UNIVERSITÀ
DEGLI STUDI
FIRENZE

Molecular methods for the analysis of microbial communities:

application to natural environments affected by anthropogenic modifications.

Francesco Vitali

UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN

Scienze e tecnologie Vegetali, Microbiologiche e Genetiche

Indirizzo: Scienze genetiche, microbiologiche e bioinformatica

CICLO XXVII

COORDINATORE Prof. Scala Aniello

Molecular methods for the analysis of microbial communities: application to natural environments affected by anthropogenic modifications.

Settore Scientifico Disciplinare BIO/19

Dottorando

Dott. Vitali Francesco

_____

Tutors

Prof. Mastromei Giorgio

_____

Prof. Casalone Enrico

_____

Coordinatore

Prof. Scala Aniello

_____

Anni 2012/2014

Contents

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

December 31th, 2014 Francesco Vitali

A copy of the thesis will be available at: www.dispaa.unifi.it

Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'Università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

31 dicembre 2014 Francesco Vitali

Una copia della tesi sarà disponibile presso: www.dispaa.unifi.it

1 Summary

Microbes are ubiquitous in the environment and are key component of Earth Biota, covering fundamental roles for the functioning of ecosystems. A comprehensive characterization of such astonishing diversity is of fundamental importance to better understand the functioning of ecosystems, and constitutes a starting point for new technological applications and for issues of natural conservation. Our understanding of the microbial world, however, has been heavily limited by the impossibility to apply classical microbiological techniques to natural populations, mostly due to our incapacity of growing them in laboratory media. In this regard, a great advance came from molecular biology and to its application to the study of molecular markers (primarily SSU rRNAs). Moreover, with the drop in costs of New Generation Sequencing (NGS) technologies, and with the consequent increase of their application to microbial ecology studies, a substantial boost in our understanding of natural microbial communities has been achieved. Molecular methods are actually the most comprehensive alternative for the study of natural microbial communities; and my PhD work has been primarily focused on the application of such methodologies to the study of microbial communities in natural environments affected by anthropogenic modifications. In this thesis, principles of the molecular analysis of microbial communities were covered, presenting the results of two case studies. Two molecular methodologies were applied: terminal restriction fragments length polymorphism (T-RFLP) and next generation sequencing (NGS) with MiSeq platform.

In one case (chapter 4, 6, and 7), activities connected to the European project "DEMETRA" were presented. General aim of the project was to provide innovative instruments for the monitoring of the risk for biodiversity connected to the (hypothetical) introduction of genetically modified (GM) crops in the environment of the regional park of Migliarino San Rossore Massacciuccoli. For this reason, a primary step was the characterization of biodiversity at various levels, including those of soil microbial communities. Subject of the study were bacterial and fungal populations in soils from. Waiting for the whole botanical/forestry, entomological and microbiological data set will be processed to determine a risk index of GM crops, we have utilized the microbiological data to evaluate the long-lasting effects of a change within a single land-use category, by comparing microbial diversity in soils from two natural forests and a forest converted to poplar plantation about thirty years ago within the Park. Results from T-RFLP and NGS approach showed high concordance. Both identified the change in land-use as a major perturbative factor for bacterial

communities (and also for fungal ones, considering T-RFLP results), individuating a decrease in richness in microbial communities from soils of the converted poplar plantation. Land-use was a stronger factor respect to vegetation cover in shaping microbial community. In fact, a higher similarity in community composition was found between the two natural forests (distant from each other in space, and characterized by a different vegetation cover), than between the converted and the natural poplar plantation (really close in space and characterized by the same vegetation cover)

In the second case (chapter 5, 7, and 9), activities connected to the European project "MAPMED" were presented. General aim of the project was to improve the environmental sustainability of tourist coastal areas in the Countries of the Mediterranean Sea Basin and to optimize, validate, and transfer tools for the sustainable management of tourist ports with regard to monitoring and reduction of marine pollution. For this reason, the diversity and structure of bacterial and archaeal communities in sediments from three ports in the Mediterranean Sea (Cagliari, Italy; El Kantaoui, Tunisia; Heraklion Greece) were characterized. Results indicate that microbial communities in sediments from the three ports were different and that port of origin constituted the main grouping factor. The overall diversity levels however, regardless of the geographic position of port, where low, indicating that the port basin represent a peculiar habitat unconnected with well-established north-south and east-west gradients of diversity and environmental conditions at the level of Mediterranean basin. A peculiar community composition was observed in sediment from a shipyard in the port of Heraklion, highly different from the rest of samples. Activities linked to shipyards are characterized by pollutants and impacts strongly different from the rest of the touristic port environment, ultimately affecting microbial communities in sediments. The use of T-RFLP on three different molecular markers (16S rRNA on bacterial communities, 16S rRNA on archaeal communities, and dsrAB) as methodology of choice for monitoring plans was evaluated and compared to results from NGS analysis to provide a mean of validation. T-RFLP was able to recapture main results from NGS fairly well, and the marker that performed better was dsrAB. For those reasons, T-RFLP could be effectively be used as the routine technique in the monitoring of marine pollution in sediments of touristic ports from the Mediterranean Sea Basin, although its effectiveness should be exceptionally corroborated by NGS analysis.

2 Riassunto

I microrganismi sono ubiquitari nell'ambiente e indubbiamente costituiscono una componente fondamentale del bioma terrestre, ricoprendo ruoli di centrale importanza per il funzionamento degli ecosistemi. L'approfondita caratterizzazione della sorprendente diversità del mondo microbico negli ambienti naturali, benché di fondamentale importanza non solo per comprendere al meglio il funzionamento degli ecosistemi, ma anche per le applicazioni biotecnologiche che ne potrebbero derivare, è stata fortemente limitata dall'impossibilità di applicare le classiche tecniche microbiologiche, principalmente a causa della nostra incapacità di allestire colture di laboratorio. A tal proposito, un forte passo in avanti è stato effettuato grazie alle tecniche di biologia molecolare, e alla loro applicazione allo studio di popolazioni microbiche naturali. L'analisi di marker molecolari (primi fra tutti gli RNA ribosomali) ha infatti permesso di svincolare l'analisi microbiologica dalla necessità di coltivazione, permettendo così di effettuare approfondite caratterizzazioni delle comunità naturali. La profondità delle analisi è ulteriormente aumentata grazie alle nuove tecniche di sequenziamento (NGS), e al conseguente aumento della loro applicazione in studi di ecologia microbica in seguito al crollo dei loro costi. Le tecniche di microbiologia molecolare rappresentano quindi il più completo mezzo per lo studio di comunità microbiche ambientali, e il mio dottorato si è principalmente focalizzato sulla loro applicazione allo studio di comunità microbiche in ambienti naturali sottoposti a modificazioni antropiche. In questa tesi, due tecniche in microbiologia molecolare verranno descritte, nello specifico terminal restriction fragments length polymorphism (T-RFLP) e next generation sequencing (NGS) con piattaforma MiSeq, e la loro applicazione a due casi studio sarà presentata.

Nei capitoli 4, 6 e 7, verranno presentate le attività connesse al progetto europeo "DEMETRA". Il progetto si è svolto nel parco regionale di Migliarino-San Rossore-Massacciuccoli con lo scopo generale di elaborare strumenti innovativi per il monitoraggio del rischio legato all'ipotetica immissione nell'ambiente di coltivazioni OGM. Al fine di elaborare un indice di rischio, un passaggio di primaria importanza, è quello della caratterizzazione della biodiversità del sito di studio a vari livelli incluso, in questo lavor, a livello delle popolazioni microbiche. Aspettando che la totalità dei dati (forestali, entomologici, microbiologici) venga processata al fine di elaborare un indice di rischio finale, abbiamo utilizzato i dati microbiologici, ottenuti tramite analisi T-RFLP e NGS, per la valutazione degli effetti a lungo termine che un cambiamento nell'utilizzo del suolo ha sulle comunità microbiche, comparando suoli prelevati da due

foreste naturali con suoli prelevati da una foresta naturale convertita a pioppeta coltivata circa trenta anni fa. I dati ottenuti dai due approcci (T-RFLP e NGS) mostrano un elevato grado di concordanza, identificando il cambiamento in utilizzo del suolo come fattore principale nell'influenzare la struttura delle comunità batteriche (e anche fungine, limitatamente all'analisi T-RFLP) e evidenziando una diminuzione nella ricchezza in specie nei suoli provenienti dalla foresta naturale convertita a pioppeta. Le comunità microbiche relative al suolo delle foreste naturali mostravano un maggior grado di similarità tra loro rispetto a quelle presenti nei suoli della pioppeta coltivata. La differenza era evidente anche nel caso della pioppeta naturale, geograficamente più vicina a quella coltivata. Questi risultati indicano che il cambiamento in utilizzo del suolo è un fattore più importante rispetto alla distanza e al tipo di vegetazione nell'influenzare le comunità microbiche.

Nei capitoli 5, 7 e 9, verranno presentate le attività connesse al progetto europeo "MAPMED". Obbiettivo generale del progetto era il miglioramento della sostenibilità delle aree costiere nei paesi del Bacino del Mar Mediterraneo, con l'ottimizzazione, validazione e trasferimento di strumenti per la gestione sostenibile dei porti turistici, con particolare riguardo al monitoraggio e riduzione dell'inquinamento marino. A tal scopo, è stata caratterizzata la diversità delle comunità batteriche e archaea nei sedimenti di tre porti turistici nel Mediterraneo (Cagliari, Italia; El Kantaoui, Tunisia; Heraklion, Grecia). La composizione delle comunità microbiche nei tre porti è risultata diversa, e il porto di origine è risultato il principale fattore nel determinare la diversità microbica. In generale, tuttavia, i livelli di diversità sono risultati bassi, indicando che i bacini portuali rappresentano un ambiente particolare, che non rispecchia i ben noti gradienti di diversità e di condizioni ambientali rilevati lungo gli assi nord-sud e est-ovest del Mediterraneo. Una composizione molto particolare, e fortemente diversa dal resto dei campioni, è stata evidenziata per quanto riguarda le comunità microbiche nei sedimenti prelevati da un cantiere navale nel porto di Heraklion. Le attività collegate ad un cantiere navale infatti, provocano un inquinamento e degli impatti molto differenti rispetto al restante ambiente in un porto turistico, influenzando in maniera peculiare le comunità microbiche nei sedimenti. La tecnica T-RFLP su tre diversi marker molecolari (16S RNA su comunità batteriche, 16S RNA su comunità archaea e dsrAB), utilizzata come tecnica di elezione nei piani di monitoraggio, è stata affiancata dall'analisi NGS. I risultati mostrano che la tecnica T-RFLP, in particolare se effettuata sul marker dsrAB, riesce a restituire bene le principali osservazioni ottenute dalla tecnica NGS. Questa validazione della T-RFLP, la renderebbe effettivamente utilizzabile come tecnica di routine nei piani di monitoraggio dell'inquinamento in sedimenti portuali, anche grazie alla semplicità di esecuzione e ai costi ridotti.. La sua efficacia in particolari situazioni potrebbe essere validata dall'analisi NGS.

Chapter 3

Introduction

3 Introduction

3.1 The microbial world

"We live in a microbial world, with microscopic organisms filling discrete ecosystems in such environments as soil, lakes and oceans, the human gut or skin, and even computer keyboards" with this sentence Foster et al. (2012) sums up the world we are living in (Bunge et al. 2014). The prokaryotic world is immense and ubiquitous. The total number of prokaryotes on Earth is estimated to be 4-6 x 10^{30}, and accounts for about 60% of total biomass (Whitman et al. 1998) and for as much as 60% of respiration in some environments (Madsen 2008). In addition, prokaryotes contain large amounts of N, P, and other essential nutrients. Most of the earth prokaryotes occur in the open ocean, in soil, and in oceanic and terrestrial subsurface, where the numbers of cells are 1.2 x 10^{29}, 2.6 x 10^{29}, 3.5 x 10^{30}, and 0.25–2.5 x 10^{30}, respectively (Whitman et al. 1998). At present however, we managed to collect detailed information only on tiny fractions of this immense world (see chapter 3.2).

We are also living in a microbial world from an historical point of view. The evolutive history of prokaryotes, being them among the eldest living organisms on Earth, is much longer than that of eukaryotes. The Earth is approximately 4.6 billion years old; at its origin, Earth could not support life, as no continents, no ozone, no atmosphere, and no oceans existed. Subsequently the prohibitive conditions of this prebiotic world changed; gases release from eruption formed the first atmosphere, the Earth cooled, and the aqueous vapor condensed forming the oceans. Leaving aside how the process took place, that is outside the scope of this introduction, life could then develop and the first life form, at least in the way we intend it now, was represented by prokaryotes. Probable prokaryotic fossil has been dated 3.5 billion years old; given that the first evidence of a eukaryotic (unicellular) fossil has been dated about 2 billion years ago, it is evident that more than 50% of the life history on our planet has been dominated by prokaryotes.

The large population size of prokaryotes implies that events that are extremely rare in the laboratory could occur frequently in nature. Prokaryotes have an enormous potential to accumulate mutations (Whitman et al. 1998), extremely high metabolic plasticity, and growth rates. These factors interacting for billions of years, allow them to easily acquire genetic diversity, and the result is the astonishing

biodiversity feature of the microbial world. Microbes are in fact capable to prosper in virtually any environments; all habitats suitable for plants and animals also harbor microbial populations. In addition, some microbes are adapted to grow under physical and chemical conditions that are too extreme for plant and animal growth (Konopa 2006), while other are adapted to a complete or partial symbiotic lifestyle (Moya et al. 2008; Moran & Wernegreen 2000), sometimes even inside Eukaryotic cells (Nowack & Melkonian 2010).

Not only the microbial world is ubiquitous; it also represents a fundamental component of the Biosphere. Microbes play key role in biogeochemical cycles, in the recycling of nutrients, in the respiration; they improve the productivity of plants and can degrade pollutants, and so on (Madsen 2008). Hence, gaining more in depth knowledge of the microbial world is extremely important not only from a microbiological or ecological point of view. Besides increasing our knowledge on animal and plant disease prevention and treatment, it would in fact allow us to better understand the processes that are at the base of the world we are living in and to exploit them for technological application, such as promoting plant grow, or removing pollutants from the environment. Moreover exploring microbial diversity is due to lead to the discovery of new enzymes, products, or genes, resulting in a whole set of biotechnological applications.

3.2 Challenges of microbial ecology

3.2.1 Cultivability of *Bacteria* from environmental sources

More than a century has passed since the introduction of growth media solidified by agar and the concept of pure culture by Koch in 1877 For more than a century cultivation, isolation, and enrichment have been our method to look at the microbial world, studying its functioning and understanding its role in the environment and its relation with human and animal health. However, the biggest limitation of the cultivation approach became soon evident. Only a fraction of the microbial cells that were visible with microscopy, especially with epifluorescent microscopy on environmental samples stained with DAPI (a fluorescent dye which selectively bind double-stranded DNA), could form colonies on agar media. It is estimated that only 1% of the microbial richness in the biosphere is actually cultivable in laboratory media (Torsvik et al. 1990; Rappé & Giovannoni 2003;

Madsen 2008). This remarkable gap is also known as “The Great Plate count anomaly” (Staley & Konopka 1985; Epstein 2013).

The magnitude of this disparity became even more evident with the introduction of molecular techniques in microbial ecology. As will be shown in chapter 3.3, those techniques completely avoid the cultivation step by amplifying and/or analyzing DNA directly extracted from the environment (frequently targeting gene encoding for SSU rRNAs). It became evident that the numbers of species known until then, defined thanks to the cultivation approach, represented a tiny fraction of the actual prokaryotic diversity in the Biosphere. Rappé & Giovannoni 2003) have analyzed the number of 16S ribosomal RNA gene sequences from Bacteria and Archaea, published in GenBank from 1993 to 2002 (Figure 3.1), searching first for sequences coming from cultured microorganisms and then searching those coming from uncultured ones.

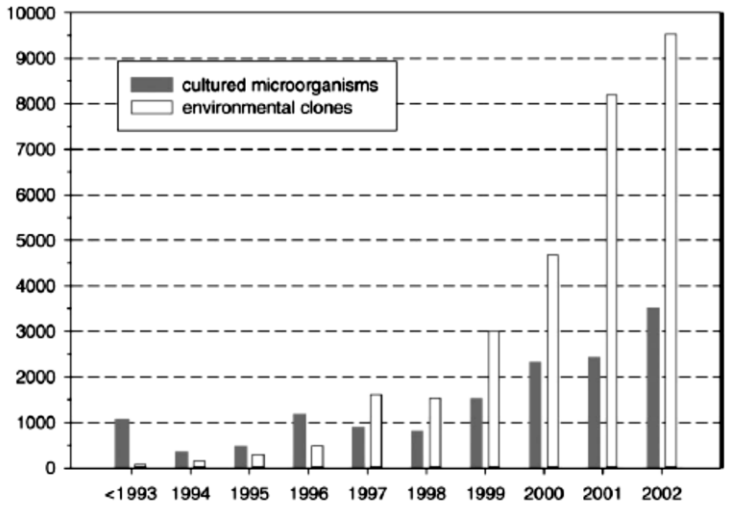


Figure 3.1: Increase in the number of 16S ribosomal RNA sequences published in GenBank since 1993 from cultivated Bacteria and Archaea and from cultivation -independent studies. Adapted from Rappè and Giovannoni 2003

Prior to the informational explosion brought by molecular techniques (before 1993), the numbers of published sequences was of the order of thousands all derived from cultivated Bacteria and Archaea. From 1993 surveys of microbial communities in natural environments exploiting cultivation-independent technologies have steadily increased, and the numbers of 16S rRNA sequences deposited in GenBank increased accordingly. The numbers of deposited sequences

from uncultured microorganisms surpassed that of cultured ones in 1997, reaching in 2002 a tenfold increase in the number of deposited sequences respect to all sequences deposited before 1993. At present, a query for cultured Bacteria and Archaea (using the following search (16S OR SSU OR small subunit) NOT (uncultured OR unknown OR unidentified) produced 1'570'818 matches, while a query for uncultured ones (using the following search (16S OR SSU OR small subunit) AND (uncultured OR unknown OR unidentified)) produced 6'179'531 matches.

Why are microbes uncultivable? Conceptually there is no such a thing as an absolutely uncultivable microorganism, assuming that it is vital. Rather we know that all microorganisms are able to grow in the environment, and most likely will grow in the laboratory as soon as we will acquire all the critical information to fully replicate the natural environmental conditions. Following this idea, uncultivable microorganisms can be divided in: (Dewi Puspita et al. 2012)

1. dividing cells (also defined "yet-to-be cultivated organisms")
2. non diving cells (dormant cells, resting cells, spores, viable but not cultivable cells)

High numbers of environmental physical conditions (temperature, pH, osmotic pressure, oxygen concentration), chemical conditions (carbon source, macro and micronutrients), plus a wealth of largely unknown growth factors and biological factors (signaling compounds, metabolites of neighbors, need for a co-culture, etc.), need to be in the right concentrations or values range for successful microbial growth. Most of those conditions are difficult to measure at the microscale at which they affect microbial cells, while some are still totally unknown. As a consequence it is difficult to make a synthetic media that accurately recreate the environmental conditions, and attempting to create it by varying all of these conditions results in an unaffordable amount of possibilities and a low rate of success. This classically represents one of the primary causes behind the "Great Plate count anomaly".

Over the past ten years the problem of microbial uncultivability has received increasing attention, and the novel cultivation methods introduced have increased the recovering of yet-to-be cultivated organisms. In 2009 Epstein proposed a new theory to explain microbial uncultivability (Epstein 2009; Epstein 2013). Natural microbial population would be composed of two entities, one small and active, while the other large and dormant, and the transition between the dormant and

the active state would be a stochastically event. In his “Scout Hypothesis” a fraction of dormant cells are thus able to randomly wake into activity and explore the environment for available resources (acting like a scout, from which the name of the hypothesis). If the environmental conditions are not permissive, the scout cells would die, causing the loss of a tiny fraction of the population, but preserving the majority of dormant cells. This random alternation of dormant and active stages continues until a scout cell is awakened in growth-permissive environmental conditions. The awakening of dormant cells into scouts would be the result of stochastic changes in, for example, the expression or repression of a master regulatory gene. At this point, alternatively, the scout cells can grow and reestablish the population, or they can produce and accumulate a growth-promoting factor that induces the dormant cells to activity, in a quorum sensing fashion.

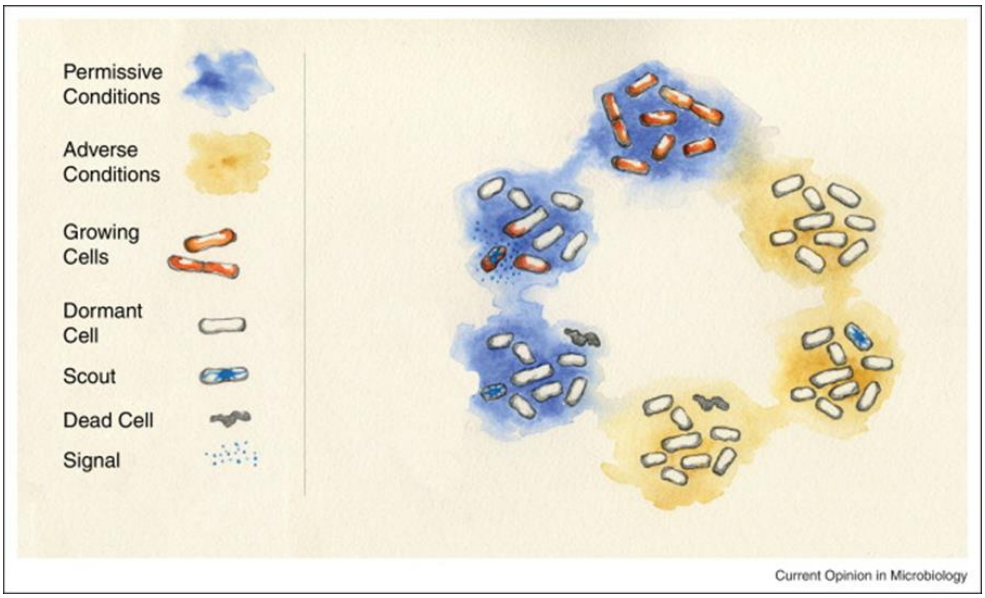


Figure 3.2: Scout model of the microbial life cycle. (I) Growth under permissive conditions, (II) dormancy under adverse conditions, (III) stochastic awakening from dormancy, in the form of ‘scout’ cells, followed by either their death (IV) or proliferation (V–VI), depending on environmental conditions, and, in case of proliferation, (VII) production and accumulation of signaling compound(s) inducing growth of the remaining dormant cells, concluding the cycle. From Epstein 2013.

The scout hypothesis can provide new insight in the mechanisms underlying microbial uncultivability. An environmental population of yet-to-be cultivated organisms would be mainly constituted by dormant cells. Considering that the awakening of dormant cells into active scouts is a stochastic and low-frequency event, the chances that active scouts cells from those populations are sampled and inoculated on a solid media, are extremely low. In those cases, the totality of cells would be in a dormant state, as such unable to grow on synthetic media, and the species would be inevitably classified as uncultured. The Epstein theory affirms that if enough time is given for the stochastic formation of scout cells, or if the initial sample size, or the numbers of replicated plates, is big enough to include active scout cells, the population should be capable of growing in any permissive synthetic media. It follows that, most of the uncultivable microorganisms are not really non-cultivable or not able to grow in the synthetic media; but rather they appear to be uncultivable for the rarity of their active form in the environment.

3.2.2 The prokaryotic species concept.

Microbial uncultivability surely represented the biggest limitation for microbial ecologist in the process of discovering the immense prokaryotic diversity. This limitation was reduced by the advent of molecular techniques and has lately received high attention, leading to the introduction of several new cultivation methodologies and to a significantly increase in the numbers of cultivable microorganisms recovered. Another problem that is classically subject of intense debate in microbial ecology, or more in general in microbiology and microbial systematics, is the prokaryotic species concept which, conversely, still lack of a wide consensus.

The definition of a species concept is a central requirement for every branch of Biology that deals with the study of diversity, its origin, and its classification, because it constitutes the principle on which we can categorize it. Some of the first attempts to define a species concept were based on the simple observation of phenotypic properties with species defined as a cluster of individuals with similar morphological characteristics. A species concept based on morphological trait was readily applicable to major organisms, plant and animals, but was not sufficient to classify prokaryotic organisms that exhibit low morphological differences. For microorganisms instead, the definition of species was based on the observation of multiple metabolic traits.

Botanists' and Zoologists' view of the fundamental properties of species was deeply affected by the Biological Species Concept, published by Ernst Mayr (1942).A species was no longer just a cluster of morphologically similar organisms; a species was now viewed as a fundamental unit of ecology and evolution, with certain dynamic properties, held together by a force of cohesion that reduce genetic divergence among its members. In particular for the biological species concept, the cohesive force was identified in the genetic exchange; as long as organisms can successfully interbreed they will remain phenotypically and genetically similar, but when they lose the ability to interbreed, they become free to diverge without bound. Mayr thus defined species as a reproductive community, a group of organisms with the potential to interbreed and produce viable and fertile offspring. (Cohan 2002). A definition of species based on breeding, was not applicable to the asexual prokaryotic world. However, some sort of cohesive evolutionary forces must also be acting to generate microbial species, because on its absence the accumulation of genetic variation would probably result in a genetic continuum and microbial species would represent artificial classification rather than natural, circumscribed biological grouping (Achtman & Wagner 2008).

The introduction of molecular technologies has provided new possibilities for defining prokaryotic species based on the properties and information contained in the DNA. The currently accepted definition of prokaryotic species follows a "polyphasic approach" (Achtman & Wagner 2008; Rosselló-Mora & Amann 2001; Vandamme et al. 1996) which combines a modern definition of species based on genomic information with the previous requirement of similarity in metabolic traits. Microbes are assigned to a common species if their reciprocal, pairwise DNA re-association values are ≥70% in whole genome DNA–DNA hybridization (DDH) experiments under standardized conditions and their ΔTm (melting temperature) is ≤5°C. In addition, all strains within a species must possess a high degree of phenotypic consistency, and species descriptions should be based on more than one type strain. A species name is only assigned if its members can be distinguished from other species by at least one diagnostic phenotypic trait.

Despite being considered the gold standard for defining prokaryotic species, the DDH approach has been criticized. It is experimentally difficult to implement and to standardize between different laboratories, it is not applicable to environmental surveys, but only on cultivable species (which, as introduced in cap 3.2.1, represent a tiny fraction of total diversity), and, most importantly, it results in too much phenotypic variation within the named species. DDH comparisons of individual strains are not necessarily symmetrical, and different values can be obtained with the same pair of strains depending on which is used as probe and which as target.

In addition, the results of this method can be intransitive: if DNA–DNA hybridization levels are >70% between strains A and B and >70% between strains B and C, they need not be >70% for strains A and C, which can lead to ambiguous or inconsistent assignments of strains to a common species (Achtman & Wagner 2008). Nevertheless, unless whole genome sequencing is routinely adopted, DDH represent the better approach for defining prokaryotic species, and finding a substitute may be difficult (Rosselló-Mora & Amann 2001).

By exploiting the information contained in the DNA, and in particular by targeting the phylogenetic information rich region encoding for the 16S rRNA (Woese 1987), another practical definition of species was created. It was observed, in fact, that microbes that had a 16S rRNA sequence with ≤98.7% identity were always member of different species, given that those levels of sequence dissimilarity correlated with values of <70% in DDH similarity. However, the opposite is not always true, distinct species have been occasionally described with 16S rRNA sequence that was >98.7% identical.

A conservative cut-off level of 3% in 16S rRNA sequence dissimilarity can thus be used to define species; this cut-off level however, was not based on any particular justification but instead was chosen to match pre-existing species definitions (Achtman & Wagner 2008). Most uncultured microbes cannot be assigned to a species because we lack information on their phenotype; nevertheless, in some cases, uncultured microbes can be assigned a provisional “Candidatus” designation if their 16S rRNA sequences are sufficiently different from those of recognized species, if experimental in situ hybridization can be used to specifically detect them and if a basic description of their morphology and biology has been provided (Murray & Stackebrandt 1995).

3.3 Methods for studying natural microbial communities.

Various molecular techniques are included in the microbial ecologist toolbox. In this chapter I will briefly outline some of the principal feature of molecular methods, while the next chapter (3.4) will provide further details on the methodologies used during the PhD.

3.3.1 Cultivation

Cultivation techniques are among the most classical tool microbiologists have used; their general purpose is the growth of microbes in laboratory synthetic media. As above outlined, microbes are extremely widespread and diverse in natural environments, hence the typical environmental sample in microbial ecology will contain a whole community of microbes. A first purpose of cultivation techniques in this case, is the growth of the higher possible proportion of microorganisms in order to enumerate them and obtain a quantitative measure of their abundance (for example the Colony Forming Unit parameter; CFU). Cultivation studies on the whole community present strong limitations (as outlined in 3.2.1), for this reason a cultivation step is often used to recover specific microbes of interest from the sample through enrichment, a techniques first introduced by Beijerinck. In this case the synthetic media is not designed to favor the growth of the higher possible proportion of microbes, but instead is tailored on the known feature of the microorganism of interest. After the enrichment, microbes of interests are recovered and isolated in pure culture. Obtaining a pure culture is the first step a researcher would take to study the metabolism, microbiology, genomics, and so on, of a specific microorganism.

3.3.2 Microscopy techniques

Another classical, and historically older, tool in the microbiologist toolbox is microscopy. Optical microscopy can be used to describe the morphology of prokaryotic cell, or the morphology of the colony formed on agar media; but it bear low information. On the contrary, microscopy has a much higher importance in the study and classification of Fungi. Nowadays, one of the most diffused applications of microscopy is Fluorescent In Situ Hybridization (FISH). FISH combines molecular techniques, using fluorescent nucleotide probes, and microscopy, in particular fluorescence microscopy and confocal microscopy. The sequence of the nucleotide probes is usually designed to be specifically complementary to the sequence of the ribosomal RNA of an organism of interest (16S for Prokaryotes or 18S for Eukaryotes). By choosing ribosomal RNAs as target the fluorescent signal can be amplified, as there will be a high numbers of target sequences for each cell (a high number of ribosomes). A sample can be stained with one, or more, fluorescent nucleotide probes; each probe will specifically bind to the ribosomes of the microbial species of interest, revealing its presence, or its specific position. The study of Bacterial biofilms often relies on modern microscopy techniques; biofilms are visualized by various kinds of microscopy including confocal laser scanning microscopy (CLSM) and scanning electron microscopy (SEM).

3.3.3 Molecular techniques

Remarkable advances in the study of environmental communities came from the advent of molecular technologies. In a molecular approach there is not a need for cultivating microorganisms in synthetic media, instead a “molecular marker” is directly isolated and purified from the environmental sample, and analyzed with a number of available analytical techniques. In this way, the limitation imposed by the uncultivability is overcome, and the yet-to-be cultured microorganisms can be included in our survey of environmental communities. By far, the most widely used molecular marker is represented by the nucleic acid, both DNA and RNA, but other suitable markers are used, for example phospholipid-derived fatty acids.

First step of any molecular technique is the direct isolation and purification of DNA or RNA from the environmental sample. Key steps in this procedure are a) the efficient release of microbial cells from the substrate (in particular soil particles), b)

the efficient lysis of microbial cells, and c) the removal of inhibitory substances. Isolation and purification of DNA is a critical process and a major source of bias; considering the high heterogeneity of environmental samples, complete detachment from the substrate and subsequent efficient lysis for all the cells, is in fact suggested to be nearly impossible. It follows that only a "window" of the true diversity is recaptured, and rigorous protocol standardization is required to reduce bias in comparative works (van Elsas & Boersma 2011). The great majority of extraction protocols are carried out with the use of commercially available kits that adopt optimized procedure and have high extraction efficiency (high yield, high quality, and low fragmentation of recovered material) and reproducibility.

Often, especially when extracting DNA from soil, a further purification step after the extraction may be required. Purified DNA in fact, may still retain some substances that can inhibit subsequent analysis. Also in this case, various commercial kits are available. A classic example is represented by the inhibition of polymerase by humic substances. Humic substances are amorphous, dark-colored organic compounds which are relatively resistant to chemical and biological degradation; they are found in soils, water, and sediments, and constitute the major components of the natural organic matter (Matheson et al. 2010). They are constituted by complex and heterogeneous mixtures of materials formed during the decay and transformation of plant and microbial remains (a process called humification). Plant lignin and its transformation products, as well as polysaccharides, melanin, cutin, proteins, lipids, nucleic acids, fine char particles, etc., are important components taking part in this process. The effect on polymerase is different depending on the inhibitory compound; some substances will inhibit the enzymatic activity while others will cause template inhibition, caused by the binding of the substance to the template, preventing it from being amplified. It has been indicated that humic substances can produce both types of inhibition (Matheson et al. 2010; Sutlovic et al. 2008).

After retrieving environmental DNA, a common step is represented by amplification, via polymerase chain reaction (PCR), of the target gene. PCR technique was first introduced by Kary Mullis in 1983 (Mullis et al. 1994) and since then it has gained a practically indispensable role in many fields of science. The technique is used in molecular biology to obtain high copy numbers of a target region, starting from one to a few of its copies in the initial material. PCR is based on: 1) a repetition of cycles of heating and cooling (thermal cycling), 2) the use of heat stable DNA polymerase, 3) the use of small oligonucleotides (the primers) whose function is to direct the reaction to the target region of the DNA, and to allow it to begin, 4) the exponential increase in copy number of the target region. A

classical PCR is carried out on a Thermocycler, an instrument that automatically performs the repetitive cycle of heating and cooling. The number of cycle, temperature, and time of most steps depends on the specific polymerase used, and suggested protocols are given by any manufacturer.

Despite being a key step in any molecular biology technique, PCR has a number of potential biases (van Elsas & Boersma 2011). First of all, most of the surveys of total microbial communities are performed using primer designed to be universal, but that in reality may be missing a considerable part of the community. Universal primers sequence is in fact designed to match conserved regions of the target, for example conserved regions in the 16S rRNA, by alignment of all the available sequences in databases. As a result, as the information in databases grow and as new taxa are added, primers designed to be complementary to the conserved regions of most prokaryotes at a given time, are not necessarily complementary to all those that exist in the database today (Baker et al. 2003), and need to be re-evaluated. It has to be considered that a truly universal primer may be impossible to design, as we actually miss the sequence information of a great part of the real diversity. Second, particular target may amplify at higher rate than other one, the so-called “differential amplification” (Reysenbach et al. 1992). Third, sequence artifact may rise from the formation of chimeric molecules, heteroduplex, or error of the polymerase (Acinas et al. 2005).

Following PCR amplification of the target gene (16S rRNA or any other phylogenetic or metabolic marker) various techniques can be implemented to retrieve information on the microbial community. We can divide them in two “families”: fingerprinting techniques and sequencing techniques (Figure 3.3).

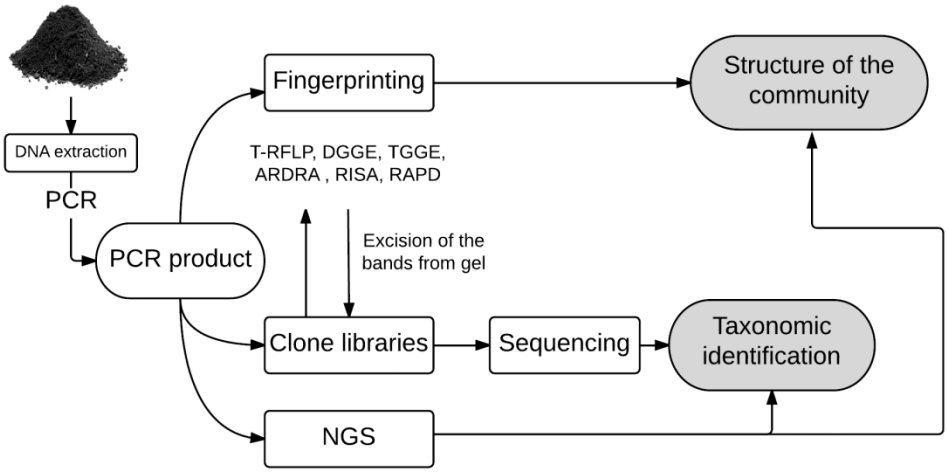


Figure 3.3: Schematic representation of the different molecular techniques that can be used in the study of microbial communities composition and diversity. Typically, genomic DNA is directly extracted from the environmental samples and the target gene is amplified with PCR. The PCR strategy will change in dependence of the subsequent methodologies. PCR products can be analyzed by fingerprinting or sequencing technologies. Fingerprinting consist in a series of methodologies, relying on electrophoretic separation of fragments, that can provide information on the genetic structure of the community. Classical sequencing methodologies involve the cloning of the PCR product, a long and laborious methodology, prior to sequencing. Similarly, some fingerprinting technologies allow the recovery of DNA fragments that can be cloned and sequenced, to provide taxonomic identification of particular populations. On the contrary, NGS methodologies rely on the direct sequencing of PCR amplicons, providing information on the taxonomy and genetic structure of the community, without the need of a cloning step. Adapted from Ranjard et al 2000.

3.3.3.1 Fingerprinting

Fingerprinting techniques are a series of application that are often used to provide a general picture of the structure and diversity of the microbial community. Those techniques offer a cost effective and a rather easy to implement method for the comparison of natural microbial communities through time and/or space, or in two particular conditions, or again before and after some kind of impact, etc. The base principle is to resolve the diversity of the amplicons resulting from the direct amplification of DNA extracted from the sample through electrophoresis. The

result is a complex profile of bands (when using gel electrophoresis) or peaks (when using capillary electrophoresis), resolved based on fragments size (for example T-RFLP) or fragments sequence (for example DGGE, TGGE), which picture the diversity of the community, just like the fingerprint on our fingers picture our identity. Each technique has different characteristics; features of T-RFLP analysis will be covered in 3.4.1.

DGGE/TGGE: Denaturing Gradient Gel Electrophoresis (DGGE) (Muyzer et al. 1993) and Temperature Gradient Gel Electrophoresis (TGGE) are fingerprinting methods based on the separation of DNA fragments (amplicons) of the same size but with different sequence. Those methodologies are based on the use of a polyacrylamide gel containing a linear gradient of a denaturing agent; the agent can be chemical, such as urea or formamide in the case of DGGE, or physical, such as temperature in the case of TGGE. Each fragment, depending on its sequence, will be denatured at a certain point along the gradient, and its migration will be arrested. The result is a complex profile of bands, which describe the diversity of the microbial community and that can be analyzed with a series of multivariate statistical techniques. Individual bands may be excised, re-amplified, and sequenced to obtain the taxonomic identification of specific populations within a community. The size of the bands (max 500 bp), however, limit the phylogenetic information obtainable; moreover, each band, or operational taxonomic unit (OTU), is likely composed by multiple sequences.

3.3.3.2 Sequencing techniques

Sequencing techniques are a valuable tool in the microbial ecologist toolbox. Most of the fingerprinting techniques, despite being capable of well recapturing community diversity, presenting a low cost, and having a good throughput, are in fact not suitable for obtaining information on the identity of the microorganisms in the community. Sequencing techniques, on the contrary, can be used to determine the species composition of a community.

Cloning and sequencing: Cloning and sequencing (using Sanger sequencing technology) represent the classical form or sequence analysis. The diversity of bacteria communities can be investigated by cloning the fragments of interest (which can be represented, for example, by 16S amplicons from the natural community or by band extracted from a DGGE gel) into plasmids followed by transformations. The fragments are then recovered by screening of the clone

library, and sequenced. Sequencing allows a fine identification of uncultured bacteria as well as an estimation of their relatedness to known cultured species. Cloning and sequencing major drawbacks are their time-consuming and cumbersome experimental procedures (a large number of clones must be analyzed to obtain good diversity estimation) and its high cost per unit (i.e. cost per base).

NGS: The last decade has witness high progress in our understanding of natural microbial communities; most of the credit goes to the application of Next Generation Sequencing technologies (NGS). The term NGS refers to a series of high throughput sequencing platform and methods that are being increasingly applied to profile microbial communities at a depth that was unattainable with the classic cloning and sequencing approach. Strength of those technologies resides in the huge amount of data (sequences) they can produce in a fraction of the time required by the previous methods (Sanger technologies), accompanied by the significantly lower per-base cost and their high multiplexing capability. Weakness of the use of NGS systems for the profiling of microbial communities lies in the short length of the sequences produced when compared to the classical Sanger sequencing. This difference impacts the phylogenetic resolution of the analysis, lowering it, as a general indication, to the order/family level.

However, this limitation can be minimized by the rational choice of the region of the 16S rRNA gene to be sequenced. The 16S rRNA gene contains regions characterized by a different variability rate; some are conserved among different species, while others are defined hypervariable (V1 to V8) and can be used to perform the phylogenetic classification (Vasileiadis et al. 2012) (Figure 3.4). Due to their high resolution power, the bacterial V4 and archaeal V3 hypervariable regions are frequently chosen for analyzing the composition of prokaryotic community(Bartram et al. 2011; Caporaso et al. 2012; Ragan-Kelley et al. 2013; Whiteley et al. 2012; Yu et al. 2008; Yu & Morrison 2004).

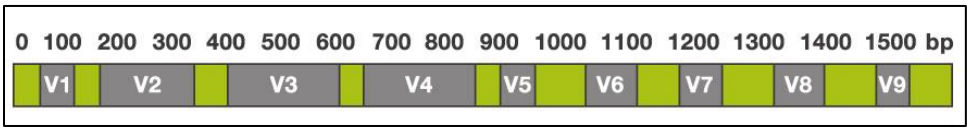


Figure 3.4: 16S rRNA gene illustrating the conserved (green) and variable (grey) regions, from http://www.alimetrics.net/en/index.php/dna-sequence-analysis

Different NGS platform are currently available on the market and each method has different strengths and weaknesses (for a comparison Loman et al. 2012; Pallen 2013). Even if the majority of NGS studies of natural microbial communities to date involved the application of 454 pyrosequencing platform (Nikolaki & Tsiamis 2013), Illumina based studies are rapidly taking the lead. High-throughput sequencing has seen extensive use in microbiology, however, until recently sequencing platforms were tailored toward large-scale applications with footprints, workflows, reagent costs and run times poorly matched to the needs of small laboratories. However, three different benchtop high-throughput sequencing instruments are currently available (Table 3.1) (Loman et al. 2012).

Table 3.1: Price comparison of benchtop instruments and sequencing runs, from Loman et al. 2012

Platform	List price	Approximate cost per run	Minimum throughput (read length)	Run time	Cost/Mb	Mb/h
454 GS Junior	$108,000	$1,100	35 Mb (400 bases)	8 h	$31	4.4
Ion Torrent PGM						
(314 chip)	$80,490[a,b]	$225[c]	10 Mb (100 bases)	3 h	$22.5	3.3
(316 chip)		$425	100 Mb[d] (100 bases)	3 h	$4.25	33.3
(318 chip)		$625	1,000 Mb (100 bases)	3 h	$0.63	333.3
MiSeq	$125,000	$750	1,500 Mb (2 × 150 bases)	27 h	$0.5	55.5

Note pricing may vary between countries and/or sales territories. Instrument prices do not include service contracts. Sample prices do not include the cost of generating the initial fragmented genomic DNA library with adaptors (an additional cost of between $75–200 depending on method used). Cost per megabase assumes one sample and one sample sequencing kit per run. Unless stated, pricing information is from the online supplement of ref. 3.
[a]Ion Torrent PGM pricing from Invitrogen US territory website (http://www.invitrogen.com/, accessed 21 February 2012).
[b]Price includes Ion Torrent PGM, server, OneTouch and OneTouch ES sample automation systems. [c]Ion Torrent PGM prices include chip and sample preparation kit. [d]Configuration used in this study.

3.4 Methods exploited during the PhD

3.4.1 Terminal restriction fragments polymorphism

Terminal restriction fragments length polymorphism (T-RFLP, Liu et al. 1997) is a reproducible, high-throughput, and high-resolution (Hartmann & Widmer 2008) cultivation independent methodology. Although currently less frequently utilized in favor of newer sequencing techniques, it remain a legitimate molecular tool in the microbial ecologist toolbox (van Dorst et al. 2014) as it allows for a rapid, simple, and inexpensive analysis of microbial community structures and for the monitoring of its changes on spatial and temporal scale (Osborn et al. 2000; Schütte et al. 2008). T-RFLP analysis is normally used for the analysis of phylogenetic marker such as the 16S rRNA gene for bacterial and archaeal population or the 18S rRNA gene and the ITS (Intergenic Transcribed Spacer) region for fungal population; however it has been also used for the analysis of metabolic marker, such as those encoding sulphate reduction

In T-RFLP a specific target gene, from the total DNA extracted from an environmental sample or a microbial isolate, is amplified using one or two fluorescently labelled primers (primers with a fluorescent dye bound at the 5'). The amplicons mixture is then digested with one or more restriction enzymes (usually "frequent cutter" that recognize a 4bp restriction sites) and the resulting fragments are separated by capillary electrophoresis on an automatic sequencer. Only the labelled fragments are detected by the detection system and the result is a spectra, called electropherogram (Figure 3.5), in which each peak represents a different fragment, characterized by a specific size (peak position along the x axis), determined by interpolation with an internal standard, and abundance (peak height or area), measured as fluorescence intensity.

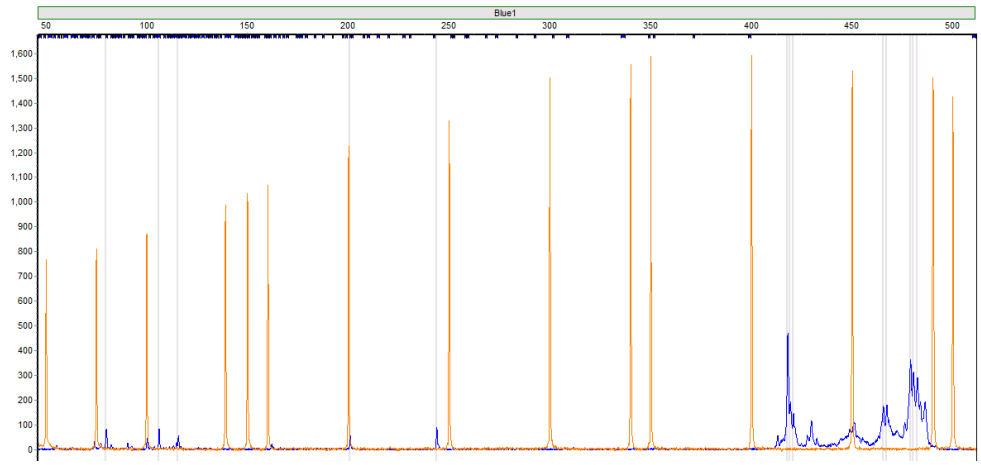


Figure 3.5: Example the electropherogram resulting from T-RFLP analysis. Orange peaks indicate the internal size standard (500-LIZ in this case), while blue peaks indicates the fragments obtained from the enzymatic digestion of the amplicons (labelled with FAM fluorochrome). X-axis indicates the size of the fragments, while y-axis indicates the intensity of the fluorescent signal (relative fluorescence units).

Each different sized fragment originates from a polymorphism on a restriction site in the target gene. For this reason T-RFLP can be used to investigate complex microbial community composition. When a phylogenetic marker is chosen as target, each peak represents a phylotipe or operational taxonomic unit (OTU). However, since this classification is only based on size and does not consider the sequence of the fragment and since several different microbial species can originate fragments of the same length, each OTU inevitably represents more than one species. Instead, if a metabolic marker is chosen as target, only the species belonging to a specific functional group, which therefore possesses the target gene on their chromosomal or plasmidic DNA, will be comprised in the analysis. In this work, T-RFLP was employed to elucidate the community structure of sulphate reducing bacteria (SRB) in surface and anoxic sediments (Pérez-Jiménez & Kerkhof 2005). This functional group, composed by at last five bacterial *phyla* and two archaeal *phyla* (Santillano et al. 2010), is involved in the decomposition and mineralization of the organic matter and oxidizes diverse carbon sources, including petroleum hydrocarbons, while using sulphate as an electron acceptor (Muyzer & Stams 2008). Hydrocarbon degradation under sulphate-reducing conditions is an important process for hydrocarbon removal in marine sediments. The dsrAB gene was chosen as genetic marker to specifically characterize SRB since this gene codes for the dissimilatory sulphite reductase which catalyzes the last step in the sulphate reduction pathway. Moreover, the dsrAB gene has been found in all the known SRB lineages (Stahl et al. 2002).

3.4.2 High throughput DNA sequencing with MiSeq platform

Illumina has released various instruments; the base principle and sequencing chemistry are the same, however there are big differences in read length, output, and run time. The platform that mostly got the attention of microbial ecologists is the MiSeq, which is becoming a standard in 16S profiling of microbial communities. This platform present a considerably lesser amount of output respect to the others, which make it very suitable for this kind of studies, and, above all, present the higher read length of 150 bp (or thanks to a recent update 300bp). The possibility of adopting a paired-end protocol further increases the length of the reads. In paired end protocols each amplicon is in fact sequenced by both ends, producing two different sequences read that overlap by a number of bases. After sequencing, the two reads can be joined producing a single, longer sequence. NGS reads are usually characterized by a gradual loss in sequence quality, producing a region of low quality towards the end of each read. The paired-end protocol can resolve this problem, as the low quality region usually corresponds to the overlap between the two reads.

Illumina technology (Sequencing by synthesis) (Metzker 2010; Mardis 2008a; Mardis 2008b) (Figure 3.6) uses nucleotides fluorescently labelled with 4 colors. Nucleotides are modified with a reversible terminator; in this way only one per cycle can be bound to the template. First step of the analysis is the immobilization of the fragments (previously ligated with adapter sequences, complementary to a "lawn" of oligonucleotides on the flow cell) to the solid support of the flow cell. Then, fragments are amplified with "bridge amplification" to form clusters of clonal sequences; this step is needed to increase the fluorescent signal. At each cycle of the subsequent sequencing reaction, just one fluorescently labelled nucleotide is bound to the DNA template, then an image of the flow cell is taken, and finally the terminator is removed and the next cycle can begin. By elaborating the images, hence by considering the color of each cluster in every cycle, the sequence of the fragments is reconstructed. At this point, in a paired-end approach, clusters are re-formed with the fragments in the opposite orientation, binding them with the other adapter. Amplification and sequencing then proceed in the same way as the first read

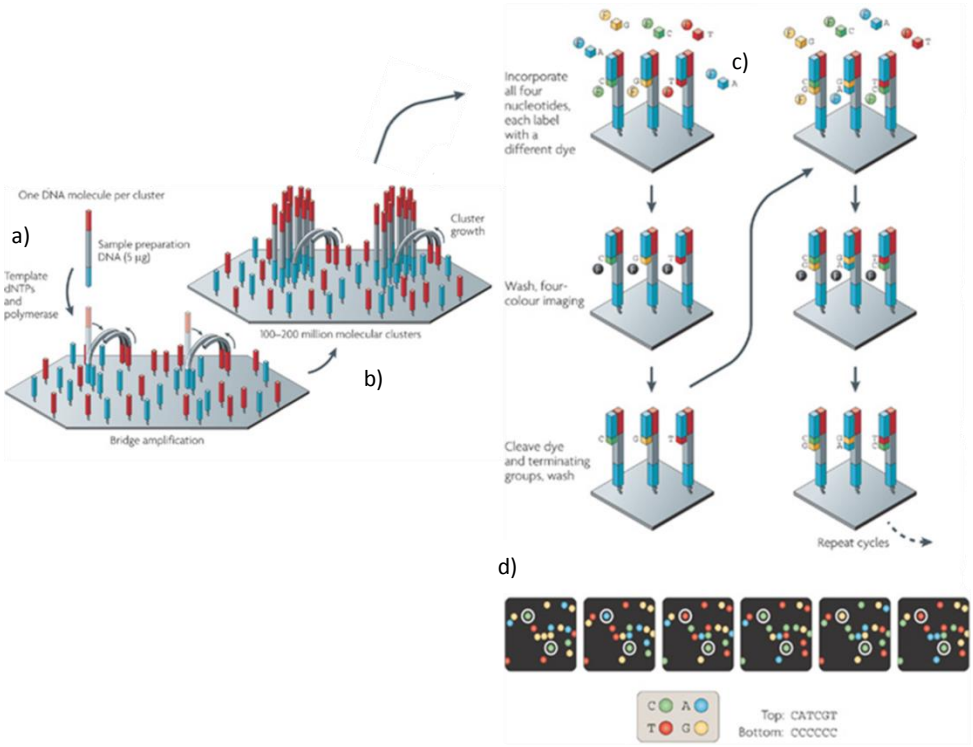


Figure 3.6: Schematic representation of ILLUMINA sequencing methodology. First step (a) is the immobilization of the template DNA to the flow cell, followed by the bridge amplification (b), using adjacent primers, to form clonal clusters. The actual sequencing (c) uses 4-color labelled bases modified with a reversible terminator. At each cycle, a single nucleotide binds to the template and, following imaging (d), the terminator and the fluorescent dye are cleaved for the next cycle.

During the last three years, our laboratory has designed and performed the characterization of the microbiome of sediment samples from the Mediterranean Sea basin. Both the bacterial and the archaeal communities have been targeted by analysing hypervariable regions (V4 for Bacteria and V3 for Archaea) of the 16S rRNA genes. The sequencing analyses were performed in outsourcing using the MiSeq platform.

We followed a custom approach similar to that used in Caporaso et al. (2011), also adopted as a standard protocol by the Earth Microbiome Project (http://www.earthmicrobiome.org/emp-standard-protocols/16s/). This custom approach guarantees to obtain the maximum information on the sequence of the hypervariable region by adopting sequencing primers designed to bind at the

amplification primers. In this way, the produced sequence spans directly the informative 16S sequence, bypassing the sequence of the amplification primer, guaranteeing higher information content. We were not interested in the abundance ratio between bacteria and Archaea community, but rather we were interested in obtaining the maximum information on the composition of both communities. For this reason the above mentioned protocol was modified, replacing the prokaryotic universal primers with kingdom-specific primers selected from literature to guarantee high coverage of the bacterial (V4 region; couple 563F/802R Claesson et al. 2010) and archaeal communities (V3 region; couple 344F/519r Yu et al. 2008).

The final amplification primers, as outlined in Figure 3.7:

- Included the Illumina flowcell adapter sequence (two different adapters, one for forward primers and one for reverse primers)
- Included a linker region
- Reverse primer included a 12-base index for multiplexing

Three sequencing primers were also designed, with the following sequence composition:

1. Sequencing primer for the first read: forward linker region + forward amplification primer
2. Sequencing primer for the second read: reverse linker region + reverse amplification primer
3. Index sequencing primer: reverse complement of reverse linker region + reverse amplification primer

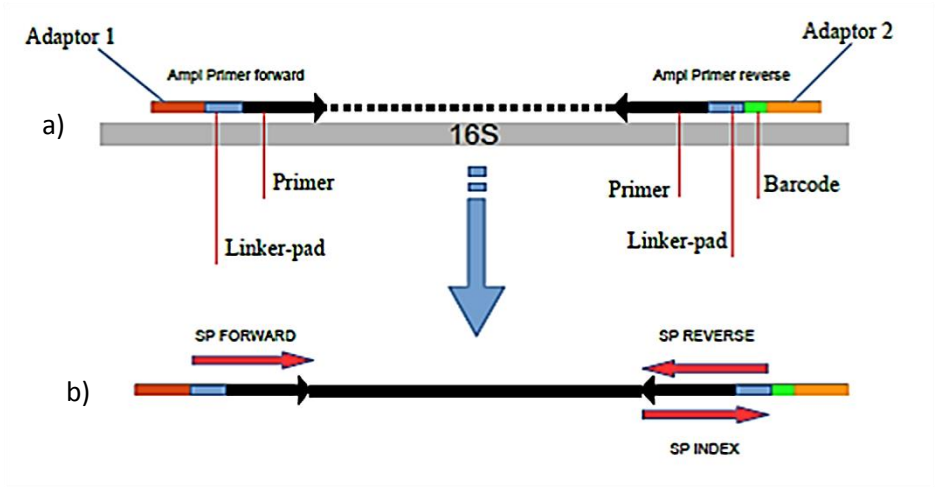


Figure 3.7: a) Structure of the amplification primers; b) Structure, and binding position on the amplicon, of the sequencing primers. "SP" stands for sequencing primer.

The linker region was designed to obtain sequencing primers (that have the same sequence of the linker + amplification primer) with melting temperature higher than 65 °C. In this way we could use a sequence different from that suggested in Caporaso et al. 2011. This expedient is needed to ensure a successful sequencing run, as the MiSeq sequencer performs a step at the temperature of 65 °C (ref. http://nextgen.mgh.harvard.edu/CustomPrimer.html,jamimmunology.blogspot.it/2012/11/custom-miseq-sequencing-primers.html, and Li et al. 2012).

We decided to adopt this kind of approach because in this way: **i)** we could obtain an higher representation (an higher number of reads) and a better description for the Archaea community which, in a previous T-RFLP investigation, resulted less abundant respect to the Bacterial one, and presented some amplifications problems **ii)** we could choose primers optimally developed for Archaea and Bacteria, respectively, instead of broad range prokaryotic primers, thus increasing the expected phylogenetic resolution **iii)** by using a different index for the bacterial and archaeal primers (for each sample we had an index corresponding to the bacterial amplicons and one index corresponding to the archaeal amplicons), we were able to split the resulting dataset into smaller datasets, one for each Kingdom, and to analyse them separately. In this way we were able to easily change the analysis pipeline, applying some different steps for the analysis of archaeal communities respect to bacterial one. In addition, splitting the read library into smaller library allowed us to perform the entire analysis pipeline on a normal laptop (Intel® Core™ i7 CPU @ 2.20 Ghz, RAM 8 GB), avoiding the need of a dedicated machine.

Giving that this approach was not a standard, there was a long phase of communication and planning between us and the sequencing facility. Major concerns regarded the capability of the system to perform well in the presence of three sequencing primers (Sequencing primer for the first read, Sequencing primer for the second read, and Sequencing prime for the PhiX control library). We performed a total of two sequencing runs and in each of them the system worked fine, except a flaw on the numbers of clusters recovered for the bacterial libraries, with high percentage of clusters not passing the filter. The adopted analysis pipeline is schematized in Figure 3.8.

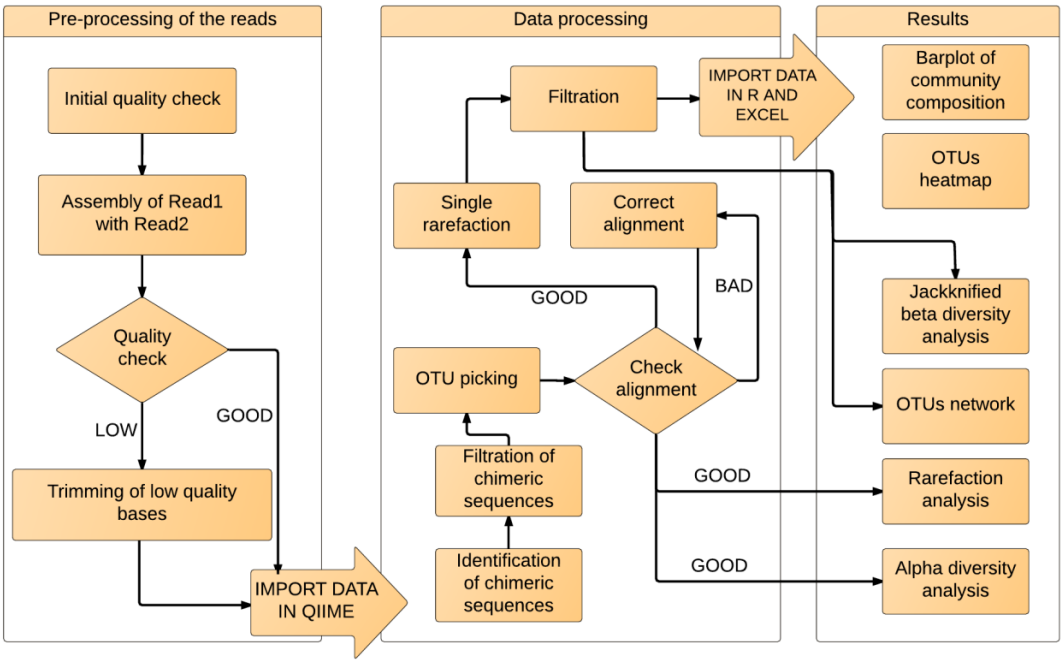


Figure 3.8: Flowchart summarizing the steps in processing and analysing NGS data.

As outlined above, a single sequencing run contained multiple samples, as well as amplicons belonging to both Bacteria and Archaea. The reads belonging to each of those sources can be divided exploiting the different barcode sequence. First step of data analysis, performed by the sequencing facility, is the division of the total reads in smaller dataset on the basis of its barcode sequence (DEMULTIPLEXING).

Following this step, for each sample we received two dataset, one containing the reads of the forward sequencing and one containing the reads of the reverse sequencing round. Next step in data analysis was joining the forward and reverse read using software PEAR (Zhang et al. 2014), quality checking before and after the joining step was performed with FastQC software (Andrews 2010). Software PEAR was chosen as merger thanks to its better way of dealing with fully overlapping reads, as it was the case for archaeal dataset. After quality check a flaw in bacteria reads was noticed; the joined reads showed a region of low quality toward the end. This was unexpected as this region corresponds to the initial part of the second read, which is supposed to have high quality. This low quality region was eliminated (TRIMMED) using StreamingTrim software (Bacci et al. 2014); a software that removes low quality bases at the end of the reads while trying to be as much conservative (low read loss) as possible. Resulting reads were considered of high quality and used for subsequent analysis.

Subsequent data processing was carried out using software QIIME (Caporaso et al. 2010). Various step of the analysis are schematically pictured in Figure 3.8. Firstly, chimeric sequences were identified with *uchime* (Edgar et al. 2011) and then filtered out from the datasets. Open reference otu picking was performed for bacterial communities, while de novo otu picking was performed for archaeal communities. Both procedures were carried out with *usearch61* (Edgar 2010). Alignments and phylogenetic trees were visually checked, and eventually edited, for bad aligned sequences. The resulting OTU table was used to perform multiple rarefaction analysis and alpha diversity analysis, and, following the suggestion in Bokulich et al. (2013), rarefied to an even depth and filtered for OTUS that were represented by a total number of reads that is < 0.005% of the total number of read, for beta diversity analysis.

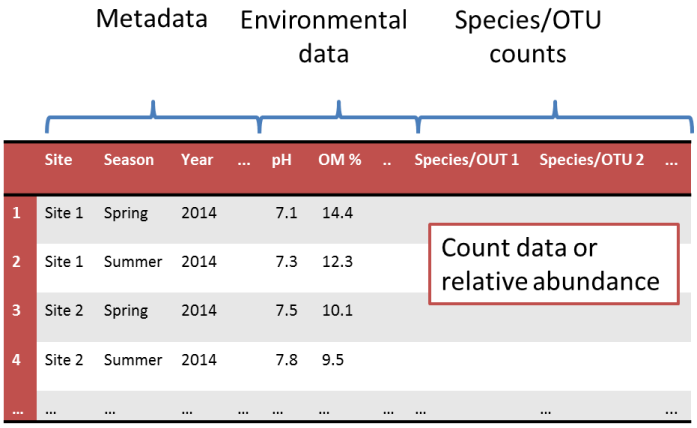
3.5 Statistical data elaboration

Microbial ecologists rely on the use of various tools which have been developed by community ecologists to work on distribution and diversity patterns of plants and animals, and that could be readily applied in microbial ecology. Typically researchers have to deal with a multivariate dataset, often in a classic sitexspecies structure (see Figure 3.9 for an example), where each row represent a sample (or more generally a single observation), while the columns usually indicate:

1. Species count. Those columns contain the observation of some variables, and constitute the "real data" upon which ordination or clustering analysis is performed. In our case those variables are represented by OTUs abundance for fingerprinting data (i.e. relative peak height or area in T-RFLP analysis) or sequence counts for NGS data.

2. Properties of the site (the metadata columns). Those columns are used to define known attributes of the samples and usually contain categorical variables that are excluded from the actual analysis. Their presence is somewhat optional but deeply suggested to keep track of important attributes of the observations. In data analysis those attributes can be used, for example, to define a priori sample grouping and to test their statistical significance (e.g. a grouping based on season in Figure 3.9) with analysis such as ANOSIM. Moreover, those attributes are used to define color or shape of the point in ordination analysis, aiding interpretation of the results.

3. Environmental data. Those columns contain a set of environmental, climatic, chemical etc. properties that have been measured. Some multivariate techniques (the constrained techniques, such as CCA) can be used to link the ordination of species and sites to those variables.



	Site	Season	Year	...	pH	OM %	..	Species/OUT 1	Species/OTU 2	...
1	Site 1	Spring	2014		7.1	14.4				
2	Site 1	Summer	2014		7.3	12.3				
3	Site 2	Spring	2014		7.5	10.1				
4	Site 2	Summer	2014		7.8	9.5				
...	...	...	...	...	...	...	...	...	...	...

Figure 3.9: Structure of a classical sitexspecies multivariate dataset.

Molecular methods are meant for recapturing a great part of the diversity. For this reason the dimensionality and complexity of the datasets will be very high, as each species or OTU recovered will add a column (a variable) to them. To statistically analyze those types of datasets it is necessary to apply multivariate statistics. Multivariate statistics refers to the extensive collection of tools available to analyze three or more variables at the same time. Objectives of molecular microbial ecology studies usually are: 1) calculating diversity indices; 2) visualizing relationships among samples using ordination analysis (e.g. nMDS, PCoA); 3) identifying samples grouping using cluster analysis (UPGMA) or testing the significance of samples grouping (ANOSIM); 4) linking differences among the microbial communities to variations observed in the environmental variables (CCA); 5) identifying the species/OTUs giving the highest contribution to the observed diversity (SIMPER)

Diversity indices: the concept of a partition of biodiversity into different sources was first introduced by Whittaker in 1972 and is illustrated in Figure 3.10. Whittaker distinguished three levels of species diversity in natural communities based on considerations on the spatial scale. The total species diversity observed in a dataset or in a landscape of interest was defined delta diversity (γ); the species diversity at local scale, considered as the smallest spatial units in which diversity was measured, was defined alpha diversity (α); and the differentiation among those local scale habitats was defined beta diversity (β).

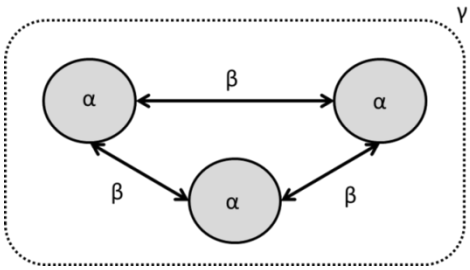


Figure 3.10: Partition of total biodiversity (γ diversity) in local scale diversity (α diversity) and in the diversity between local-scale habitats (β diversity)

In this thesis, diversity indices were calculated as follows:

Table 3.2: Measure of diversity used in this thesis and their calculation.

	Index	Calculation
Alpha diversity	Species richness (S)	Calculated as the sums of different species or OTUs
	Shannon index (H)	$H = -\sum_{i=1}^{S} p_i \log_b p_i$
	Simpson index (D)	$D_1 = 1 - \sum_{i=1}^{S} p_i^2$
	Pielou's Evenness (E)	Calculated as E = H/ log(S)
	Chaol richness estimation	$S_p = S_o + \frac{f_1^2}{2f_2}\frac{N-1}{N}$
Beta diversity	Bray-Curtis index	$d_{jk} = \frac{A+B-2J}{A+B}$
	Sørensen index	$\beta = \frac{a+b+c}{(2a+b+c)/2} - 1 = \frac{b+c}{2a+b+c}$
	Unifrac distance	Measures the difference between two environments in terms of the branch length that is unique to one environment or the other

Where: $A = \sum_{i=1}^{N} x_{ij}$ $B = \sum_{i=1}^{N} x_{ik}$ $J = \sum_{i=1}^{N} \min(x_{ij}, x_{ik})$

a = number of shared species in two sites is
b and c = the numbers of species unique to each site

Visualizing relationships among samples: Ordination methods are used to visualize the dissimilarities among microbial communities. All ordination methods reduce the dimensionality of multivariate data to a few number of axis, which are then plotted in two or three dimensions (Schütte et al. 2008) in order to represent data structure as best as possible. The result is usually a biplot in which each point represents a site; their position is determined by the community composition and the vicinity of point means a similarity in community composition. In this thesis, two methods were used:

NMDS: The non-metric multidimensional scaling (nMDS) algorithm ranks distances between objects, and uses these ranks to map the objects nonlinearly onto a simplified, two-dimensional ordination space so as to preserve their ranked differences, and not the original distances (Ramette 2007). Result of an nMDS ordination is a map where each point represents a sample; the position of each point is determined by its distance from all other points (Rees et al. 2004) and the proximity of points correspond to their similarity. Since it involves an iterative algorithm a "goodness of fit" (the stress of the plot) estimate is available, which indicates how different the rank in the ordination space are from the ranks in the original distance matrix. A stress value greater than 0.2 indicates that the plot is close to random, stress less than 0.2 indicates a useful 2 dimensional picture and less than 0.1 corresponds to an ideal ordination with no real prospect of misinterpretation (Rees et al. 2004). Most ordination methods are analytical and have a unique solution to a set of data, while nMDS is a numerical technique that iteratively seeks a solution and stops computation when an acceptable solution has been found, or it stops after some specified number of attempts. nMDS ordination can be used with any distance measure. NMDS is non-parametric, free of assumption and does not suffer of many pitfalls, such as the arch effect, affecting other ordination techniques, e.g. PCA.

PCoA: principal coordinates analysis (PCoA) was used for the ordination of NGS data as it is the default ordination method used by QIIME, and a diffused ordination method for those data in literature. It uses a linear (Euclidean) mapping of the distance or dissimilarities between objects onto the ordination space and the algorithm attempts to explain most of the variance in the original data set. It can be used with any dissimilarity measure. Eigenvalues are used to measure how much variance is accounted for on each PCoA synthetic axis. (Ramette 2007)

Identifying samples grouping: Cluster Analysis (CA) has the task of assigning a set of objects into groups called clusters so that objects in the same cluster are more similar to each other than to those in other clusters. CA is based on pairwise distances among those objects and on the use of a clustering algorithm. There are a

lot of different algorithms showing different characteristics, and with the possibility to use different measures of distance. The appropriate algorithm and parameter settings are chosen depending on the setting of individual data. In our case the UPGMA method was used: in UPGMA clustering the distance between two clusters is calculated as the average distance between all pairs of objects in the two different clusters. UPGMA is efficient when objects form natural distinct groups; it also performs well with elongated, chain-type clusters.

Samples grouping can be tested with analysis of similarity (ANOSIM) routine. ANOSIM test the null hypothesis that the average rank similarity between objects within group is the same as the average rank similarity between object between groups. The significance is tested with permutations, and an R statistic is produced. R range from -1 to +1; if the object are more dissimilar between groups than within groups, the R statistic will have a value greater than 0, indicating that the samples are effectively clustered based on the grouping we were testing. An R value of 0 indicates that the null hypothesis is true.

Linking differences to environmental variables: Canonical correspondence analysis (CCA) can be used to link the observed ordination of microbial communities to differences in environmental conditions. The method is based on the same algorithm as correspondence analysis (CA), but the ordination is constrained to a second data matrix containing environmental factors. It is free of assumption of normality on the data, but suffers of arching effect when applied to gradient data.

Defining the species/OTUs that contribute most to diversity: Similarity percentage (SIMPER) analysis can be used to calculate the average contribution of each species/OTU to diversity. The average contribution is calculated from the decomposition of the Bray-Curtis dissimilarity index.

All data elaborations and analysis were carried out using R Statistical software version 2.15.1 (R Core Team 2013) and Qiime version 1.8 (Caporaso et al. 2010). Principal R packages used were: package Vegan (Oksanen et al. 2008) for data analysis and statistics on T-RFLP results, Phyloseq (McMurdie & Holmes 2013) for data analysis and statistics on sequencing results, and ggplot (Wickham 2011) for graphics.

3.6 Bibliography

Achtman, M. & Wagner, M., 2008. Microbial diversity and the genetic nature of microbial species. Nature reviews. Microbiology, 6(6), pp.431–40.

Acinas, S.G. et al., 2005. PCR-Induced Sequence Artifacts and Bias : Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. , 71(12), pp.8966–8969.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bacci, G. et al., 2014. StreamingTrim 1.0: a Java software for dynamic trimming of 16S rRNA sequence data from metagenetic studies. Molecular Ecology Resources, 14(2), pp.426–434.

Baker, G.C., Smith, J.J. & Cowan, D. a., 2003. Review and re-analysis of domain-specific 16S primers. Journal of Microbiological Methods, 55(3), pp.541–555.

Bartram, A.K. et al., 2011. Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. Applied and environmental microbiology, 77 (11), pp. 3846-3852 .

Bokulich, N. et al., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature methods, 10(1), pp.57–9.

Bunge, J., Willis, A. & Walsh, F., 2014. Estimating the Number of Species in Microbial Diversity Studies. Annual Review of Statistics and Its Application, 1(1), pp.427–445.

Caporaso, J.G. et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America, 108 Suppl., pp.4516–4522.

Caporaso, J.G. et al., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7, pp.335–336.

Caporaso, J.G. et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal, pp.1–4.

Claesson, M.J. et al., 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic acids research, 38(22), p.e200.

Cohan, F.M., 2002. What are bacterial species? Annual review of microbiology, 56, pp.457–87.

Dewi Puspita, I. et al., 2012. Are Uncultivated Bacteria Really Uncultivable? Microbes and Environments, 27(4), pp.356–366.

Van Dorst, J. et al., 2014. Community fingerprinting in a sequencing world. FEMS microbiology ecology, 89(2), pp.316–30.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics (Oxford, England), 26(19), pp.2460–1.

Edgar, R.C. et al., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics (Oxford, England), 27(16), pp.2194–200.

Van Elsas, J.D. & Boersma, F.G.H., 2011. A review of molecular methods to study the microbiota of soil and the mycosphere. European Journal of Soil Biology, 47(2), pp.77–87.

Epstein, S.S., 2009. Microbial awakenings. Nature, 457(February), p.2009.

Epstein, S.S., 2013. The phenomenon of microbial uncultivability. Current opinion in microbiology, 16(5), pp.636–42.

Foster, J.A. et al., 2012. Measuring the microbiome: Perspectives on advances in DNA-based techniques for exploring microbial life. Briefings in Bioinformatics, 13, pp.420–429.

Hartmann, M. & Widmer, F., 2008. Reliability for detecting composition and changes of microbial communities by T-RFLP genetic profiling. FEMS microbiology ecology, 63(2), pp.249–60.

Konopa, A., 2006. Microbial Ecology: Searching for Principles. Microbe, 1(4), pp.175–179.

Li, J.-W. et al., 2012. SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics (Oxford, England), 28(9), pp.1272–3.

Liu, W. et al., 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Applied and environmental microbiology, 63.

Loman, N.J. et al., 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnology, 30(5), pp.434–439.

Madsen, E.L., 2008. Environmental Microbiology: From Genomes to Biogeochemistry, Blackwell Publishing.

Mardis, E.R., 2008a. Next-generation DNA sequencing methods. Annual review of genomics and human genetics, 9, pp.387–402.

Mardis, E.R., 2008b. The impact of next-generation sequencing technology on genetics. Trends in genetics : TIG, 24(3), pp.133–41.

Matheson, C.D. et al., 2010. Assessing PCR Inhibition from Humic Substances. The Open Enzyme Inhibition Journal, 3(1), pp.38–45.

Mayr, E., 1942. Systematics and the origin of species, from the viewpoint of a zoologist, Harvard University Press.

McMurdie, P.J. & Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one, 8(4), p.e61217.

Metzker, M.L., 2010. Sequencing technologies - the next generation. Nature reviews. Genetics, 11(1), pp.31–46.

Moran, N. a. & Wernegreen, J.J., 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends in Ecology & Evolution, 15(8), pp.321–326.

Moya, A. et al., 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nature reviews. Genetics, 9(3), pp.218–29.

Mullis, K.B., Ferré, F. & Gibbs, R.A., 1994. The polymerase chain reaction, Birkhauser Boston Inc.

Murray, R.G.E. & Stackebrandt, E., 1995. Taxonomic Note: Implementation of the Provisional Status Candidatus for Incompletely Described Procaryotes. International Journal of Systematic Bacteriology, 45(1), pp.186–187.

Muyzer, G. & Stams, A.J.M., 2008. The ecology and biotechnology of sulphate-reducing bacteria. Nature reviews. Microbiology, 6(6), pp.441–54.

Muyzer, G., Waal, E.C.D.E. & Uitierlinden, A.G., 1993. Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S rRNA. , 59(3), pp.695–700.

Nikolaki, S. & Tsiamis, G., 2013. Microbial diversity in the era of omic technologies. BioMed Research International, 2013.

Nowack, E.C.M. & Melkonian, M., 2010. Endosymbiotic associations within protists. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 365(1541), pp.699–712.

Oksanen, A.J. et al., 2008. The vegan Package.

Osborn, A.M., Moore, E.R.B. & Timmis, K.N., 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environmental Microbiology, 2(1), pp.39–50.

Pallen, M.J., 2013. Reply to Updating benchtop sequencing performance comparison. Nature biotechnology, 31(4), p.296.

Pérez-Jiménez, J.R. & Kerkhof, L.J., 2005. Phylogeography of Sulfate-Reducing Bacteria among Disturbed Sediments, Disclosed by Analysis of the Dissimilatory Sulfite Reductase Genes (dsrAB). Applied and environmental microbiology, 71(2), pp.1004–1011.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing.

Ragan-Kelley, B. et al., 2013. Collaborative cloud-enabled tools allow rapid, reproducible biological insights. The ISME journal, 7(3), pp.461–4.

Ramette, A., 2007. Multivariate analyses in microbial ecology. FEMS microbiology ecology, 62(2), pp.142–60.

Rappé, M.S. & Giovannoni, S.J., 2003. The uncultured microbial majority. Annual review of microbiology, 57, pp.369–94.

Rees, G.N. et al., 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms : application of multivariate statistics. , pp.339–347.

Reysenbach, A. et al., 1992. Differential Amplification of rRNA Genes by Polymerase Chain Reaction. , 58(10), pp.3417–3418.

Rosselló-Mora, R. & Amann, R., 2001. The species concept for prokaryotes. FEMS microbiology reviews, 25(1), pp.39–67.

Santillano, D., Boetius, A. & Ramette, A., 2010. Improved dsrA-based terminal restriction fragment length polymorphism analysis of sulfate-reducing bacteria. Applied and environmental microbiology, 76(15), pp.5308–11.

Schütte, U.M.E. et al., 2008. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Applied Microbiology, pp.365–380.

Stahl, D.A. et al., 2002. Origins and diversification of sulfate-respiring microorganisms. , pp.189–195.

Staley, J.T. & Konopka, A., 1985. Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. Annual Review of Microbiology, 39(1), pp.321–346.

Sutlovic, D. et al., 2008. Interaction of humic acids with human DNA: proposed mechanisms and kinetics. Electrophoresis, 29(7), pp.1467–72.

Torsvik, V. et al., 1990. High diversity in DNA of soil bacteria. Applied and environmental microbiology, 56(3), pp.782–7.

Vandamme, P. et al., 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiological Reviews, 60(2).

Vasileiadis, S. et al., 2012. Soil bacterial diversity screening using single 16S rRNA gene V regions coupled with multi-million read generating sequencing technologies. PloS one, 7(8), p.e42671.

Whiteley, A.S. et al., 2012. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. Journal of microbiological methods, 91(1), pp.80–88.

Whitman, W.B., Coleman, D.C. & Wiebe, W.J., 1998. Perspective Prokaryotes : The unseen majority. Proceedings of the National Academy of Sciences of the United States of America, 95(June), pp.6578–6583.

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics, 3, pp.180–185.

Woese, C.R., 1987. Bacterial evolution. Microbiological Reviews, 51(2), pp.221–271.

Yu, Z. et al., 2008. Evaluations of different hypervariable regions of archaeal 16S rRNA genes in profiling of methanogens by Archaea-specific PCR and denaturing gradient gel electrophoresis. Applied and environmental microbiology, 74(3), pp.889–93.

Yu, Z. & Morrison, M., 2004. Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. Applied and environmental microbiology, 70(8), pp.4800–6.

Zhang, J. et al., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics (Oxford, England), 30(5), pp.614–20.

Chapter 4

The DEMETRA project

4 The DEMETRA project

The first commercialization of a genetically modified organisms (GMO) crop dates back to 1996; since then the number of commercialized GMOs has increased, involving nowadays 27 different crops, with most of them seeing multiple commercialized varieties (https://isaaa.org/gmapprovaldatabase/cropslist/default.asp). Worldwide, an increasing number of farmers choose to cultivate GMO as crops. The total amount of hectares cultivated with GMO crops exceeds 170 million (James 2014), involving nearly every country in North and South America, Australia, and a large number of counties in South Asia. Europe has completed the legislative framework for the use of GMO crops in 2003, but the number of countries which approved and implied their cultivation is still low (Spain, Portugal, Czech Republic, Romania, and Slovakia).

The DEMETRA project (Development of a quick Monitoring index as a tool to assess Environmental impacts of TRAnsgenic crops) was funded under the LIFE+ initiative of the European Community. Specific objective of the project was the development of a new instrument for monitoring the environmental impact of transgenic crops. In particular, the project aimed at the development of a quick monitoring index (QMI) that, integrated with a Geographic Information System (GIS), could assist those Public Authorities (Regions in the case of Italy) in deciding where and when implementing monitoring efforts, in the case that GMO crops would be released on their territory. DEMENTRA project did not employed GMO crops, and relied uniquely on bibliographic information to carry out the risk assessment and to construct the QMI. To do so, a thoroughly study of the biodiversity of the study site is of fundamental importance. Biodiversity was characterized at various levels: plant and macroinvertebrates biodiversity, and finally soil microbial biodiversity. During my PhD, our group carried out the characterization of soil microbial diversity; results are reported in chapter 7 and 8.

The diversity of soil inhabiting microorganisms is enormous; one gram of soil can contain between 10^9 and 10^{11} bacterial cells (Torsvik et al. 2002; Watt et al. 2006) and up to 200 m of fungal hyphae (Leake et al. 2004). However, from the microbes' point of view, not all the soil is equal. Most of this diversity, in fact, is found in a specific environment of soils, where the interactions between microbes and plant root can takes place, the so called rizosphere. Up to more than 30.000 bacterial and archaeal operational taxonomic units (OTUs) were detected in the rhizosphere microbiome (Mendes et al. 2011). Plant roots continuously modify the soil environment by rhizodeposition, a process that, by releasing a wide range of low and high-molecular

weight compounds (Hinsinger et al. 2005), creates selective pressure on the local microbial communities (Nunes da Rocha et al. 2009; Koranda et al. 2011; Lambers et al. 2009; Paterson et al. 2007; Pinton et al. 2007) and shape the composition and activity of microbial populations (Nunes da Rocha et al. 2009). Vegetation exert direct interactions on soil microbial communities through root exudates and, vice-versa, soil microbial community exert direct interaction on vegetation, ad example promoting plant growth.

It is important to recognize possible modifications that human activities can determine on soil microbial communities in natural environments. A loss of biodiversity, a modification in the interaction between microbes, or between microbes and plants, have to be regarded as negative effects, both for issues of nature conservation and to put into practice an high yielding and balanced agriculture. Various soil properties have been proven to significantly impact the structure of microbial communities, such as soil type (Bach et al. 2010), salinity (Lozupone & Knight 2007) and pH (Fierer & Jackson 2006). It is clear that human activities can deeply influence microbial communities in soil habitat, as changes in any of these factors can influence physical and metabolic niche diversity, thereby resulting in different composition of microbial communities (Jangid et al. 2011). Land-use, especially cultivation, has a strong and lasting impact on soil communities, and can represent a stronger determinant of the composition of microbial communities than aboveground vegetation and soil properties. (Jangid et al. 2011). Moreover, agricultural systems are often characterized by external inputs, applied with the ultimate goal of maximizing productivity and economic returns, and that can have various effects of microbial populations in soil. Those inputs include mineral fertilisers such as urea, ammonium nitrate, sulfates, and phosphates; organic fertilisers such as animal manures, composts, and biosolids; various other organic products such as humic acids and microbial inoculants, and pesticides including herbicides, insecticides, nematicides, fungicides, veterinary health products, and soil fumigants (Bünemann et al. 2006).

In this respect, of particular relevance is the forecasted increase in genetically modified (GM) crops over the last decades (Carpenter 2011). Concerns have been raised that some traits carried by GMP may negatively affect the environment and the beneficial soil biota, potentially leading to alterations in soil functioning. Among these environmental concerns, the unintentional impact that GMPs might have on soil-associated microbes, especially rizosphere-inhabiting bacteria, is still a poorly studied and understood area (Filion 2008; Kowalchuk et al. 2003). Insecticidal proteins from transgenic crops expressing the insecticidal *Bacillus thuringiensis* protein (Bt) are exemplary of the possible impact of GMPs on soil microorganisms. Bt proteins released by the root exudates of Bt-maize and by Bt-plant biomass during and after

harvest (Saxena & Stotzky 2001) and represent a potential exposure hazard to non-target organisms (de Vaufleury et al. 2007). Further examples are toxic agents as antibiotics, herbicides or drugs produced by selectable marker genes introduced in transgenic plants together with the transgene. Although data on their effects on microbial diversity are limited, these agents could exhibit activity against a broad range of bacteria and fungi and could potentially affect natural plant-associated bacterial and fungal populations. GMP transgenic products can directly or indirectly affect unintentional target organisms, but other risks are associated also to the use of GMP: (1) the mobilization of the transgene in the environment by pollen-mediated gene flow, from one plant into another (Husken et al. 2010); and (2) the horizontal gene transfer (HGT), between plant and other organisms; microorganisms are the most likely final hosts of a HGT event but also they function as intermediate hosts in the further transfer of the transgenic gene to other (prokaryotic and eukaryotic) organisms (Brigulla & Wackernagel 2010). Overall, pollen mediated gene flow and HGT are able to move in time and space the eventual toxic effects of transgenic products on natural biodiversity.

For the above reported reasons, environmental risk assessment of GM crops remains the subject of many studies (Sparrow 2010). Any evaluation of the environmental risk by GM crops must take account, among other things, of the actual biodiversity at the site where GMPs should be cultivated, before cultivation starts.

4.1 Description of the study site

The study was conducted inside the Migliarino - San Rossore – Massacciuccoli regional park; located along the coast of the Tyrrhenian Sea in the Tuscany region. The park is characterized by high ecosystems diversity comprising coastal area, such as dune ecosystem and wetlands, and forest areas adjacent to lands dedicated to cultivation. The presence of protected areas, with minimal human intervention, along with cultivated areas, that potentially could be grown with GMO crops, in close proximity is of particular interest for the objective of DEMETRA project. The project was conducted in three main areas: Massacciuccoli lake area, Arno river area, and Serchio river area. For the purpose of this thesis only Serchio river and Arno river areas, were the microbial diversity assessment was carried out, will be described. Figure 4.1 shows their geographic position and, inside them, the position of the trees and sites interested by this work.

Serchio river study area: the area in the vicinity of Serchio river, is located in the central part of the park. In this area two study sites have been selected. The first (Site 1) was a flat, easy accessible, natural poplar forest large approx. 7400 m^2. The forest was of mixed type, characterized by the presence of White poplar (*Populus alba L*.) and Grey poplar (*Populus x canescens* ((Aiton) Sm.)), plus other tree species. In this site, fifty poplars were georeferenced and the soils associated to seven of these were subject of the microbial diversity assessment. The second (Site 2) was located in close proximity to the natural poplar forest (Site 1) towards the north, along the bank of Serchio river. It was a mature poplar plantation (hybrid Triplo clone *Populus nigra* x *Populus deltoids*) turned from a natural mixed forest about 30 years ago, between 1978 and 1982. In this site four poplars were georeferenced and the soils associated with them were subject of the microbial diversity assessment.

Arno river study area: the area in the vicinity of Arno river is located in the southern part of the park. In this area, a natural broadleaf forest of approx. 118.000 m^2 was chosen as study site (Site 3). The forest was characterized by mixed vegetation, widely varying in dimension, and with a considerable amount of elm leaf blackberries and common hawthorn constituting its undergrowth. In this site, 267 individuals of maple (*Acer campestre)* were georeferenced and the soils associated to eleven of these were subject of the microbial diversity assessment.

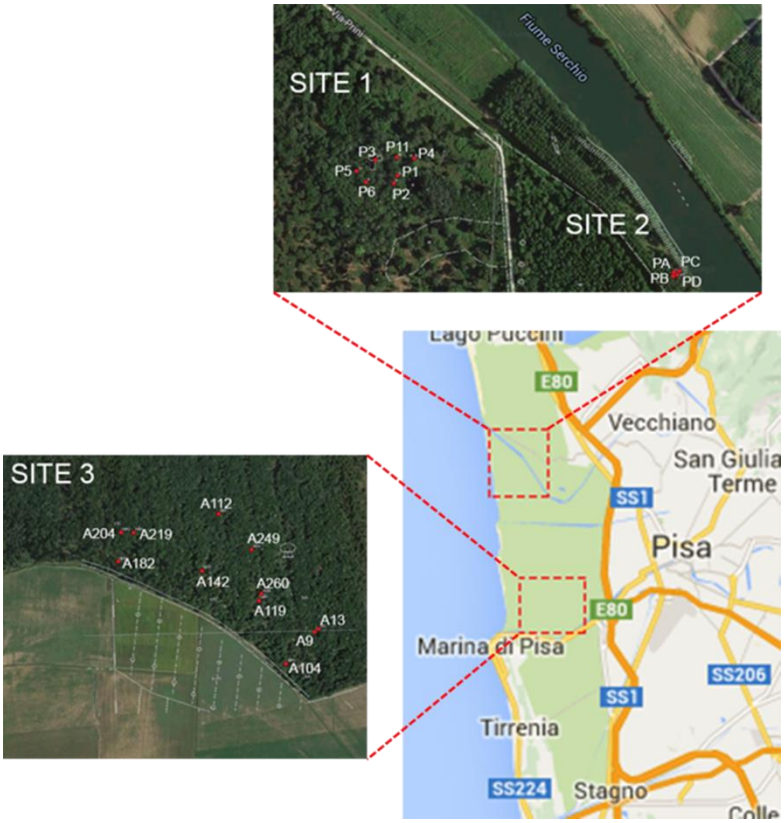


Figure 4.1: Location of the sampling site inside the Migliarino-San Rossore–Massacciuccoli regional park, and location of each of the georeferenced trees considered in this thesis

4.2 Bibliography

Bach, E.M. et al., 2010. Soil texture affects soil microbial and structural recovery during grassland restoration. Soil Biology and Biochemistry, 42, pp.2182–2191.

Brigulla, M. & Wackernagel, W., 2010. Molecular aspects of gene transfer and foreign DNA acquisition in prokaryotes with regard to safety issues. Applied Microbiology and Biotechnology, 86, pp.1027–1041.

Bünemann, E.K., Schwenke, G.D. & Van Zwieten, L., 2006. Impact of agricultural inputs on soil organisms—a review. Australian Journal of Soil Research, 44(4), p.379.

Carpenter, J.E., 2011. Impact of GM crops on biodiversity. GM crops, 2(1), pp.7–23.

Fierer, N. & Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America, 103(3), pp.626–31.

Filion, M., 2008. Do transgenic plants affect rhizobacteria populations? Microbial biotechnology, 1(6), pp.463–75.

Hinsinger, P. et al., 2005. Rhizosphere geometry and heterogeneity arising from root-mediated physical and chemical processes. The New phytologist, 168(2), pp.293–303.

Husken, A., Prescher, S. & Schiemann, J., 2010. Evaluating biological containment strategies for pollen-mediated gene flow. Environ. Biosafety Res., 9, pp.67–73.

James, C., 2014. EXECUTIVE SUMMARY BRIEF 46 Global Status of Commercialized Biotech / GM Crops : 2013. , (46).

Jangid, K. et al., 2011. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. Soil Biology and Biochemistry, 43(10), pp.2184–2193.

Koranda, M. et al., 2011. Microbial processes and community composition in the rhizosphere of European beech - The influence of plant C exudates. Soil biology & biochemistry, 43(3), pp.551–558.

Kowalchuk, G. a., Bruinsma, M. & van Veen, J. a., 2003. Assessing responses of soil microorganisms to GM plants. Trends in Ecology & Evolution, 18(8), pp.403–410.

Lambers, H. et al., 2009. Plant-microbe-soil interactions in the rhizosphere: an evolutionary perspective. Plant and Soil, 321(1-2), pp.83–115.

Leake, J. et al., 2004. Networks of power and influence: the role of mycorrhizal mycelium in controlling plant communities and agroecosystem functioning. Canadian Journal of Botany, 82, pp.1016–1045.

Li, J.-W. et al., 2012. SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics (Oxford, England), 28(9), pp.1272–3.

Lozupone, C.A. & Knight, R., 2007. Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences of the United States of America, 104, pp.11436–11440.

Mendes, R. et al., 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science (New York, N.Y.), 332, pp.1097–1100.

Paterson, E. et al., 2007. Rhizodeposition shapes rhizosphere microbial community structure in organic soil. The New phytologist, 173(3), pp.600–10.

Pinton, R., Varanini, Z. & Nannipieri, P., 2007. The rhizosphere: biochemistry and organic substances at the soil-plant interface, CRC press.

Da Rocha, U.N., van Overbeek, L. & van Elsas, J.D., 2009. Exploration of hitherto-uncultured bacteria from the rhizosphere. FEMS microbiology ecology, 69(3), pp.313–28.

Saxena, D. & Stotzky, G., 2001. Bacillus thuringiensis (Bt) toxin released from root exudates and biomass of Bt corn has no apparent effect on earthworms, nematodes, protozoa, bacteria, and fungi in soil. Soil Biology and Biochemistry, 33(9), pp.1225–1230.

Sparrow, P.A.C., 2010. GM risk assessment. Molecular Biotechnology, 44, pp.267–275.

Torsvik, V. & Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology, 5(3), pp.240–245.

De Vaufleury, A. et al., 2007. Exposure and effects assessments of Bt-maize on non-target organisms (gastropods, microarthropods, mycorrhizal fungi) in microcosms. Pedobiologia, 51(3), pp.185–194.

Watt, M. et al., 2006. Numbers and locations of native bacteria on field-grown wheat roots quantified by fluorescence in situ hybridization (FISH). Environmental microbiology, 8, pp.871–884.

Chapter 5

The MAPMED project

5 The MAPMED project

The Mediterranean has become one of the most demanded destinations for organized touristic routes (increasing cruising market) and "self-made" ones (boating and chartering). Despite the strict link between tourism development in coastal areas and port facilities, the sector growth has not been sustained neither by an adequate expansion nor by an adaptation of the ports. Ports have a decisive role in the development of coastal areas and the risk of impact of infrastructures construction and maritime traffic on the coastal zone is high. Balancing between environmental and social-economic issues (sustainable development) is not an easy task. In fact, if on one hand the development of infrastructures has to take into account environmental impact and restoration to ecological standards, on the other hand, denial of port expansion for environmental reasons may favour other competing ports. Due to their strategic location between sea and land, ports are particularly critical environments as they can receive pollution coming from land, ships and from the port facilities themselves. Furthermore, ports are not closed systems and their pollution may impact large parts of the adjacent coastal areas.

The overall objective of MAPMED project (Management of Port areas in the Mediterranean Sea Basin) is to improve the environmental sustainability of tourist coastal areas in the Countries of the Mediterranean Sea Basin through the promotion of a long term cooperation between Institutional Authorities and the scientific Community and, at a more specific level, to optimize, validate and transfer tools to guide Institutional Authorities in the sustainable management of tourist ports/harbours with regard to monitoring and reduction of marine pollution.

Due to their high growth rates, prokaryotes exhibit a fast response to environmental changes and perturbations. In the last decades, the characterization of prokaryotic communities has been recognised as a valuable tool for assessing anthropogenic impact, including exposure to hydrocarbon pollutants, on marine ecosystems (Paerl et al. 2003). Bacteria are an important component of biota in marine environments where they play a fundamental role in element cycling and functioning of the ecosystems. Moreover, many microorganisms evolved the ability to use hydrocarbons as carbon and energy sources for their growth. Hydrocarbon degraders and particularly the obligate hydrocarbonoclastic bacteria carry out a fundamental and global activity in biological removal of petroleum hydrocarbons and natural cleaning processes in marine habitats (Yakimov et al. 2007). Moreover, distinct bacterial

populations specifically degrade structural different classes of petroleum hydrocarbons (Head et al. 2006).

In contrast, and despite their abundance, diversity and significant role in marine ecosystems, archaeal community has been less investigated. Archaea domain has been demonstrated to have a fundamental role in key steps of the element cycles under oxic and anoxic conditions. However, the response of archaeal communities to hydrocarbon pollution has been less extensively examined as compared to the bacterial components. In oxic environments, Archaea seems to have a limited role in hydrocarbon degradation showing a high sensitivity to oil exposure (Head et al. 2006). However, a contribution in PAH degradation of methanogenic community has been recently suggested in anoxic sediments (Kim et al. 2008).

5.1 Description of the study sites

The case study sites of MAPMED project were three tourist ports located across the Mediterranean Sea Basin: Cagliari (Sardinia, Italy) in the western part, El Kantaoui (Tunisia) in the central part, and Heraklion (Crete, Greece) in the eastern part. Beside their geographic position, port selection was also based on other different elements, such as categories of maritime traffic, port dimension and existing information on pollution. The maritime traffic inside the three tourist ports is represented by recreational boats, passenger ships and fishing vessels. In addition to the marine traffic, the major pollution sources related to the three port areas are wastewater discharges into the sea, river mouth, fuelling stations, and fishing activities.

Figure 5.1: Position of the three selected touristic ports in the Mediterranean Sea. C: Cagliari (Sardinia, Italy), H: Heraklion (Crete, Greece), E: El Kantaoui (Tunisia).

For each port, sampling stations were chosen according to their characteristic position within the port area and their related use (i.e. touristic, industrial, cargo area) (Table 5.1).

Table 5.1: Sampling stations in the three port areas.

Station	Cagliari (C)	El Kantaoui (E)	Heraklion (H)
1	Leisure boats	Leisure boats	Leisure boats
2	Intermediate	Fuel station	-
3	Passenger ships	Port entrance	Passenger ships
4	Cargo ships	-	Cargo ships
5	Port entrance	-	Shipyard

The Port of Cagliari (Figure 5.2 a) is located in the Island of Sardinia (Italy). Founded by the Phoenicians, then influenced by the Carthaginians and the Romans, it has been a fundamental commercial and strategic crossroad for more than 2000 years, thanks to its central position in the Mediterranean Sea. Actually, it is extended up to 5.800 meters of quay and it is used for the commercial traffic, Roll on/Roll off (RO/RO), passenger ships, yachting, fishing and cruising; the last one especially, thanks to the collaboration with local energies, is developing very quickly.

The Port of El Kantaoui (Tunisia) (Figure 5.2 b) was built in 1979 specifically as a tourist centre around a large artificial harbour extending over an area of 4 ha, which provides mooring with 550 berths for luxury yachts. Giving its purely touristic nature, it hosts a variety of activities, like sporting activities and several golf courses.

The Port of Heraklion (Greece) (Figure 5.2 c) is the main and most modern gate of entrance for passengers and cargo to the island of Crete. It has an exceptionally strategic position, since it is located in the centre of the south-eastern Mediterranean Basin, interconnecting three continents. The passenger port of Heraklion is the third one in passenger traffic within Greece and serves about 2 million passengers every year and more than 300.000 vehicles. The Port of Heraklion also hosts a cargo zone, which also includes the transportation of fuels and cement.

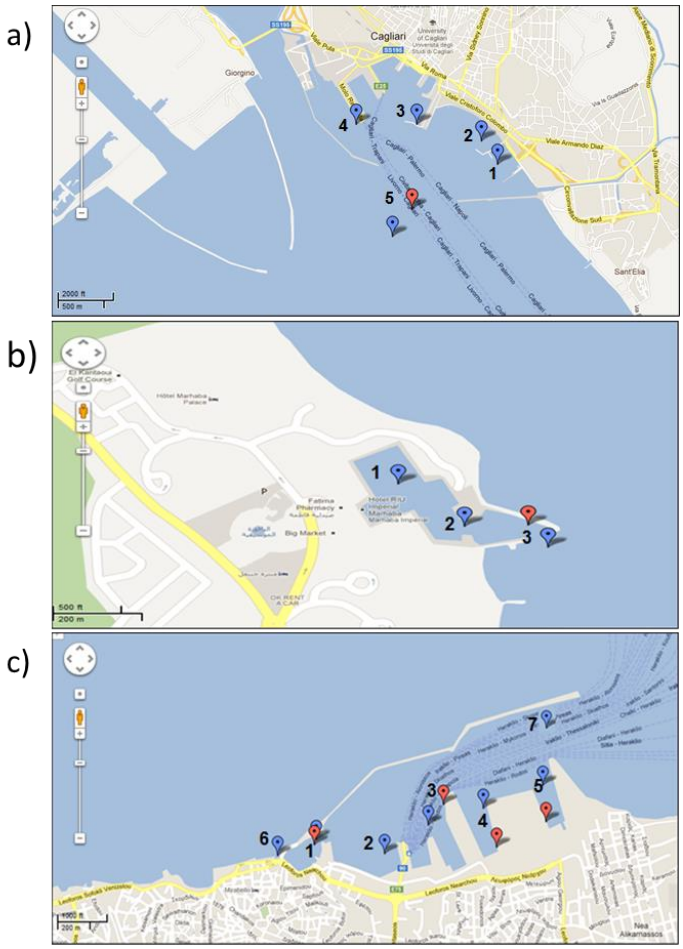


Figure 5.2: Position of sampling stations in the port area. The blue marks indicate the position where water and sediment samples were collected. In cases where the collection point of sediment samples was slightly different than the collection point of water samples an additional red mark is used.

5.2 Bibliography

Head, I.M., Jones, D.M. & Röling, W.F.M., 2006. Marine microorganisms make a meal of oil. Nature reviews. Microbiology, 4(3), pp.173–82.

Kim, M. et al., 2008. Monitoring nutrient impact on bacterial community composition during bioremediation of anoxic PAH-contaminated sediment. Journal of Microbiology, 46, pp.615–623.

Li, J.-W. et al., 2012. SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics (Oxford, England), 28(9), pp.1272–3.

Mardis, E.R., 2008. Next-generation DNA sequencing methods. Annual review of genomics and human genetics, 9, pp.387–402.

Metzker, M.L., 2010. Sequencing technologies - the next generation. Nature reviews. Genetics, 11(1), pp.31–46.

Paerl, H.W. et al., 2003. Microbial indicators of aquatic ecosystem change: current applications to eutrophication studies. FEMS microbiology ecology, 46(3), pp.233–46.

Yakimov, M.M., Timmis, K.N. & Golyshin, P.N., 2007. Obligate oil-degrading marine bacteria. Current opinion in biotechnology, 18(3), pp.257–66.

Chapter 6

Conversion from natural forest to poplar plantation alters the structure and causes a decrease in the diversity of soil bacterial and fungal communities in a Mediterranean natural park.

6 Conversion from natural forest to poplar plantation alters the structure and causes a decrease in the diversity of soil bacterial and fungal communities in a Mediterranean natural park.

Francesco Vitali, Giorgio Mastromei, Giuliana Senatore, Cesarea Caroppo, Enrico Casalone*

Department of Biology, University of Florence,

Via Madonna del Piano 6, 50019, Sesto Fiorentino, Florence, Italy.

* Corresponding author:

Enrico Casalone, enrico.casalone@unifi.it, +39 055 4574735

Submitted to: Science of the Total Environment

6.1 Abstract

This study wanted to evaluate the long-lasting effects of a change within a single land-use category, specifically the conversion from natural forest to forest plantation, on soil microbial communities. With this aim, we compared microbial diversity in soils from two natural forests and a forest converted to poplar plantation about thirty years ago. To minimize climatic influences and human impacts other than land-use, we chose a local-scale study area within a Natural Park. Soil bacterial and fungal communities were investigated by T-RFLP analysis of the 16S-rRNA and the ITS-rDNA regions, respectively, and their structure was related to the soil characteristics. The results showed that bacterial and fungal communities from soils of the cultivated poplars had different structure and lower α-diversity values compared to the natural forest sites. Fungi appeared more sensitive than bacteria as indicator of the perturbative effects of forest conversion. β-diversity values showed that microbial communities from the two natural forests were much more similar to each other than they are to communities from the poplar plantation. Even if a differentiation of microbial communities on a seasonal base was not clear, our data suggested that land-use influenced the seasonal fluctuations of fungal and bacterial α-diversity in a different way. pH and organic matter content are confirmed as important drivers of soil microbial diversity. Differences between soils related to their association with different tree type, may also have contributed to shape the community microbial structure in the three sampling sites.

Keywords: T-RFLP, forest soil, land-use, microbial community, diversity

6.2 Introduction

Soil microbial communities are a key component of the forest ecosystem; they are involved in fundamental processes, like decomposition and nutrient cycling, and perform a link role between plants and ecosystem functions [1–3]. With 200 million ha (10 million of which in Italy), forests represent the major natural vegetation cover in West Europe (31.5% of the land area, 5% of the world's forests); a quarter of these forests are of the mixed type. Despite a positive reforestation trend, risks for European forest health and vitality seem on the increase, mainly due to anthropic impact [4]. At risk is a wealth of soil microbial diversity composed of bacteria [5, 6] and fungi [7, 8], and the enormous and still untapped pool of biological resources they constitute, especially in hot spot of plant diversity, such as the Mediterranean area [9].

Changes in land-use have to be regarded as potentially affecting the diversity of natural habitats [10], even if their effects on biodiversity are still poorly understood [11]. Despite the importance of the microbial community for the functioning of ecosystem, very few studies have investigated the effects of forest management and forest conversion on microbial diversity [12]. Therefore, studies of microbial communities diversity in response to land-use changes, from a preserved natural environment to an agriculture management, are fundamental for issues of nature conservation and biodiversity preservation. Deforestation of Amazon rainforest [13, 14] or of pristine forests in the Pampa biome [15], to obtain pasture sites, has been reported to alter microbial diversity and community structure of soil microorganisms. Very recently, two studies reported that changes in land-use from natural forest to tree plantation, in the Pampa biome [16] and in southeast Asian tropical forest [17], can alter the below-ground ecosystem, and ultimately affect the microbial communities resident in the soil. The influence of the tree type on microbial community structure and function was supported by a number of different reports [9, 18–23] In addition to land-use and plant type, several abiotic factors, among which soil type seems one of the most important, influence structural and functional diversity of microbial communities and plant-soil microbial community relationships [9, 18, 20–22, 24, 25]). The differential and quantitative contribution that land-use, on one hand, and soil type and plant cover, on the other hand, have on microbial structure and diversity might not be so easy to evaluate [20].

In turn, belowground-living microorganisms have been demonstrated to influence, directly or indirectly, the productivity, diversity and composition of plant communities [1, 25]. Soil microorganisms can also influence or select each other. The finding that ectomycorrhizal fungi favor specific bacterial communities with contrasting functional

characteristics from the surrounding soil [26] stresses the importance to study both bacteria and fungi in the soil, to get a more complete understanding of the dynamics and interactions of microbial communities in natural environments.

Although soil microbial communities have been investigated mainly for changes across space, their temporal variability has received far less attention [27]. Against a large number of temporal variability studies in microbially driven biogeochemical processes, those looking to temporal variability of composition and/or function of microbial communities in forest soil are scarce; and long-term seasonal dynamics studies of these communities are even rarer. Of further interest is the question whether observed temporal patterns in soil microbial communities may vary across different land-use [27].

The present study investigated, within the Migliarino–San Rossore–Massaciuccoli Regional Park (Tuscany, Italy), the differences in bacterial and fungal community structures in soils from two natural plain forests and an adjacent poplar plantation, along with seasonal variations. The aim of the study is to gain more insight on the effects that the conversion from natural forest to tree cultivation have on soil inhabiting microbial communities. In order to minimize the impact that climatic differences and human interventions, other than the land-use, may have on soil microbial communities, we have chosen as a study area a local landscape, confined within a natural park. T-RFLP analysis, a cost-effective and rapid method, widely used for studying microbial community structure and dynamics [28] and that has been recently re-evaluated for the estimation of microbial community diversity [29], was used to obtain fingerprinting data on microbial communities structure and to evaluate α-diversity and β-diversity for comparison purpose. The relationships that the microbial community structure had with the physicochemical properties of the soils and with seasonality have been analyzed with the aim to recognize the contribution that these factors have on forest microbial diversity beyond land-use.

6.3 Materials and methods

Site description and soil sampling

The study area is located inside the Migliarino–San Rossore–Massaciuccoli Regional Park (latitude 43°35'-43°51', longitude 10◦15'-10°22', approximately; mean altitude 4 m), which ranges along the Tyrrhenian Sea between the cities of Viareggio and Livorno (Tuscany, Italy), and belongs to the Mediterranean climate type. The study was conducted in three different field sites within the Park (Table 1). Sites 1 (an

approximately 7,000 m^2 large meso-hygrophilic/hygrophilic deciduous forest) and site 3 (an approximately 118,000 m^2 meso-hygrophilic deciduous forest) were natural mixed-deciduous forests older than 100 years with minimum intervention management; they are 8,300 m apart each other, in the north and south of the Park, respectively. Site 2 (a 28,000 m^2 planted forest) was a mature 15 years old (in 2010) poplar plantation; it is 325 m apart from site 1 toward north. Until mid 1970's forests in site 1 and 3 were subjected to controlled logging, after that they were left undisturbed. Site 2 was originally a natural mixed-deciduous forest, converted to poplar plantation; aerial photos of this site (including also site 1) placed the first establishment of a poplar plantation between 1978 and 1982 (Figure 1).

Table 1: Characteristics of the sampling sites.

Site	Area (m^2)	Forest management	Mean distances between area (m)	Mean distances between trees (m)	Sampled trees	N° of samples	Sampling density[b]
1	7,000	Natural	1-2 (325)	51	PN	7	23.3 %
2	27,745	Cultivated	2-3 (8250)	7	PC	4	n.d.
3	118,000	Natural	3-1 (8310)	217	A	11	4.12 %

[a] PN, *Populus alba*; PC*, Populus nigra* x *Populus deltoids*; A, *Acer campestris*

[b] percentage of sampled trees on the total number of tree from the same species in that site

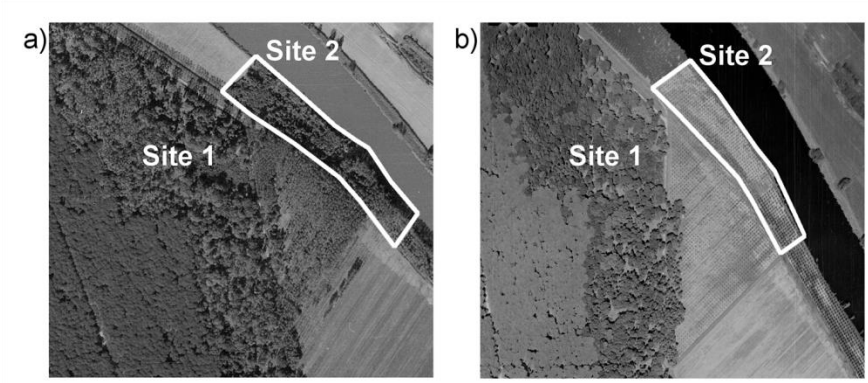


Figure 1: Aerial photography of Site 2 in a) 1978 and b) 1982, executed by Rossi-Brescia: year 1978, folder c0117, swipe 29, frame 412; and year 1982, folder c0233, swipe 3a, frame 335, respectively. Images Courtesy of the General Cartographic archive of the Tuscany Region.

Plants in the three sites were identified and some of them have been georeferenced with the aim of tracing them for future sampling of tree-associated soils. From Autumn 2010 to Summer 2012, individual soil samples were collected from a number of georeferenced trees, using a bulb planter (10 cm wide x 15 cm depth), at about 20 cm from the trunk at different points around it at different seasons. A total of 156 soil samples were collected from soils associated with seven natural poplars (*Populus alba*: P1, P2, P3, P4, P6; *Populus canescens* P5, P11) in site 1, four cultivated poplars (hybrid Triplo clone *Populus nigra* x *Populus deltoids*: PA, PB, PC, PD) in site 2, and eleven maples (*A. campestre*: A9, A13, A104, A112, A119, A142, A182, A204, A249, A260) in site 3. Individual soil samples were placed separately in sterile plastic bags and immediately stored at 5°C; the same day the samples were brought to the lab. Each individual sample was sequentially sieved through 5 mm and 2 mm pores size stainless steel sieves. The sieved soil from each sample was split into four aliquots that were stored at 4 °C, for total microbial counts, moisture content, pH determination and loss on ignition measure; a further aliquot was stored at -80 °C for molecular analysis.

Rain and air temperature values, measured by two meteorological stations placed as part of the LIFE08 NAT/IT/00342 – DEMETRA project, have been averaged over a 7 days period before the sampling dates.

Analysis of soil chemistry

Sampling sites were classified as follow: site 1, Humic Eutrudepts, coarse sand, mixed, thermic; site 2, sandy loam, mixed, thermic; site 3, Fragic Hapludalfs, fine, mixed, thermic. Data for site 1 and 3 were from map units dated 2002, site 2 was not in these maps and was defined at the time of the sampling campaign.

For physicochemical analysis, all the soil samples from each site in each season were pooled to make a single composite sample; a total of 21 composite soil samples were analysed. Gravimetric water content, organic matter (OM) content and pH were determined on composite soil samples. Gravimetric water content was determined as weight loss after over-drying the freshly sieved soil at 105°C for 24 h. Weight percentage OM was determined on over-dried soils by the loss on ignition (LOI) procedure in a muffle furnace at 550 °C for 24 h [30]. pH was measured on sieved air-dried soil samples mixed to deionized water at ratio of 1:2.5 (w/v); the mixture was shaken to form a slurry and left undisturbed for 15 min prior to withdraw 70 µL of supernatant to be analysed by a microelectrode (Ross® pH microelectrode, Thermo Scientific; Beverly, MA, USA).

Viable counts

To determinate the number of viable bacteria and fungi in the soil, five grams of each individual sieved soil sample were placed in a separate sterile plastic bag with 50 mL of saline solution and processed by Stomacher® 400 Circulator (Seward, UK) for 3 minutes at 260 rpm to ensure the detachment of microorganisms from soil particles. After 15 min sedimentation, the suspension-supernatant from each soil was serially diluted and plated in triplicate (0.1 ml) on Soil Agar with cycloheximide (1 µg/ml) and on Rose Bengal Agar with chloramphenicol (0.1 µg/ml), for selective growth of bacteria and fungi, respectively. Plates were incubated for three days at 37°C and 30°C, respectively; only plates containing between 30 and 300 colonies were taken in consideration to calculate viability as Colony Forming Units (CFU/g^{-1} of dry soil).

Soil DNA extraction and purification

Total DNA was extracted from 250 mg aliquots of each of the 156 individual soil samples using NucleoSpin Soil kit (Macherey Nagel, Düren, Germany), with Lysis Buffer SL2 and 150 µL of Enhancer SX,. Extracted DNAs were further purified by Power Clean® DNA Clean-Up Kit (Macherey Nagel, Düren, Germany).

PCR amplification of genomic DNA from soil

Total DNA extracted from each of the 156 individual soil samples was used for PCR amplification of the bacterial 16S rDNA gene and the ITS1 and ITS2 internally transcribed spacers (ITS region) of fungi. PCR reactions were performed with 10 ng of template DNA in a final reaction volume of 25 µL with Dream Taq Buffer (containing 20 mM $MgCl_2$ - Thermo Fisher Scientific GmbH, Karlsruhe, Germany), 0.2 mM of each dNTP (EuroClone, Milano, Italy), 1 µM of each primer, 1U of Dream Taq polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). Bacterial 16S rDNA genes were amplified using the fluorescently labelled 27F-FAM (5'-[6FAM]-AGAGTTTGATCCTGGCTCAG-3') and the 1525R (5'-AAGGAGGTGWTCCARCC-3') primers [31]; fungi ITS regions were amplified with the fluorescently labelled ITS1f-(5'-[6FAM]-CTTGGTCATTTAGAGGAAGTAA – 3') and ITS4r (5'-[PET]- TCCTCCGCTTATTGATATGC- 3') primers [32]. Bacterial amplification reaction consisted of an initial denaturation step at 95 °C for 5 min, 5 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 72 °C for 2 min, then 5 cycles at 95 °C for 30 sec, 55 °C for 30 sec, 72 °C for 2 min, followed by 25 cycles at 95 °C for 30 sec, 52 °C for 30 sec, 72 °C for 2 min and a final extension at 72 °C for 10 min. Fungal amplification reaction consisted of an initial denaturation step at 94 °C for 5 min; 34 cycles at 94 °C for 1 min, 52 °C for 1 min, 72 °C for 2 min and a final extension at 72 °C for 10 min. 16SrDNA (approximately 1500 bp) and ITS region (variable sizes) PCR amplicons were purified by Wizard® SV Gel and PCR Clean-Up Kit (Promega, Madison, Wisconsis, USA) from gel after electrophoresis on a 0.8% agarose gel or directly from the amplification reaction, respectively.

Genetic profiling of microbial communities

The diversity of the dominant members of bacterial and fungal domains was characterized by genetic profiling using terminal restriction fragment length polymorphism (T-RFLP). 16S rRNA PCR amplicons (0.6 – 1 µg) were digested with 20 unit of *Rsa*I restriction enzyme (Thermo Fisher Scientific, GmbH, Karlsruhe, Germany) in a final volume of 20 µl, at 37°C for 4 hours; the digestion was terminated by heating at 80 °C for 20 min. ITS region PCR amplicons (0.1 µg) were digested with 6 unit of *Hinf*I restriction enzyme (Roche, Basilea, Switzerland) in a final volume of 20 µl, at 37°C for 4 hours; the digestion was terminated by heating at 65 °C for 20 min. The length of the fluorescently labelled terminal restriction fragments (T-RFs) was determined with an Applied Biosystems® 3500 Series Genetic Analyzer automated sequencer using LIZ 500 (Applied Biosystems, Foster City, California, USA) size standard as a dimensional standard. T-RFs profiles were analysed with GeneMapper software (Applied Biosystems, Foster City, California, USA). Fixed thresholds of 50 and 100 RFU were used to remove baseline noise for blue channel (6FAM-labelled primers) and red

channel (PET-labelled primer), respectively. Alignment of T-RFs peaks was automatically performed by the software and manually checked.

Data analysis and statistical methods

All data elaborations and analysis were carried out using R Statistical software version 2.15.1 [33] with the package *Vegan* [34] *PerformanceAnalytics* [35] and ggplot [36].

For the analyses of the T-RFLP profiles of bacterial and fungal communities in the soil samples, only T-RFs with height above the fixed threshold were considered. Biodiversity analysis was performed on normalized T-RFLP profiles. Normalization was done by measuring the relative abundance of each individual T-RF in a profile and dividing its peak area by the total peak area of the profile.

The overall bacterial and fungal diversity inside each study site (α-diversity) was evaluated measuring phylotype richness (S; number of T-RFs), evenness (E; Shannon evenness index; equality of phylotype abundance in a community), and Shannon-Wiener index (H') [37] and Simpson's index [38] of diversity (D). β-diversity between sites was calculated using Bray-Curtis dissimilarity index comparing cumulative T-RFLP profiles obtained by calculating the mean area of each T-RF in the T-RFLP profiles of all the samples in each site; the cumulative profiles were then normalized (see above) and T-RFs with a relative area <0.01 were eliminated.

To estimate true phylotype richness in each site, individual T-RFLP profiles were further inspected with Chao1 index [39] and phylotypes (T-RFs) accumulation curves [40]. Ordination analysis of T-RFLP profiles took into account only qualitative information about presence/absence of T-RFs with height above the fixed height threshold. A dissimilarity matrix was computed using Sørensen index on the presence/absence matrix and used as input for ordination analysis of microbial communities with Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering and with Non-parametric Multi-Dimensional Scaling (nMDS) with 100 random starts. Environmental variables (cumulative rain and air temperature), soil chemistry variables (OM, pH, and relative humidity), and microbial abundance data ($\log_{10}$ of viable counts) were fitted onto the nMDS ordination using *envfit* function in Vegan package. The significance of fitted vectors was tested using 999 permutations. The ANalysis Of SIMilarities (ANOSIM) was used to test spatial and seasonal variability of the soil microbial communities. Spearman correlations between α-diversity (richness and Shannon's index) and pH, OM, relative humidity, cumulative rain, air temperature, and viable counts, were obtained with command "*chart.Correlation"* in the *PerformanceAnalytics* R package.

6.4 Results

Physicochemical properties and microbial abundance in soils across land-use.

Results of physicochemical characterization of composite seasonal pools of soils from the three sites were reported in Figure 2 as mean values plus standard deviation (SD). In site 1, pH was 7.24 (SD= 0.13), OM was 14.40% (SD= 1.28), and relative humidity was 26.78% (SD= 8.24); in site 2, pH was 7.70 (SD= 0.14), OM was 10.23% (SD= 1.14), and relative humidity was 13.81% (SD= 7.73); in site 3, pH was 6.97 (SD= 0.21), OM was 15.45% (SD= 0.97), and relative humidity was 21.33% (SD= 6.89). The soil from poplar plantation (site 2) differed from natural forest soils (site 1 and 3) more than they differed from each other. Soils in site 2 showed the highest pH values and the lowest organic matter content and humidity values. Differences among sites were tested with two-sample T-test (Figure 2). pH was the only parameter that showed statistically significant differences in all comparisons, whereas differences in OM content were (highly) significant only when comparing site 2 with the other two sites, and soil relative humidity was significantly different only between site 2 and 1.

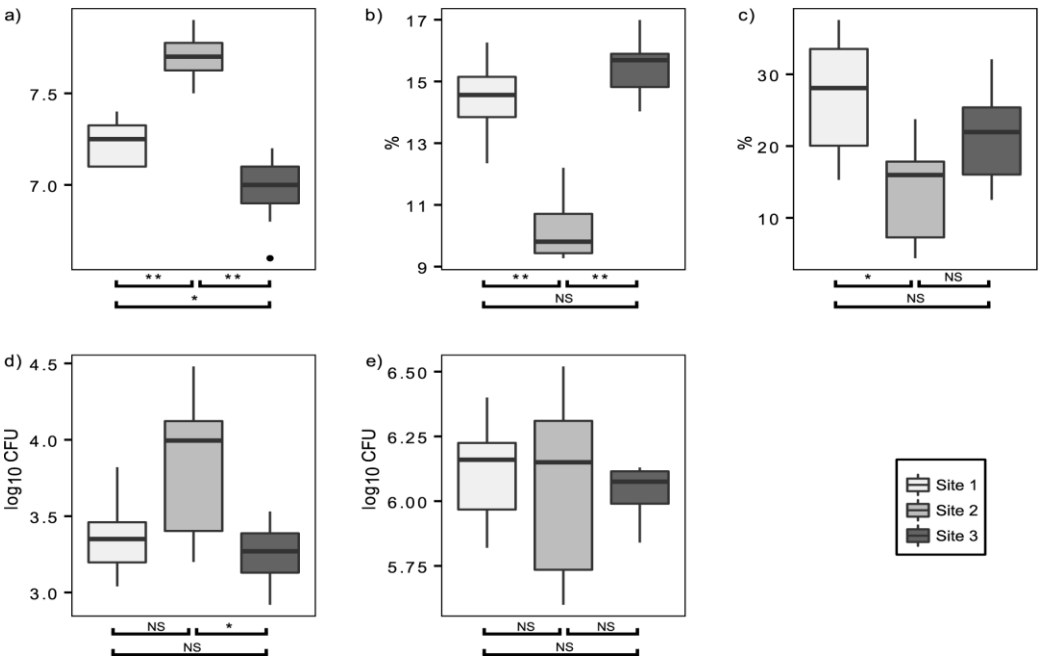


Figure 2 Physicochemical characterization and microbial abundance of soils from the three study sites. a) pH b) Relative organic matter content c) Relative humidity d) Fungal viable counts e) Bacterial viable counts. Data in a, b, and c are measured on pooled soil, while data in d and e are mean seasonal values. Boxes represent interquartile range (IQR), median values are indicated by the black lines, and point represent outliers. Significance of T-Test between sites is reported under each plot: ** P<0.01; * P<0.05; NS not significant

Microbial abundance, evaluated by viable counts in individual soil samples from the three sites, was reported in Figure 2 as mean values plus standard deviation (SD). The results showed that the number of cultivable bacteria were always two-three orders of magnitude greater (mean= 1.39 x 10^6 CFU/g dry soil) than cultivable fungi (mean= 5.38 x 10^3 CFU/ g dry soil). Bacterial viable counts were similar in the three sites. Fungal viable counts in site 1 and 3, similar, to each other, were lower than in site 2. Viable count values in site 2 always showed the greatest variability, suggesting higher quantitative instability of the culturable microbial components in poplar plantation (site 2) compared to natural forest (site 1 and 3) soils. As demonstrated by Kruskal-Wallis test (Table S1), quantitative fluctuations of the culturable components of microbial communities in three sites were not linked to spatial factors (different trees) but, with the exception of bacterial communities in site 3, to temporal factors (seasons).

Microbial community analysis

Ordination analysis

The similarity of individual T-RFLP profiles of bacterial and fungal communities in each sampling site was analyzed by nMDS ordination. ANOSIM analysis (not shown) was used to test the significance of grouping the samples on the basis of time (season and year) and space (individual trees within the site). As regard bacteria, nMDS ordination plots were always significant (0.19 > stress values > 0.11), and showed a tendency of the samples to form seasonal clusters (Figure S1) in the 2D-ordination space, even if ANOSIM analysis did not support any tested grouping. Unlike bacteria, nMDS ordination plots of fungal communities (Figure S2) were close to random (stress value > 0.2), but ANOSIM analysis of fungal communities showed statistical significance when fungi samples were grouped on the basis of belonging to the same individual tree (site1: R=0.58, p=0.001; site2: R=0.16, p=0.024; site3: R=0.55, p=0.001), or, as in site 3 only, on year of sampling (R=0.09, p=0.003).

The above outlined microbial diversity frame pointed out a high heterogeneity of microbial communities over space and time. This seemed to be particularly true for fungi populations. High within-site heterogeneity may reduce the resolution power of the analysis of similarity of microbial community structure between different sites, which represented the main objective of this work. For this reason, in each of the three sites, the T-RFLP profiles of the microbial communities from individual samples in a given season were put together in silico and averaged to obtain a total of 21 cumulative profiles. In this way, inter alia, community profile data were more directly comparable to the data obtained from analysis of soil chemistry, which were performed on 21 corresponding seasonal composite of soils. ANOSIM analysis was

performed on both individual and cumulative T-RFLP profiles for comparison. In both cases, differences between sites were significant, but cumulative profile analysis always explained a larger part of site variability (Table 2). Cumulative T-RFLP profiles were used in subsequent analyzes.

Table 2 Comparison of similarity (ANOSIM) results in testing the "site" grouping factor on cumulative and individual T-RFLP profiles data. Cumulative data were averaged site profiles.

	ANOSIM			
T-RFLP profiles	Bacteria		Fungi	
	R	P	R	P
Individual	0.07	**	0.45	**
Cumulative	0.53	**	0.92	**

Significance: ** P<0.01

Figure 3 reports nMDS ordination plot and UPGMA cluster analysis of cumulative T-RFLP profiles of bacterial communities from the three sites. The ordination was significant (stress value = 0.107), and overall the results highlighted a separation of soil bacterial communities in the poplar cultivation in site 2 from those of the natural mixed forests in site 1 and 3, which largely overlapped with each other forming mixed clusters. nMDS ordination analysis of fungal communities (Figure 4) was significant (stress value= 0.112), and showed a more clear separation between the three sites, which corresponding cumulative T-RFLP profiles formed well separated and defined clusters. Environmental variables (OM, pH, relative humidity, cumulative rain, and air temperature) and bacterial and fungal viable counts were fitted onto the ordination space of the nMDS plots in Figure 3 and Figure 4. The environmental descriptors that best explained differences in assemblage structure were pH, OM and relative humidity, both in bacteria (pH: r^2= 0.583, p= 0.003; OM: r^2= 0.673, p= 0.001; relative humidity: r^2= 0.556, p= 0.007) and in fungi (pH: r^2= 0.654, p= 0.003; OM: r^2= 0.854, p= 0.001; relative humidity: r^2= 0.654, p= 0.001). In addition, the fitting of the fungal viable count parameter in the fungi ordination was also significant (r^2= 0.519, p= 0.005).

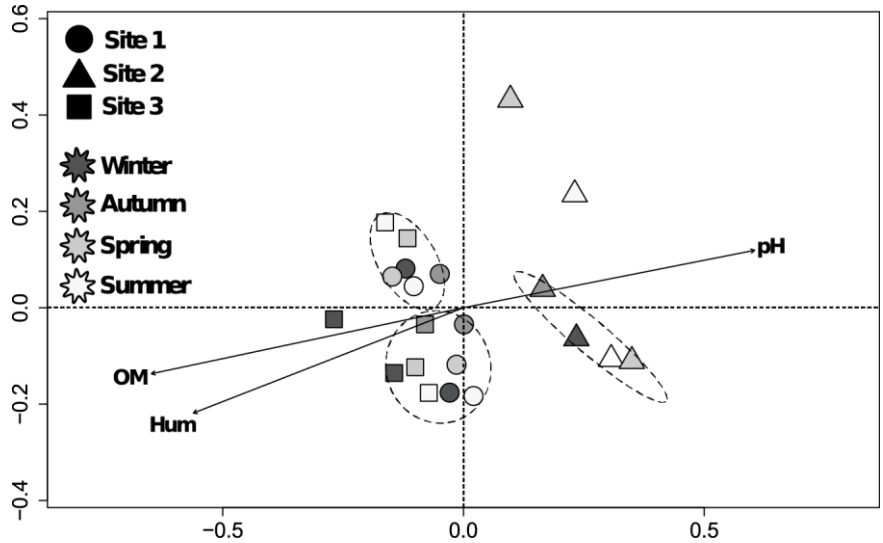


Figure 3 Non-metric multidimensional scaling of T-RFLP-based composition of bacterial communities in soils from Site 1(circles), Site 2 (triangles), and Site 3 (squares). Each point represent a pool of T-RFLP profiles from the same site, season, and year. Dotted ellipses show results of UPGMA analysis and highlight samples clustered at >75% similarity, based on Sørensen dissimilarity index. Vectors indicate only environmental variables that were significantly correlated with the ordination ($p<0.01$)

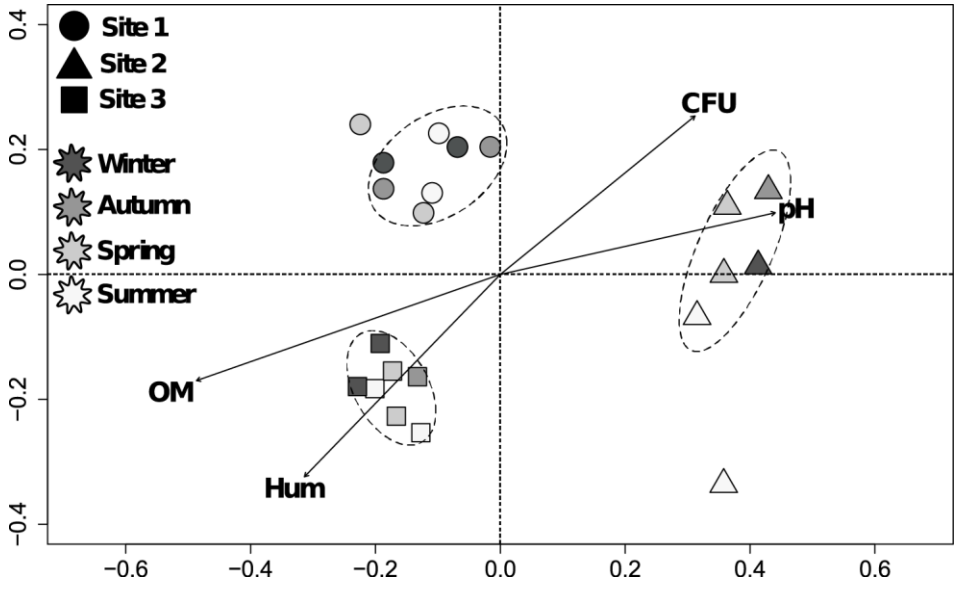


Figure 4 Non-metric multidimensional scaling ordination of T-RFLP-based of fungal communities in soils from Site 1(circles), Site 2 (triangles), and Site 3 (squares). Each point represent a pool of T-RFLP profiles from the same site, season, and year. Dotted ellipses show results of UPGMA analysis and highlight samples clustered at >50% similarity, based on Sørensen dissimilarity index. Vectors only indicate environmental variables that were significantly correlated with the ordination ($p<0.01$)

Microbial diversity analysis

As shown in Table 3, bacterial communities from soils of the cultivated poplars in site 2 had lower α-diversity values (richness, S=18; Shannon Index, H'=2.31) compared to sites 1 (S=22; H'=2.51) and 3 (S=21; H'=2.39), whereas Eveness values were similar (E=0.82, E=0.80, and E=0.79 in site 1, 2 and 3, respectively), indicating equal contributions of dominant species to bacterial α-diversity in the three sites. β-diversity values indicated that bacterial communities from the two natural-mixed forest (sites 1 and 3) were more similar to each other (β= 0.21) than they are to communities from the poplar plantation in site 2. Interestingly, bacterial communities from the poplar plantation were more similar to communities from the natural-mixed forest in sites 1 (β= 0.26), where soils associated to natural poplars were sampled, than to those in site 3 (β= 0.33), where soils associated to natural maples were sampled.

Table 3 Mean values of α-diversity and β-diversity of bacterial communities in soil samples from the three sampling sites. In brackets standard deviation values.

	Alpha diversity				Beta diversity	
	Richness (S)	Evenness (E)	Simpson Index (D)	Shannon Index (H')	Site 1	Site 2
Site 1	22 (4)	0.82 (0.05)	0.87 (0.05)	2.51 (0.27)	-	-
Site 2	18 (3)	0.80 (0.04)	0.85 (0.04)	2.31 (0.23)	0.26	-
Site 3	21 (5)	0.79 (0.06)	0.84 (0.06)	2.39 (0.32)	0.21	0.33

Table 4 Mean values of α-diversity and β-diversity of fungal communities in soil samples from the three sampling sites. In brackets standard deviation values.

	Alpha diversity				Beta diversity	
	Richness (S)	Evenness (E)	Simpson Index (D)	Shannon Index (H')	Site 1	Site 2
Site 1	12 (4)	0.86 (0.12)	0.82 (0.14)	2.12 (0.50)	-	-
Site 2	9 (3)	0.86 (0.09)	0.80 (0.09)	1.91 (0.36)	0.80	-
Site 3	11 (4)	0.87 (0.10)	0.81 (0.12)	2.06 (0.47)	0.66	0.88

Similarly to bacteria, fungal communities in site 2 had the lowest values of α-diversity, whereas Eveness values in the three sites were comparable (Table 4). β-diversity values indicate that site 2 differed from site 1 and 3 (0.80 and 0.88, respectively) more than site 1 and 3 differed from each other (0.66). In comparing bacteria and fungi diversity in the three sites (Table 3 data vs. Table 4 data), it was apparent that bacteria richness was always higher, but bacteria to fungi richness ratio were comparable (always near two). β-diversity behaved opposite, indicating a higher heterogeneity of the fungal communities compared to bacterial.

Figure 5 and Figure 6 show richness estimation for bacterial and fungal communities, respectively. Accumulation curves estimated almost identical numbers of bacterial and fungal phylotypes in the two natural forests (sites 1 and 3) and a lower number in site 2, indicating that soils from the cultivated forest had lower microbial richness compared to natural forests. The actual difference between sites was corroborated by the observation that, at a sample size of 24, the error bars (representing 2 times the standard deviation) of the phylotype (T-RF) accumulation curves of site 2 never overlapped with those of site 1 and 3; this, although not being the result of a proper statistical test, was a good indication of significant differences in the microbial richness between site 2 and the other two sampling sites. However, microbial communities generally appeared undersampled; in fact, the accumulation curves, with the only exception of bacterial phylotypes in site 2, never approached the ChaoI index values. The estimated numbers of fungal phylotypes was always higher than bacterial, a result that apparently conflicted with the higher richness of bacteria compared to fungi (Table 3 and 4), but that could be explained by the higher heterogeneity of fungal communities in the three sites (see also ANOSIM analysis in Table 2).

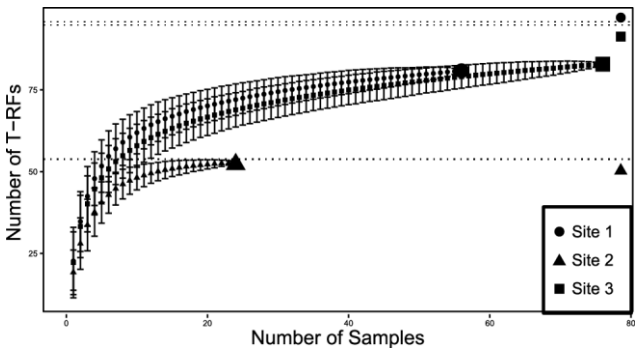


Figure 5 Species accumulation curves of bacterial communities in soils from Site 1(circles), Site 2 (triangles), and Site 3 (squares). The error bars are 95% confidence interval. Horizontal lines represent Chao1 richness estimator values

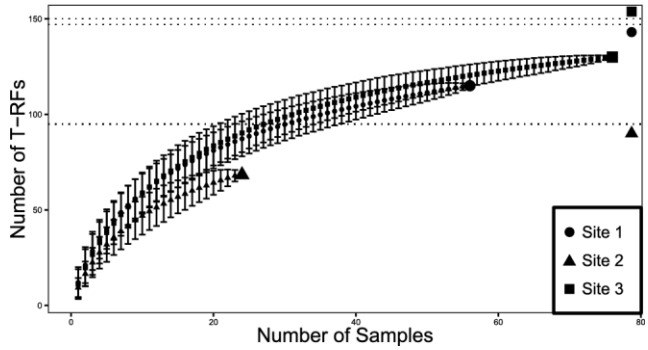


Figure 6 Species accumulation curves of fungal communities in soils from Site 1(circles), Site 2 (triangles), and Site 3 (squares). The error bars are 95% confidence interval. Horizontal lines represent Chao1 richness estimator values

The site-distribution of phylotypes (T-RFs), corresponding to dominant members of bacterial and fungal domains, is shown in Figure 7. The number of T-RFs shared among all the sites, was much lower in fungal communities (10/99, 10%) than in bacterial ones (31/60, 52%). However, each site had a number of site-exclusive bacterial and fungal T-RFs. Site 2 had the lower number of bacterial (5.0 %) and the higher number of fungal (24.7 %) site-exclusive T-RFs, and always shared a greater number of T-RFs with site 1 than with site 3. Site 2 and 3 shared a lower numbers of bacterial and fungal T-RFs compared to site1 and 2, and site 1 and 3. These results clearly demonstrated that the three sites differentiated not only for the soil physicochemical characteristics but also for their microbial components, and that site 2 (poplar plantation) differed from the other two sites more than they differed from each other, but differed from site 1 (where soils associated to natural poplars were sampled) less than from site 3 (where soils associated to natural maples were sampled).

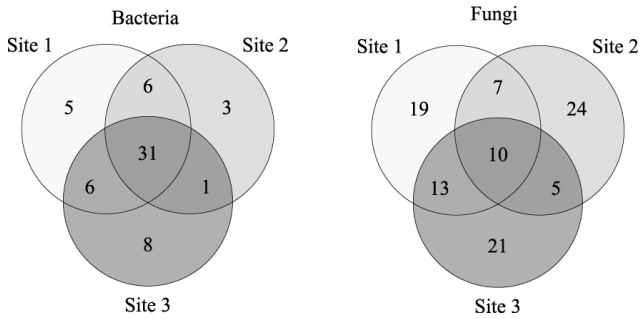


Figure 7 Venn diagram illustrating the numbers of unique and shared T-RFs in the pooled bacterial (a) and fungal (b) T-RFLP profiles. Only T-RFs with a relative height > 0.01 are considered

Correlation of α-diversity with environmental variables

The results of correlation analysis of bacteria and fungi α-diversity, measured using seasonal cumulative T-RFLP profiles, with different environmental parameters and with viable counts is shown in Figure S3. The estimated slope and the correlation coefficient r always share the same sign. Correlations with pH and OM were always significant, confirming how these parameters are among the main drivers of microbial diversity in soil. However, a more complex picture was revealed when the correlation analysis was performed for each site separately (Figure S4). In bacteria the sign of the correlations in the three sites appeared to be essentially random and the correlations were never significant (with exception of OM in site 1). In fungi the correlation between α-diversity and relative humidity of soil, cumulative rain, and air temperature, showed that estimated slope and correlation coefficient r in site 2 always had opposite-sign compared with the other two sites. Such differences between site 2 and the other two sites have been observed also for the correlation between fungal α-diversity and the viable count parameter; furthermore, in site 2 we observed high values of correlation and statistical significance between these two parameters. The fungi α-diversity values had a positive correlation with pH in all the three sites (in site 2 with very high values of correlation and statistical significance), whereas a concordant (positive) correlation of α-diversity with OM was seen only in site 1 and 3.

Temporal fluctuation analysis of microbial community diversity.

Given the long-term nature of our study, we were also interested in the effect that seasonality have on microbial communities diversity and in understanding how sites with a different land-use respond to seasonal changes. As above reported, a differentiation of microbial communities on a seasonal base was not supported by ANOSIM analysis; nevertheless, nMDS ordinations showed some degree of seasonal clusterization for bacterial communities (Figure S1). In order to better clarify the relationship between seasonality and α-diversity in each sampling site, we have calculated a coefficient of variation as the ratio between standard deviation and the mean value of different α-diversity parameters (Table 5). The bacterial and fungal microbial communities in soils from the poplar plantation in site 2 always exhibited lower temporal α-diversity variability (lower levels of variation) than natural forests (site 1 and 3). However, plotting the mean values of the different α-diversity parameters against time (season succession) in the three sites (Figure 8), we were not

able to identify any cyclic trend in the seasonal fluctuations of bacterial and fungal α-diversity. Data in Figure 8 have been used to calculate, for each site, a rate of trend concordance (TC) of seasonal fluctuation of α-diversity, where TC is the number of times in which the direction (increase or decrease) of an α-diversity parameter in two consecutive seasons was concordant (Table 6). TC values for fungi α-diversity in the two natural forests were always higher respect to the cultivated forest, indicating how fungal communities in natural forests (site 1 and 3) varied in time in a more coordinated way respect to the cultivated forest in site 2. For bacteria community, α-diversity parameters did not show any particular TC pattern, suggesting that land-use influenced the seasonal fluctuations of fungal and bacterial α-diversity in a different way.

Table 5 Coefficient of variation (CV) of α-diversity over time. CV was reported as percentage ratio between standard deviation and the mean value of the diversity parameter.

	Bacteria				**Fungi**			
	Richness	Evenness	Simpson index	Shannon index	Richness	Evenness	Simpson index	Shannon index
Site 1	19.8 %	6.5 %	5.7 %	10.7 %	32.6 %	13.7 %	17.0 %	23.5 %
Site 2	18.9 %	5.1 %	4.5 %	10.0 %	27.7 %	9.8 %	11.1 %	18.9 %
Site 3	25.8 %	7.8 %	7.6 %	13.4 %	35.3 %	10.9 %	14.8 %	22.9 %

Table 6 Rate of trend concordance of bacteria and fungi α-diversity parameters between the three sites. In bold the higher value for each parameter

		Bacteria		**Fungi**	
		Site 1	Site 2	Site 1	Site 2
Richness	Site 2	0.40	-	0.40	-
	Site 3	0.67	0.20	0.67	0.60
Evenness	Site 2	0.20	-	0.60	-
	Site 3	0.33	0.80	0.67	0.40
Shannon's Index	Site 2	0.60	-	0.40	-
	Site 3	0.33	0.40	0.67	0.60

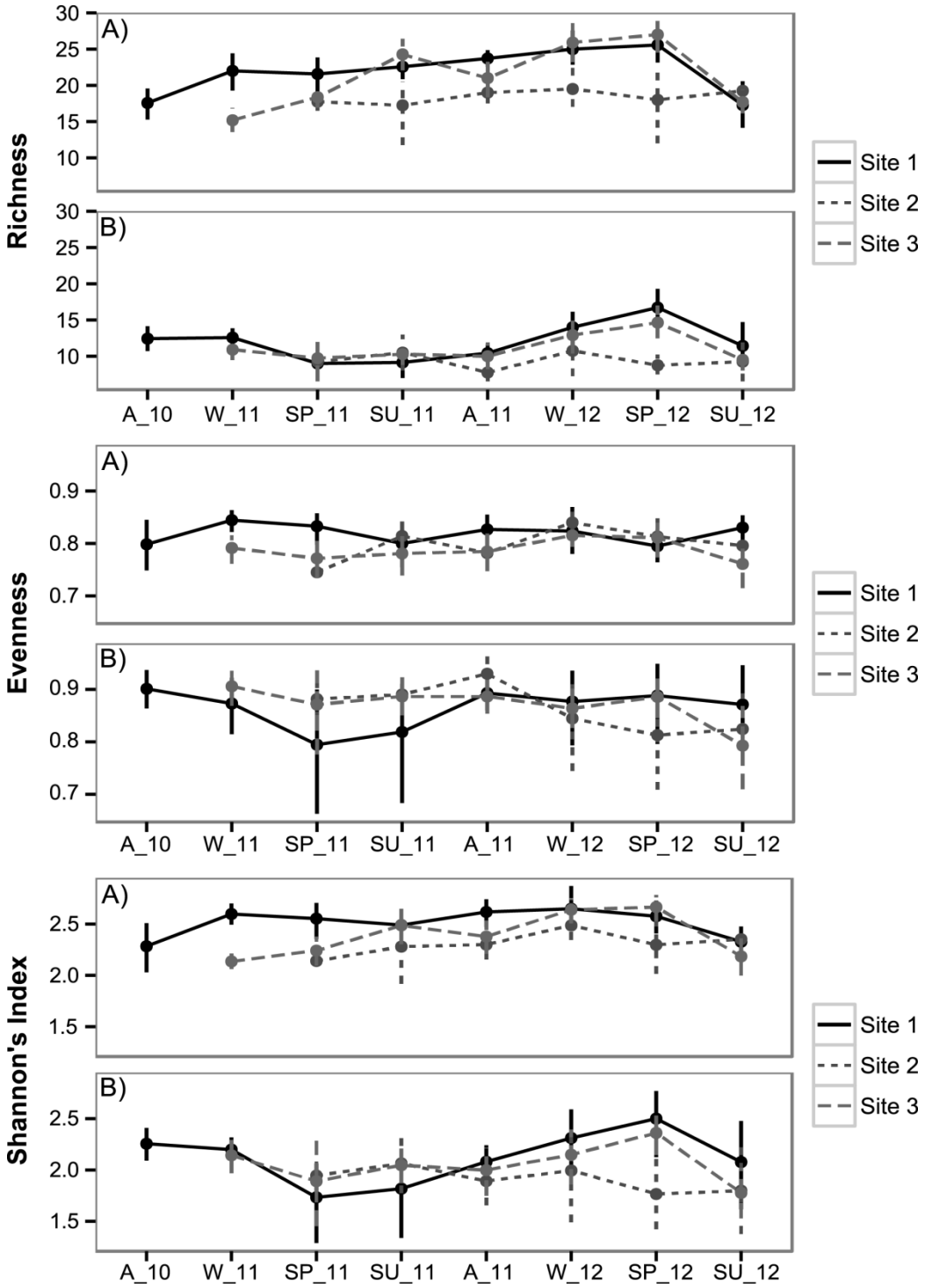


Figure 8 Temporal variations of Richness, Evenness, and Shannon's index in bacterial (A) and fungal (B) communities in the three sites.

6.5 Discussion

In this study, we wanted to evaluate long-term effects that a conversion from natural to planted forest have had on microbial community diversity and its seasonal dynamics. With the aim to minimize the effects of general climatic differences and those of unknown and unwanted anthropic impacts, we have chosen a local scale study area within a natural park for a comparative study of two natural forests (site 1 and 3) and a natural forest turned to tree plantation about 30 years ago (site 2). Indeed, pH and organic matter content, and to a lesser extent relative humidity, in the soil of the three sites appeared to be quite stable; this chemical stability was indicative of low human inputs, as it would be expected in a natural park, also in the mature poplar plantation in site 2. However, site 2 significantly differed from sites 1 and 3 for the physicochemical characteristics of the soil (pH, humidity and organics matter content) and in microbial abundance, and their seasonal fluctuations (Figure 2 and Figure S4). Overall, site 2 differed from the other two sites more than they differed from each other, suggesting that the conversion from natural to planted forest may have influenced, directly or through modifications of the soil physicochemical characteristics, its microbial components.

In all the three sampling sites, the ordination analysis of individual T-RFLP profiles of soil bacterial and fungi communities over time (different seasons) and space (different trees within each sampling site) showed generally high spatial heterogeneity of microbial communities, especially fungi, and no clear ordination according to seasons (Figure S1 and Figure S2). To mitigate the effects of this high heterogeneity on the resolving power of the similarity analysis of microbial communities, we have put together individual samples from different seasons in each site and constructed 21 cumulative T-RFLP profiles, corresponding to the 21 seasonal samples of composite soils, on which the physicochemical analysis were made. UPGMA cluster analysis and nMDS ordination plots clearly differentiated the bacterial and fungal cumulative T-RFLP profiles of site 2 (poplar plantation) from those of sites 1 and 3 (natural forests) along the first ordination axis (Figure 3 and Figure 4). Only with regard to the fungi, sites 1 (soil samples associated to natural poplars) and site 3 (soil samples associated to natural maples) separated along the second ordination axis. In any case, the fungal ordination was always sharper than bacterial. By interpreting these results, we speculate that the first axis was strongly correlated to soil land-use (poplar plantation in site 2 vs. natural forest in site 1 and 3), while the second axis could be correlated, more weakly and only as regard fungi, to the tree type (poplar in site 1 and 2 vs.

maples in site 3). From the direction and angle of the vectors resulting from the fitting of soil physicochemical characteristics on nMDS plots, pH and organic matter (OM), and to a lesser extent soil humidity, were correlated to the first axis. Overall, these data would suggest that land-use, directly or indirectly, through its influence on the soil physicochemical characteristics, has a major role in determining the differences in microbial communities derived from the conversion from natural forest to tree plantation. The role that this conversion would have had in driving the microbial structure in site 2 toward a structure different from those of the other two sites would be confirmed by data of diversity analysis. In fact, even if the contribution of T-RFLP phylotypes to bacterial and fungi α-diversity in the three sites were comparable (similar Eveness and Simpson Index values), α-diversity (Richness and Shannon's index) in the poplar plantation in site 2 was always lower compared to that of the natural forest in site 1 and 3. Moreover, bacterial and fungal β-diversity values between the poplar plantation in site 2 and the two natural-mixed forests were always higher than those between the last two. Interestingly, bacterial and fungal β-diversity values between site 2 (poplar plantation) and the natural-mixed forest in site 1 (where soils associated to natural poplars were sampled) were lower than between site 2 and site 3 (where soils associated to natural maples were sampled). This finding further support our hypothesis, formulated on the basis of the results of nMDS analysis (Figure 3 and Figure 4) and site-distribution of phylotypes (Figure 7), that the tree type also may have contributed to shape the community microbial structure in the three sampling sites. This hypothesis was supported by several studies reporting significant plant specific effects on microbial community structure [9, 18–22]. Instead, the contribution that genotype-specific plants may have on microbial diversity and structure [23] may help to explain the differences in the microbial populations in soils sampled in the vicinity of poplar trees in site 1 (Populus alba and Populus canescens) and site 2 (Populus nigra x Populus deltoids hybrid), in addition to other environmental factors.

An interesting but complex picture came from the analysis of the seasonal cumulative T-RFLP profiles in the search of a correlation between bacterial and fungal α-diversity and different environmental parameters (Figure S3). Overall, these data showed that, when the three sampling sites were analyzed together, the strongest and significant α-diversity correlations were with pH (negative correlation), and OM and relative humidity of soil (positive correlation). When the seasonal cumulative T-RFLP profiles were analyzed per site (Figure S4), site 2, and only as regard fungi, differentiated from the other two sites for all the correlations but pH, which always lead towards an increase in α-diversity, and OM, which was positively correlated with fungi diversity in site 1 and 2 only. This was a confirmation of the results of a number of scientific reports that stated that pH and OM are the main general drivers of microbial diversity

[41–43]. As our results have been obtained at a very local scale, the role of pH and OM in shaping soil microbial diversity seems to be confirmed regardless of the wideness of the study area. Nevertheless, limited to fungal α-diversity, other environmental factors (cumulative rain and air temperature), not directly linked to soil, and could have affected differentially the three sites.

Our data highlighted other distinctive traits between site 2 and the other two sites, but interestingly, in most cases bacterial and fungal communities behaved differently. We have found a strong significant negative correlation between fungal α-diversity and microbial (bacterial and fungal) viable counts in site 2; whereas, of opposite sign, but not significant, positive correlations were found in the other two sites (Figure S4). The correlation between bacterial α-diversity and the microbial viable counts in the same site was less strong, and as regard site 2, of opposite sign (Figure S4). A further element of differentiation between site 2 and the other two sampling sites was the lower estimated bacterial and fungal phylotypes richness of the former compared to the natural forests (Figure 5 and Figure 6). In site 2, the sampled fungal richness never approached the estimated true phylotype diversity (ChaoI index). Fungal communities, compared to bacteria, showed higher number of site-exclusive T-RFs and lower number of T-RFs shared among all sites (Figure 7); but site 2, showed the higher and the lower numbers of, respectively, fungal and bacterial exclusive phylotypes. Furthermore, fungal communities, always showed lower site richness (Table 3 and 4), even if the estimated number of fungal phylotypes was higher (Figure 5 and Figure 6). Overall, these results, as reported by other authors [44, 45], can be explained by assuming a higher intra-site spatial heterogeneity of fungal species compared to bacterial.

The study of the variability of α-diversity over time (Figure 8 and Tables 5 and 6) produced results that were not readily interpretable, but that did not seem to define any particular relationship between seasonality and α-diversity. Nevertheless, this analysis contributed to delineate further distinctive traits between bacterial and fungal communities in different forest soils. First of all, α-diversity of both bacteria and fungi communities in site 2 had narrower temporal fluctuations compared to site 1 and 3, indicating higher temporal stability of microbial communities within the poplar plantation soil. Moreover, even if bacteria α-diversity did not show any similarity in trend concordance between sites, fungal α-diversity in site 1 and site 3 did, differentiating the natural forests from the poplar plantation.

From the above, fungi, compared to bacteria, would be a more sensitive indicator of the effects of microbial perturbation caused by the change of land-use occurred in site 2.

In conclusion, our T-RFLP fingerprinting analysis highlighted important site-effects that differentiated the microbial communities of poplar plantation in site 2 from those of the two natural forests in sites 1 and 3. As the three forests sites were placed within a natural park, protected from unknown anthropogenic impact, and on a very local spatial scale, which guarantees extremely homogenous climatic characteristics, we concluded that the differences in microbial structure and diversity were mainly determined by the conversion of site 2 from natural forest to poplar plantation, occurred about thirty years ago. The differences in the physicochemical characteristics of soil between the poplar plantation and the two natural forests would be the consequence of this conversion and indirectly a differentiating factor of microbial communities. However, forests are highly heterogeneous natural systems, where microbial communities could be definitely influenced by a multitude of factors. So, besides the reported role of land-use and of physical-chemical properties in shaping and driving microbial diversity, a multitude of covariant factors, not easy to identify, are more likely influencing the diversity estimators. From the results of nMDS analysis (Figure 3 and Figure 4) and site-distribution of phylotypes (Figure 7), we have tentatively identified the tree type (poplars in site 1 and 2 and maples in site 3) as one of the factors affecting the microbial community structure. The influence of the tree type on microbial community structure and function was supported by a number of different reports [9, 18–23]

This study also represents an implementation of the knowledge of the Migliarino–San Rossore–Massaciuccoli Regional park biodiversity. In fact, although the diversity of fauna and flora of the park has been studied fairly well, information about microbial diversity are very scant [26, 46]. This study will be useful to predict consequences of future changes and to identify microbial predictors, with the aim to preserve the microbial diversity of the natural forests of this park, and more generally of natural environments.

Acknowledgments:

The authors would like to thank: Alessio Mengoni (Dept. of Biology, University of Florence) for his review of the work and for his precious suggestions, Cristina Indorato (Dept. of Biology, University of Florence) for her technical assistance, Cristina Vettori (IGV-FI, CNR) for assistance in T-RFLP analysis, Emiliano Fratini (Dept. of Chemistry, University of Florence) for soil chemistry analysis, Davide Travaglini and Francesca Bottalico (GESAAF, University of Florence) for plants georeferencing, Alessandro

Materassi and Gianni Fasano (IBIMET CNR, Florence) for metereological data, Francesca Logli (Migliarino San Rossore Massaciuccoli regional park) for useful forestry information about the sampling sites within the park. This work was founded by the European Community under the LIFE+ programme (grant n° LIFE08 NAT/IT/00342 – DEMETRA project). Giuliana Senatore and Cesarea Caroppo have been financially supported by grant n° LIFE08 NAT/IT/00342.

6.6 References

1. Van der Heijden MG a, Bardgett RD, van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol Lett 11:296–310. doi: 10.1111/j.1461-0248.2007.01139.x

2. Bardgett RD, Wardle DA (2010) Aboveground-belowground linkages: biotic interactions, ecosystem processes, and global change. Oxford University Press Oxford

3. Zak DR, Holmes WE, White DC, et al. (2003) Plant diversity, soil microbial communities, and ecosystem function: Are there any links? Ecology 84:2042–2050. doi: 10.1890/02-0433

4. MCPFE, UNECE, FAO (2007) State of Europe's forests 2007. MCPFE Rep Sustain For Manag Eur 1–263.

5. Ward BB (2002) How many species of prokaryotes are there? Proc Natl Acad Sci U S A 99:10234–10236. doi: 10.1073/pnas.162359199

6. Torsvik V, Ovreas L, Thingstad TF (2002) Prokaryotic diversity - Magnitude, dynamics, and controlling factors. Science (80-) 296:1064–1066. doi: 10.1126/science.1071698

7. Gams W (2007) Biodiversity of soil-inhabiting fungi. Biodivers Conserv 16:69–72. doi: 10.1007/s10531-006-9121-y

8. Buée M, Reich M, Murat C, et al. (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytol 184:449–456. doi: 10.1111/j.1469-8137.2009.03003.x

9. Myers RT, Zak DR, White DC, Peacock A (2001) Landscape-Level Patterns of Microbial Community Composition and Substrate Use in Upland Forest Ecosystems. Soil Sci Soc Am J 65:359. doi: 10.2136/sssaj2001.652359x

10. Chapin III FS, Zavaleta ES, Eviner VT, et al. (2000) Consequences of changing biodiversity. Nature 405:234–42. doi: 10.1038/35012241

11. Halme P, Toivanen T, Honkanen M, et al. (2010) Flawed meta-analysis of biodiversity effects of forest management. Conserv Biol 24:1154–6. doi: 10.1111/j.1523-1739.2010.01542.x

12. Birkhofer K, Schöning I, Alt F, et al. (2012) General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. PLoS One 7:e43292. doi: 10.1371/journal.pone.0043292

13. Rodrigues JLM, Pellizari VH, Mueller R, et al. (2013) Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proc Natl Acad Sci U S A 110:988–93. doi: 10.1073/pnas.1220608110

14. Da C Jesus E, Marsh TL, Tiedje JM, de S Moreira FM (2009) Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J 3:1004–11. doi: 10.1038/ismej.2009.47

15. Suleiman AKA, Manoeli L, Boldo JT, et al. (2013) Shifts in soil bacterial community after eight years of land-use change. Syst Appl Microbiol 36:137–44. doi: 10.1016/j.syapm.2012.10.007

16. Lupatini M, Jacques RJS, Antoniolli ZI, et al. (2013) Land-use change and soil type are drivers of fungal and archaeal communities in the Pampa biome. World J Microbiol Biotechnol 29:223–33. doi: 10.1007/s11274-012-1174-3

17. McGuire KL, D'Angelo H, Brearley FQ, et al. (2014) Responses of Soil Fungi to Logging and Oil Palm Agriculture in Southeast Asian Tropical Forests. Microb Ecol. doi: 10.1007/s00248-014-0468-4

18. Hackl E, Zechmeister-Boltenstern S, Bodrossy L, Sessitsch A (2004) Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. Appl Environ Microbiol 70:5057–5065. doi: 10.1128/AEM.70.9.5057-5065.2004

19. Bastias BA, Anderson IC, Xu Z, Cairney JWG (2007) RNA- and DNA-based profiling of soil fungal communities in a native Australian eucalypt forest and adjacent Pinus elliotti plantation. Soil Biol Biochem 39:3108–3114. doi: 10.1016/j.soilbio.2007.06.022

20. Berg G, Smalla K (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiol Ecol 68:1–13. doi: 10.1111/j.1574-6941.2009.00654.x

21. Wang M, Qu L, Ma K, Yuan X (2013) Soil microbial properties under different vegetation types on Mountain Han. Sci China Life Sci 56:561–570. doi: 10.1007/s11427-013-4486-0

22. Wubet T, Christ S, Schöning I, et al. (2012) Differences in Soil Fungal Communities between European Beech (Fagus sylvatica L.) Dominated Forests Are Related to Soil and Understory Vegetation. PLoS One 7:e47500. doi: 10.1371/journal.pone.0047500

23. Schweitzer JA, Bailey JK, Fischer DG, et al. (2008) Plant-soil microorganism interactions: heritable relationship between plant genotype and associated soil microorganisms. Ecology 89:773–81. doi: 10.1890/07-0337.1

24. Buckley DH, Schmidt TM (2002) Exploring the biodiversity of soil—a microbial rain forest. Biodivers. Microb. Life. Wiley-Liss, Inc

25. Wagg C, Husband BC, Green DS, et al. (2011) Soil microbial communities from an elevational cline differ in their effect on conifer seedling growth. Plant Soil 340:491–504. doi: 10.1007/s11104-010-0621-x

26. Uroz S, Courty PE, Pierrat JC, et al. (2013) Functional profiling and distribution of the forest soil bacterial communities along the soil mycorrhizosphere continuum. Microb Ecol 66:404–15. doi: 10.1007/s00248-013-0199-y

27. Lauber CL, Ramirez KS, Aanderud Z, et al. (2013) Temporal variability in soil microbial communities across land-use types. ISME J 7:1641–50. doi: 10.1038/ismej.2013.50

28. Osborn AM, Moore ERB, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ Microbiol 2:39–50. doi: 10.1046/j.1462-2920.2000.00081.x

29. Van Dorst J, Bissett A, Palmer AS, et al. (2014) Community fingerprinting in a sequencing world. FEMS Microbiol Ecol 89:316–30. doi: 10.1111/1574-6941.12308

30. Heiri O, Lotter AF, Lemcke G (2001) Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. J Paleolimnol 25:101–110. doi: 10.1023/A:1008119611481

31. Osborne CA, Galic M, Sangwan P, Janssen PH (2005) PCR-generated artefact from 16S rRNA gene-specific primers. FEMS Microbiol Lett 248:183–187. doi: 10.1016/j.femsle.2005.05.043

32. Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycprrhizae and rusts. Mol Ecol 2:113–118.

33. R Core Team (2013) R: A Language and Environment for Statistical Computing. R Found. Stat. Comput.

34. Oksanen J, Blanchet GL, Kindt R, et al. (2013) vegan: Community Ecology Package. Community Ecol. Packag.

35. Carl P, Brian G, Peterson GB (2013) PerformanceAnalytics: Econometric tools for performance and risk analysis.

36. Wickham H (2011) ggplot2. Wiley Interdiscip Rev Comput Stat 3:180–185. doi: 10.1002/wics.147

37. Shannon, C. E. and Weaver W. (1948) A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423 and 623–656.

38. Simpson EH (1949) Measurement of diversity. Nature 163:688. doi: 10.1038/163688a0

39. Chao A (1984) Nonparametric Estimation of the Number of Classes in a Population. Scand J Stat 11:265–270.

40. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol 67:4399–4406. doi: 10.1128/AEM.67.10.4399

41. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 103:626–31. doi: 10.1073/pnas.0507535103

42. Lauber CL, Strickland MS, Bradford MA, Fierer N (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem 40:2407–2415. doi: 10.1016/j.soilbio.2008.05.021

43. Rousk J, Bååth E, Brookes PC, et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 4:1340–51. doi: 10.1038/ismej.2010.58

44. Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM (2004) A taxa-area relationship for bacteria. Nature 432:750–753. doi: 10.1038/nature03073

45. Girvan MS, Bullimore J, Ball AS, et al. (2004) Responses of active bacterial and fungal communities in soils under winter wheat to different fertilizer and pesticide regimens. Appl Environ Microbiol 70:2692–2701. doi: 10.1128/AEM.70.5.2692-2701.2004

46. Zechmeister-Boltenstern S, Michel K, Pfeffer M (2010) Soil microbial community structure in European forests in relation to forest type and atmospheric nitrogen deposition. Plant Soil 343:37–50. doi: 10.1007/s11104-010-0528-6

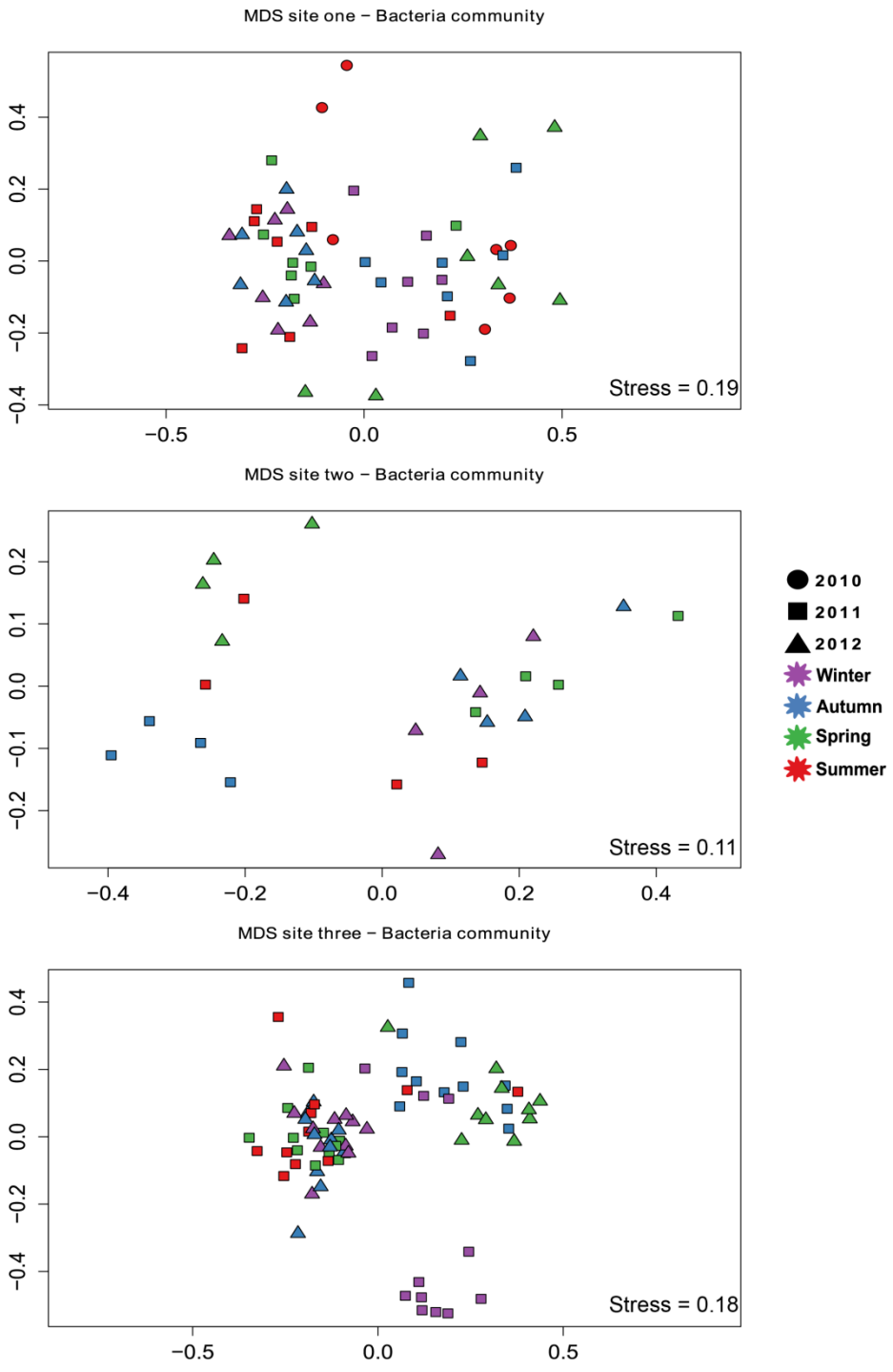


Fig. S1 Non-metric multidimensional scaling of T-RFLP-based composition of bacterial communities in soils from the three sites. Points represent individual T-RFLP profiles of the bacterial community in a different season and year. Shape of points represent sampling year, while colors represent sampling season.

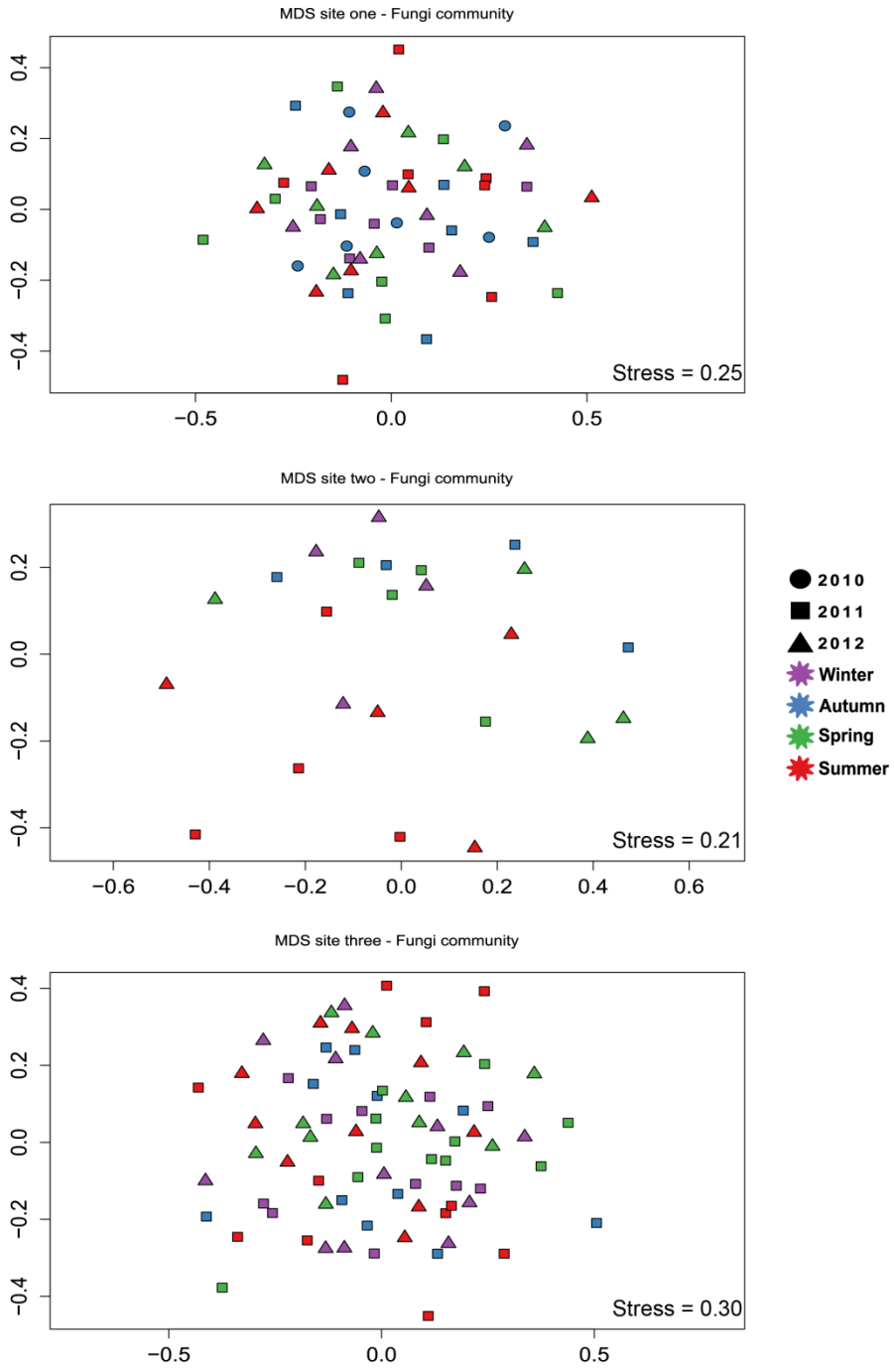


Fig. S2 Non-metric multidimensional scaling of T-RFLP-based composition of fungal communities in soils from the three sites. Points represent an individual T-RFLP profiles of the fungal community in a different season and year. Shape of points represent sampling year, while colors indicate sampling season.

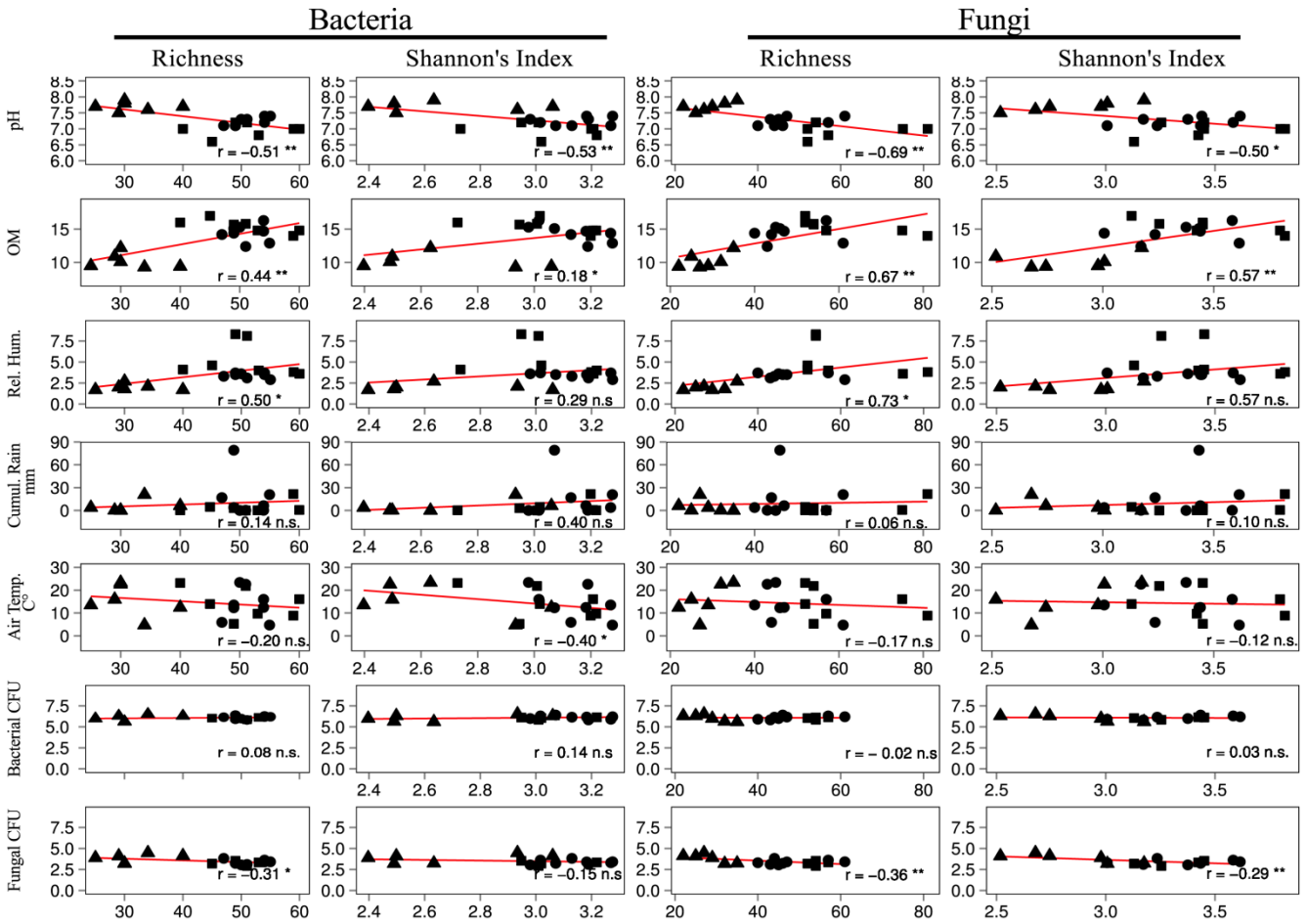


Fig. S3 Cumulative Spearman's correlation of microbial α-diversity (Richness and Shannon's index) with physicochemical characteristics of soil and soil microbial abundance in the three sites. Shape of the points indicate sampling site: circles, site 1; triangles, site 2; squares, site 3. Significance: ** P<0.01; * P<0.05

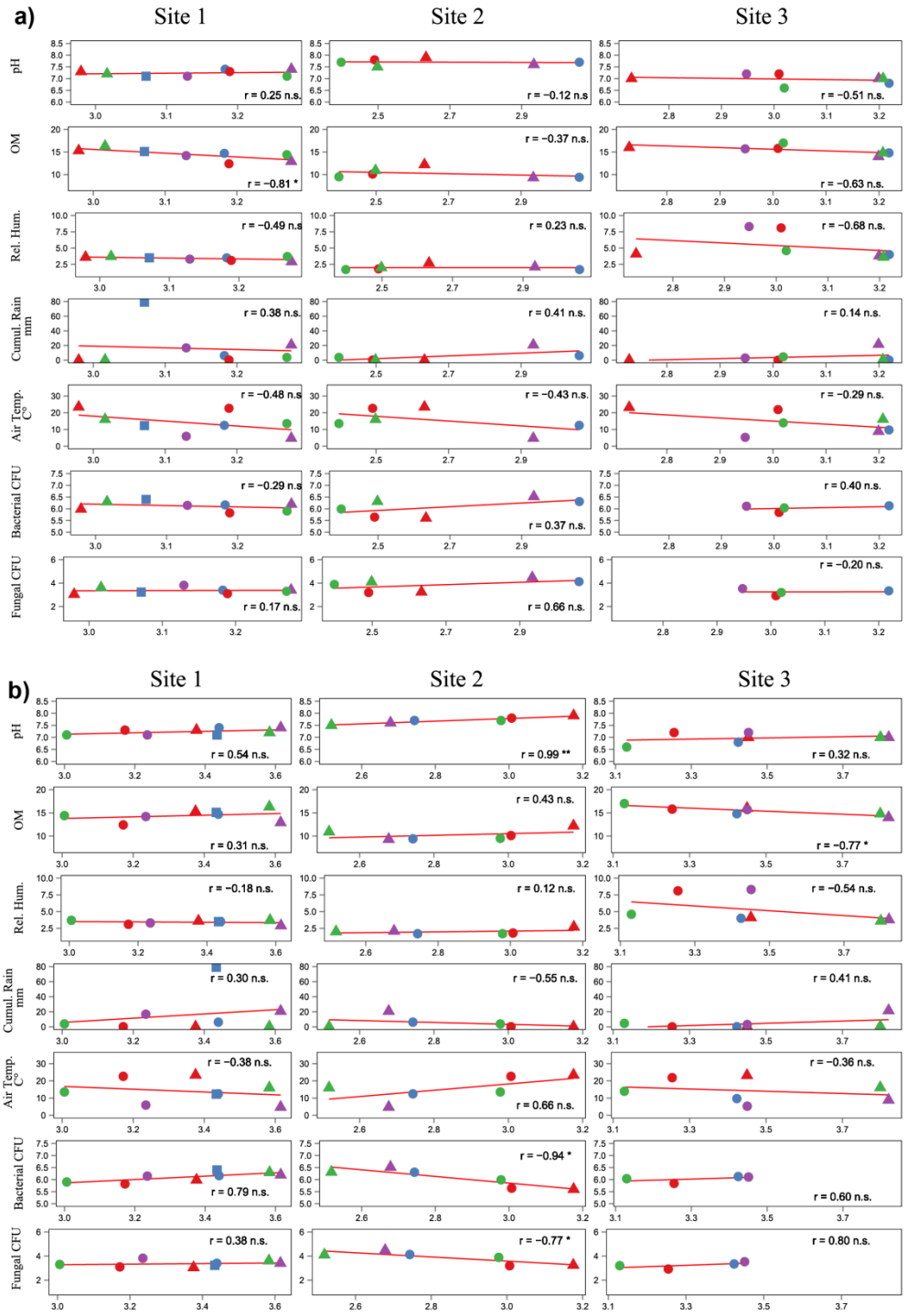


Fig. S4 Per site Spearman's correlation of a) bacterial and b) fungal Richness and Shannon's index with physicochemical characteristics of soil and soil microbial abundance.
Shape of the points indicate sampling year: squares, 2010; circles, 2011; triangles, 2012.
Colors indicate sampling season: purple, Winter; green, Spring; red, Summer; blue, Autumn. Significance: ** P<0.01; * P<0.05

Supplementary Table S1: Results of Kruskal-Wallis test. Within each site, we tested the dependency of the culturable component of microbial community (CFU values) with temporal and spatial factor.

		Bacteria			Fungi		
	Factor	Chi-squared	DF	P-Value	Chi-squared	DF	P-value
Site 1	Season	21.932	7	0.003	15.282	7	0.032
	Tree	4.685	6	0.585	5.568	6	0.473
Site 2	Season	17.230	5	0.004	14.190	5	0.014
	Tree	2.220	3	0.528	3.327	3	0.344
Site 3	Season	6.089	3	0.107	14.801	3	0.002
	Tree	16.106	10	0.097	15.018	10	0.131

Chapter 7

Application of Next Generation Sequencing to the ecological characterization of bacterial communities in forest soils from a natural park.

7 Application of Next Generation Sequencing to the ecological characterization of bacterial communities in forest soils from a natural park.

The DEMETRA project originally only included the use of T-RFLP as molecular approach for the study of microbial diversity in forest soils. The planning and realization of a sequencing project in the context of the MAPMED project, provided the opportunity to deepen our investigation on the soil microbiome. For this reason the sequencing is poorer than the experimental setting of the T-RFLP approach (i.e. lower number of samples).In this chapter, the results of the sequencing analysis on 12 pooled samples, representing each site in the 4 seasons, will be presented. Insight in community structure from T-RFLP analysis (chapter 6) will be compared to the results from the sequencing approach, in order to evaluate the two techniques.

7.1 Material and methods

Site description, soil sampling, and DNA extraction

For details on the sampling procedure and DNA extraction see chapter 6. Of all the soil samples collected during the two-year long activity, only those from Spring 2011, Summer 2011, Autumn 2011, and Winter 2012 were chosen for subsequent microbiome analysis. For each season the number of samples were seven in site 1, four in site 2, and eleven in site 3; for a total of 88 samples.

Site	Autumn 2010	Winter 2011	Spring 2011	Summer 2011	Autumn 2011	Winter 2012	Spring 2012	Summer 2012
	Sampling							
Site 1	✓	✓	✓	✓	✓	✓	✓	✓
Site 2	X	✓	✓	✓	✓	✓	✓	✓
Site 3	X	X	✓	✓	✓	✓	✓	✓

Samples used in microbiome analysis

DNA pooling, amplification, and sequencing

Total genomic DNA was extracted as described in chapter 6. Fixed quantity of DNA (40 ng for soil samples from site 1 and site 3, while, owing to the lower concentrations, 30 ng for site 2) were mixed to obtain four seasonal pools for each site. 10 ng of the pooled DNA was used as template in three replicate reactions. PCR reactions were performed in a final reaction volume of 25 µL with Dream Taq Buffer (containing 20 mM MgCl2 - Thermo Fisher Scientific GmbH, Karlsruhe, Germany), 0.2 mM of each dNTP (EuroClone, Milano, Italy), 1 µM of each primer, 1U of Dream Taq polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). The hypervariable V4 region of the bacterial 16S rRNA was amplified using forward amplification primer 563f (5'-AYTGGGYDTAAAGNG) and reverse amplification primer 802r (5'-TACNGGGTATCTAATCC) (Claesson et al. 2010). Amplification primers were modified for use in the MiSeq as described in chapter 3.4.2. Reverse primers with different barcodes were used for each different pool. The amplification reaction consisted of an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 45 sec, 50 °C for 1 minute, 72 °C for 1.5 min, and a final extension at 72 °C for 10 min.

Amplicons were checked for dimension on agarose gel and purified using the QIAquick Gel extraction kit (Qiagen). Purified DNA was quantified using the QUBIT fluorimeter (Life Technologies) and diluted at 4 nM. For sequencing, 10 µL of diluted DNA from each triplicate reaction were pooled in order to minimize stochastic PCR bias. The pool was sent in dry ice to the sequencing facility (Source Bioscience, Nottingham, UK), together with the custom sequencing primers. The sequencing facility performed an initial quality check and subsequent sequencing.

7.2 Results and Discussion

As stated above, the 12 DEMETRA pooled samples were analysed on the MiSeq with samples from the MAPMED project (see chapter 8). After sequencing, the sequencing facility performed the demultiplexing (i.e. division of the total library of the reads in smaller libraries, one for each index); Table 7.2 reports some parameter of the obtained libraries. The "reads" column reports the number of reads contained in each library; that is, the number of reads passing the filtering procedure performed by the MiSeq. Column "final reads" reports the number of reads remaining after the initial step of the analysis: I) paired end joining II) trimming of low quality end. (See chapter 3.4). The final, high quality, reads represents the effective input for the subsequent analysis (table 7.2)

Table 7.2: Number of reads in each dataset before and after the pre-processing step

Sample	Reads	Final reads
Site1_W	65015	45440
Site1_SP	65741	45430
Site1_SU	34289	23806
Site1_A	24244	15706
Site2_W	15753	10370
Site2_SP	17419	11276
Site2_SU	17413	11401
Site2_A	15964	11020
Site3_W	67296	46230
Site3_SP	58678	40562
Site3_SU	66683	42314
Site3_A	63059	41288

As highlighted in Table 7.2, the 12 samples had rather different library size, ranging from a minimum of 10370 sequences to a maximum of 46230.This is a general feature of NGS technologies; total reads per sample (library size) can vary by orders of magnitude within a single sequencing run. The number of OTUs is dependent upon the dimension of the library (i.e. the higher the number of sequences, the more richness is supposed to be recovered). As a consequence, observed diversity between samples could be greatly dependent upon unequal library size, rather than on real and biologically supported differences (Gihring et al. 2012). To account for this, it is a diffused practice to perform in silico normalization by applying rarefaction analysis.

Rarefaction is done by random subsampling (without replacement) the OTU table (i.e. the table containing the identified OTUs and counts representing the number of times each OTU is observed in each sample) at a certain sequencing depth, usually down to the lowest depth across samples. For example, when OTU tables are rarefied to 1000 sequence per sample, a new OTU table for each sample is created by random subsampling 1000 sequences from the original library. When each sample is rarefied to the same sequencing depth, bias-free diversity measure can be calculated. Moreover, if this procedure is repeated with increasing sequencing depth, diversity

measure can be calculated at each different depth, and we can observe the "growth" of the calculated measure plotting the results in an accumulation curve.

Fig 7.1 reports the result of this rarefaction analysis, performed to an even depth of 10000 sequences per sample on various alpha diversity measures. Bacterial community in the three sites, did not exhibit differences in Observed species, Phylogenetic diversity, and Simpson index, and the resulting rarefaction curves are almost overlapping. On the contrary, differences were observed for Shannon's index, ChaoI index, and Equitability. In particular Bacterial community from site 2, the natural forest turned into a poplar plantation, showed lower ChaoI index, higher Shannon's index, and higher Equitability values respect to the other two sites (natural forests). Bacterial communities from site 1 and 3 resulted highly similar, showing a low level of difference only for ChaoI index. Error bars, representing standard deviation, were calculated from 10 permutation of the diversity measure at each sampling depth. The fact that they do not overlap is a good indication of a significant difference, even if not the result of a proper statistical test.

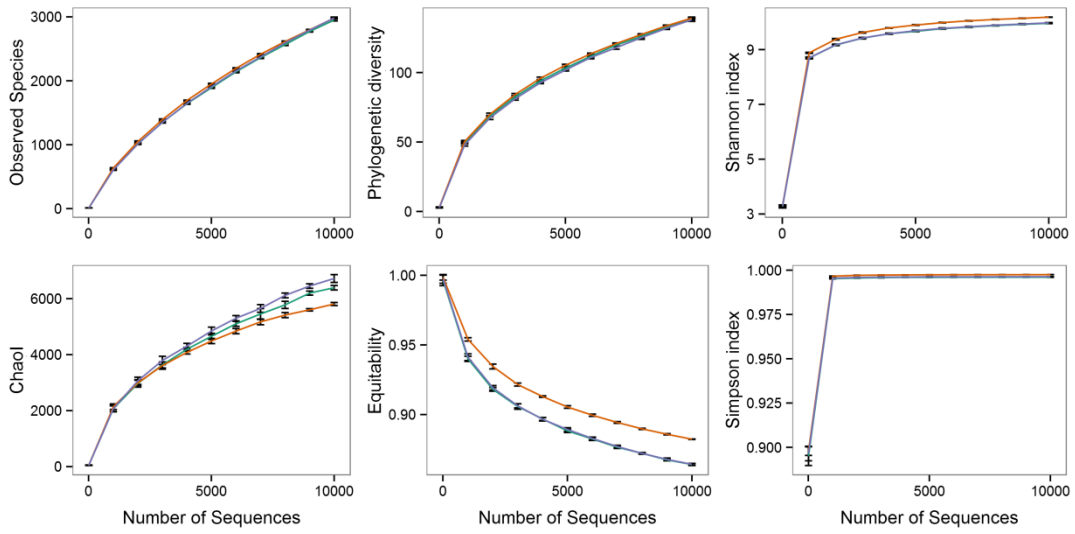


Figure 7.1 Rarefaction plots of six α-diversity parameters for bacterial communities in Site 1 (green), Site 2 (orange), and Site 3 (purple). Error bars represent standard deviation.

Community composition

Bacterial community composition in the soils from the three study sites was first evaluated at the *phylum* level (Figure 7.2). Communities from the three sites were always dominated by *phyla Acidobacteria* and *Proteobacteria*, accounting for more than 50% of the diversity. Other well represented *phyla* were *Actinobacteria*, *Gemmatimonadetes*, and *Verrucomicrobia*. As it is visible, there are not great differences in composition between different seasons in the same site (left panel of the figure). However, investigating community composition in the three sites as a whole (right panel of the figure), highlight some differences. *Phyla* composition in site 1 and site 2 appear fairly similar; mayor differences can be found in *phyla Proteobacteria*, more abundant in communities from site 2, and in *phyla Verrucomicrobia*, *WS3*, and *Nitrospira* more abundant in communities from site 1. *Phyla* composition in site 3, on the other end, resulted highly different from that in site 1 and 2. Some *phyla* were more represented, in particular *Acidobacteria* and *Verrucomicrobia*, while others were less represented, in particular *Proteobacteria* and *Gemmatimonadetes*.

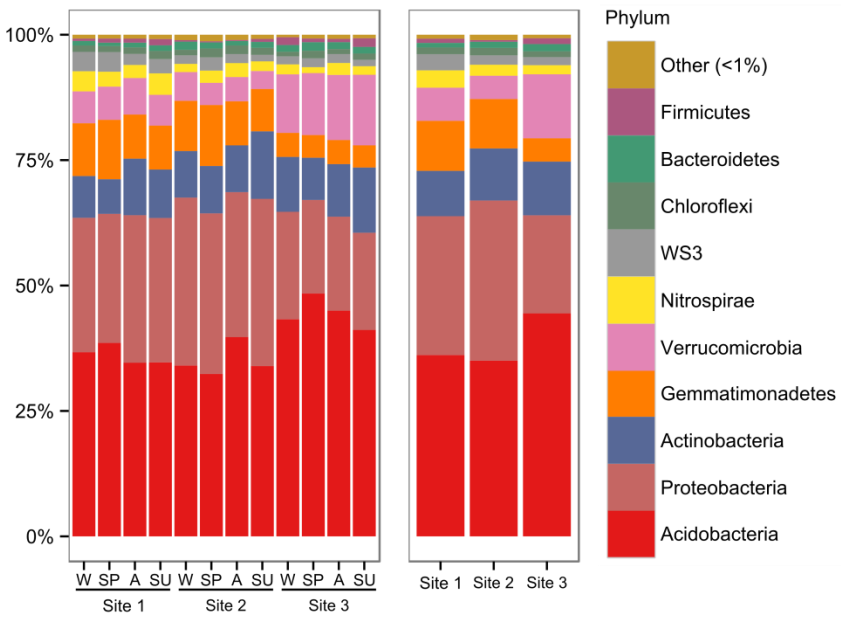


Figure 7.2: Bar plots representing the *phylum*-level composition of bacterial communities in each dataset (left) and the mean *phylum*-level composition of bacterial communities in each site (right).

To further investigate the community composition, OTUs were collapsed at the species level (i.e. all the OTUs identified as the same species were joined, and the counts were added up) and their distribution was pictured with heatmap and venn diagrams (Figure 7.3). Heatmap was realized with package *phyloseq* (McMurdie & Holmes 2013) which does not apply the classical hierarchical clustering method for organising columns and rows, but use ordination methods instead (in this case NMDS) providing easily interpretable order of elements (Rajaram & Oono 2010). Heatmap representation suggests a particular pattern of OTUs distribution in the three sites. Bacterial communities from site 1 and 3 appear rather similar, sharing, at high abundance values, OTUs in the upper and lower part of the heatmap. Site 2, on the contrary, appear to have a somewhat intermediate composition, sharing different OTUs with both sites. However, community composition in site 2 also exhibits some peculiarities, which differentiate it from the other two sites. OTUs in the lower and upper part of the heatmap, shared between site 1 and 3, have in fact low abundance in site 2, while a group of OTUs in the centre of the heatmap is overrepresented in site 2 respect to the other two sites. Venn diagrams in Figure 7.3 indicate the site-distribution of all the identified species (upper) and of the 200 most abundant only (lower). As can be seen, the proportion of shared species between couples of sites remained almost constant between the two sets of data; on the contrary the proportion of site-exclusive and shared between all sites ("core") species had great variations. In particular, the 200 most abundant species dataset mainly presented a site-specific distribution, with high amount of site-specific species. The dataset comprising the whole set of species, on the contrary, showed an opposite picture with high amount of "core" species. Those results indicate that bacterial community composition exhibited a high rate of site specificity and that only a small proportion of core species can be identified among the most abundant members, while most of them are found among the rarest component of the community. Considering the 200 most abundant species, a higher proportion of sharing was observed between site 1 and 3 than between site 1 and 2 and between site 2 and 3; with the latter two sharing the lowest proportion of species. This observation indicate that microbial community composition was most similar between the two natural forests, despite their spatial distance and despite the fact that soils associated to different plant types (poplars in site 1 and maples in site 3) were sampled.

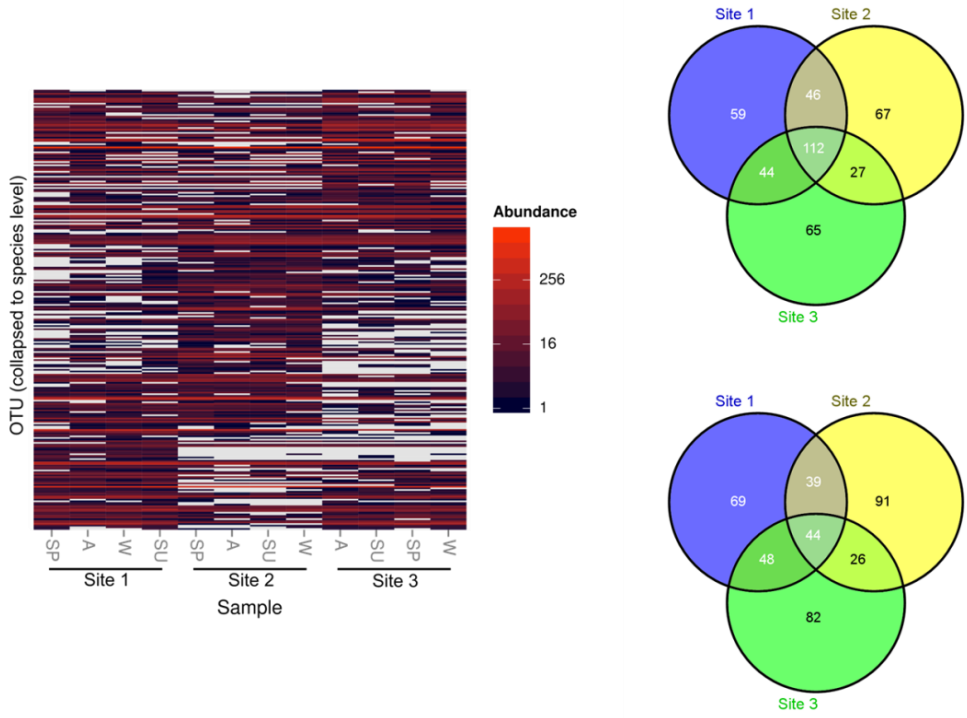


Figure 7.3: Left: Heatmap depicting bacterial OTUs distribution and abundance among the samples. Colours from blue to red represent OTU abundance (sequence count); grey colours represent OTUs with 0 counts. OTUs were collapsed to species level. Right: Venn-diagrams representing the distribution of OTUs among sites. Upper diagram was obtained by considering all the OTUs, while lower diagram was obtained by considering only the 200 most abundant OTUs. In both cases OTUs were collapsed to species level.

Ordination analysis

The higher similarity of bacterial communities in site 1 and 3 respect to site 2, as deduced by T-RFLP analysis (bacterial communities from site 1 and site 3 were overlapped in NMDS ordination in chapter 6 Figure 3), and that was one of the results on which conclusions on the effect of land-use change were drawn, is less evident in the ordination resulting from NGS data (Figure 7.4), placing the two approaches in apparent disagreement. Ordinations were done with PCoA analysis, using weighted unifrac (Lozupone & Knight 2005) as distance measure. It must be noted that the simple spatial difference is not enough to explain the PCoA ordination in Figure 7.4 Site 1 and 2, in fact, are really close in space while site 3 is far from both. If space is assumed to be the principal factor shaping the bacterial community, samples from site

1 and 2 would be close in the ordination space, or at least closer respect to site 3. The same ordination would also be seen if assuming tree type as principal factor. Those observations suggest that, as expected, NGS analysis can better recapture natural diversity, being able to highlight the differences between the community in site 1 and 3 that were not seen with T-RFLP approach, and that can be attributed to spatial and tree type factors. However, also in this case, the separate clustering of samples from site 2 suggests that the conversion from natural forest to poplar planation had an impact on the diversity of bacterial communities.

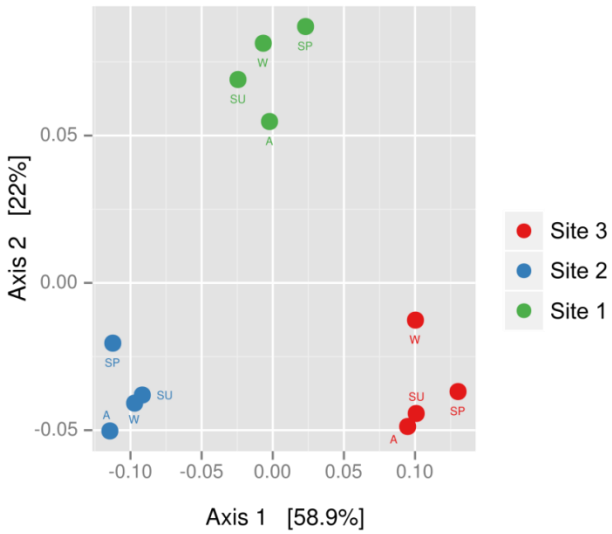


Figure 7.4: PCoA ordination of bacterial community in soils from the three sites, based on Weighted UniFrac distance. The percentage of the variation in the samples described by the plotted principal coordinates is indicated on the axes.

PCoA ordination (Figure 7.4) shows a clear clusterization of the samples on the basis of sampling site, indicating that the diversity of bacterial communities in the study sites is mainly dependent on space, while time (sampling season) does not have any influence. Those observations are confirmed by Analysis of Similarity (ANOSIM), as reported in table 7.3.

Table 7.3: ANOSIM results comparing bacterial community composition grouped based on site of origin or season of sampling

Factor	R	p-value
SITE	1	0.001
SEASON	-0.284	0.955

7.3 Conclusions

T-RFLP fingerprinting analysis (cap 6) highlighted important site-effects that differentiated the microbial communities of poplar plantation in site 2 from those of the two natural forests in sites 1 and 3. After exploring the results from NGS analysis on a subset of the same samples, we can affirm that similar conclusions can be drawn. In particular, rarefaction analysis (Figure 7.1) indicate that the bacterial communities in soil from site 2 were characterized by lower richness and higher evenness values, suggesting that the transformation from natural to planted forest had a strong impact on the bacterial community, lowering its diversity (the number of species predicted by ChaoI index) and increasing it evenness (equitability). Consequently we can speculate that this change mostly impacted the rare biosphere, selecting for a lower number of species that were dominant, and even distributed, in the community. In addition, the analysis of community composition at the species level (Figure 7.3), in particular the 200 most abundant species (see Venn diagram), further suggested a higher similarity between the two natural forests sites.

A big difference between T-RFLP and NGS approach was found with regard to ordination analysis. T-RFLP approach, in fact, failed to separate the samples from site 1 and 3, corroborating the hypothesis of a strong effect of the conversion from natural to planted forest. In NGS data, on the contrary, all the samples resulted grouped in site specific, strongly separated, clusters, better recapturing the difference attributable to spatial distance among the samples and to the difference in tree type. This divergence can be attributed to the great disparity in analysis depth between the two approaches; in other words, T-RFLP approach was not able to differentiate bacterial communities between site 1 and site 3 because too few T-RFs were generated (low information content) and/or because the set of species corresponding to each operational taxonomic unit (represented by each T-RFs in T-RFLP profiles) was too broad.

7.4 Bibliography

Claesson, M.J. et al., 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic acids research, 38(22), p.e200.

Gihring, T.M., Green, S.J. & Schadt, C.W., 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. Environmental microbiology, 14(2), pp.285–90.

Lozupone, C. & Knight, R., 2005. UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. Society, 71(12).

McMurdie, P.J. & Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one, 8(4), p.e61217.

Rajaram, S. & Oono, Y., 2010. NeatMap--non-clustering heat map alternatives in R. BMC bioinformatics, 11(1), p.45.

Chapter 8

Spatial variability of microbial communities in sediments from ports of the Mediterranean Sea Basin

8 Spatial variability of microbial communities in sediments from ports of the Mediterranean Sea Basin

8.1 Material and methods

Sampling

The case study sites of the present work were three tourist ports located across the Mediterranean Sea Basin: Cagliari (Sardinia, Italy) in the western part, El Kantaoui (Tunisia) in the central part, and Heraklion (Crete, Greece) in the eastern part. Port selection was also based on elements such as maritime traffic, port dimension, and existing information on pollution. The maritime traffic inside the three tourist ports is represented by recreational boats, passenger ships and fishing vessels. In addition to the marine traffic, the major pollution sources are wastewater discharges into the sea, river mouth, fuelling stations and fishing activities. The selected sampling stations were chosen according to their position within the port area and their use (i.e. touristic, industrial, cargo area) (Table 8.1). Samplings were performed in 2012 after the tourism period (Heraklion port 14-16 Sept.; Cagliari port 19-22 Sept.; El Kantaoui port 24-25 Sept.).

Sediments samples were collected using a specially designed box corer (dimensions: 13.5 x 13.5 x 16 cm) operated by hand from a boat. The box corer sampler was equipped with a modular expandable handle, consisting of 8 poles of 2 m length each attached to each other which could allow for sampling down to about 14 m. Sediment subsamples were extracted on board from the box corer using three ethanol treated plastic corers (4.4 cm Ø). Once on the shore, each sediment core was aseptically sliced into sections: surface sediments (SS) were collected from the first centimeter whereas anoxic sediments (AS) were collected in the black zone (evidence of redox conditions), discharging the intermediate sediments (2-4 cm). (Figure 8.1) Sediment samples from each core were transferred into sterile plastic bags and accurately mixed. Surface and anoxic sediment samples were used for DNA extraction for molecular analyses. Sediment sampling and processing were operated by the team of the Hellenic Centre for Marine Research and the team of microbiologists of the University of Cagliari.

Table 8.1: Sampling stations in the three port areas.

Station	Cagliari (C)	El Kantaoui (E)	Heraklion (H)
1	Leisure boats	Leisure boats	Leisure boats
2	Intermediate	Fuel station	-
3	Passenger ships	Port entrance	Passenger ships
4	Cargo ships	-	Cargo ships
5	Port entrance	-	Shipyard

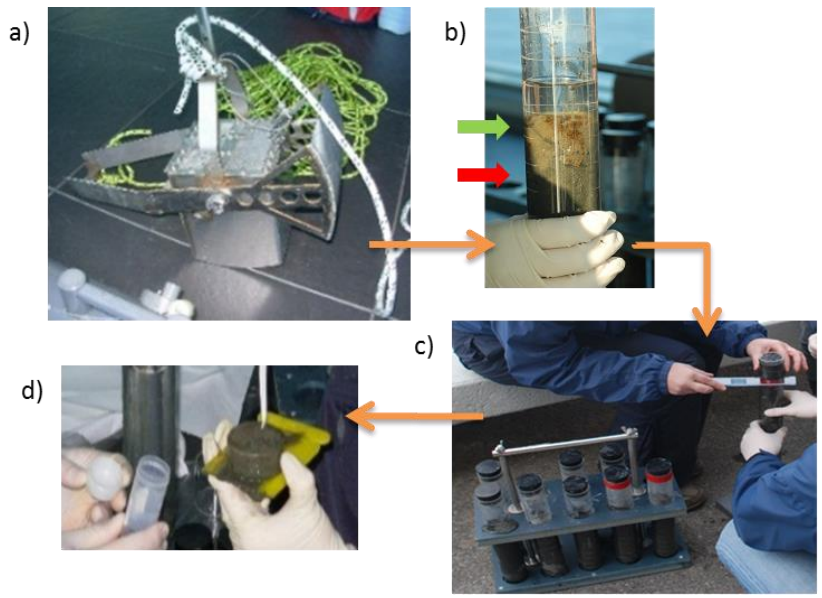


Figure 8.1: Sediment sampling and processing. a): box corer, b): sediment core (green: superficial sediments, red: anoxic sediment), c): slicing of cores, d): sediment sample.

DNA extraction from sediment samples

Sediment subsamples were transferred into sterile vials using sterile volumetric spatulas, stabilized by adding an equal volume of RNAlater® reagent (SIGMA), and stored at 4 °C until shipment to the laboratory at room temperature (Foti et al. 2008). Subsamples (250 mg wet weight) were obtained by aseptically combining an equal amount of sediments from the three cores collected per station and stored in RNAlater reagent (1:1 v/v). The storing reagent was removed prior to DNA extraction through centrifugation. Genomic DNA was extracted using the NucleoSpin® Soil kit (Macherey Nagel), following the manufacturer protocol with minor revisions. DNA concentration was determined on agarose gel. DNAs obtained from duplicate extractions were combined.

PCR amplification and sequencing

For bacterial community, 10 ng of the pooled DNA was used as template in three replicate reactions. PCR reactions were performed in a final reaction volume of 25 µL with Dream Taq Buffer (containing 20 mM MgCl2 - Thermo Fisher Scientific GmbH, Karlsruhe, Germany), 0.2 mM of each dNTP (EuroClone, Milano, Italy), 1 µM of each primer, 1U of Dream Taq polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). The hypervariable V4 region of the 16S rRNA was amplified using bacterial-specific forward amplification primer 563f (5'-AYTGGGYDTAAAGNG) and bacterial-specific reverse amplification primer 802r (5'-TACNGGGTATCTAATCC)(Claesson et al. 2010). Amplification primers were modified for use in the MiSeq as described in chapter 7.1. A reverse primer with a different barcode was used for each different pool. The amplification reaction consisted of an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 45 sec, 50 °C for 1 minute, 72 °C for 1.5 min, and a final extension at 72 °C for 10 min.

The amplification of 16S rRNA gene with archaeal-specific primers was inefficient, resulting in low yields, probably owing to the low quantities of archaeal cells within the prokaryotic community in marine samples. For this reason, a nested PCR approach was adopted performing two amplification rounds. For the first amplification round, 10 ng of total DNA were used as template with archaeal-specific primers 21F/958R, designed to amplify the entire 16S (DeLong 1992). PCR reactions were performed in a final reaction volume of 25 µL with Dream Taq Buffer (containing 20 mM MgCl2 - Thermo Fisher Scientific GmbH, Karlsruhe, Germany), 0.2 mM of each dNTP (EuroClone, Milano, Italy), 1 µM of each primer, 1U of Dream Taq polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany), and 0.1 mg/mL of BSA. 1 µL of the PCR

product was used as template for the second amplification round. The hypervariable V3 region of the 16S rRNA was amplified using archaeal-specific forward amplification primer 344f (5'-ACGGGGYGCAGCAGCAGGCGCGA) and archaeal-specific reverse amplification primer 519r (5'-GWATTACCGCGGCKGCTG)(Yu et al. 2008). PCR reactions were performed in a final reaction volume of 25 µL with Dream Taq Buffer (containing 20 mM MgCl2 - Thermo Fisher Scientific GmbH, Karlsruhe, Germany), 0.2 mM of each dNTP (EuroClone, Milano, Italy), 1 µM of each primer, 1U of Dream Taq polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). The amplification reaction consisted of an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 45 sec, 55 °C for 1 minute, 72 °C for 1.5 min, and a final extension at 72 °C for 10 min.

The PCR products from three replicate reactions were combined in order to minimize stochastic PCR bias. Amplicons were checked for dimension on agarose gel and purified using the QIAquick Gel extraction kit (Qiagen). Purified DNA was quantified using the QUBIT fluorimeter (Life Technologies) and diluted at 4 nM. For sequencing, 10 µL of each diluted amplicons were pooled. The pool, containing both bacterial and archaeal amplicons, was sent in dry ice to the sequencing facility (Source Bioscience, Nottingham, UK), together with the three custom sequencing primers. The sequencing facility performed the initial quality check on each pool and subsequent sequencing.

8.2 Results and Discussion

After sequencing, the sequencing facility performed the demultiplexing (i.e. division of the total library of the reads in smaller libraries, one for each index); Table 8.1 reports some parameter on the obtained libraries. The "reads" column reports the number of reads contained in each library; that is the number of reads passing the filtering procedure performed by the MiSeq. Column "final reads" reports the number of reads remaining after the initial step of the analysis: I) paired end joining II) trimming of low quality end. (See chapter 3.4). Those high quality reads represents the effective input for the subsequent analysis (table 8.1)

Table 8.1: Number of reads in each dataset before and after the pre-processing step

Samples		Bacteria		Archaea	
Station	**Matrix**	**Reads**	**Final reads**	**Reads**	**Final reads**
C1	SS	29170	23114	197883	177850
C2	SS	75076	61412	180183	163667
C3	SS	88660	70651	186495	174216
C4	SS	67887	56115	167683	155347
C5	SS	77291	62308	224938	209683
C1	AS	39633	31055	267862	245496
C2	AS	54331	42697	223658	206398
C3	AS	53966	44692	233448	217955
C4	AS	64689	52600	220727	200183
C5	AS	58953	46877	190653	177724
E1	SS	94328	74323	241875	227319
E2	SS	73543	61388	247845	228022
E3	SS	101033	83057	197417	184035
E1	AS	51374	42086	209236	193848
E2	AS	63985	50802	146840	132736
E3	AS	55099	47320	180514	166459
H1	SS	62930	51717	196950	180888
H3	SS	67619	52673	206559	186566
H4	SS	71321	56913	135654	124221
H5	SS	53844	42410	102741	89256
H1	AS	36962	30141	210129	195131
H3	AS	44679	35247	251702	222102
H4	AS	19139	15847	206217	191537
H5	AS	28563	23408	167744	155214

Samples grouping

At first, grouping of samples on the basis of predictable experimental factors were tested with ANOSIM analysis, using the Bray-Curtis index as a measure of distance (table 8.2). ANOSIM tests the null hypothesis that the average rank similarity between objects within a group is the same as the average rank similarity between objects between groups (Rees et al. 2004). The significance is tested with permutations and an R statistic, with values ranging from -1 to +1, is produced, indicating the strength of the separation between groups. Negative values of R statistic represent greater variation of sample within groups than between groups.

All tested grouping were significant, with the exception of the matrix grouping in Archaea community. Despite being significant, the R values resulted generally low (average of 0.28) indicating almost no separation between the groups. The sole case that might lead to a real separation of samples, is the port grouping showing a mid-level R value (R=0.445; p-value=0.001). These results suggest that microbial communities in port sediments are mainly influenced by the spatial position of the port in the Mediterranean Sea basin. Moreover, the lack of significance of matrix grouping for Archaea (R= 0.071; p-value= 0.082) suggest that superficial and anoxic sediments are characterized by fairly similar communities, which cannot be successfully divided in groups.

ANOSIM analyses were also performed with single port datasets, testing matrix and use as grouping factors inside each port. Differences were not significant, or showed low significance (table 8.2). Matrix resulted a significant grouping factor only for the port of Cagliari both for bacterial (R=0.556; p-value=0.008) and archaeal (R=0.396; p-value=0.044) communities. On the contrary, station use resulted a significant grouping factor only for the port of Heraklion both for bacterial (R=0.458; p-value=0.042) and archaeal (R=0.687; p-value=0.016) communities. In those cases, R had mid-high values, indicating a real division among groups, though with some overlaps.

Table 8.2: ANOSIM results comparing bacterial and archaeal community composition grouped based on port of origin, matrix, and station use. Firstly (upper part), the analysis was conducted on the entire dataset, then (lower part) it was repeated with single port datasets.

	Bacteria		Archaea	
Factor	**R**	**p-value**	**R**	**p-value**
Port *(C,E,H)*	0.445	0.001	0.396	0.001
Matrix *(AS,SS)*	0.165	0.009	0.071	0.082
Station Use *(table 8.1)*	0.222	0.010	0.250	0.004
Matrix				
Cagliari	0.556	0.008	0.292	0.044
El Kantaoui	-0.037	0.700	-0.185	0.800
Heraklion	0.012	0.473	-0.104	0.764
Station Use				
Cagliari	-0.220	0.851	0.320	0.090
El Kantaoui	0.611	0.067	0.667	0.067
Heraklion	0.458	0.042	0.687	0.016

Diversity analysis

ANOSIM analysis provided an indication on which factors were meaningful for dividing the community in sample groups; to further investigate these results α-diversity was compared considering the significant factors identified by ANOSIM. As introduced in chapter 7, diversity comparisons between samples can only be made if the sampling effort (in our case the number of reads) is homogeneous between the compared samples. For this reason, diversity comparisons were performed by rarefaction analysis.

Comparison between different ports: Port of origin resulted the main grouping factor in ANOSIM analysis. Figure 8.2 and Figure 8.3 report rarefaction analysis (to an even depth of 15000 reads per sample) for bacteria and archaeal communities, respectively. Species richness was evaluated by enumerating the observed species (panel a) and by Chao1 index (panel b); diversity in each site was evaluated with Shannon's index (panel c) and with Phylogenetic diversity (panel d). The difference among port-groups was low; major differences among port-groups were visible for Archaea community (greater division among curves) Samples from Heraklion port (orange) showed lower values for all the considered diversity indices, while samples from Cagliari (red) and El Kantaoui ports (blue) exhibited higher similarity, especially in the case of bacteria community, and the curves overlapped.

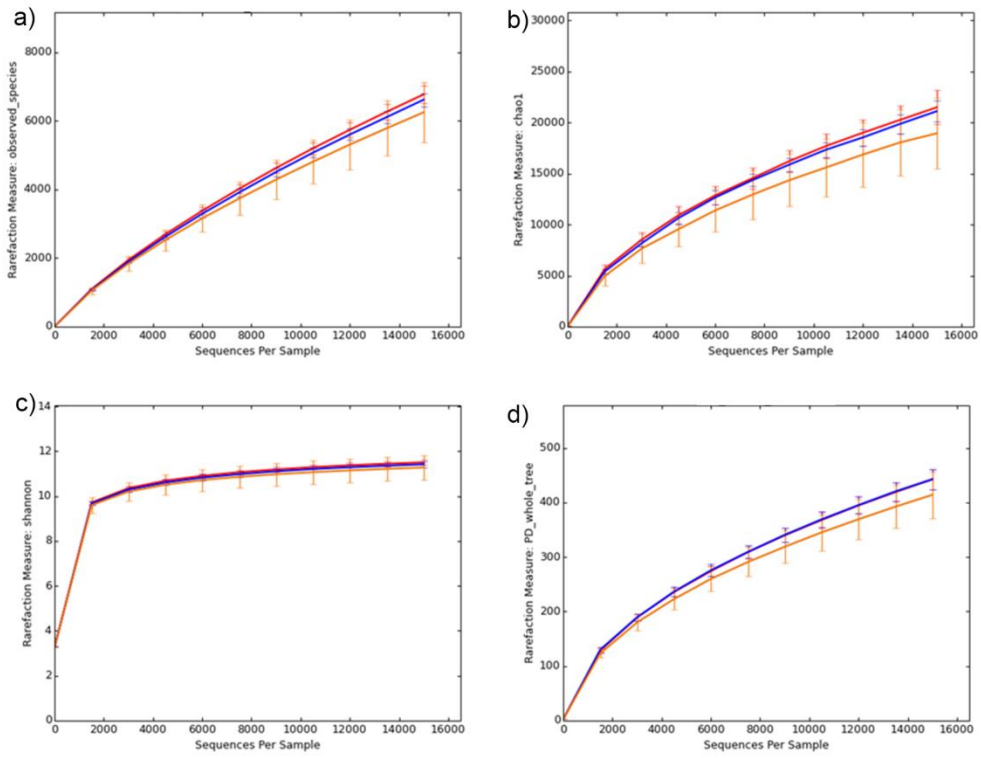


Figure 8.2: Rarefaction plots for observed richness (a), Chao1 index (b), Shannon's index (c), and Phylogenetic diversity (d) for bacterial communities in sediments from the three ports. Curves are colored on the basis of port: Cagliari (red), El Kantaoui (blue), and Heraklion (orange).

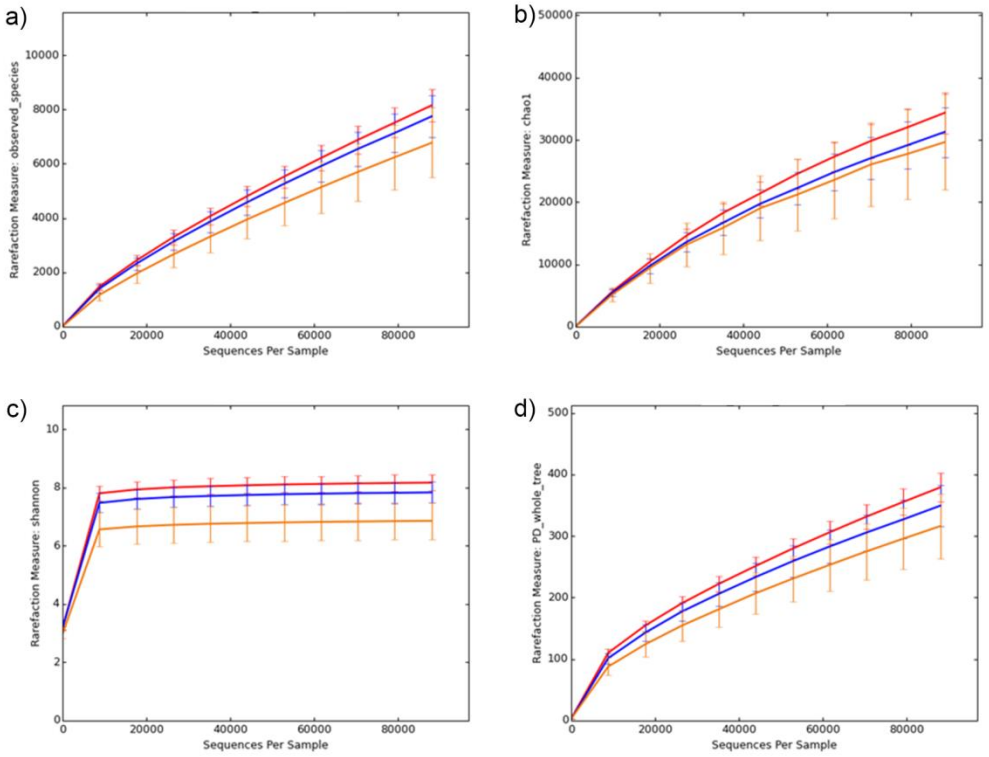


Figure 8.3: Rarefaction plots for observed richness (a), ChaoI index (b), Shannon's index (c), and Phylogenetic diversity (d) for archaeal communities in sediments from the three ports. Curves are colored on the basis of port: Cagliari (red), El Kantaoui (blue), and Heraklion (orange).

Comparison between different station uses: Station use resulted an important grouping factor for the port of Heraklion in ANOSIM analysis. Figure 8.4 and Figure 8.5 report rarefaction analysis (to an even depth of 15000 reads per sample) for bacteria and archaeal communities, respectively; as above reported, major differences were found for archaea communities. The most notable difference, however, was observed in the group with use "Shipyard" (yellow curve) which showed remarkably lower observed (panel b) and estimated (ChaoI index; panel a) richness values as well as remarkably lower diversity (Shannon's index and Phylogenetic diversity; panel c and d) respect to all other station-use. A site with shipyard use was only present in port of Heraklion (station number five in port of Heraklion; H5 hereafter) (table 8.1). Data analysis suggested a strong impact of the activities carried out in this station on bacterial and archaeal communities.

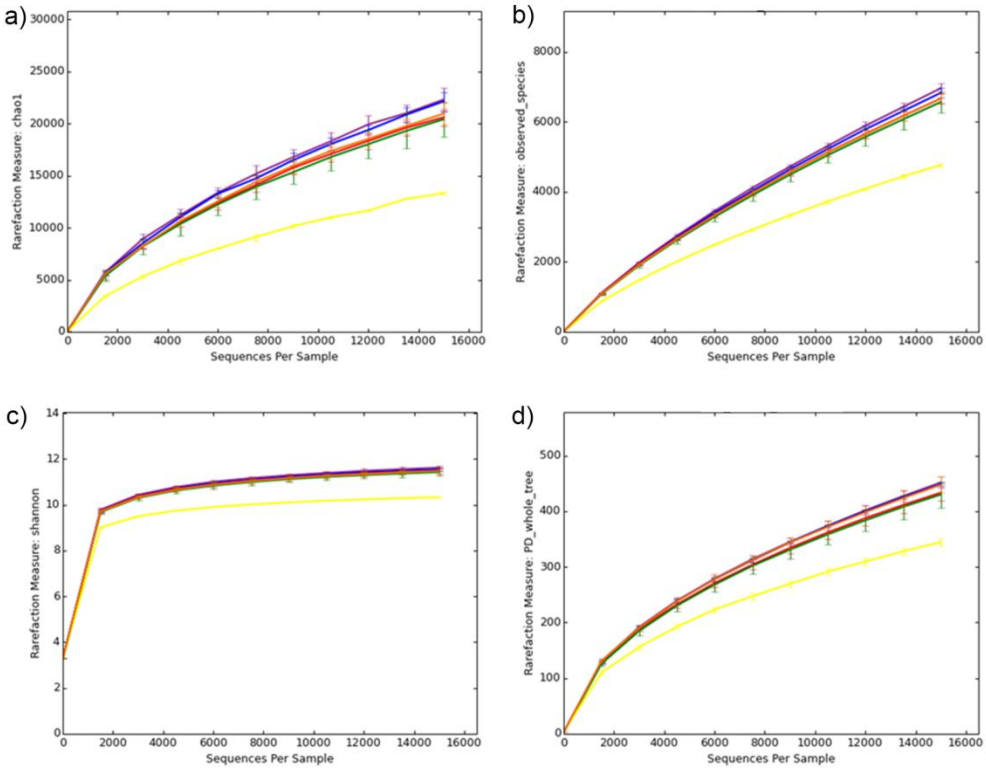


Figure 8.4: Rarefaction plots for observed richness (a), Chaol index (b), Shannon's index (c), and Phylogenetic diversity (d) for bacterial communities in sediments from the three ports. Curves are colored on the basis of station use: cargo (red), port entrance (blue), intermediate (orange), leisure boat (green), passenger boat (purple), and shipyard (yellow)

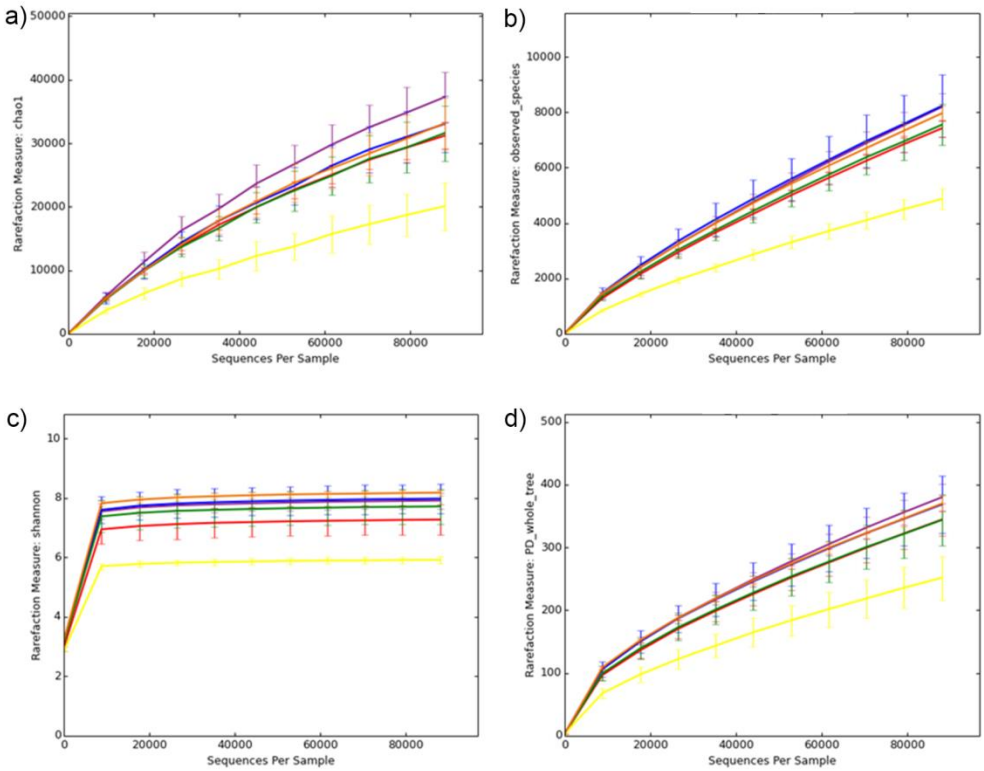


Figure 8.5: Rarefaction plots for observed richness (a), Chaol index (b), Shannon's index (c), and Phylogenetic diversity (d) for archaeal communities in sediments from the three ports. Curves are coloured on the basis of station use: cargo (red), port entrance (blue), intermediate (orange), leisure boat (green), passenger boat (purple), and shipyard (yellow)

Community composition

Figure 8.6 reports the *phylum* composition of bacterial communities in superficial and anoxic sediments samples from the three ports. Proportion of unassigned sequences varied from 2.6% to 5.4%; bacterial communities were always dominated by *phylum Proteobacteria*, *Acidobacteria*, and *Actinobacteria*, which together constituted more than 50% of the communities. *Phyla Proteobacteria,* was mainly composed by class *Deltaproteobacteria*, ranging from 53% to 80%, followed by *Gammaproteobacteria*, ranging from 15% to 29%, and *Alphaproteobacteria*, from 3% to 16%.

The composition of bacterial communities in sediments from the three ports was fairly similar, although some peculiarities can be identified. For example, proportion of

phylum Chloroflexi was higher in anoxic sediments from the port of Cagliari, while proportion of *phylum Nitrospira* was higher in superficial sediments from the port of *Heraklion*, and absent in superficial sediments of the port of Cagliari. *Phylum Firmicutes* was not found in sediments from the port of Cagliari and only in low abundance in those from the port of El Kantaouoi, but it constituted 6% of the community in sediments from the port of Heraklion. *Phylum OP1* was only found in sediments from the port of *Heraklion*, and principally in the superficial ones. Finally, high proportion of *phylum Hyd24-12* was found in superficial sediments from the port of Heraklion. Also the *Class* composition of *Proteobacteria phyla* showed some peculiarities; in particular, in superficial sediments from the port of Heraklion, and to a lesser extent in anoxic sediments, a substantial decrease in *Deltaproteobacteria* and a correspondent increase in *Gamma* and *Alphaproteobacteria* was observed.

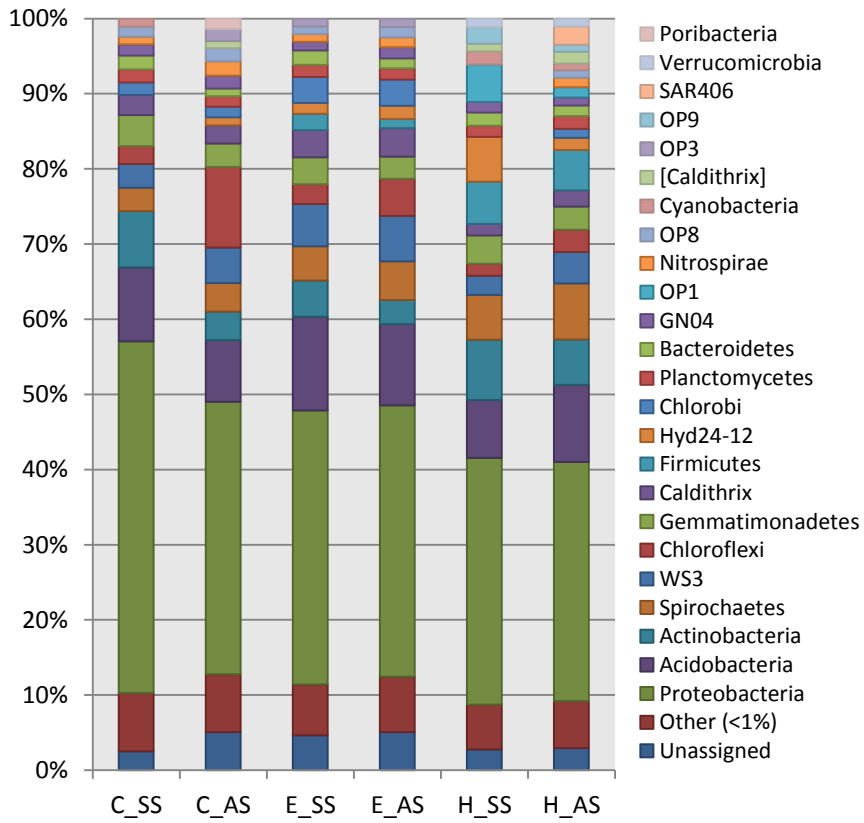


Figure 8.6: Bar plots representing the *phylum*-level composition of bacterial communities in superficial and anoxic sediments samples from the three ports.

Figure 8.7 reports the class composition of archaeal communities in superficial and anoxic sediments samples from the three ports. Proportion of unassigned sequences varied from 5% to 10%. Archaeal communities in sediment samples from the port of Cagliari and El Kantaoui were fairly similar; they were dominated by classes *Thermoplasmata* and *MCG*, which together constituted more than 80% of the communities. Samples from the port of Heraklion were characterized by a strongly different community; here classes *Thermoplasmata* and *MCG* were less abundant and classes *Methanomicrobia* and *Thaumarchaeota* (constituting around 20% of the community) were exclusively found.

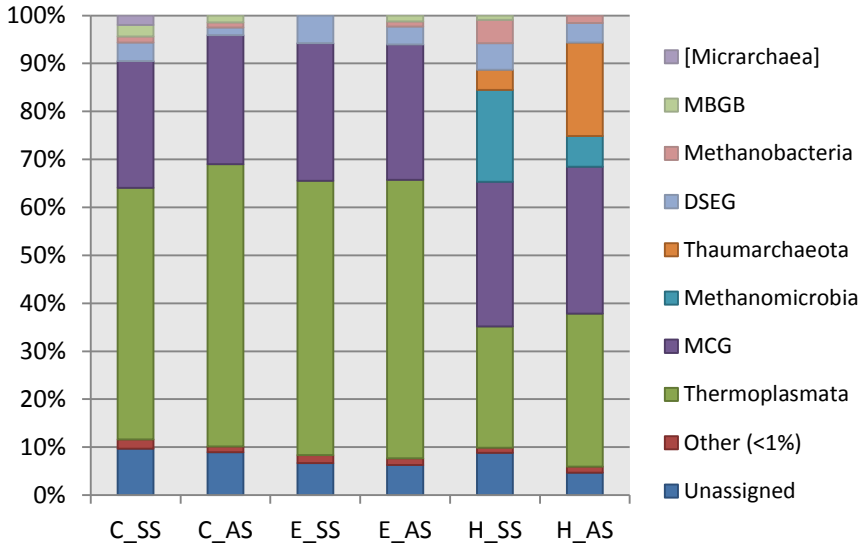


Figure 8.7: Bar plots representing the class-level composition of archaeal communities in superficial and anoxic sediments samples from the three ports.

The above reported observation are mainly based on visual evaluation of the bar plots in Figure 8.6 and 8.7. To test the most important *phyla*, or classes, in determining the diversity between sites, SIMPER analysis was applied. Similarity percentage (SIMPER) (Clarke 1993) is based on the decomposition of Bray-Curtis dissimilarity index: the contribution of each variable (in our case the *phyla* or the classes in the community) to the overall Bray-Curtis distance between two samples is calculated, and the most important variables for each pair of groups are identified and displayed. SIMPER was performed with function *simper* in the package *VEGAN* (Oksanen et al. 2013), which by

default displays the variables contributing to at least 70% of the Bray-Curtis dissimilarity. Results of this analysis are reported in table 8.3 and 8.4.

Table 8.3: SIMPER analysis identifying which bacterial *phylum* contributed for at least the 70% of the dissimilarity between couples of ports. *Phylum* are ordered based on their contribution. Overall av. diss. is the Bray-Curtis dissimilarity between couple of ports. For each *phyla*, the contribution to the overall average dissimilarity and the cumulative contribution is reported.

Comparison	Overall av. diss.	*Phylum*	Contribution	Cum. contrib.
Cagliari vs. El Kantaoui	0.163	*Proteobacteria*	0.036 (21.8%)	21.8%
		Chloroflexi	0.024 (14.9%)	36.7%
		Acidobacteria	0.013 (8.2%)	44.9%
		Actinobacteria	0.012 (7.6%)	52.5%
		Chlorobi	0.011 (6.6%)	59.1%
		WS3	0.010 (6%)	65.1%
		Firmicutes	0.009 (5.7%)	70.8%
Cagliari vs. Heraklion	0.239	*Firmicutes*	0.032 (13.5%)	13.5%
		Proteobacteria	0.030 (12.6%)	26.1%
		Chloroflexi	0.024 (10%)	36.1%
		Spirochaetes	0.021 (8.9%)	45.0%
		Hyd24-12	0.019 (8.1%)	53.1%
		OP1	0.019 (7.9%)	61.0%
		Actinobacteria	0.014 (5.8%)	66.8%
		OP9	0.009 (3.9%)	70.7%
El Kantaoui vs. Heraklion	0.210	*Firmicutes*	0.024 (11.5%)	11.5%
		Actinobacteria	0.02 (10.1%)1	21.6%
		OP1	0.019 (9.1%)	30.7%
		Spirochaetes	0.015 (7%)	37.7%
		Chlorobi	0.014 (6.9%)	44.6%
		Hyd24-12	0.014 (6.8%)	51.4%
		WS3	0.010 (4.8%)	56.2%
		OP9	0.009 (4.5%)	60.7%
		Acidobacteria	0.009 (4.3%)	65.0%
		Caldithrix	0.008 (4%)	69.0%
		Chloroflexi	0.008 (3.8%)	72.8%

Table 8.4: SIMPER analysis identifying which archaea classes contributed for at least the 70% of the dissimilarity between couples of ports. Classes are ordered based on their contribution. Overall av. diss. is the Bray-Curtis dissimilarity between couple of ports. For each class, the contribution to the overall average dissimilarity and the cumulative contribution is reported.

Comparison	Overall av. diss.	Class	Contribution	Cum. contrib
Cagliari vs. El Kantaoui	0.09	*DSEG*	0.031 (32%)	32.0%
		MCG	0.022 (23.3%)	55.3%
		MBGB	0.019 (19.7%)	75.0%
Cagliari vs. Heraklion	0.43	*Methanomicrobia*	0.138 (32.4%)	32.4%
		Thaumarchaeota	0.130 (30.4%)	62.8%
		MCG	0.086 (20.1%)	82.9%
El Kantaoui vs. Heraklion	0.39	*Methanomicrobia*	0.138 (35.5%)	35.5%
		Thaumarchaeota	0.129 (33.3%)	68.8%
		MCG	0.072 (18.6%)	87.4%

A deeper look at the port of Heraklion

The above reported results indicate that microbial communities in sediments from station H5 were highly different both from sediments in other stations of the same port, and from sediments of the other ports.

In this section, a detailed analysis on the community composition inside the port of Heraklion will be presented, with the aim to unravel the peculiarity of station H5.

Figure 8.8 reports the relative abundance of the *phyla* identified in the superficial sediments from the port of Heraklion. Stations H1, H3, and H4 had a very similar composition; the bacterial community was dominated by *phylum Proteobacteria*, *Acidobacteria*, and *Actinobacteria.* As already suggested by ordination and diversity analysis, bacterial community in samples collected from station H5 had a *phylum* composition markedly different from the other stations, presenting a lower proportion of Acidobacteria and Proteobacteria, but a higher proportion of *Spirochaetes* and *Firmicutes.*

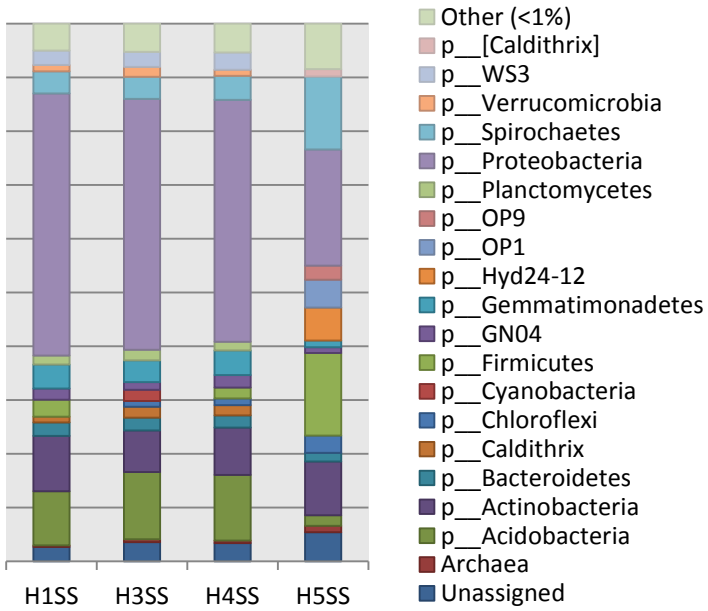


Figure 8.8: *Phylum* composition of bacterial communities in water samples from the port of Heraklion. Data are reported as relative abundance; bacterial *Phyla* with relative abundance < 1% were summed up and merged in category "Other".

Figure 8.9 reports the relative abundance of the classes identified in the superficial sediments from the port of Heraklion. The archaeal community collected from superficial sediments of station H1 (Leisure boats) and H4 (Cargo ships) were fairly similar, being dominated by class *MCG* (*phylum Crenarchaeaota*), *Thermoplasmata* (*phylum Euryarchaeota*) and *DSEG* (*phylum Euryarchaeota*), while the archaeal community collected from superficial sediments of station H3 showed a similar composition, but presented higher proportion of *Thermoplasmata* (*phylum Euryarchaeota*). The archaeal community in samples collected from station H5, on the contrary, showed an highly different class composition, with lower proportion of class *MCG* (*phylum Crenarchaeaota*) and *Thermoplasmata* (*phylum Euryarchaeota*), with higher proportion of class *Thaumarchaeota* (*phylum Crenarchaeota*) and *Methanobacteria* (*phylum Euryarchaeota*), and, finally, being the sole sample containing OTUs belonging to the *Methanomicrobia* class (*phylum Euryarchaeota*)

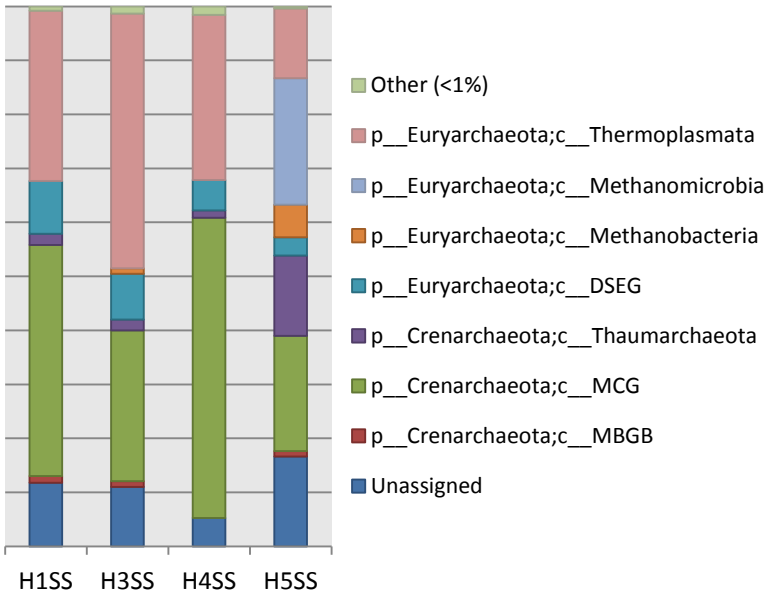


Figure 8.9: Class composition of archaeal communities in superficial sediment samples from the port of Heraklion. Data are reported as relative abundance; archaeal Classes with relative abundance < 1% were summed up and merged in category "Other".

Figure 8.10 reports the relative abundance of the *phyla* identified in the anoxic sediments from the port of Heraklion. Stations H1, H3, and H4 had a very similar composition; the bacterial community was dominated by *phylum Proteobacteria*, *Acidobacteria*, and *Actinobacteria.* Again, bacterial community in samples collected from station H5 had a *phylum* composition markedly different from the other stations, presenting lower proportion of Acidobacteria and Proteobacteria, but higher proportion of *Spirochaetes* and *Firmicutes.*

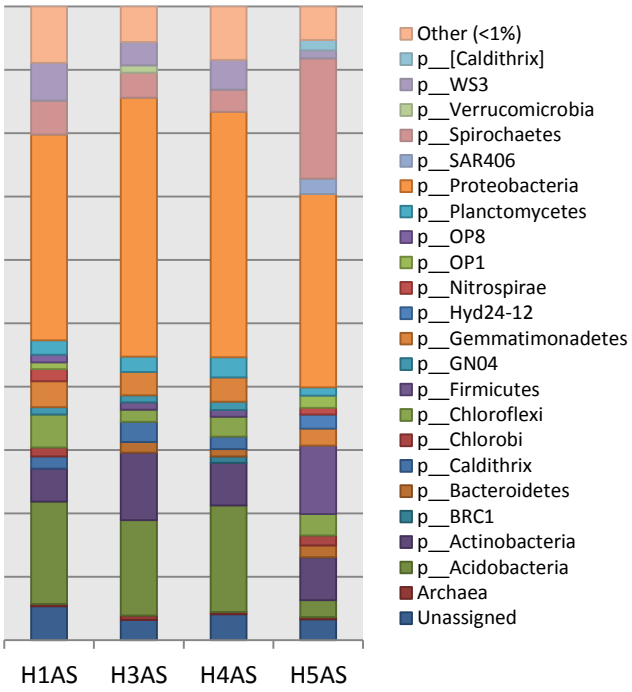


Figure 8.10: *Phylum* composition of bacterial communities in anoxic sediment samples from the port of Heraklion. Data are reported as relative abundance; bacterial *Phyla* with relative abundance < 1% were summed up and merged in category "Other".

Figure 8.11 reports the relative abundance of the classes identified in the anoxic sediments from the port of Heraklion. The archaeal communities collected from anoxic sediments of station H1 (Leisure boats), H3 (Passenger ships), and H4 (Cargo ships) were fairly similar, being dominated by class *MCG* (*phylum Crenarchaeaota*) and *Thermoplasmata* (*phylum Euryarchaeota*). The archaeal communities in samples

collected from station H5 (Shipyard), on the contrary, showed an highly different class composition, presenting lower proportion of class *MCG* (*phylum Crenarchaeaota*) and *Thermoplasmata* (*phylum Euryarchaeota*), being dominated by class *Thaumarchaeota* (*phylum Crenarchaeota*), and, finally, being the sole sample containing OTUs belonging to the *Methanomicrobia* class (*phylum Euryarchaeota*)

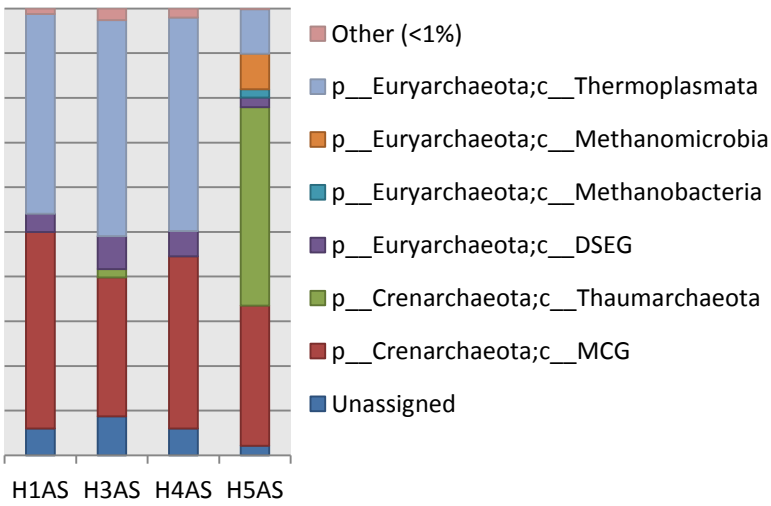


Figure 8.11: Class composition of archaeal communities in anoxic sediment samples from the port of Heraklion. Data are reported as relative abundance; archaeal Classes with relative abundance < 1% were summed up and merged in category "Other".

Ordination analysis

Figure 8.12 and Figure 8.13 report the results of PCoA ordination analysis on bacterial and archaeal communities, respectively. At first ordination was performed including all the samples (fig 8.12. and fig 8.13 panel a). As above reported, samples from station H5 showed high dissimilarity respect to all other samples. This observation is confirmed in ordination analysis, were they formed a clearly separated cluster in the ordination space. In this regard, they can be considered outliers and were removed to improve the ordination of the other samples (fig 8.12 and fig 8.13 panel b). Samples formed loose, separated clusters on the basis of port of origin, which showed some overlapping, especially in archaea community. Inside each port-cluster, samples exhibited a certain level of separation on the basis of matrix; this is particularly evident in the case of bacterial communities in the port of Cagliari, as already pointed out by ANOSIM analysis.

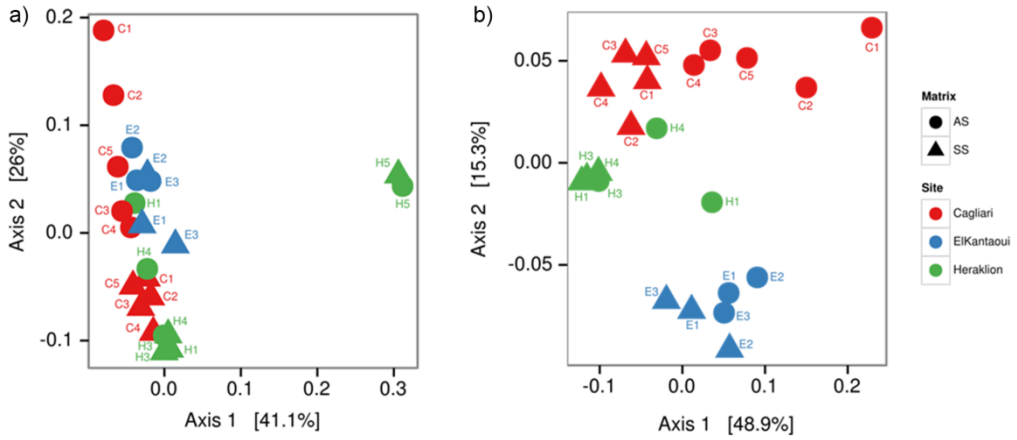


Figure 8.12: a) PCoA ordination of bacterial communities in sediments from the three ports, based on Weighted UniFrac distance. The percentage of the variation in the samples described by the plotted principal coordinates is indicated on the axes. b) PCoA ordination of bacterial community in sediments from the three ports after removal of samples from station H5

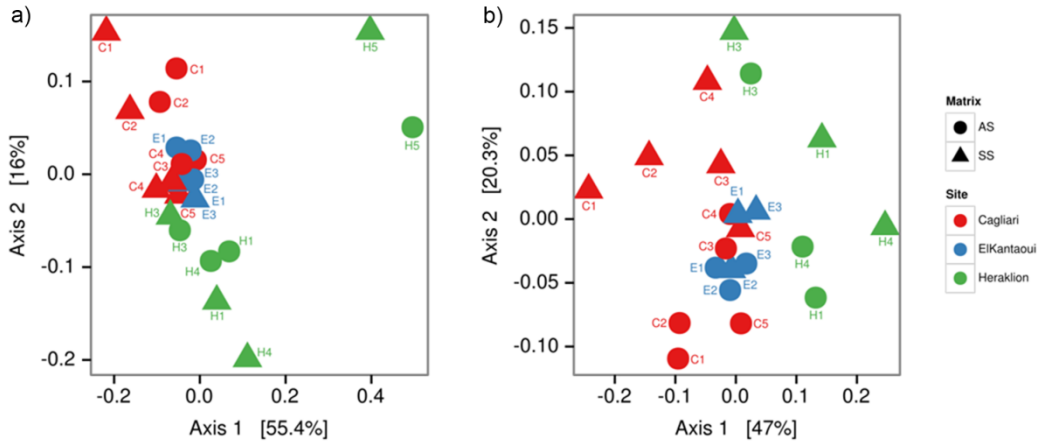


Figure 8.13: a) PCoA ordination of archaeal communities in sediments from the three ports, based on Weighted UniFrac distance. The percentage of the variation in the samples described by the plotted principal coordinates is indicated on the axes. b) PCoA ordination of bacterial community in sediments from the three ports after removal of samples from station H5

8.3 Conclusions

The Mediterranean Sea is a well-studied ecosystem. It has been proposed as a biodiversity hot spot, hosting about 17 000 marine species belonging to the Eukarya domain and a number, at present impossible to estimate, of Bacteria and Archaea species (Coll et al. 2010; Mapelli et al. 2013). It is a semi-enclosed basin (e.g. the typical water retention time in Mediterranean is of the order of 70–90 years) (Daffonchio et al. 2012) divided into two main sub-basins by the Sicilian Channel, the western and eastern Mediterranean Sea, with extremely complex water mass circulation and dynamics (Coll et al. 2010; Mapelli et al. 2013). Those particular conditions give rise to a multitude of habitats across an east-west and north-south distribution, which have to be taken into account when attempting to evaluate its biodiversity. Nevertheless, the particular environments studied in this work, the port basins, are partially closed habitats, characterized by peculiar conditions and particular pollution sources that represent a major driving force for the composition and diversity of microbial communities respect to basin-level patterns of variability.

Our results indicate that microbial communities in sediments from the three ports were different (ordination analysis in Figure 8.12 and 8.13) and that port of origin constituted the main grouping factor (table 8.2). However, the overlapping of the port-clusters in the ordination, the low values of Bray-Curtis dissimilarity (table 8.3 and 8.4), and the low values of R statistic in ANOSIM analysis indicate that the overall level of diversity was low. Nevertheless, ports separated from each other in the ordination space after removing of samples from station H5. The removing of station H5 from the analysis was done because within port analysis (Figures 8.8 - 8.11) showed that bacterial and archaeal communities composition in superficial and anoxic sediments from this station, was very different from that of the other stations in Heraklion port. This difference was attributed to the peculiar use of the H5 station as a Shipyard. The new ordination could reflect some very general feature of the ports. In fact, despite their geographic position, the dimension and type of maritime traffic between port of Cagliari and Heraklion were surely more similar respect to the port of El Kantaoui (a small harbor with only leisure boat). These finding supports the observation that port basin represent a peculiar habitat unconnected with basin-level patterns of variability in the Mediterranean Sea. On the contrary, archaeal communities behaved differently and the ordination showed a less evident clusterization, with broad overlapping between samples from port of Cagliari and El Kantaoui.

Bacterial communities compositions from sediments of the three study sites were largely in accordance to that reported by other works (Chiellini et al. 2013; Todorova et al. 2014; Sun et al. 2013; Zhang et al. 2008). In particular, our results are similar to that in Sun et al. (2013) and Zhang et al. (2007), which identified a similar composition,

dominated by Proteobacteria (with Gammaproteobacteria being the most represented, followed by Deltaproteobacteria and Alphaproteobacteria) and with Actinobacteria, Acidobacteria, Bacteroidetes, Chloroflexi, and Firmicutes being the other prevalent phyla. Initially, analysis of community composition (fig 8.6 and 8.7) suggested that microbial communities in sediments from the port of Heraklion were highly different from those in other ports. This seemed particularly true for Archaea. Further analysis on this port showed that most of the differences could be attributed to the peculiar community found in sediments from station H5, rather than to a whole port level. When removing samples from station H5 in fact, the community composition (data not shown) in port of Heraklion was fairly similar to that in the other two ports. Moreover, microbial communities in sediments from this station showed a significantly lower level of diversity (fig 8.4 and 8.5). These observations imply that bacterial and archaeal communities in sediments from station H5 were heavily impacted by the activities carried out in this station.

Station H5 was a shipyard, and the activities linked to this use are characterized by pollutants and impacts strongly different from the rest of the touristic port environment (http://toxtown.nlm.nih.gov/text_version/locations.php?id=54). A shipyard is a facility for shipbuilding, repair, maintenance, and shipbreaking. Activities carried out include painting, surface preparation, tank cleaning, and all aspects of marine construction. A variety of toxic chemicals are used in shipbuilding, including chromium, copper, nickel, and lead. Moreover, routine ship engine maintenance requires handling engine fluids such as oil, antifreeze, and hydraulic fluids, and, before repairs, a ship's ballast tank may be emptied of dirty water into the surrounding waters. Ship surfaces are treated with antifoulant materials to clean barnacles and other organisms from the hull and coated to preserve their steel. Cleaning and coating activities use chemicals that include heavy metals, solvents, copper, and hazardous or flammable materials. Because shipyards are located on the water, pollution created by those activities can fall into the water directly or be carried in by runoff, and ultimately exert their toxic effects on marine microbial communities.

8.4 Bibliography

Chiellini, C. et al., 2013. Bacterial communities in polluted seabed sediments: A molecular biology assay in leghorn harbor. The Scientific World Journal, 2013.

Claesson, M.J. et al., 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic acids research, 38(22), p.e200.

Clarke, K.R., 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology, 18, pp.117–143.

Coll, M. et al., 2010. The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PloS one, 5(8), p.e11842.

Daffonchio, D. et al., 2012. ULIXES, unravelling and exploiting Mediterranean Sea microbial diversity and ecology for xenobiotics' and pollutants' clean up. Reviews in Environmental Science and Bio/Technology, 11(3), pp.207–211.

DeLong, E.F., 1992. Archaea in coastal marine environments. Proceedings of the National Academy of Sciences of the United States of America, 89(12), pp.5685–9.

Foti, M.J. et al., 2008. Bacterial diversity and activity along a salinity gradient in soda lakes of the Kulunda Steppe (Altai, Russia). Extremophiles : life under extreme conditions, 12(1), pp.133–45.

Li, J.-W. et al., 2012. SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics (Oxford, England), 28(9), pp.1272–3.

Mapelli, F. et al., 2013. Biogeography of planktonic bacterial communities across the whole Mediterranean Sea. Ocean Science, 9(4), pp.585–595.

Oksanen, J. et al., 2013. Package "vegan." R package ver. 2.0–8, p.254.

Rees, G.N. et al., 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms : application of multivariate statistics. , pp.339–347.

Sun, M.Y. et al., 2013. Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. Environmental Microbiology, 15, pp.2517–2531.

Todorova, N.H., Mironova, R.S. & Karamfilov, V.K., 2014. Comparative molecular analysis of bacterial communities inhabiting pristine and polluted with polycyclic aromatic hydrocarbons Black Sea coastal sediments. Marine pollution bulletin, 83(1), pp.231–40.

Yu, Z. et al., 2008. Evaluations of different hypervariable regions of archaeal 16S rRNA genes in profiling of methanogens by Archaea-specific PCR and denaturing gradient gel electrophoresis. Applied and environmental microbiology, 74(3), pp.889–93.

Zhang, W., Ki, J.S. & Qian, P.Y., 2008. Microbial diversity in polluted harbor sediments I: Bacterial community assessment based on four clone libraries of 16S rDNA. Estuarine, Coastal and Shelf Science, 76, pp.668–681.

Chapter 9

Investigating the use of T-RFLP technique for monitoring microbial populations in sediments from ports of the Mediterranean Sea Basin.

9 Investigating the use of T-RFLP technique for monitoring microbial populations in sediments from ports of the Mediterranean Sea Basin

Main objective of MAPMED project was to evaluate, optimize, and transfer tools for the sustainable development of tourist port areas, with regard to monitoring and reduction of marine pollution, especially the one from hydrocarbons. Final stakeholder of those activities will be a series of subjects which, at various levels, are involved in the management and development of port areas, or that are economically correlated to it (i.e. public bodies, such as port authorities, national or regional authorities, coast guard; and private bodies, such as shipping companies and touristic activities). To accomplish those objectives, a suitable monitoring technique must be identified. Despite the significant drop in cost during the latest years, and despite its great potential in analyzing environmental communities, the NGS analysis with the MiSeq platform was not considered the best choice for this scope for two reasons:

1) A single run on the MiSeq can produce an incredibly high amount of sequences (up to 50 million of paired-end reads). Microbial ecologists have taken advantage of those high yields to simultaneously analyze a high number of samples (multiplexing); in our case, for example, 60 samples were multiplexed in a single run. In this way, the overall cost-effectiveness of the analysis was improved, while maintaining a high depth (number of sequences) for each sample. During a monitoring plan, however, a limited number of stations in each port could be sampled with a certain temporal frequency (e.g. monthly or weekly), and, as the NGS analysis need to be completed soon after any sampling campaign, the use of the MiSeq for monitoring purpose could result in a low cost-effectiveness, as each analysis would include only a low number of samples.

2) NGS analysis on the MiSeq is not an easy task and requires expertise both in molecular laboratories practices and in bioinformatics. The machine itself is still not widespread between sequencing facilities, making it difficult to find a National service provider. Taken together those facts make a monitoring plan based on sequencing, difficult to be efficiently transferred to the port authorities.

On the other hand, T-RFLP, as a well-established DNA profiling technique, recently re-evaluated and validated, offers a valuable alternative sequencing technologies (van Dorst et al. 2014), allowing for a rapid and cost-effective characterization of microbial communities. The shortcomings highlighted for the use of NGS analysis in monitoring plans are not a problem for T-RFLP analysis. T-RFLP, in fact, is a rather easy and inexpensive technique, which does not need particular bioinformatics expertise. Systems for capillary electrophoresis are normal equipment of sequencing facilities.

In chapter 8, the ecological characterization of bacterial and archaeal communities in the tree ports was presented, along with an evaluation of their between and within ports variations. Exploiting the high amount of data that NGS technologies can yield, chapter 8 provided a deep picture of the diversity, identifying the port of origin as the main grouping factor, differentiating, in some cases, communities in superficial from anoxic sediments, and spotting the peculiarities in station H5. In this chapter, the results of T-RFLP analysis on three markers (16S of bacterial communities, 16S of archaeal communities, and dsrAB gene) will be presented with the aim of identifying the most suitable one in picturing spatial and temporal pattern of diversity in the three ports. The patterns presented in chapter 8 were obtained through NGS analysis, a technique which resolution can reach the level of microbial species and that can survey the community at a definitely higher depth respect to T-RFLP. For those reasons, the conclusion drawn in the previous chapter will be considered a better representation of reality, and will be compared to those obtained from the T-RFLP approach.

9.1 Material and methods

Sampling

The case study sites of the present work were three tourist ports located across the Mediterranean Sea Basin: Cagliari (Sardinia, Italy) in the western part, El Kantaoui (Tunisia) in the central part, and Heraklion (Crete, Greece) in the eastern part. Port selection was also based on the different elements, such as categories of maritime traffic, port dimension and existing information on pollution. The maritime traffic inside the three tourist ports is represented by recreational boats, passenger ships and fishing vessels. In addition to the marine traffic, the major pollution sources related to the three port areas are wastewater discharges into the sea, river mouth, fuelling stations and fishing activities.

Three sampling campaigns were organized during winter, at the beginning of the tourism period, and at the end of the tourism period (Table 9.1). The selected

sampling stations were chosen according to their characteristic position within the port area and their related use (i.e. touristic, industrial, cargo area) (Table 9.2). Samples were labelled with a common nomenclature: the first number in sample labels refers to sampling campaign, the letter refers to the sampling site (C: Cagliari, E: El Kantaoui, H: Heraklion), the second number refers to the sampling station.

The sediment samples were collected using a specially designed box corer (dimensions: 13.5 x 13.5 x 16 cm) operated by hand from a boat. The box corer sampler was equipped with a modular expandable handle, consisting of maximally 8 poles of 2 m length attached to each other which could allow for sampling down to about 14 m. Sediment subsamples were extracted on board from the box corer sample using three ethanol treated plastic corers (4.4 cm Ø). One the shore, each sediment core was aseptically sliced into sections: surface sediments (SS) were collected from the first centimeter whereas anoxic sediments (AS) were collected in the black zone (evidence of redox conditions), discharging the intermediate sediments (2-4 cm). (Figure 9.1) Sediment samples from each core were transferred into sterile plastic bags and accurately mixed. Surface and anoxic sediment samples were used for DNA extraction for molecular analyses. Sediment sampling and processing were operated by the team of the Hellenic Centre for Marine Research and the team of microbiologists of the University of Cagliari.

Table 9.1: Sampling campaigns in the three port areas.

Id	Campaign (Year 2012)	Cagliari Sardinia, Italy	El Kantaoui Tunisia	Heraklion Crete, Greece
1	**Winter**	13-15 Feb	17-18 Feb	03-06 Mar
2	**Spring** **before tourism period**	28-31 May	04-05 Jun	19-21 May
3	**Summer** **after tourism period**	19-22 Sep	24-25 Sep	14-16 Sep

Table 9.2: Sampling stations in the three port areas.

Station	Cagliari (C)	El Kantaoui (E)	Heraklion (H)
1	Leisure boats	Leisure boats	Leisure boats
2	Intermediate	Fuel station	-
3	Passenger ships	Port entrance	Passenger ships
4	Cargo ships	-	Cargo ships
5	Port entrance	-	Shipyard

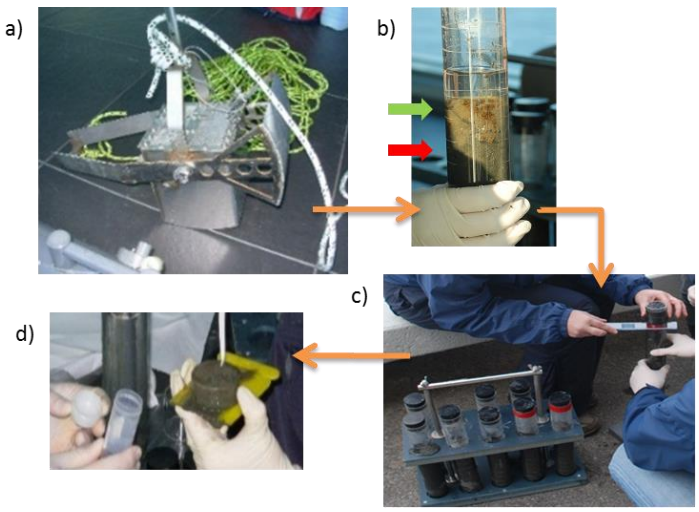


Figure 9.1: Sediment sampling and processing. a): box corer, b): sediment core (green: superficial sediments, red: anoxic sediment), c): slicing of cores, d): sediment sample.

DNA extraction from sediment samples

Sediment subsamples were transferred into sterile vials using sterile volumetric spatulas, stabilized by adding an equal volume of RNAlater® reagent (SIGMA), and stored at 4 °C until shipment to the laboratory at room temperature (Foti et al., 2008). Subsamples (250 mg wet weight) were obtained by aseptically combining an equal amount of sediments from the three cores collected per station and stored in RNAlater reagent (1:1 v/v). The stored reagent was removed prior to DNA extraction through centrifugation. Genomic DNA was extracted using the NucleoSpin® Soil kit (Macherey Nagel), following the manufacturer protocol with minor revisions. DNA concentration was determined on agarose gel. DNA obtained from duplicate extractions was combined.

T-RFLP analysis of 16S rRNA genes in sediments

PCR amplification of 16S rRNA genes was done using the Bacteria-specific primer pair 27f-FAM/1492r (Moeseneder et al. 1999) and the Archaea-specific primer pair 21f-NED/958r (DeLong 1992). The amplification reaction was performed as previously described by DeLong (1992). PCR product was checked for dimension on 0.8% agarose gel and extracted using the QIAquick Gel extraction kit (Qiagen). A web-based analysis tool (E.R.P.A.) provided by the MiCA3 software (Shyu et al. 2007) was used to select the optimal restriction enzymes for T-RFLP analysis. Bacterial 16S rRNA gene amplicons were digested with RsaI and AluI restriction enzymes, while archaeal 16S rRNA gene amplicons were digested with HhaI and AluI. Terminal restriction fragments (T-RFs) were separated by capillary electrophoresis on an ABI 310 genetic analyzer using LIZ 500 as size standard. T-RFs with sizes ranging from 35 to 500 bp and peak heights of more than 35 RFU (Relative Fluorescence Unit) were included in the analysis.

T-RFLP analysis of dsrAB genes

Total genomic DNA was extracted and purified from the sediment samples (250 mg wet weight) as described in paragraph above. Primers DSR1F, labelled at the 5' end with 6-carboxyfluorescein, and DSR4R were used to amplify ~ 1.9 kb of the dsrAB gene as previously described by Pérez-Jiménez & Kerkhof (2005). After amplification, 600 ng of each dsrAB amplicon were digested with MboI (isoschizometer of NdeII). Labelled T-RFs were separated by capillary electrophoresis in an ABI 310 genetic analyzer using LIZ 500 as size standard. Peak normalization was performed as described by Pérez-Jiménez & Kerkhof (2005)

9.2 Results and Discussion

T-RFLP profiles

A comparison between T-RFLP profiles obtained from the three markers is reported in Figure 9.2. Panels on the left indicate the relative area of each T-RF, while panels on the right indicate the number of time each T-RF was found. T-RFLP profiles of the three markers showed variable number of T-RFs. 16S profile of Bacteria communities had a total of 73 T-RFs, 40 of which were observed only in 1 sample (singletons), while 16S profile of Archaeal communities had a total of 90 T-RFs, 40 of which were singletons. dsrAB profiles showed the higher number of T-RFs (130) and the lowest number of singletons (26). Overall, T-RFs had a lower mean relative area in dsrAB profiles, indicating a more even distribution of the eighth/area of the peaks in each profile. Moreover, dsrAB T-RFs appear more uniformly distributed along the full analysis range, from 0 to 500 bp.

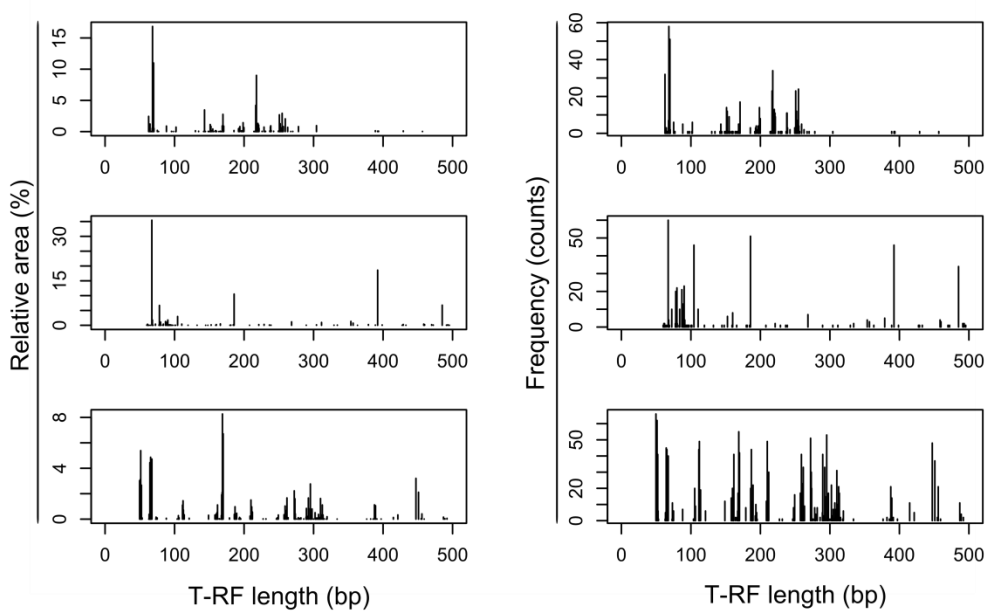


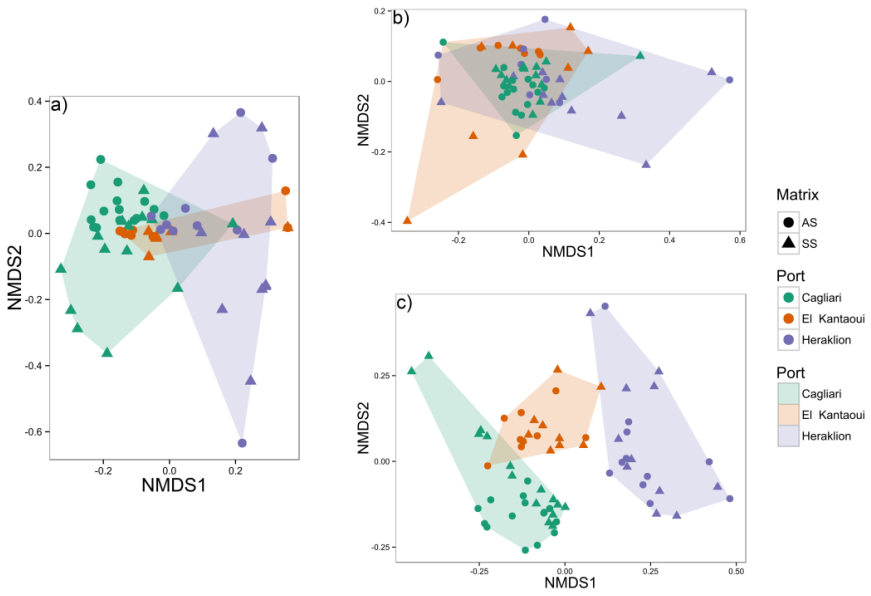
Figure 9.2: Comparison of profiles obtained from T-RFLP analysis on 16S on bacterial communities (top), 16S on archaeal communities (mid), and dsrAB (bottom). Panels on the left illustrate the relative area of each T-RF, while panels on the right indicate the number of time each T-RF was found.

Table 9.3 reports results of ANOSIM analysis, using the Bray-Curtis index as a measure of distance, on the T-RFLP datasets obtained from the three markers. ANOSIM (Rees et al. 2004) is used to test the strength and the significance of the division of the samples in groups; samples were grouped on the basis of predictable experimental factors (i.e. sampling campaign, port of origin, matrix of origin, and station use). Most of the tested grouping resulted significant, however only when samples were grouped on the basis of port of origin, ANOSIM analysis produced R statistic values high enough to support a real separation of the groups. In ANOSIM analysis, in fact, despite the significance, a low value in R statistic suggests a small division of the groups and indicates a low effect of the tested factor.

Table 9.3: ANOSIM results comparing microbial communities from the three markers, grouped based on sampling campaign, port origin, matrix, and station use.

	16S Bacteria		**16S Archaea**		**drsAB**	
Factor	**R**	**p-value**	**R**	**p-value**	**R**	**p-value**
Campaign *(1,2,3,)*	0.100	0.001	0.084	0.002	0.007	0.297
Port *(C,E,H)*	0.297	0.001	0.322	0.001	0.617	0.001
Matrix *(AS,SS)*	0.092	0.001	0.090	0.004	0.057	0.015
Station Use *(table 9.2)*	0.064	0.102	0.082	0.054	0.150	0.001

Figure 9.3 and 9.4 report NMDS analysis on the T-RFLP profile obtained from the three markers. Samples are colored on the basis of port of origin, while symbols represent matrix. Convex hull areas (i.e. the smallest convex polygon formed connecting all the point from a particular environment, treatment, or condition) were drown on the ordination space to highlight samples from the same port (figure 9.3) or the same campaign (Figure 9.4).



Figure

9.3: NMDS ordinations of T-RFLP profiles obtained from 16S marker on bacterial communities (b), 16S marker on archaeal communities (a), and dsrAB marker (c) in sediments from the three ports, based on Bray-Curtis distance. Convex hull highlight samples collected from the same port. a) Stress= 0.124; b) Stress= 0.144; c) Stress= 0.186

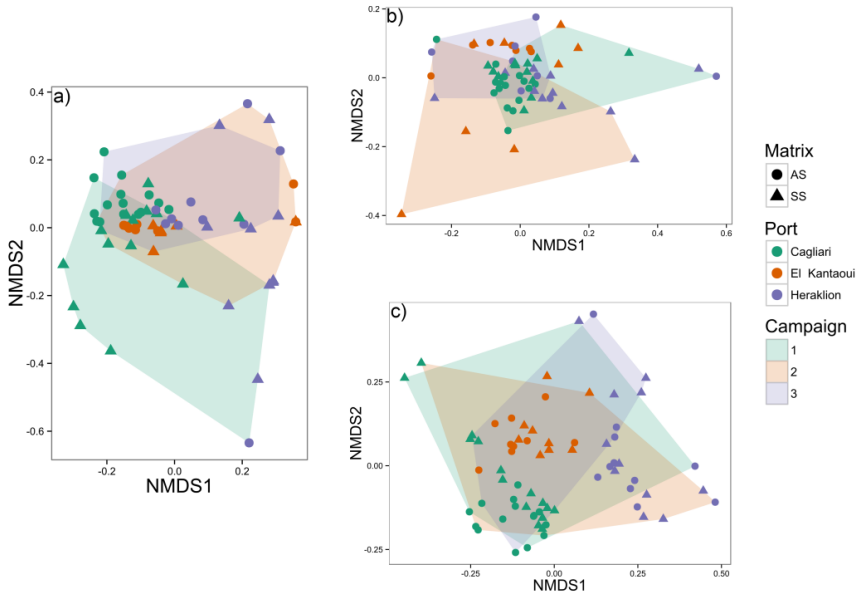


Figure 9.4: NMDS ordinations of T-RFLP profiles obtained from 16S marker on bacterial communities (b), 16S marker on archaeal communities (a), and dsrAB marker (c) in sediments from the three ports, based on Bray-Curtis distance. Convex hull highlight samples collected during the same campaign. a) Stress= 0.124; b) Stress= 0.144; c) Stress= 0.186

As already indicated by ANOSIM analysis, none of the ordination showed a clustering on the basis of sampling campaign, consequently the convex hulls in Figure 9.4 amply overlap. On the contrary, a clustering of samples on the basis of port of origin can be seen in Figure 9.3, as indicated by the separation of port-groups identified by the convex hulls. This division is really poor for the T-RFLP profiles of 16S marker of bacterial community (panel b), but very clear for the T-RFLP profiles of drsAB marker (panel c). Those observations are in complete agreement with ANOSIM analysis results, where 16S marker on bacterial communities showed the lowest R statistic value, while drsAB marker showed the highest.

In comparison with results obtained from NGS data, the ordination of archaeal communities was similar, showing a good separation of clusters of samples from Cagliari and Heraklion ports (green and purple hull in Figure 9.3) and an overlap of the cluster of samples from El Kantaoui and Cagliari ports. On the contrary, the ordination of bacterial community did not show any evident clusterization of samples on the basis of port of origin. However, the biggest difference among ordination from NGS and T-RFLP analysis can be found in the samples from station H5. In the ordination of NGS data, in fact, samples from station H5 exhibited an extremely high diversity, so much that in the ordination space they formed a cluster distant from all other samples. In NMDS ordination of T-RFLP data, on the contrary, samples from station H5 still tend to cluster apart, but the resulting cluster is not so distant from the rest of the samples. Those observations hold true also if the ordination is performed with PCoA instead of NMDS (data not shown).

Dendrogram clusterization

Figure 9.5 shows UPGMA dendrogram clusterization of T-RFLP profiles obtained from the analysis of 16S on bacterial communities (a), 16S on archaeal communities (b), and dsrAB (c) markers. As already observed, samples were clustered mainly on the basis of port of origin (color of the squares). This was particularly evident in the case of profiles obtained from dsrAB.

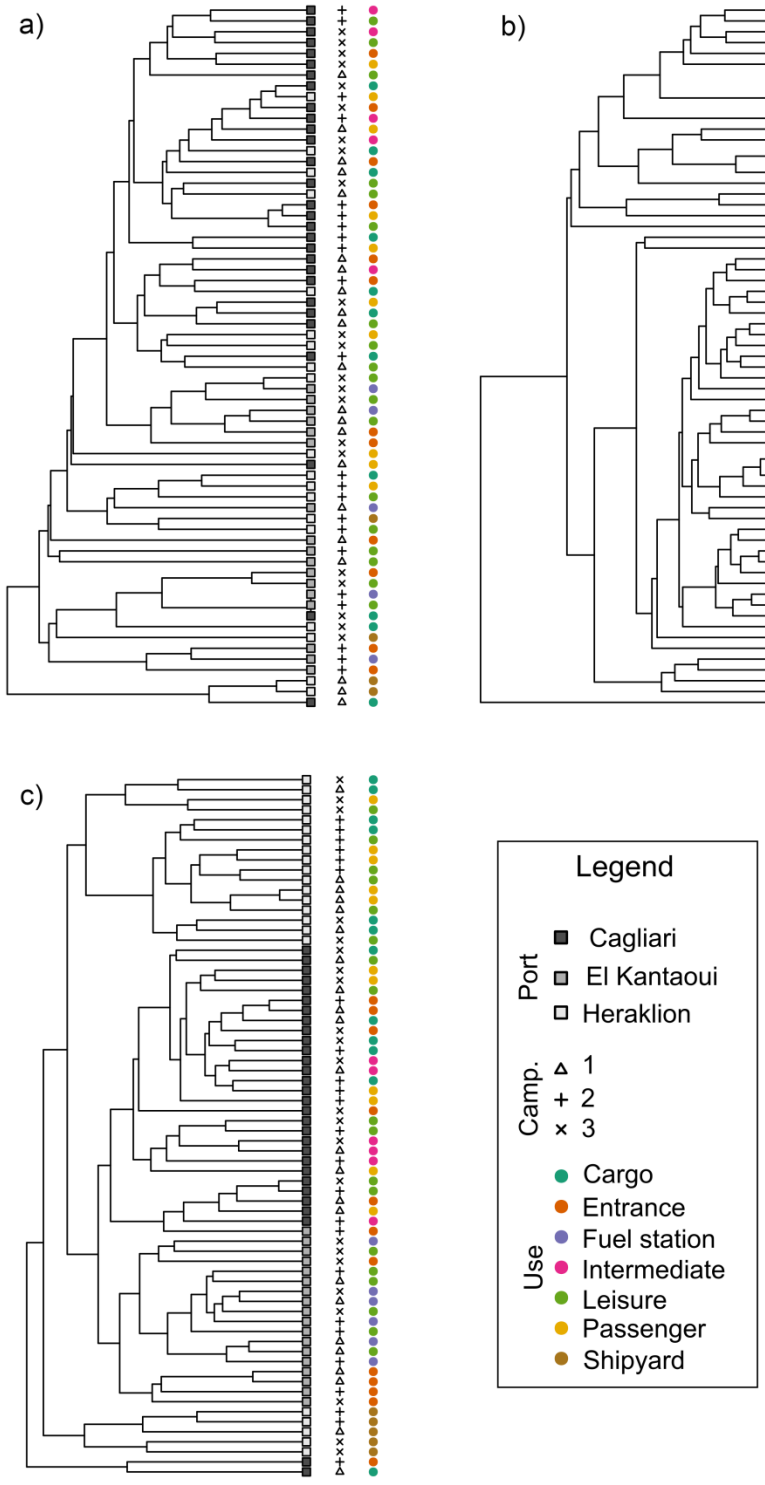


Figure 9.5: UPGMA dendrogram clusterization of T-RFLP profiles obtained from the analysis of 16S on bacterial communities (a), 16S on archaeal communities (b), and dsrAB (c) markers

Other, less evident, levels of clusterization can also be identified. T-RFLP profiles in many cases form clusters from the same port and the same sampling campaign, indicating that, inside the same port, samples from the same sampling campaign are more similar to each other. Differently to ordination analysis, UPGMA clusterization was also able to highlight the peculiarity of archaeal and, to a lesser extent, bacteria communities in station H5 (the station with shipyard use, indicated by brown color of the dots) which was observed with NGS analysis. Samples from this station tend to form separate clusters in the dendrograms. The peculiarity of microbial communities in station H5 is also highlighted in UPGMA analysis of profiles obtained from dsrAB marker. In this case the cluster containing all the samples from station H5 is, in fact, very well defined and separated from the rest of the samples. This result is in accordance with ANOSIM analysis, where station use resulted a significative factor only for dsrAB marker, even if with a low value of R statistic.

Temporal differentiation

To better explore the temporal differentiation of microbial communities, ordination analysis were repeated at the level of single port. Figures 9.6, 9.7, and 9.8 show NMDS ordinations of T-RFLP profiles obtained from the analysis of 16S on bacterial and archaeal communities, and dsrAB markers, respectively. To better visualize temporal clusterization, convex hulls were drawn on the ordination space to highlight samples from the same campaign; numbers indicate the sampling campaign, empty squares indicate AS matrix while black squares SS matrix.

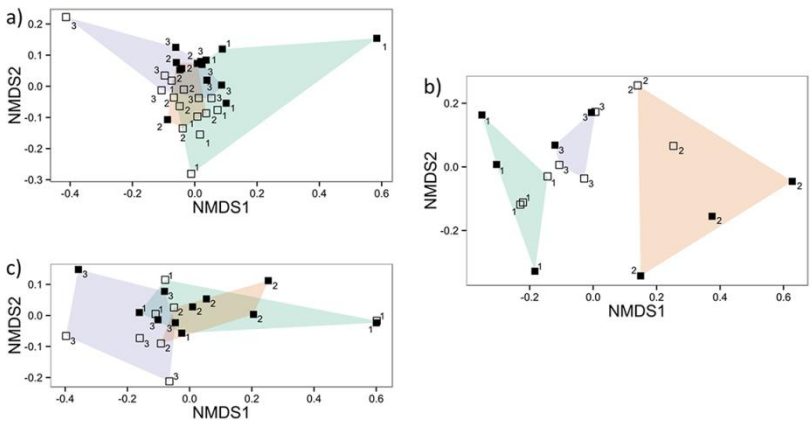


Figure 9.6: NMDS ordinations of T-RFLP profiles obtained from 16S marker on bacterial communities in sediments from the port of Cagliari (a), El Kantaoui (b), and Heraklion (c), based on Bray-Curtis distance. Convex hull highlight samples collected from the same campaign. a) Stress= 0.119; b) Stress= 0.118; c) Stress= 0.074

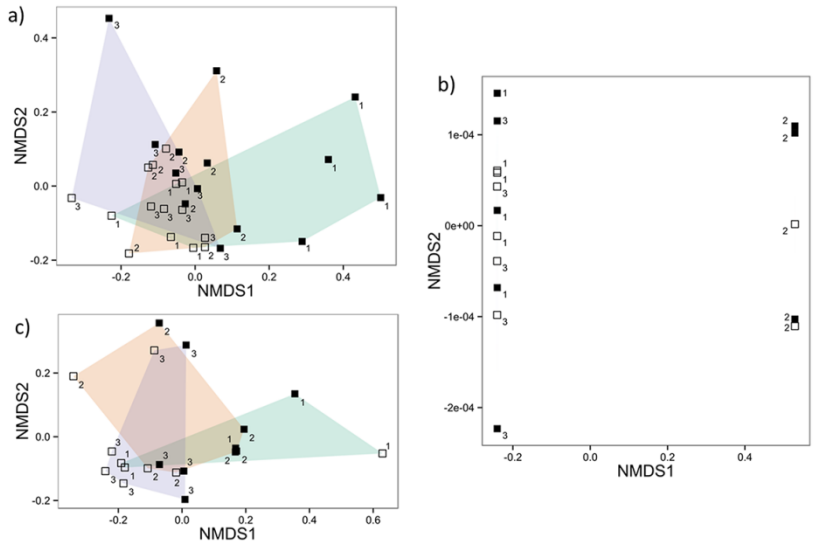


Figure 9.7: NMDS ordinations of T-RFLP profiles obtained from 16S marker on archaeal communities in sediments from the port of Cagliari (a), El Kantaoui (b), and Heraklion (c), based on Bray-Curtis distance. Convex hull highlight samples collected from the same campaign. a) Stress= 0.133; b) Stress= 0.000; c) Stress= 0.114

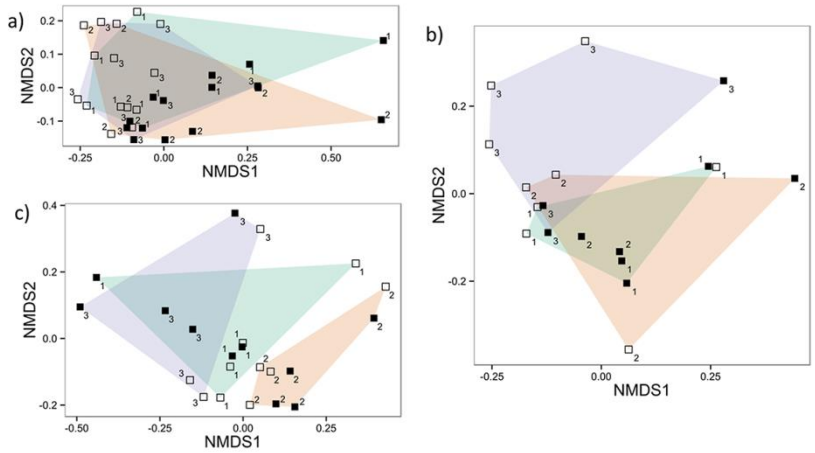


Figure 9.8: NMDS ordinations of T-RFLP profiles obtained from dsrAB marker in sediments from the port of Cagliari (a), El Kantaoui (b), and Heraklion (c), based on Bray-Curtis distance. Convex hull highlight samples collected from the same campaign. a) Stress= 0.129; b) Stress= 0.149; c) Stress= 0.104

The presence of a temporal clusterization of samples, varied from port to port. Microbial communities in sediments from the port of Cagliari (panels a in Figures 9.6, 9.7, and 9.8) never clustered on the basis of sampling campaign; while port of Heraklion (panels c in Figures 9.6, 9.7, and 9.8) showed a temporal clusterization only for dsrAB marker (panel c in Figure 9.8). On the contrary, a net temporal clusterization was visible in the port of El Kantaoui (panel b in Figures 9.6 and 9.7) for 16S marker on bacterial and archaeal communities, but not for dsrAB marker (Figure 9.8). Ordination plots on single port also showed that, generally, the separation between microbial community in superficial and anoxic sediments was fairly clear.

Correlation between T-RFLP and NGS data

Bray-Curtis distance matrices obtained from T-RFLP analysis were compared to those obtained from NGS analysis using Mantel test, which is a statistical test of the correlation between two distance matrices. To do so, Bray-Curtis distance for NGS and T-RFLP data were calculated and, to obtain comparable matrices, the samples number and order was uniformed. Chapter 7 presented results of NGS data analysis; in that chapter PCoA ordinations were performed using Weighted Unifrac as the measure of distance. When, for comparison, PCoA analysis was repeated using Bray-Curtis distance the ordination led to the same conclusions (data not shown). Bray-Curtis distance was the basis on which ordination techniques and clustering analysis of T-RFLP data were constructed; for this reason, measuring their correlation can give further insight on how T-RFLP technique can recapture NGS results.

Results of Mantel test are reported in table 9.4.

Table 9.4: Mantel test with 9999 permutations showing the correlations among Bray-Curtis distance matrices obtained from NGS and T-RFLP analysis.

Comparison	r	p
Bacteria NGS vs Bacteria T-RFLP	0.3564	0.0033
Archaea NGS vs Archaea T-RFLP	0.5495	0.0001
Bacteria NGS vs dsrAB T-RFLP	0.5036	0.0001
Archaea NGS vs dsrAB T-RFLP	0.6088	0.0001

Overall, the correlation was lower for bacterial communities and higher for archaeal ones. The lowest values of r statistics was seen when comparing the distance matrix of bacterial communities from NGS and from T-RFLP (r=0.356, p=0.0033), while the

highest correlation was seen between the distance matrix of archaeal communities from NGS and the distance matrix of drsAB (r=0.6088, p=0.0001). The correlation between T-RFLP data obtained from dsrAB marker and NGS data, both for bacterial and archaeal communities, was higher respect to the correlations obtained between T-RFLP data from 16S marker and the corresponding NGS dataset.

9.3 Conclusion

This work represents a direct follow-up from the one illustrated in chapter 8. In this case, results from the NGS analysis are used to evaluate the effectiveness of T-RFLP approach, and to identify which marker performed better. As already introduced, the T-RFLP approach could represent a better solution for the monitoring of bacterial and archaeal communities in sediments inside a single port, with low numbers of samples (three to five sampling points), thanks to its cost effectiveness, flexibility, and methodological simplicity.

Results were in agreement with what suggested by NGS analysis, and indicated that port of origin was the main factor shaping sediment's microbial communities. However, the three markers behave differently in highlighting this difference, and a clusterization on the basis of port of origin is clearly visible only for dsrAB marker (ordination plot in Figure 9.3 and R values in ANOSIM analysis in table 9.3). Moreover, dsrAB marker gave the best correlation with NGS results. The peculiarity of microbial communities in station H5, highlighted with NGS analysis, was generally not well caught by T-RFLP analysis, although in dendrogram of dsrAB marker those samples where clearly clustered together. This discrepancy could be directly related to the resolution of the analysis technique; apparently the peculiarity of microbial communities from station H5 unfolds at a phylogenetic level (e.g. the species level) that is not resolved by T-RFLP analysis. Dendrogram clusterization and ordination analysis on single port suggests that, at least for the port of El Kantaoui, sampling campaign was also an important factor. Those results could reflect some very general feature of the ports. In fact, type of maritime traffic between port of Cagliari and Heraklion was similar, and hypothetically more stable through the year. On the contrary, port of El Kantaoui is exclusively a touristic port; for this reason heavy changes in boat traffic, and in the impacts related to it, are expected during the tourism period. Moreover, results from this port, highlighted a higher similarity in microbial communities from sampling campaign in winter (1, sampled in February) and after the tourism period (3, sampled in September), respect to microbial community in

spring (2, sampled in June). The fact that the second campaign for the port of El Kantaoui was conducted in June, when the tourism period was supposedly already started, could explain those results.

In conclusion, T-RFLP was able to recapture main results from NGS fairly well, and the Mantel test showed that the two analyses were always significantly correlated. In this regard, the marker that performed better was dsrAB, which showed the highest correlation to NGS, and the most clear ordination and clusterization on the basis of port of origin. For those reasons, T-RFLP could be effectively be used as the routine technique in the monitoring of marine pollution in sediments of touristic ports from the Mediterranean Sea Basin, although its effectiveness should be exceptionally corroborated by NGS analysis.

9.4 Bibliography

DeLong, E.F., 1992. Archaea in coastal marine environments. Proceedings of the National Academy of Sciences of the United States of America, 89(12), pp.5685–9.

Van Dorst, J. et al., 2014. Community fingerprinting in a sequencing world. FEMS microbiology ecology, 89(2), pp.316–30.

Moeseneder, M.M. et al., 1999. Optimization of Terminal-Restriction Fragment Length Polymorphism Analysis for Complex Marine Bacterioplankton Communities and Comparison with Denaturing Gradient Gel Electrophoresis Optimization of Terminal-Restriction Fragment Length Polymorphism Analysis. Applied and environmental microbiology, 65, pp.3518–3523.

Pérez-Jiménez, J.R. & Kerkhof, L.J., 2005. Phylogeography of Sulfate-Reducing Bacteria among Disturbed Sediments , Disclosed by Analysis of the Dissimilatory Sulfite Reductase Genes (dsrAB). Applied and environmental microbiology, 71(2), pp.1004–1011.

Rees, G.N. et al., 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms : application of multivariate statistics. , pp.339–347.

Shyu, C. et al., 2007. MiCA: a web-based tool for the analysis of microbial communities based on terminal-restriction fragment length polymorphisms of 16S and 18S rRNA genes. Microbial ecology, 53(4), pp.562–70.

Chapter 10

General conclusions and future perspectives.

10 General conclusions and future perspectives

The application of molecular biology techniques to microbial ecology has surely widened our capability to describe, analyze, and monitor microbial communities in natural environments. Nowadays, as technical advance in the field of next generation sequencing accumulate at a fast pace, we are able to analyze microbial communities at a previously unimaginable depth. In this work, the MiSeq sequencing was carried out with an unusual protocol, which envisioned the simultaneous use of two different custom primers designed to specifically amplify hypervariable regions of bacterial and archaeal 16S rRNA. Ultimately, the peculiarity of this approach was the presence of two different sequencing primers during each of the two reads on the MiSeq platform. The success achieved in this work is per se of great interest in the field of microbial ecology, as it allowed the simultaneous phylogenetic characterization of bacterial and archaeal communities, on a high number of samples, using kingdom-specific primers. The effective low "universality" of the so-called universal primers, used in PCR to recover the totality of bacterial and archaeal diversity, is in fact a well-known bias in any technique involving an amplification step. For this reason, whenever the bacterial/archaeal abundance ratio is not of primary interest, the use of kingdom-specific primers, that are designed to match a lower number of 16S rRNA sequences in databases, could better represent real diversity. Future research could further extend this concept, and verify if the numbers of simultaneously analyzed traits of the community could be increased. In this way a researcher would be able to perform a broader phylogenetic (Bacteria, Archaea, and Fungi) and metabolic simultaneous characterization of a community. High multiplexing capability would be guaranteed by the use of a different index for each marker-sample combination. The possibility of simultaneously characterizing multiple phylogenetic and metabolic traits is of great interest to obtain detailed information on microbial communities in a particular moment, for example a certain point in time, or after some sort of stimulation of the microbial community.

In this thesis, some principles of molecular analysis of microbial communities were covered, presenting the application of two molecular techniques and the results of two case studies of microbial communities in natural environments affected by anthropogenic modifications. Being a ubiquitous presence in the environments, and a fundamental part in ecosystem functioning, microbial communities are likely affected by any anthropogenic activity, and monitoring their response to anthropic impacts is of primary interest. Results showed that in both case studies the diversity of microbial communities was indeed affected by human activities. In the first case study, the

change in land use from natural to cultivated forest, determined a modification in the structure and a reduction in α-diversity of microbial communities. In this regard Fungi were more sensitive than bacteria as indicator of forest conversion effects. In the second case, the low level of diversity in microbial communities among different ports, led to the speculation that port basin represent a peculiar habitat unconnected with Mediterranean basin-level patterns of variability. Moreover, a particular and highly diverse microbial community was found in sediments from a shipyard in the port of Heraklion, indicating a strong impact of the activities carried out in shipyards on microbial communities.

In both cases, microbial communities were analyzed with both T-RFLP and NGS. In comparing results from the two approaches, NGS technique was considered to better represent reality, due to its higher analysis depth. Generally, T-RFLP led to conclusions similar to that derived from NGS, and for this reason was considered a valuable technique in the monitoring of microbial communities in touristic port basin. This finding will be of great importance during the upcoming phases of MAPMED project, in which port of El Kantaoui was selected as case study site for a field scale demonstration of (bio)remediation technology in the water compartment. Subject of MAPMED project, in fact, was the characterization of microbial communities from the three ports both in sediments and water compartment, the latter not covered in this thesis. Moreover, physic-chemical data on waters and sediments were collected (e.g. metal and hydrocarbons concentration), but could not be used in this thesis due to copyright agreement inside the MAPMED consortium. Future works will explore the correlation of microbial communities data with the actual pollutant in sediments and waters.

11 Funding

The presented PhD thesis falls within the framework of project "**MA**nagement of **P**ort areas in the **MED**iterranean Sea Basin (MAPMED)", funded by ENPI CBC MED *Cross-Border Cooperation in the Mediterranean* Programme. This publication has been produced with the financial assistance of the European Union under the ENPI CBC Mediterranean Sea Basin Programme. The contents of this document are the sole responsibility of UNIFI and can under no circumstances be regarded as reflecting the position of the European Union or of the Programme's management structures.

MAPMED pursues its objectives through an integrated multidisciplinary approach based on the skills and know-how of the scientists, technicians, socio-economic and legal experts involved in the implementation of the activities. University of Cagliari, beneficiary of the funding, participates with the Dept. of Civil-Environmental Engineering and Architecture (DICAAR), Dept. of Biomedical Sciences (DiSB), and Dept. of Law (DIPGIUR). Partners of the project are also University of Florence - Dept. of Biology, Institute of Marine Biology, Biotechnology and Aquaculture - Hellenic Centre for Marine Research, Autonomous Region of Sardinia – Head Office Regional Agency of the Sardinian River basin district, Institute of Graduate Studies and Research - Alexandria University, Faculty of Science of Tunis.

This PhD thesis was also held as part of research activities of the University of Florence – Department of Biology within the LIFE08/NAT/IT/342 DEMETRA (**DE**velopment of a quick **M**onitoring index as a tool to assess **E**nvironmental impacts of **TRA**nsgenic crops) project, founded by the LIFE programme of the European Community. Project coordinator Dr. Cristina Vettori (IBBR UOS FI – CNR).

AKNOWLEDGMENTS

First I would like to thank my supervisors, Prof. Giorgio Mastromei and Prof. Enrico Casalone who gave me the opportunity of this PhD, supporting and supervising me through the planning and realization of the experimental work, and through the writing of this thesis with extreme interest and attention. My gratitude also goes to Cristina Indorato, for teaching all the laboratory stuff through my Master degree internship and my PhD, and to my lab mates in the first years of this journey, Giuliana and Cesarea.

I would like to thank all the other PhD students and researchers at the Department of Biology (well… not the entire department, but the guys at Prof. Fani's and Prof. Bazzicalupo's lab) for hours of entertainment and science during lunch break and in office.

My deep gratitude also goes to Alessio Mengoni and Elena Tamburini, for their willingness to clarify my doubts, and for enlightening science discussion.

Last but not least I would like to thank my parents, without whom all of this could not have been possible in the first instance; for their constant support and belief, and for giving me the possibility to live the life I wanted, but that where never thanked enough. Finally, a warm thanks goes to Carolina for staying by my side during the last year, supporting (and suffering) me especially during the writing of this thesis.