ASSESING DIFFERENT DIVERSITY METRICS TO EXPRESS ARCHAEAL AND BACTERIAL ASSEMBLAGE STRUCTURE

Evangelia Smeti

Department of Marine Sciences, University of the Aegean, University Hill, 81100 Mytilene, Greece

ABSTRACT

Molecular methods developed over the last years have given a boost to the study of prokaryotic diversity. Both Archaeal and Bacterial assemblages were taken into consideration in the present study, in order to investigate diversity patterns within same habitat and among different environments. Data were provided from previous studies of fresh water, thermal springs and mud volcano environments. A number of 26 metrics including diversity, evenness, and dominance indices were assessed in order to select the most efficient to express microbial assemblage structure. Relative Abundance Distributions (RADs) were additionally employed to explore the structure and diversity of different assemblages. Overall, Bacterial diversity was higher than Archaeal diversity with respect to species richness, diversity, and evenness. This trend was best reflected by the Brillouin diversity, the Evenness E2, and the McNaughton dominance indices. This trend was visualized using RADs, which clearly showed that Bacterial assemblages presented more smooth and “speciose” species-abundance curves, compared to Archaeal assemblages that were characterized by steep, species-poor distributions. Moreover, marked differences were observed for Bacteria among contrasting environments. Although Archaea did not present significant differences among different environments, their diversity was higher in thermal springs, where Bacterial diversity was found to be significantly lower. It is hypothesized that those differences are due to the more efficient and flexible way Bacteria exploit available energy for growth, whereas the increased Archaeal diversity of thermal springs could be due to their capacity to survive under chronic stress.

Key words: thermal springs, fresh water, mud volcanoes, diversity indices, relative abundance distributions**.**

INTRODUCTION

Microbial communities probably constitute the majority of the earth's biodiversity and catalyze processes that are critical for sustaining life on earth (Van der Gucht et al., 2007). Molecular methods revealed a vast number of different organisms not yet known by classic cultivation methods (Bohannan & Hughes, 2003), whereas new studies have reported an unsuspected dominance of groups that were previously unknown or considered relatively rare (Kemp & Aller, 2004a). Until the development of molecular techniques, prokaryotic diversity remained basically unexplored, since classic systematics could not provide the necessary information to ecologists. This molecular approach offers crucial information towards the study of prokaryotic community structure and diversity, providing an estimate of both the richness of different phylotypes and their abundances in a sample.

Data from microbial communities have a peculiarity: the notion of species, as it is known, does not exist. Since most of the microbes are not cultured, their morphology and physiology is unknown, preventing taxonomical identification. Therefore the basic unit of microbial biodiversity studies is not the species but rather the 16S rRNA gene sequences, called operational taxonomic units (OTU) or phylotypes (Olsen et al., 1986). Moreover, the high cost of the method does not enable the acquisition of replicates, restricting the analysis to a small number of samples from each site. This sampling limitation leads to under-sampling (Curtis et al., 2006), which needs to be considered when assessing the overall microbial richness and diversity. Nowadays, the accumulation of data from contrasting environments enables comparative studies of microbial diversity. However, for this comparison to be valuable, the libraries should represent similar sampling efforts, thus capture a large enough

fraction of microbial richness (>70%). This is usually assessed using both the coverage of the libraries and species richness estimators (Kemp & Aller, 2004a; Kemp & Aller, 2004b).

So far prokaryotic communities have been investigated mainly with phylogenetic approaches (Lozupone & Knight, 2008) including α-diversity (Faith, 1992; Martin, 2002; Martin, 2002) and β-diversity measures (Lozupone & Knight, 2005; Lozupone et al., 2007; Martin, 2002; Singleton et al., 2001; Schloss et al., 2004; Pavoine et al., 2004). Other common approaches not taking gene sequences into consideration include the species richness estimators Chao 1 and ACE (Aller & Kemp, 2008; Kemp & Aller, 2004a). Sample diversity is usually estimated using the Shannon diversity (Hill et al., 2003) and the Margalef indices (Lehours 2007). Curtis et al. (2002) suggest that microbial communities could be represented by the log normal species-abundance distribution and that a possible estimation of their species richness could be obtained by the area under the curve. However, the lack of abundance data kept this suggestion on a theoretical basis. Sloan et al. (2006) created a simple neutral model using Hubbell's model as basis and modifying it in order to best fit microbial data. During the last decades, the wide range of methodological approaches developed in the field of terrestrial and marine ecology could be applied to explore patterns in microbial diversity (Magurran, 2004). More recently, Bunge (2009) discusses the advantages and disadvantages of parametric, nonparametric, and coverage-based estimation models and suggests that all different approaches should be used in order to conclude over the uncultivated microbial diversity.

Prokaryotic communities incorporate both Bacteria and Archaea which however maintain important ecological differences and constitute quite differentiated assemblages based on their genetic, biochemical, and evolutionary adaptations (Woese & Fox, 1977). Valentine

(2007) proposes that the primary factor differentiating Archaeal and Bacterial ecology is the adaptation of the former to withstand chronic energy stress due to their lipid-membrane composition and their mechanisms of energy conservation. On the other hand, Bacteria maintain a broad range of genetic, metabolic, and physiological capacities that allow a high degree of adaptability and metabolic diversification into numerous ecological niches and habitats. These capacities allow Bacteria to dominate Archaea in many environments, especially those in which energy stress is not chronic (Valentine, 2007). Research within specific areas over the past years, focused separately on Bacteria (Moffett et al., 2003) and Archaea (Bintrim et al., 1997), whereas few studies have incorporated both (Madrid et al., 2001; Kormas et al., 2003; Lehours et al., 2007). Such research provides the necessary data for comparative studies of prokaryotic diversity (Aller & Kemp, 2008; Hill et al., 2003). Although Bacteria show higher diversity than Archaea in most studied environments (Aller & Kemp, 2008), further insights into community diversity and structure of these two assemblages are necessary in order to improve our understanding of their ecology, occurrence, and interaction.

The aim of the present study was to explore potential differences in the diversity of Archaeal and Bacterial assemblages among contrasting environments such as lakes, thermal springs, and mud volcanoes. To this purpose, a large number of ecological indices representing diversity, evenness, and dominance was assessed in order to select those which best reflect differences between the two assemblages and among different environments. Additionally, further investigation of Archaeal and Bacterial structure is attempted using Relative Abundance Distributions (RADs) which are considered as the most detailed representation of an assemblage (Magurran, 2004).

MATERIALS AND METHODS

Data compilation

Data collected from different environments in Eastern Mediterranean (Table 1) were used in order to compare diversity of Bacterial and Archaeal assemblages. Data from thermal springs correspond to five Greek geothermal springs all over the country: Polihnitos, Edipsos, Thermopiles, Eleftheres, and Lagadas (Kormas et al., 2009). Freshwater environments are represented by data from Marathonas Lake (Lymperopoulou et al. in prep). More information on Marathonas Lake and the sampling methodology are provided in Lymperopoulou et al. (2010). Mud Volcanoes are represented by two datasets, the Kazan Mud Volcano (Kormas et al., 2008; Pachiadaki et al., 2010) and the Amsterdam Mud Volcano (Pachiadaki et al. in prep), in Anaximander Mountains, Eastern Mediterranean Sea. Reliability of the size of the libraries, and thus of the results, was already tested in the above studies, where Good's coverage estimator showed that the coverage was sufficiently high to suppose that at least the most prevalent Archaeal and Bacterial groups in each clone library have been recorded.

Data processing

Biodiversity indices were used to compare different assemblages. These non-parametric measures take into consideration species richness, abundance, and distribution of individuals to species. They are characterized as diversity, evenness, or dominance indices, according to their mathematical formula weighting more to the species richness or evenness components of diversity (Magurran, 2004). In this study 17 diversity, seven evenness, and two dominance

indices commonly applied in community ecology (Washington, 1984; Karydis & Tsirtsis, 1996) were considered (Table 2). Indices shown on table 2 were calculated with a specially developed Fortran code and the results were statistically treated using the Statgraphics package, version XV. The non-parametric Kruskal-Wallis test and the least significant difference (LSD) method were used in order to select the index which best describes the data. Species richness estimators S_{Chao1} and S_{ACE} were calculated using the EstimateS package v.8.2.

RESULTS

Indices that best revealed differences in community diversity among contrasting environments were the Brillouin diversity, the McNaughton dominance, and the E2 evenness indices, based on the Kruskal-Wallis (Table 3) and LSD tests. The Brillouin index showed similar and correlated estimates of diversity with the Shannon index, producing lower values in the same dataset than the latter and higher sensitivity in discriminating among different environments and assemblages. In general, Archaeal assemblages were characterized by low phylotype richness and abundance, high dominance, and medium evenness. The contradicting finding in freshwater Archaea, where high dominance co-occurs with high evenness, is due to the fact that in some samples Archaea were represented by only one species. On the other hand, Bacterial assemblages were characterized by higher phylotype richness and abundance, lower dominance, and medium evenness with the exception of Bacteria from thermal springs, which presented high dominance and low evenness (Fig.1).

Species richness estimators Chao1 and ACE (Table 4) suggested a high number of phylotypes in Bacteria assemblages and in Archaea from thermal springs, while lake and mud volcano Archaea were characterized by relatively low phylotype richness. These estimators additionally revealed a vast number of unobserved phylotypes, especially in the samples with higher richness. Chao1 estimator indicated higher phylotype richness than ACE, with the exception of Bacteria from thermal springs, probably because of an elevated number of rare phylotypes.

Archaeal assemblages from contrasting environments seem to have similar diversity, as expressed by most of the indices, using the LSD test (Table 5). Bacterial diversity on the

other hand, was habitat-dependent, presenting lower values in thermal springs, similar to those of the Archaeal assemblages. It was observed that in this environment Bacterial diversity decreases as Archaeal diversity increases.

RADs revealed differences in assemblage structure between Archaea and Bacteria (Fig. 2). Overall, Bacterial assemblages are characterized by smoother (linear regression slope ranging from -0.102 to -0.349) and more speciose distributions compared to Archaeal assemblages characterized by steeper (linear regression slope ranging from -2.106 to -5.469) and less speciose distributions. Thus Archaea present assemblages with extremely high dominance and very low evenness, while Bacteria constitute more even and rich assemblages.

Data from Thermopiles thermal spring were removed from the Archaea RAD plot, due to their higher species number, derived from the high number of singletons phylotypes, which resulted to a different distribution. For the same reason, data from the surface of Amsterdam sediment were removed from the Bacteria RAD. The higher species richness in this sample could be explained by the occurrence of Bacteria from the overlying water column due to proximity with the sediment-water interface.

DISCUSSION

The results of the present study are in agreement with previous works demonstrating the higher diversity of Bacteria compared to Archaea (Aller & Kemp, 2008). However, these studies suggested that thermal springs and methanogenic environments in general may support a greater Archaea richness, whereas in the present study both assemblages supported similar diversity. This trend was better expressed using the Brillouin diversity index, which showed slightly higher sensitivity than the most commonly applied Shannon index (Hill et al., 2003). The Shannon index, even though widely criticized over the years (Camargo, 1993; Magurran, 2004) is broadly used to depict diversity, mainly because, according to Magurran (2004), ecologists feel happier about adopting a measure with long tradition of use. However, the Brillouin index seems more sensitive to express the structural differences of microbial assemblages and according to Laxton (1978) it is mathematically superior to Shannon, while Pielou (1969) recommends its use when a collection is made.

Microbial diversity expresses predominantly the metabolic diversity (Reeve & Schmitz, 2005); therefore a variety of specific substrates and electron donors and acceptors is crucial to support microbial growth and richness. The lower diversity of Archaea, presenting fewer phylotypes, lower abundance, and higher dominance could be explained by the fact that the energetic costs of metabolic processes carried out by Archaea, in at least some environments, may be great and in expense to phylogenetic diversification, in comparison to Bacteria (Aller & Kemp, 2008). It is also suggested that chronic energy stress is the primary selective pressure governing the evolution of Archaea (Valentine, 2007). Archaea's adaptation to chronic energy stress is illustrated by their lipid-membrane composition. Archaeal membranes are less permeable to ions than Bacterial membranes, reducing the amount of

energy loss at the cellular level. Secondary adaptations, including catabolic pathways and mechanisms of energy conservation contribute to the above ability. This adaptation could explain the higher diversity of Archaea in thermal springs, which could be a supplementary indication to the hypothesis of a hyperthermophilic last common Archaeal ancestor (Gribaldo & Brochier-Armanet, 2006). If this is the case, then the low temperature environments could be considered as extreme to Archaea and thus explain partially the corresponding lack of diversity.

Direct competition between Archaea and Bacteria is not indicated in literature. Aller & Kemp (2008) suggest that Archaea may perceive and make use of the environment in ways that are more restrictive compared to Bacteria. For example, Archaea might live in microniches while Bacteria exploit a wider space or expand in different microniches within the same macroenvironment. Nearly any sample collected for analysis of prokaryotic diversity will contain a multitude of microenvironments and in a sense Archaea and Bacteria may not truly coexist even if they are collected in the same sample. However this hypothesis could not be easily tested and even perceived since prokaryotes live in a scale beyond the one of human experience.

Species-abundance distributions of Archaea and Bacteria show a very different structure of these two assemblages. Bacteria constitute more even assemblages, with more phylotypes than Archaea. Archaea's curve appears to be of geometric series-type, whereas Bacteria curves are more representative of a log normal RAD. The geometric series model usually describes species-poor environments, which are often harsh or perturbated (Magurran, 2004). It is also referred that it could be representative of very early stages of succession (Whittaker, 1972). On the other hand, the log normal distribution is theoretically considered to be the

most suitable to describe microbial data (Curtis et al., 2002). These observations are in agreement with Curtis (2006) using data from Godon (1997), who showed that geometric series best described Archaeal assemblages in an anaerobic digestor, while Bacteria formed a hypothesized log normal curve. However, the lack of an efficient number of samples in the present and in previous studies prevents the statistical test of goodness-of-fit of these models to the natural assemblages. Another interesting observation from the RADs of Archaea and Bacteria is that within an assemblage, curves from contrasting environments show very similar distribution. Thus, the distribution of species is not likely to be a simple function of the suitability of the environment (Curtis et al., 2006) and it is likely to depend more on self-organization processes (Spatharis et al., 2009). Finally, Kemp & Aller (2004) in their attempt to find out which library size is large enough to depict the total phylotype richness observed that under-sampled libraries are usually represented by the geometric series distribution. This could not be the case in the present study, since the sampling could not be selectional in favor of Bacteria, excluding Archaea.

An innovative look in microbial diversity is the simple neutral model (Sloan et al., 2006) created using Hubbell's model as basis and modifying it in order to best fit microbial data. Sloan et.al. suggest that neutral models can best describe microbial communities; therefore such an approach could be very promising for their understanding. However, no matter how vast the number of techniques is and how promising they look, the key to investigate more profoundly the prokaryotes would be the more exhaustive sampling. This will lead to the development of more reliable models and to a better understanding of the microbial world.

References

Aller, J. Y.& P. F. Kemp, 2008. Are Archaea inherently less diverse than Bacteria in the same environments?, Fems Microbiology Ecology 65: 74-87.

Berger, W. H.& F. L. Parker, 1970. Diversity of Planktonic Foraminifera in Deep-Sea Sediments, Science 168: 1345-&.

Bintrim, S. B., T. J. Donohue, J. Handelsman, G. P. Roberts & R. M. Goodman, 1997. Molecular phylogeny of archaea from soil, Proceedings of the National Academy of Sciences of the United States of America 94: 277-282.

Bohannan, B. J. M.& J. Hughes, 2003. New approaches to analyzing microbial biodiversity data, Current Opinion in Microbiology 6: 282-287.

Bunge J., 2009. Statistical Estimation os Uncultivated Microbial Diversity, Microbial Monogr.

Camargo, J. A., 2008. Revisiting the Relation Between Species Diversity and Information Theory, Acta Biotheoretica 56: 275-283.

Camargo, J. A., 1993. Must Dominance Increase with the Number of Subordinate Species in Competitive Interactions, Journal of Theoretical Biology 161: 537-543.

Chazdon, R. L., R. K. Colwell, J. S. Denslow & M. R. Guariguata, 1998. Forest biodiversity research,monitoring and modeling:conceptual background and old world case studies. In F.Dallmeier & J.A.Comiskey (ed), Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of northeastern Costa Rica. Parthenon Publishing, Paris: 285-309.

Curtis, T. P., I. M. Head, M. Lunn, S. Woodcock, P. D. Schloss & W. T. Sloan, 2006. What is the extent of prokaryotic diversity?, Philosophical Transactions of the Royal Society B-Biological Sciences 361: 2023-2037.

Curtis, T. P., W. T. Sloan & J. W. Scannell, 2002. Estimating prokaryotic diversity and its limits, Proceedings of the National Academy of Sciences of the United States of America 99: 10494-10499.

Faith, D. P., 1992. Conservation Evaluation and Phylogenetic Diversity, Biological Conservation 61: 1-10.

Gribaldo, S.& C. Brochier-Armanet, 2006. The origin and evolution of Archaea: a state of the art, Philosophical Transactions of the Royal Society B-Biological Sciences 361: 1007-1022.

Hill, T. C. J., K. A. Walsh, J. A. Harris & B. F. Moffett, 2003. Using ecological diversity measures with bacterial communities, Fems Microbiology Ecology 43: 1-11.

Hurlbert, S. H., 1971. Nonconcept of Species Diversity - Critique and Alternative Parameters, Ecology 52: 577-&.

Karydis, M.& G. Tsirtsis, 1996. Ecological indices: A biometric approach for assessing eutrophication levels in the marine environment, Science of the Total Environment 186: 209-219.

297 Kemp, P. F.& J. Y. Aller, 2004a. Bacterial diversity in aquatic and other environments: what 16S rDNA
298 libraries can tell us, Fems Microbiology Ecology 47: 161-177.

299 Kemp, P. F.& J. Y. Aller, 2004b. Estimating prokaryotic diversity: When are 16S rDNA libraries large
300 enough?, Limnology and Oceanography-Methods 2: 114-125.

301

302 Kormas, K. A., A. Meziti, A. Dahlmann, G. J. De Lange & V. Lykousis, 2008. Characterization of
303 methanogenic and prokaryotic assemblages based on mcrA and 16S rRNA gene diversity in
304 sediments of the Kazan mud volcano (Mediterranean Sea), Geobiology 6: 450-460.

305 Kormas, K. A., D. C. Smith, V. Edgcomb & A. Teske, 2003. Molecular analysis of deep subsurface
306 microbial communities in Nankai Trough sediments (ODP Leg 190, Site 1176), Fems Microbiology
307 Ecology 45: 115-125.

308 Kormas, K. A., H. Tamaki, S. Hanada & Y. Kamagata, 2009. Apparent richness and community
309 composition of Bacteria and Archaea in geothermal springs, Aquatic Microbial Ecology 57: 113-122.

310 Laxton, R. R., 1978. The measure of diversity, Journal of Theoretical Biology 51-67.

311 Lehours, A. C., P. Evans, C. Bardot, K. Joblin & F. Gerard, 2007. Phylogenetic diversity of archaea and
312 bacteria in the anoxic zone of a meromictic lake (Lake Pavin, France), Applied and Environmental
313 Microbiology 73: 2016-2019.

314 Lozupone, C.& R. Knight, 2005. UniFrac: a new phylogenetic method for comparing microbial
315 communities, Applied and Environmental Microbiology 71: 8228-8235.

316 Lozupone, C. A., M. Hamady, S. T. Kelley & R. Knight, 2007. Quantitative and qualitative beta
317 diversity measures lead to different insights into factors that structure microbial communities,
318 Applied and Environmental Microbiology 73: 1576-1585.

319 Lozupone, C. A.& R. Knight, 2008. Species divergence and the measurement of microbial diversity,
320 Fems Microbiology Reviews 32: 557-578.

321 Ludwig, A. J. R. J. F., 1988. Statistical Ecology: a Primer on Methods and Computing. New York.

322 Lymperopoulou D., K. A. Kormas, M. Moustaka-Gouni & A. D. Karagouni, 2010. Diversity of
323 cyanobacterial phylotypes in a Mediterranean drinking water reservoir (Marathonas,Greece),
324 Environmental Monitoring and Assessment.

325 Madrid, V. M., G. T. Taylor, M. I. Scranton & A. Y. Chistoserdov, 2001. Phylogenetic diversity of
326 bacterial and archaeal communities in the anoxic zone of the Cariaco Basin, Applied and
327 Environmental Microbiology 67: 1663-1674.

328 Magurran, 2004. Measuring BiologicalDiversity. Blackwell Publishing.

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

330 Martin, A. P., 2002. Phylogenetic approaches for describing and comparing the diversity of microbial
331 communities, Applied and Environmental Microbiology 68: 3673-3682.

McIntosh, R. P., 1967. An index of diversity and the relation of certain concepts to diversity., Ecology 392-404.

McNaughton, J., 1967. Relationship among functional properties of California grassland., Nature 216: 168-169.

Menhinick, E. P., 1964. A comparison of some species-individuals diversity indices applied to samples of field insects., Ecology 45: 859-861.

Moffett, B. F., F. A. Nicholson, N. C. Uwakwe, B. J. Chambers, J. A. Harris & T. C. J. Hill, 2003. Zinc contamination decreases the bacterial diversity of agricultural soil, Fems Microbiology Ecology 43: 13-19.

Odum, H. T. C. J. E. K. L. S., 1960. An organizational hierarchy postulate for the interpretation of species-individuals distribution, species entropy and ecosystem evolution and the meaning of a species variety index., Ecology 41: 395-399.

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

Pachiadaki, M. G., V. Lykousis, E. G. Stefanou & K. A. Kormas, 2010. Prokaryotic community structure and diversity in the sediments of an active submarine mud volcano (Kazan mud volcano, East Mediterranean Sea), Fems Microbiology Ecology 72: 429-444.

Pavoine, S., A. B. Dufour & D. Chessel, 2004. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis, Journal of Theoretical Biology 228: 523-537.

Pielou, E. C., 1969. An introduction to mathematical ecology. Wiley, New York.

Pielou, E. C., 1975. Ecological diversity. Wiley InterScience, New York.

Reeve, J. N.& R. A. Schmitz, 2005. Biology, biochemistry and the molecular machinery of Archaea, Current Opinion in Microbiology 8: 627-629.

Schloss, P. D., B. R. Larget & J. Handelsman, 2004. Integration of microbial ecology and statistics: a test to compare gene libraries, Applied and Environmental Microbiology 70: 5485-5492.

Shannon, C. E. W. W., 1949. The mathematical theory of communication. University of Illinois Press.

Singleton, D. R., M. A. Furlong, S. L. Rathbun & W. B. Whitman, 2001. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples, Applied and Environmental Microbiology 67: 4374-4376.

Sloan, W. T., M. Lunn, S. Woodcock, I. M. Head, S. Nee & T. P. Curtis, 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure, Environmental Microbiology 8: 732-740.

Spatharis, S., D. Mouillot, T. D. Chi, D. B. Danielidis & G. Tsirtsis, 2009. A niche-based modeling approach to phytoplankton community assembly rules, Oecologia 159: 171-180.

Valentine, D. L., 2007. Adaptations to energy stress dictate the ecology and evolution of the Archaea, Nature Reviews Microbiology 5: 316-323.

Van der Gucht, K., K. Cottenie, K. Muylaert, N. Vloemans, S. Cousin, S. Declerck, E. Jeppesen, J. M. Conde-Porcuna, K. Schwenk, G. Zwart, H. Degans, W. Vyverman & L. De Meester, 2007. The power of species sorting: Local factors drive bacterial community composition over a wide range of spatial scales, Proceedings of the National Academy of Sciences of the United States of America 104: 20404-20409.

Washington, H. G., 1984. Diversity, Biotic and Similarity Indexes - A Review with Special Relevance to Aquatic Ecosystems, Water Research 18: 653-694.

Whittaker, R. H., 1972. Evolution and measurement of species diversity., Taxon 21: 213-251.

Woese, C. R.& G. E. Fox, 1977. Phylogenetic Structure of Prokaryotic Domain - Primary Kingdoms, Proceedings of the National Academy of Sciences of the United States of America 74: 5088-5090.

384 Table 1. Data used in the present study.
385

Environment	Site	No of samples	Year	Sampling	References
Fresh water	Marathonas lake	4	2007	4 sampling points	Lymperopoulou et al. in prep
		4	2008		
Mud volcanoes	Kazan	7	2003	0-30cm every	Kormas et al. (2008); Pachiadaki et al. (2010)
	Amsterdam	7	2003	5cm depth	Pachiadaki et al. in prep
Thermal springs	Polihnitos	1	2005	1 sample	Kormas et al. (2009)
	Eleftheres	1	2005	every spring	
	Edipsos	1	2005		
	Thermopiles	1	2005		
	Lagadas	1	2005		

386

387

388 Table 2. The 26 indices applied in the current study including 17 diversity, 7 evenness and 2
389 dominance indices.
390

Index type	Index	Formula	Reference
Diversity indices	Margalef	$D = \frac{S-1}{\ln N}$	(Margalef, 1958)
	Gleason	$D = \frac{S}{\ln N}$	(Ludwig, 1988)
	Menhinick	$D = \frac{S}{\sqrt{N}}$	(Menhinick, 1964)
	Simpson's	$D = \frac{\sum_{i=1}^{S} n_i \times (n_i - 1)}{n \times (n-1)}$	Ludwig and Reynolds (1988)
	Shannon H'	$H' = -\sum_{i=1}^{S} \frac{n_i}{n} \times \ln \frac{n_i}{n}$	(Shannon, 1949)
	Shannon D'	$D' = S - (S \times d')$	(Camargo, 2008)
	Brillouin	$HB = \frac{\ln N! - \sum \ln ni!}{N}$	Pielou (1969,1975)
	Hill N_0	$N_0 = S$	Ludwig and Reynolds (1988)
	Hill N_1	$N_1 = \exp(H')$	Ludwig and Reynolds (1988)
	Hill N_2	$N_2 = 1/\text{Simpson's D}$	Ludwig and Reynolds (1988)
	Odum	$O = \frac{S \times 1000}{N}$	(Odum, 1960)
	Kothe	$D = \frac{S_{max} - S_i}{S_{max}}$	(Pielou, 1975)
	Keefe	$TU = 1 - \frac{n}{n-1} \times \left(\sum_{i=1}^{s} pi^2 - \frac{1}{n} \right)$	Keefe and Bergensen (1977)
	Hurlbert	$PIE = \left(\frac{N}{N-1} \right) \times \left(1 - \sum_{i=1}^{S} p_i{}^2 \right)$	(Hurlbert, 1971)
	McIntosh	$M = \frac{n - \sqrt{\sum_{i=1}^{S} n_i{}^2}}{n - \sqrt{n}}$	(McIntosh, 1967)

	Chao 1	$S = Sobs + \frac{F1^2}{2F2}$	Chazdon et al. (1998)
	ACE	$S = Sabund + \frac{Srare}{CACE} + \frac{F1}{CACE}\gamma ACE^2$	Chazdon et al. (1998)
Evenness indices	Evenness E_1	$E_1 = \frac{H'}{\ln S}$	Pielou, 1975
	Evenness E_2	$E_2 = \frac{\exp(H')}{S}$	Sheldon, 1969
	Evenness E_3	$E_3 = \frac{\exp(H') - 1}{S - 1}$	Ludwig and Reynolds (1988)
	Evenness E_4	$E_4 = \frac{1/\text{Simpson's D}}{\exp(H')}$	Ludwig and Reynolds (1988)
	Evenness E_5	$E_5 = \frac{(1/\text{Simpson's D}) - 1}{\exp(H') - 1}$	Ludwig and Reynolds (1988)
	Evenness E_6	$E_6 = 1 - d'$	Camargo (2008)
	Redundancy	$R = \frac{H'_{max} - H'}{H'_{max} - H'_{min}}$	Patten (1962)
Dominance indices	Berger-Parker	$B = n_1/n$	(Berger & Parker, 1970)
	McNaughton	$\alpha = (n_1 + n_2)/n$	(McNaughton, 1967)

391

392 The terms used in the formulas are given below:

393 S = the number of species in a sample or a population

394 N = the number of individuals in a population or community

395 N_i = the number of individuals in species i of a population or community

396 n = the number of individuals in a sample from a population

397 n_i = the number of individuals in a species i of a sample from a population

398 $p_i = n_i/n$ = the fraction of a sample of individuals belonging to species i

399 S_{max} = the maximum number of species in a sample

400 n_1, n_2 = the number of individuals in the two most abundant species

401 Fi=the number of species with i individuals

402 Srare=the number of rare species (≤10 individuals)

403 Sabund=the number of abundant species(>10 individuals)

404 Nrare=the total number of individuals in rare species

405 $H'_{max} = \ln S$ (Pielou, 1975), $H'_{min} = \frac{1}{N} \ln \frac{N!}{(N-S+1)!}$ (Pielou, 1975), and $d' = \frac{\sum_{i \neq j}^{K} |pi - pj|}{S}$

406 (Camargo, 2008)

407 $C_{ACE} = 1 - \frac{F1}{Nrare}$, $\gamma_{ACE}{}^2 = max\left\{\frac{Srare}{CACE} \frac{\sum_{i=1}^{10} i(1-i)Fi}{(Nrare)(Nrare-1)} - 1,0\right\}$ (Chazdon et al., 1998)

408

409 Table 3. Results of the Kruskal-Wallis non-parametric statistical test. The test statistic
410 decides the more suitable index to depict prokaryotic diversity. n is the number of samples
411 used in the calculation of each index.

412

		Kruskal-Wallis	
Index	**n**	**Test statistic**	**p-value**
N	53	41.42**	0.00
Hill N0	53	32.14**	0.00
Hill N1	53	31.39**	0.00
Hill N2	53	28.22**	0.00
Margalef	53	29.45**	0.00
Gleason	53	29.46**	0.00
Menhinick	53	28.50**	0.00
Simpson's	53	28.22**	0.00
Shannon H'	53	31.39**	0.00
Shannon D'	53	28.31**	0.00
Brillouin	53	33.56**	0.00
Hurlbert	53	28.22**	0.00
McIntosh	53	25.65**	0.00
Keefe	53	28.22**	0.00
Kothe	53	8.01	0.16
Odum	53	22.86**	0.00
Evenness E1	51	17.33**	0.00
Evenness E2	53	18.86**	0.00
Evenness E3	51	15.99**	0.01
Evenness E4	53	17.74**	0.00
Evenness E5	51	16.94**	0.00
Evenness E6	53	16.84**	0.00
Redundancy	51	12.81*	0.03
Berger-Parker	53	24.98**	0.00
McNaughton	53	30.07**	0.00

413

414 * Statistically significant difference at the 0.05 level

415 ** Statistically significant difference at the 0.01 level

416

417

418

Table 4: Observed (S_{obs}) and calculated species richness with the Chao 1 (S_{Chao1}) and ACE (S_{ACE}) species richness estimators. Abbreviations: afw: Archaea freshwater , amv: Archaea mud volcanoes, ats: Archaea thermal springs, bfw: Bacteria freshwater, bmv: Bacteria mud volcanoes, bts: Bacteria thermal springs

Index	afw	amv	ats	bfw	bmv	bts
S_{obs}	31	48	85	184	330	121
S_{Chao1}	37	77	387	370	818	405
S_{ACE}	38	76	417	358	767	319

Table 5: Groups formed for the selected indices according to the results of LSD test. Abbreviations A_{fw}: Archaea freshwater, A_{mv}: Archaea mud volcanoes, A_{ts}: Archaea thermal springs, B_{fw}: Bacteria freshwater, B_{mv}: Bacteria mud volcanoes, B_{ts}: Bacteria thermal springs.

Hill N0	A_{fw}	A_{mv}	A_{ts} A_{ts}	B_{ts} B_{ts}	B_{mv}	B_{fw}
N	A_{fw}	A_{mv}	A_{ts}	B_{fw}	B_{mv}	B_{ts}
Brillouin	A_{fw}	A_{mv}	A_{ts} A_{ts}	B_{ts}	B_{mv}	B_{fw}
Evenness E2	B_{ts}	A_{ts}	A_{mv}	B_{mv}	B_{fw} B_{fw}	A_{fw}
McNaughton	B_{fw}	B_{mv}	A_{fw}	A_{mv}	A_{ts}	B_{ts}

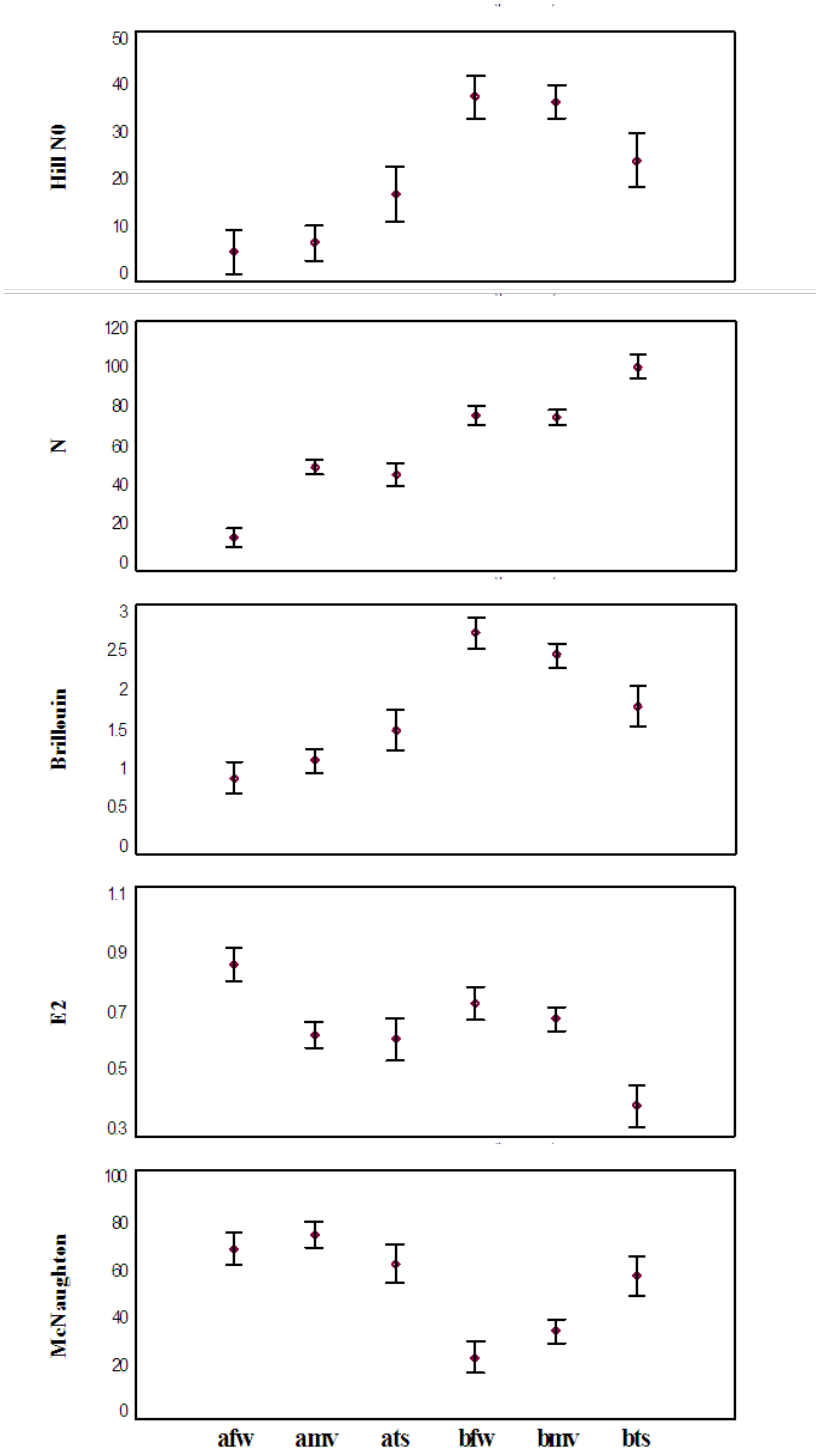
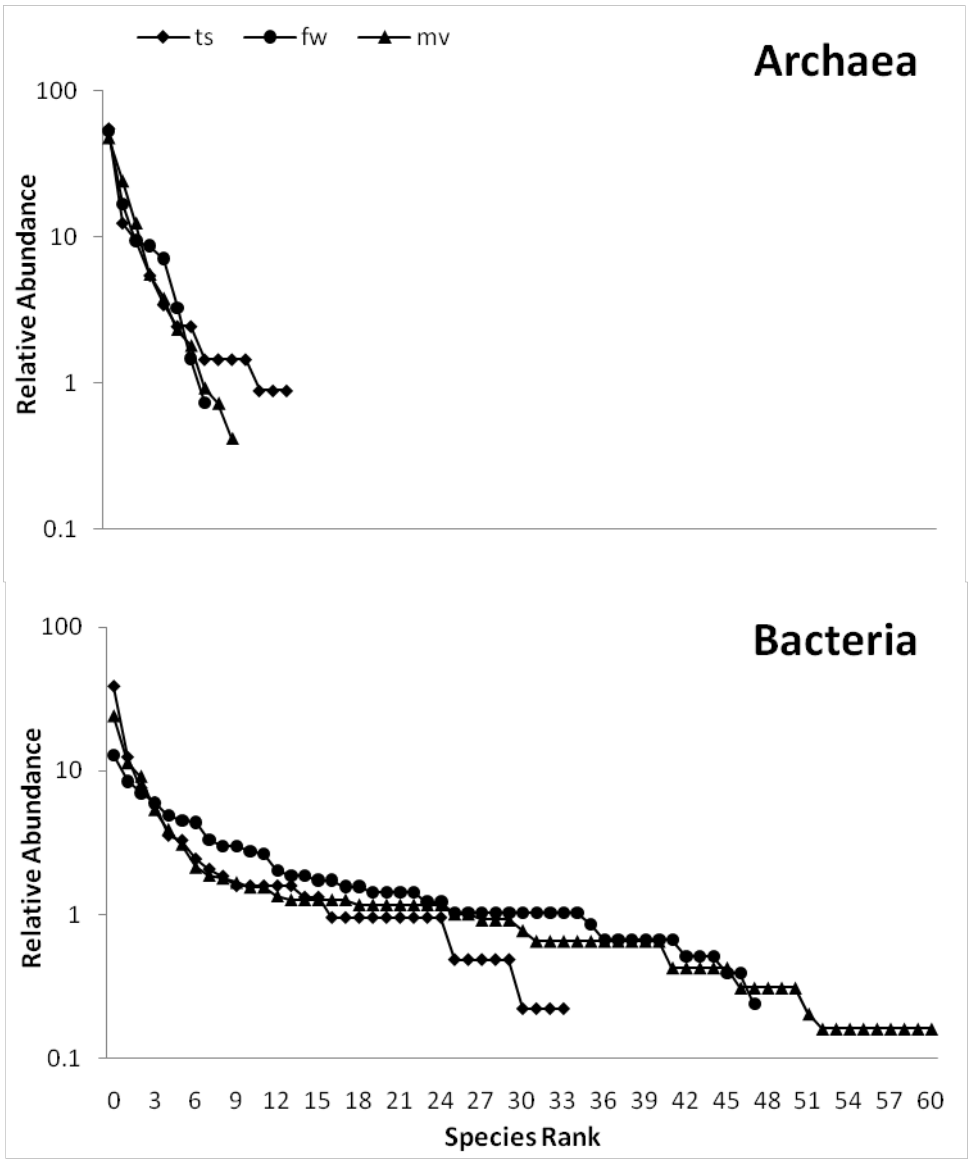


Fig.1. Mean plots of indices (species richness Hill N0, abundance N, Brillouin diversity, evenness E2, and McNaughton dominance), that best reveal differences in microbial diversity among contrasting environments: afw- Archaea freshwater , amv- Archaea mud volcanoes, ats-Archaea thermal springs, bfw- Bacteria freshwater, bmv- Bacteria mud volcanoes, bts-Bacteria thermal springs.



440

441 Fig.2 Rank abundance plots of Archaea and Bacteria assemblages in contrasting

442 environments. Abbreviations: ts- thermal springs, fw- fresh water, mv- mud volcanoes.