry. Philos. Trans. R. Soc. Lond. A. 355: 369-386.

Chase, R.L., J.R. Delaney, J.L. Karsten, H.P. Johnson, S.K. Juniper, J.E. Lupton, S.D. Scott, V. Tunnicliffe, S.R. Hammond, and R.E. McDuff. 1985. Hydrothermal vents on an axis seamount of the Juan de Fuca ridge (Canadian American Seamount Expedition). Nature 313: 212-214.

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

Cooksey, K.E., and B. Wigglesworth-Cooksey. 1995. Adhesion of bacteria and diatoms to surfaces in the sea: a review. Aquat. Microb. Ecol. 9: 87-96.

Corliss, J.B., J. Dymond, L.I. Gordon, J.M. Edmond, R.P von Herzen, R.D. Ballard, K. Green, D. William, A. Bainbridge, K. Crane, and T.H. van Andel. 1979. Submarine thermal springs on the Galapagos rift. Science 203: 1073-1083.

Corre, E., A.L. Reysenbach, and D. Prieur. 2001. Epsilon-proteobacterial diversity from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge. FEMS Microbiol. Lett. 205: 329-335.

Cowen, J.P., R. Shackelford, D. McGee, P. Lam, E.T. Baker, and E. Olson. 1999. Microbial biomass in the hydrothermal plumes associated with the 1998 Axial Volcano eruption. Geophys. Res. Lett. 26: 3637-3640.

Dang, H., and C.R. Lovell. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. Appl. Environ. Microbiol. 66: 467-475.

Dang, H., and C.R. Lovell. 2002. Numerical dominance and phylotype diversity of marine *Rhodobacter* species during early colonization of submerged surfaces in coastal marine waters as determined by 16S ribosomal DNA sequence analysis and fluorescence in situ hybridization. Appl. Environ. Microbiol. 68: 496-504.

Davis, E. 2000. Volcanic action at Axial Seamount. Nature 403: 379-380.

Delong, E.F., D.G. Franks, and A.L. Alldredge. 1993. Phylogenetic diversity of aggregate-attached versus free-living marine bacterial assemblages. Limnol. Oceanogr. 38: 924-934.

Dunbar, J., S. White, and L. Forney. 1997. Genetic diversity through the looking glass: Effect of enrichment bias. Appl. Environ. Microbiol. 63: 1326-1331.

Dunbar, J., L.O. Ticknor, and C.R. Kuske. 2000. Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment length analysis. Appl. Environ. Microbiol. 66: 2943-2950.

Dunbar, J., S.M. Barns, L.O. Ticknor, and C.R. Kuske. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. Appl. Environ. Microbiol. 68: 3035-3045.

Dykhuizen, D.E. 1998. Santa Rosalia revisited: Why are there so many species of bacteria? Antonie van Leeuwenhoek 73: 25-33.

Dziak, R.P., and C.G. Fox. 1999. The January 1998 earthquake swarm at Axial Volcano, Juan de Fuca Ridge: Hydroacoustic evidence of seafloor volcanic activity. Geophys. Res. Lett. 26: 3429-3432.

Embley, R., and E. Baker. 1999. Interdisciplinary group explores seafloor eruption with remotely operated vehicle. EOS Trans. Am. Geophys. Union 80: 213.

Embley, R.W., W.W. Chadwick Jr., D. Clague, and D. Stakes. 1999. 1998 eruption of Axial Volcano: multibeam anomalies and seafloor observations. Geophys. Res. Lett. 26: 3425-3428.

Emerson, D.E. , and C.L. Moyer. 2002. Neutrophilic Fe-oxidizing bacteria are abundant at the Loihi Seamount hydrothermal vents and play a major role in Fe oxide deposition. Appl. Environ. Microbiol. 68: 3085-3093.

Feely, R.A., E.T. Baker, G.T. Lebon, J.F. Gendron, and J.A. Resing. 1999. Evidence for iron and sulfur enrichments in hydrothermal plumes at Axial Volcano following the January-February 1998 eruption. Geophys. Res. Lett. 26: 3649-3652.

Fernandez, A., S. Huang, S. Seston, J. Xing, R. Hickey, C. Criddle, and J. Tiedje. 1999. How stable is stable? Function versus community composition. Appl. Environ. Microbiol. 65: 3697-3704.

Grassle, J.F., C.J. Berg, J.J. Childress, J.P. Grassle, R.R. Hessler, H.W. Jannasch, D.M. Karl, R.A. Lutz, T.J. Mickel, D.C. Rhoads, H.L. Sanders, K.L. Smith, G.N. Somero, R.D. Turner, J.H. Tuttle, P.J. Walsh, and A.J. Williams. 1979. Galapagos '79: Initial findings of a deep-sea biological quest. Oceanus 22: 2-10.

Grimm, V., E. Schmidt, and C. Wissel. 1992. On the application of stability concepts in ecology. Ecol. Model. 63: 143-161.

Haddad, A., F. Camacho, P. Durand, and S.C. Cary. 1995. Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. Appl. Environ. Microbiol. 61: 1679-1687.

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

Hugenholtz, P., B.M. Goebel, and N.R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 180: 4765-4774.

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. 5: 299-314.

Jannasch, H.W. 1983. Microbial processes at deep sea hydrothermal vents. *In* P.A. Rona, K. Boström, L. Laubier, and K.L. Smith Jr. (eds.), Hydrothermal processes at seafloor spreading centers. Plenum Press, New York, pp. 677-709.

Jannasch, H.W., and M.J. Mottl. 1985. Geomicrobiology of deep-sea hydrothermal vents. Science 229: 717-725.

Jannasch, H.W., and C.O. Wirsen. 1985. The biochemical versatility of chemosynthetic bacteria at deep-sea hydrothermal vents. Biol. Soc. Wash. Bull. 6: 325-334.

Jannasch, H.W. 1995. Microbial interactions with hydrothermal fluids, p. 273-296. *In* S.E. Humphris, R.A. Zierenberg, L.S. Mullineaux, and R.E. Thomson (eds.), Seafloor hydrothermal systems: Physical, chemical, biological, and geological interactions. Geophysical monograph 91, American Geophysical Union, Washington DC.

Jeanthon, C. 2000. Molecular ecology of hydrothermal vent microbial communities. Antonie Van Leeuwenhoek 77: 117-133.

Johnson, H.P., and R.W. Embley. 1990. Axial seamount: an active ridge axis volcano on the central Juan de Fuca Ridge. J. Geophys. Res. 95: 12689-12696.

Karl, D.M. 1995. Ecology of free-living, hydrothermal vent microbial communities, p. 35-124. *In* D.M. Karl (ed.), The microbiology of deep-sea hydrothermal vents. CRC Press, Boca Raton, FL.

Kim, N.W., R. Lombardi, H. Bingham, E. Hani, H. Louie, D. Ng, and V.L. Chan. 1993. Fine mapping of three rRNA operons on the updated genomic map of *Campylobacter jejuni* TGH9011 (ATCC 43431) J. Bacteriol. 175: 7468-7470.

Kleinkauf, P.S. 2000. Bacterial community structure of hydrothermal vents at Guaymas Basin, Mexico as determined by amplified ribosomal DNA restriction analysis. Master's thesis, Western Washington University.

Kolenbrander, P.E. 2000. Oral microbial communities: biofilms, interactions, and genetic systems. Annu. Rev. Microbiol. 54: 413-437.

Kuske, C.R., L.O. Ticknor, M.E. Miller, J.M. Dunbar, J.A. Davis, S.M. Barns, and J. Belnap. 2002. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. Appl. Environ. Microbiol. 68: 1854-1863.

Lane, D.J. 1991. 16S/23S rRNA sequencing, p. 115-175. *In* E. Stackebrandt, and M. Goodfellow (eds.), Nucleic acid techniques in bacterial systematics. John Wiley and Sons, Ltd., London.

Li, L., C. Kato, and K. Horikoshi. 1999. Microbial diversity in sediments collected from the deepest cold-seep area, the Japan Trench. Mar. Biotechnol. 1: 391-400.

Liu, W., T.L. Marsh, H. Cheng and L.J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63: 4516-4522.

Logan, J.M.J., A.P. Burnens, D. Linton, A.J. Lawson, and J. Stanley. 2000. *Campylobacter lanienae* sp. nov., a new species isolated from workers in an abattoir. Int. J. Syst. Evol. Microbiol. 50: 865-872.

Longnecker, K., and A.L. Reysenbach. 2001. Expansion of the geographic distribution of a novel lineage of *epsilon-Proteobacteria* to a hydrothermal vent site on the Southern East Pacific Rise. FEMS Microbiol. Ecol. 35: 287-293.

Lonsdale, P. 1977. Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. Deep-Sea Res. 24: 857-863.

Lynch, K.S. 2000. Bacterial community structure and phylogenetic diversity of hydrothermal vents at Axial Volcano, Juan de Fuca Ridge. Master's Thesis, Western Washington University.

Maidak, B.L., J.R. Cole, T.G. Lilburn, C.T. Parker Jr., P.R. Saxman, R.J. Farris, G.M. Garrity, G.J. Olsen, T.M. Schmidt, and J.M. Tiedje. 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. 29: 173-4.

McCollum, T.M., and E.L. Shock. 1997. Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. Geochim. Cosmochim. Acta 61: 4375-4391.

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.

Moyer, C.L., F.C. Dobbs, and D.M. Karl. 1995. Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. 61: 1555-1562.

Moyer, C.L. 2001. Molecular phylogeny: applications and implications for marine microbiology. Methods Microbiol. 30: 375-394.

Muyzer, G., A. Teske, C.O. Wirsen, and H.W. Jannasch. 1995. Phylogenetic-relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel-electrophoresis of 16S rDNA fragments. Arch. Microbiol. 164: 165-172.

Nilsson, W. B., and M. S. Strom. 2002. Detection and identification of bacterial pathogens of fish in kidney tissue using terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes. Dis. Aquat. Organ. 48: 175-185.

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

Pedros-Alio, C. 1993. Diversity of bacterioplankton. Trends Ecol. Evol. 8: 86-90.

Polz, M.F., and C.M. Cavanaugh. 1995. Dominance of one bacterial phylotype at a mid-Atlantic Ridge hydrothermal vent site. Proc. Nat. Acad. Sci.USA 92: 7232-7236

Qiu, X., L. Wu, H. Huang, P.E. McDonel, A.B. Palumbo, J.M. Tiedje, and J. Zhou. 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol. 67: 880-887.

Rath, J., K.Y. Wu, G.J. Herndl, and E.F. DeLong. 1998. High phylogenetic diversity in a marine-snow-associated bacterial assemblage. Aquat. Microb. Ecol. 14: 261-269.

Reysenbach, A.L., K. Longnecker, and J. Kirshtein. 2000. Novel bacterial and archaeal lineages from an *in situ* growth chamber deployed at a mid-Atlantic Ridge hydrothermal vent. Appl. Environ. Microbiol. 66: 3798-3806.

Ruby, E.G., C.O. Wirsen, and H.W. Jannasch. 1981. Chemolithotrophic sulfur-oxidizing bacteria from the Galapagos Rift hydrothermal vents. Appl. Environ. Microbiol. 42: 317-324.

Sessitsch, A., A. Weilharter, M.H. Gerzabek, H. Kirchmann, and E. Kandeler. 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl. Environ. Microbiol. 67: 4215-4224.

Seyfried, W.E. Jr., and M.J. Mottl. 1995. Geologic setting and chemistry of deep-sea hydrothermal vents, p. 1-34. *In* D.M. Karl (ed.), The microbiology of deep-sea hydrothermal vents. CRC Press, Boca Raton, FL.

Sievert, S.M., J. Kuever, and G. Muyzer. 2000. Identification of 16S Ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos Island (Greece). Appl. Environ. Microbiol. 66: 3102-3109.

Taylor, C.D., C.O. Wirsen, and F. Gaill. 1999. Rapid microbial production of filamentous sulfur mats at hydrothermal vents. Appl. Environ. Microbiol. 65: 2253-2255.

Teske, A., K. Hinrichs, V. Edgcomb, A.V. Gomez, D. Kysela, S.P. Sylva, M.L. Sogin, and H.W. Jannasch. 2002. Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. Appl. Environ. Microbiol. 68: 1994-2007.

Tilman, D. 1996. Biodiversity: Population versus ecosystem stability. Ecology 77: 350-363.

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

Tunnicliffe, V., S.K. Juniper, and M.E. de Burgh. 1985. The hydrothermal vent community on Axial Seamount, Juan De Fuca Ridge. Biol. Soc. Wash. Bull. 6: 453-464.

Tuttle, J.H., C.O. Wirsen, and H.W. Jannasch. 1983. Microbial activities in the emitted hydrothermal waters of the Galápagos Rift vents. Mar. Biol. 73: 293-299.

Van Dover, C.L. 2000. The ecology of deep-sea hydrothermal vents. Princeton University Press, Princeton, NJ.

Vandamme, P., E. Falsen, R. Rossau, B. Hoste, P. Segers, R. Tytgat, and J. DeLey. 1991. Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: emendation of descriptions and proposal of *Arcobacter* gen. nov. Int. J. Syst. Bacteriol. 41: 88-103.

Ward, D.M., M.M. Bateson, R. Weller, and A.L. Ruff-Roberts. 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. Adv. Microbiol. Ecol. 12: 219-286.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. J. American Stat. Assoc. 58: 236-245.

Wirsen, C.O., T. Brinkhoff, J. Kuever, G. Muyzer, S. Molyneaux, and H.W. Jannasch. 1998. A new *Thiomicrospira* strain from the Mid-Atlantic ridge compared to known hydrothermal vent isolates. Appl. Environ. Microbiol. 64: 4057-4059.

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

APPENDIX I: QBASIC CODE FOR T-RFLP APPLICATION

```
'This program allows the user to search a library of PCR amplicons and
to size the 5' terminal fragments.

'Clear the screen and declare the variables

CLS
DIM FPrimer$
DIM RPrimer$
DIM LittleString$
DIM SpName$
DIM BigString$
DIM TRFrag AS INTEGER

'Request user to input primers and restriction enzymes

STARTOVER:
INPUT "What is the 4-base restriction enzyme"; LittleString$
LET FPrimer$ = "UNANACAUGCAAGUCGNNCG"
LET RPrimer$ = "AAGUCGUAACAAGGUANCN"

'Verify that primers and restriction enzymes were entered correctly.
PRINT "To verify, your 4 bp restriction enzyme is -> "; LittleString$;
"."
PRINT "To verify, your 5' primer is -> "; FPrimer$; "."
PRINT "To verify, your 3' primer is -> "; RPrimer$; "."
INPUT "Is this correct (1=yes; 2=no)"; verify
IF verify = 2 THEN GOTO STARTOVER

'Open output file

OPEN "C:\OUTFILE.TXT" FOR APPEND AS #3
      PRINT #3, LittleString$
CLOSE #3

'A loop that reads the library file one line at a time.
'A loop searches for primers and removes bordering sequences.
'Another loop within the loop searches for the restriction sequence
'within the larger string.

'Open library file

CLS
OPEN "C:\rdplib2.txt" FOR INPUT AS #1


'Setup a loop to go through each sequence in the library file
DO WHILE NOT EOF(1)
LET SpCount = SpCount + 1
'variable to keep track of the analysis
PRINT SpCount
TRYAGAIN:
MISS = MISS + 1
PRINT , , MISS
```

```
I = 0
'resetting the counters
K = 0
LINE INPUT #1, SpName$
IF LEN(SpName$) > 20 THEN
GOTO TRYAGAIN
END IF
LINE INPUT #1, BigString$
'Look for the reverse primer
REVSTARTOVER:
I = I + 1
RevAmbig = 0

'Exclude sequences that have no reverse primer
IF I = (LEN(BigString$) - 15) THEN
GOTO TRYAGAIN
END IF

IF MID$(RPrimer$, 1, 1) = "N" THEN
GOTO REVNUCLEO2
END IF
IF MID$(BigString$, I, 1) = MID$(RPrimer$, 1, 1) THEN
GOTO REVNUCLEO2
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO2
END IF
GOTO REVSTARTOVER


REVNUCLEO2:
IF MID$(RPrimer$, 2, 1) = "N" THEN
GOTO REVNUCLEO3
END IF
IF MID$(BigString$, (I + 1), 1) = MID$(RPrimer$, 2, 1) THEN
GOTO REVNUCLEO3
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO3
END IF
GOTO REVSTARTOVER


REVNUCLEO3:
IF MID$(RPrimer$, 3, 1) = "N" THEN
GOTO REVNUCLEO4
END IF
IF MID$(BigString$, (I + 2), 1) = MID$(RPrimer$, 3, 1) THEN
GOTO REVNUCLEO4
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO4
END IF
```

```
GOTO REVSTARTOVER

REVNUCLEO4:
IF MID$(RPrimer$, 4, 1) = "N" THEN
GOTO REVNUCLEO5
END IF
IF MID$(BigString$, (I + 3), 1) = MID$(RPrimer$, 4, 1) THEN
GOTO REVNUCLEO5
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO5
END IF
GOTO REVSTARTOVER

REVNUCLEO5:
IF MID$(RPrimer$, 5, 1) = "N" THEN
GOTO REVNUCLEO6
END IF
IF MID$(BigString$, (I + 4), 1) = MID$(RPrimer$, 5, 1) THEN
GOTO REVNUCLEO6
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO6
END IF
GOTO REVSTARTOVER


REVNUCLEO6:
IF MID$(RPrimer$, 6, 1) = "N" THEN
GOTO REVNUCLEO7
END IF
IF MID$(BigString$, (I + 5), 1) = MID$(RPrimer$, 6, 1) THEN
GOTO REVNUCLEO7
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO7
END IF
GOTO REVSTARTOVER


REVNUCLEO7:
IF MID$(RPrimer$, 7, 1) = "N" THEN
GOTO REVNUCLEO8
END IF
IF MID$(BigString$, (I + 6), 1) = MID$(RPrimer$, 7, 1) THEN
GOTO REVNUCLEO8
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO8
END IF
GOTO REVSTARTOVER
```

```
REVNUCLEO8:
IF MID$(RPrimer$, 8, 1) = "N" THEN
GOTO REVNUCLEO9
END IF
IF MID$(BigString$, (I + 7), 1) = MID$(RPrimer$, 8, 1) THEN
GOTO REVNUCLEO9
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO9
END IF
GOTO REVSTARTOVER


REVNUCLEO9:
IF MID$(RPrimer$, 9, 1) = "N" THEN
GOTO REVNUCLEO10
END IF
IF MID$(BigString$, (I + 8), 1) = MID$(RPrimer$, 9, 1) THEN
GOTO REVNUCLEO10
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO10
END IF
GOTO REVSTARTOVER


REVNUCLEO10:
IF MID$(RPrimer$, 10, 1) = "N" THEN
GOTO REVNUCLEO11
END IF
IF MID$(BigString$, (I + 9), 1) = MID$(RPrimer$, 10, 1) THEN
GOTO REVNUCLEO11
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO11
END IF
GOTO REVSTARTOVER


REVNUCLEO11:
IF MID$(RPrimer$, 11, 1) = "N" THEN
GOTO REVNUCLEO12
END IF
IF MID$(BigString$, (I + 10), 1) = MID$(RPrimer$, 11, 1) THEN
GOTO REVNUCLEO12
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO12
END IF
GOTO REVSTARTOVER
```

```
REVNUCLEO12:
IF MID$(RPrimer$, 12, 1) = "N" THEN
GOTO REVNUCLEO13
END IF
IF MID$(BigString$, (I + 11), 1) = MID$(RPrimer$, 12, 1) THEN
GOTO REVNUCLEO13
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO13
END IF
GOTO REVSTARTOVER


REVNUCLEO13:
IF MID$(RPrimer$, 13, 1) = "N" THEN
GOTO REVNUCLEO14
END IF
IF MID$(BigString$, (I + 12), 1) = MID$(RPrimer$, 13, 1) THEN
GOTO REVNUCLEO14
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO14
END IF
GOTO REVSTARTOVER

REVNUCLEO14:
IF MID$(RPrimer$, 14, 1) = "N" THEN
GOTO REVNUCLEO15
END IF
IF MID$(BigString$, (I + 13), 1) = MID$(RPrimer$, 14, 1) THEN
GOTO REVNUCLEO15
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO15
END IF
GOTO REVSTARTOVER

REVNUCLEO15:
IF MID$(RPrimer$, 15, 1) = "N" THEN
GOTO REVNUCLEO16
END IF
IF MID$(BigString$, (I + 14), 1) = MID$(RPrimer$, 15, 1) THEN
GOTO REVNUCLEO16
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO16
END IF
GOTO REVSTARTOVER

REVNUCLEO16:
IF MID$(RPrimer$, 16, 1) = "N" THEN
```

```
GOTO REVNUCLEO17
END IF
IF MID$(BigString$, (I + 15), 1) = MID$(RPrimer$, 16, 1) THEN
GOTO REVNUCLEO17
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO17
END IF
GOTO REVSTARTOVER

REVNUCLEO17:
IF MID$(RPrimer$, 17, 1) = "N" THEN
GOTO REVNUCLEO18
END IF
IF MID$(BigString$, (I + 16), 1) = MID$(RPrimer$, 17, 1) THEN
GOTO REVNUCLEO18
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO18
END IF
GOTO REVSTARTOVER

REVNUCLEO18:
IF MID$(RPrimer$, 18, 1) = "N" THEN
GOTO REVNUCLEO19
END IF
IF MID$(BigString$, (I + 17), 1) = MID$(RPrimer$, 18, 1) THEN
GOTO REVNUCLEO19
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO19
END IF
GOTO REVSTARTOVER

REVNUCLEO19:
IF MID$(RPrimer$, 19, 1) = "N" THEN
GOTO REVNUCLEO20
END IF
IF MID$(BigString$, (I + 18), 1) = MID$(RPrimer$, 19, 1) THEN
GOTO REVNUCLEO20
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO REVNUCLEO20
END IF
GOTO REVSTARTOVER

REVNUCLEO20:
IF MID$(RPrimer$, 20, 1) = "N" THEN
GOTO Rprim.Match
END IF
IF MID$(BigString$, (I + 19), 1) = MID$(RPrimer$, 20, 1) THEN
GOTO Rprim.Match
```

```
END IF
LET RevAmbig = RevAmbig + 1
IF RevAmbig < 5 THEN
GOTO Rprim.Match
END IF
GOTO REVSTARTOVER

'Make a  new big string that removes all characters to the right of
revprimer
Rprim.Match:
LET BigString$ = LEFT$(BigString$, (I + LEN(RPrimer$) - 1))

'Look for the forward primer
FORSTARTOVER:
ForAmbig = 0
K = K + 1
'Exclude sequences that have no forward primer
IF K = (LEN(BigString$) - 50) THEN
GOTO TRYAGAIN
END IF

IF MID$(FPrimer$, 1, 1) = "N" THEN
GOTO FORNUCLEO2
END IF
IF MID$(BigString$, K, 1) = MID$(FPrimer$, 1, 1) THEN
GOTO FORNUCLEO2
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO2
END IF
GOTO FORSTARTOVER

FORNUCLEO2:
IF MID$(FPrimer$, 2, 1) = "N" THEN
GOTO FORNUCLEO3
END IF
IF MID$(BigString$, (K + 1), 1) = MID$(FPrimer$, 2, 1) THEN
GOTO FORNUCLEO3
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO3
END IF
GOTO FORSTARTOVER

FORNUCLEO3:
IF MID$(FPrimer$, 3, 1) = "N" THEN
GOTO FORNUCLEO4
END IF
IF MID$(BigString$, (K + 2), 1) = MID$(FPrimer$, 3, 1) THEN
GOTO FORNUCLEO4
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO4
```

```
END IF
GOTO FORSTARTOVER

FORNUCLEO4:
IF MID$(FPrimer$, 4, 1) = "N" THEN
GOTO FORNUCLEO5
END IF
IF MID$(BigString$, (K + 3), 1) = MID$(FPrimer$, 4, 1) THEN
GOTO FORNUCLEO5
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO5
END IF
GOTO FORSTARTOVER

FORNUCLEO5:
IF MID$(FPrimer$, 5, 1) = "N" THEN
GOTO FORNUCLEO6
END IF
IF MID$(BigString$, (K + 4), 1) = MID$(FPrimer$, 5, 1) THEN
GOTO FORNUCLEO6
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO6
END IF
GOTO FORSTARTOVER

FORNUCLEO6:
IF MID$(FPrimer$, 6, 1) = "N" THEN
GOTO FORNUCLEO7
END IF
IF MID$(BigString$, (K + 5), 1) = MID$(FPrimer$, 6, 1) THEN
GOTO FORNUCLEO7
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO7
END IF
GOTO FORSTARTOVER

FORNUCLEO7:
IF MID$(FPrimer$, 7, 1) = "N" THEN
GOTO FORNUCLEO8
END IF
IF MID$(BigString$, (K + 6), 1) = MID$(FPrimer$, 7, 1) THEN
GOTO FORNUCLEO8
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO8
END IF
GOTO FORSTARTOVER

FORNUCLEO8:
```

```
IF MID$(FPrimer$, 8, 1) = "N" THEN
GOTO FORNUCLEO9
END IF
IF MID$(BigString$, (K + 7), 1) = MID$(FPrimer$, 8, 1) THEN
GOTO FORNUCLEO9
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO9
END IF
GOTO FORSTARTOVER

FORNUCLEO9:
IF MID$(FPrimer$, 9, 1) = "N" THEN
GOTO FORNUCLEO10
END IF
IF MID$(BigString$, (K + 8), 1) = MID$(FPrimer$, 9, 1) THEN
GOTO FORNUCLEO10
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO10
END IF
GOTO FORSTARTOVER

FORNUCLEO10:
IF MID$(FPrimer$, 10, 1) = "N" THEN
GOTO FORNUCLEO11
END IF
IF MID$(BigString$, (K + 9), 1) = MID$(FPrimer$, 10, 1) THEN
GOTO FORNUCLEO11
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO11
END IF
GOTO FORSTARTOVER

FORNUCLEO11:
IF MID$(FPrimer$, 11, 1) = "N" THEN
GOTO FORNUCLEO12
END IF
IF MID$(BigString$, (K + 10), 1) = MID$(FPrimer$, 11, 1) THEN
GOTO FORNUCLEO12
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO12
END IF
GOTO FORSTARTOVER

FORNUCLEO12:
IF MID$(FPrimer$, 12, 1) = "N" THEN
GOTO FORNUCLEO13
END IF
IF MID$(BigString$, (K + 11), 1) = MID$(FPrimer$, 12, 1) THEN
```

```
GOTO FORNUCLEO13
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO13
END IF
GOTO FORSTARTOVER

FORNUCLEO13:
IF MID$(FPrimer$, 13, 1) = "N" THEN
GOTO FORNUCLEO14
END IF
IF MID$(BigString$, (K + 12), 1) = MID$(FPrimer$, 13, 1) THEN
GOTO FORNUCLEO14
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO14
END IF
GOTO FORSTARTOVER

FORNUCLEO14:
IF MID$(FPrimer$, 14, 1) = "N" THEN
GOTO FORNUCLEO15
END IF
IF MID$(BigString$, (K + 13), 1) = MID$(FPrimer$, 14, 1) THEN
GOTO FORNUCLEO15
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO15
END IF
GOTO FORSTARTOVER

FORNUCLEO15:
IF MID$(FPrimer$, 15, 1) = "N" THEN
GOTO FORNUCLEO16
END IF
IF MID$(BigString$, (K + 14), 1) = MID$(FPrimer$, 15, 1) THEN
GOTO FORNUCLEO16
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO16
END IF
GOTO FORSTARTOVER

FORNUCLEO16:
IF MID$(FPrimer$, 16, 1) = "N" THEN
GOTO FORNUCLEO17
END IF
IF MID$(BigString$, (K + 15), 1) = MID$(FPrimer$, 16, 1) THEN
GOTO FORNUCLEO17
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
```

```
GOTO FORNUCLEO17
END IF
GOTO FORSTARTOVER

FORNUCLEO17:
IF MID$(FPrimer$, 17, 1) = "N" THEN
GOTO FORNUCLEO18
END IF
IF MID$(BigString$, (K + 16), 1) = MID$(FPrimer$, 17, 1) THEN
GOTO FORNUCLEO18
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO18
END IF
GOTO FORSTARTOVER

FORNUCLEO18:
IF MID$(FPrimer$, 18, 1) = "N" THEN
GOTO FORNUCLEO19
END IF
IF MID$(BigString$, (K + 17), 1) = MID$(FPrimer$, 18, 1) THEN
GOTO FORNUCLEO19
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO19
END IF
GOTO FORSTARTOVER

FORNUCLEO19:
IF MID$(FPrimer$, 19, 1) = "N" THEN
GOTO FORNUCLEO20
END IF
IF MID$(BigString$, (K + 18), 1) = MID$(FPrimer$, 19, 1) THEN
GOTO FORNUCLEO20
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO FORNUCLEO20
END IF
GOTO FORSTARTOVER

FORNUCLEO20:
IF MID$(FPrimer$, 20, 1) = "N" THEN
GOTO Fprim.Match
END IF
IF MID$(BigString$, (K + 19), 1) = MID$(FPrimer$, 20, 1) THEN
GOTO Fprim.Match
END IF
LET ForAmbig = ForAmbig + 1
IF ForAmbig < 5 THEN
GOTO Fprim.Match
END IF
GOTO FORSTARTOVER
```

```
'Make a  new big string that removes all characters to the left of
forPrimer
Fprim.Match:
LET BigString$ = RIGHT$(BigString$, (LEN(BigString$) - K + 1))

'Look for the Restriction fragment and calculate the length of the
terminal fragment
FOR TRFrag = 1 TO LEN(BigString$)
      IF MID$(BigString$, TRFrag, 4) = LittleString$ THEN GOTO CALCSTEP
NEXT TRFrag

CALCSTEP:
'Paste results into output file
OPEN "C:\OUTFILE.TXT" FOR APPEND AS #2
      PRINT #2, SpName$, TRFrag
CLOSE #2

LOOP

CLOSE #1
BEEP
END
```