

Microbial diversity in the municipal composting process and development of detection methods

Jenni Hultman

Department of Ecological and Environmental Sciences
Faculty of Biosciences
and
Institute of Biotechnology
and
Viikki Graduate School in Biosciences
University of Helsinki

Academic Dissertation in Environmental Ecology

To be presented for public examination with the permission of the Faculty of Biosciences of the University of Helsinki in the Auditorium of Lahti Science and Business Park, Niemenkatu 73, Lahti, on 23.1.2009 at 12 o'clock noon.

Supervisors: Docent Petri Auvinen
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

Professor Martin Romantschuk
Department of Ecological and Environmental Sciences
University of Helsinki
Lahti, Finland

Reviewers: Academy professor Kaarina Sivonen
Department of Applied Microbiology and Chemistry
University of Helsinki
Helsinki, Finland

Academy professor Jari Valkonen
Department of Applied Biology
University of Helsinki
Helsinki, Finland

Opponent: Professor Heribert Insam
Department of Microbiology
University of Innsbruck
Innsbruck, Austria

Custos: Professor Martin Romantschuk
Department of Ecological and Environmental Sciences
University of Helsinki
Lahti, Finland

ISSN 1239-1271
ISBN 978-952-10-5182-1 (paperback)
ISBN 978-952-10-5183-8 (PDF, <http://ethesis.helsinki.fi>)

e-mail: jenni.hultman@helsinki.fi

Layout: Essi Lindahl
Printed: Yliopistopaino 2009, Helsinki

CONTENTS

LIST OF ORIGINAL ARTICLES

ABSTRACT

ABBREVIATIONS

1. INTRODUCTION	1
1.1 Microbial diversity.....	1
1.2 Ribosomal RNA genes (rRNA).....	1
1.3 Diversity of Bacteria and Fungi.....	2
1.4 Composting.....	4
1.5 Microbes in the composting process.....	5
1.6 Methods of studying microbial diversity.....	6
1.6.1 DNA sequencing.....	6
1.6.2 DNA fingerprinting methods.....	7
1.6.3 Diagnostic microarrays.....	7
1.6.4 Ligation detection reaction (LDR).....	8
2. AIMS OF THE STUDY	10
3. MATERIALS AND METHODS	11
3.1 Study sites and sampling.....	11
3.2 Experimental methods.....	12
3.2.1 DNA extraction, PCR and clone library construction.....	12
3.2.2 Sequencing the clones with high-throughput sequencing pipeline.....	12
3.2.3 Sequence and phylogenetic analysis.....	13
3.2.4 Probe design for macroarray and array hybridisation.....	13
3.2.5 LDR probe design, ligation and hybridisation.....	13
3.2.6 LDR data-analysis.....	13
3.2.7 Additional sequence similarity analysis.....	14
4. RESULTS	15
4.1 Fungal and bacterial clone libraries and sequencing.....	15
4.2 Fungal diversity in samples (I, III, IV).....	16
4.3 Fungal community composition and similarity between communities.....	16
4.4 Fungal succession in the composting processes.....	18
4.5 Bacterial diversity and succession (II).....	19
4.6 Macroarray (III).....	21
4.7 Microarray (IV).....	22
4.7.1 Specificity.....	22
4.7.2 Sensitivity.....	22
4.7.3 Application of the microarray to environmental samples.....	23
5. DISCUSSION	24
6. CONCLUSIONS	30
7. ACKNOWLEDGEMENTS	31
8. REFERENCES	31

LIST OF ORIGINAL ARTICLES

The thesis is based on the following articles, which are referred to in the text by their Roman numerals.

- I** **Jenni Hultman**, Tuija Vasara, Pasi Partanen, Jukka Kurola, Merja Kontro, Lars Paulin, Petri Auvinen, Martin Romantschuk (2009). Determination of fungal succession during municipal solid waste composting using a cloning-based analysis. Manuscript submitted.
- II** Pasi Partanen, **Jenni Hultman**, Lars Paulin, Petri Auvinen and Martin Romantschuk (2009). Bacterial diversity at different stages of the composting process based on analysis of the 16S rRNA gene. Manuscript submitted.
- III** **Jenni Hultman**, Miia Pitkäranta, Martin Romantschuk, Petri Auvinen and Lars Paulin (2008). Probe-based negative selection for underrepresented phylotypes in large environmental clone libraries. *Journal of Microbiological Methods* 75:457-463.
- IV** **Jenni Hultman**, Jarmo Ritari, Martin Romantschuk, Lars Paulin and Petri Auvinen (2009). Universal Ligation-Detection-Reaction Microarray applied for Compost Microbes. *BMC Microbiology* *in press*.

Author's contribution to each publication

- I** JH took the samples, prepared and sequenced the clone libraries and did the data analysis. JH wrote the manuscript.
- II** JH participated in data-analysis and writing the manuscript. PP prepared and sequenced the clone libraries and did the data analysis. He also drafted the manuscript.
- III** JH did the laboratory experiment and wrote the manuscript with contributions from MP.
- IV** JH participated in the design of the study and coordinated the clone library preparation, did part of the microarray experiments, analysed the data and drafted the manuscript. JR wrote the R-script for the microarray analysis, did part of the microarray experiments, analysed the data and drafted the manuscript.

ABSTRACT

Composting refers to aerobic degradation of organic material and is one of the main waste treatment methods used in Finland for treating separated organic waste. The composting process allows converting organic waste to a humus-like end product which can be used to increase the organic matter in agricultural soils, in gardening, or in landscaping. Microbes play a key role as degraders during the composting-process, and the microbiology of composting has been studied for decades, but there are still open questions regarding the microbiota in industrial composting processes. It is known that with the traditional, culturing-based methods only a small fraction, below 1%, of the species in a sample is normally detected. In recent years an immense diversity of bacteria, fungi and archaea has been found to occupy many different environments. Therefore the methods of characterising microbes constantly need to be developed further.

In this thesis the presence of fungi and bacteria in full-scale and pilot-scale composting processes was characterised with cloning and sequencing. Several clone libraries were constructed and altogether nearly 6000 clones were sequenced. The microbial communities detected in this study were found to differ from the compost microbes observed in previous research with cultivation based methods or with molecular methods from processes of smaller scale, although there were similarities as well. The bacterial diversity was high. Based on the non-parametric coverage estimations, the number of bacterial operational taxonomic units (OTU) in certain stages of composting was over 500. Sequences similar to *Lactobacillus* and Acetobacteria were frequently detected in the early stages of drum composting. In tunnel stages of composting the bacterial community comprised of *Bacillus*, *Thermoactinomyces*, Actinobacteria and *Lactobacillus*.

The fungal diversity was found to be high and phylotypes similar to yeasts were abundantly found in the full-scale drum and tunnel processes. In addition to phylotypes similar to *Candida*, *Pichia* and *Geotrichum* moulds from genus *Thermomyces* and *Penicillium* were observed in tunnel stages of composting. Zygomycetes were detected in the pilot-scale composting processes and in the compost piles. In some of the samples there were a few abundant phylotypes present in the clone libraries that masked the rare ones. The rare phylotypes were of interest and a method for collecting them from clone libraries for sequencing was developed. With negative selection of the abundant phylotypes the rare ones were picked from the clone libraries. Thus 41% of the clones in the studied clone libraries were sequenced.

Since microbes play a central role in composting and in many other biotechnological processes, rapid methods for characterization of microbial diversity would be of value, both scientifically and commercially. Current methods, however, lack sensitivity and specificity and are therefore under development. Microarrays have been used in microbial ecology for a decade to study the presence or absence of certain microbes of interest in a multiplex manner. The sequence database collected in this thesis was used as basis for probe design and microarray development. The enzyme assisted detection method, ligation-detection-reaction (LDR) based microarray, was adapted for species-level detection of microbes characteristic of each stage of the composting process. With the use of a specially designed control probe it was established that a species specific probe can detect target DNA representing as little as 0.04% of total DNA in a sample. The developed microarray can be used to monitor composting processes or the hygienisation of the compost end product.

A large compost microbe sequence dataset was collected and analysed in this thesis. The results provide valuable information on microbial community composition during industrial scale composting processes. The microarray method was developed based on the sequence database collected in this study. The method can be utilised in following the fate of interesting microbes during composting process in an extremely sensitive and specific manner. The platform for the microarray is universal and the method can easily be adapted for studying microbes from environments other than compost.

TIIVISTELMÄ (abstract in Finnish)

Kompostoinnilla tarkoitetaan biologisesti hajoavan orgaanisen jätteen hajotusta ja se on yksi tärkeimmistä biojätteen käsittelytavoista Suomessa. Mullan kaltaista lopputuotetta voidaan käyttää lannoitteena maataloudessa ja puutarhanhoidossa sekä materiaalina viherrakentamisessa ja maisemoinnissa. Täten jätteenkäsittelymenetelmän ohella kompostointi on myös kierrätysmenetelmä. Kompostointiprosessissa mikrobit ovat tärkeässä roolissa orgaanisen aineen hajottajina. Vaikka kompostoinnin mikrobiologiaa onkin tutkittu jo vuosikymmeniä, mikrobisto teollisessa suuren mittakaavan kompostointilaitoksessa on suurelta osin tuntematon.

Tässä väitöskirjatyössä selvitettiin sienten ja bakteerien esiintymistä teollisessa kompostointilaitoksessa ja pienemmässä tutkimuskäyttöön rakennetussa kompostointirummussa molekyylibiologisin menetelmin. Aiemmin on havaittu, että maljakasvatuksiin perustuvilla menetelmillä saadaan selville vain noin 1 % näytteen mikrobiyhteisöstä. Yli 6000 mikrobien tunnistusalueiden DNA-fragmenttia sekvensoitiin ja tuloksena oli yli 500 erilaista bakteerien tunnistussekvenssiä ja 127 erilaista sienten sekvenssiä. Näitä sekvenssejä verrattiin kansainvälisiin sekvenssitietokantoihin. Sekä bakteerien että sienten havaittiin eroavan aiemmissä tutkimuksissa löydettyistä mikrobeista, vaikka yhteneväisyyksiäkin löytyi. Myös aiemmin tunnistamattomia mikrobisekvenssejä oli runsaasti. Tutkimuksessa havaittiin, että mikrobiyhteisöt teollisen- ja tutkimusmittakaavan laitoksella erosivat selvästi.

Kloonikirjastoissa on usein monta kappaletta samaa sekvenssiä. Väitöskirjatyössä kehitettiin menetelmä, jolla runsaslukuiset sekvenssit merkitään ennen sekvensointia ja vain harvalukuiset sekvensoidaan. Tällöin tietyt yleiset sekvenssityypit eivät estä harvalukuisempien sekvenssityyppien havainnointia. Menetelmän avulla havaittiin, että 41 % kloonikirjastojen klooneista oli harvinaisia ja vain ne sekvensoitiin. Säästö sekvensoinnissa ja käsiteltävän datan määrässä oli täten huomattava.

Kompostin mikrobien esiintymisessä havaittiin eroja prosessin eri vaiheiden välillä. Koska kaikkien mikrobien kloonaaminen ja sekvensointi on aikaa vievää ja kiinnostuksen kohteena on usein vain muutama sekvenssityyppi, kehitettiin näiden sekvenssien havainnointia varten mikrosirudiagnostiikkaan perustuva menetelmä. Mikrosiruja on käytetty lajiston selvittämiseen jo vuosikymmeniä, mutta käytetyt menetelmät ovat olleet puutteellisia herkkyydeltään ja tarkkuudeltaan. Väitöskirjassa kehitetyn mikrosirumenetelmän avulla saavutettiin tarkkuus, jolla kaksi lähisukuista mikrobilajia voitiin erottaa toisistaan. Menetelmä perustuu kahden DNA-koettimen liittämiseen ja työssä suunnitellun kontrollikoettimen käyttöön. Menetelmä osoittautui hyvin herkäksi ja myös pieninä määrinä esiintyvät mikrobit voitiin havainnoida. Mikrosirua voidaan siksi hyödyntää kompostointiprosessin seuraamisessa tai kompostin lopputuotteen hygieenisyyden varmentamisessa.

Tämän väitöskirjatyön tulokset tuovat tärkeää tietoa teollisen kompostointiprosessin mikrobiyhteisöstä. Suunniteltua mikrosirumenetelmää voidaan soveltaa myös muiden ympäristöjen mikrobien tutkimiseen.

ABBREVIATIONS

ARDRA	Amplified ribosomal DNA restriction analysis
ARISA	Automated ribosomal intergenic spacer analysis
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
ITS	Internal transcribed spacer. A non-coding sequence between coding ribosomal RNA gene subunits
HGT	Horizontal gene transfer
LDR	Ligase detection reaction
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PLFA	Phospholipid fatty acid analysis
qPCR	Quantitative polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rRNA	A gene coding form of ribosomal RNA
SSCP	Single strand-conformation polymorphisms
Sp	Species
t-RFLP	Terminal restriction fragment polymorphism

1. INTRODUCTION

1.1 Microbial diversity

Microbes exist everywhere on earth. It has been estimated that the number of microbes on earth is as much as $4\text{-}6 \times 10^{30}$ (Whitman et al. 1998) and the estimation for the number of different bacterial species in a gram of soil is 10^7 (Gans et al. 2005). They affect our health and agriculture, and have key roles in most of the planet's geochemical cycles (Curtis & Sloan 2005). The high diversity of microbes has been revealed through the rapid development of sequencing techniques. The human gut has been found to be coated with microbes essential to human health (Eckburg et al. 2005). Vast microbial diversity has been found in the seas where a huge number of novel genes have been detected (Venter et al. 2004). Likewise, great diversity and many novel sequences have been discovered everywhere from farm soil to whale falls, i.e. whale carcasses that have fallen to the ocean floor (Tringe et al. 2005), as well as in the air (Brodie et al. 2007). Recently, the basaltic lava in the oceanic lithosphere has been studied and the microbial diversity there (Santelli et al. 2008) was comparable to terrestrial soil. Terrestrial soil, itself, is considered to be extremely diverse because of its spatial heterogeneity, multiphase nature and complex chemical and biological properties (Daniel 2005), which provide growth conditions for diverse groups of microbes.

In many studies the number of analysed clones is usually small, from tens to hundreds while the number of individual microbes in the sample can be billions (reviewed in Whitman et al. 1998). It is now widely accepted that with cultivation based methods less than 1% of microbes are detected because the growth conditions do not represent the growth conditions of all the microbes present (Amann 1995). In other words, there is no growth condition or culture media that is general enough to allow growth of all microbes in the studied samples. For example, their nutrient, pH and

temperature requirements may be strict, while anaerobic microbes do not form colonies in aerobic conditions, and slow growing microbes can be out-grown by fast growers, which produce large colonies. In many cases it is thus possible to design the conditions so that a specific group of organisms is targeted, however in practise this is not a viable approach if the total microbial diversity of an environment needs to be revealed. Furthermore, microbes can be mutualistic, symbiotic and parasitic and for these microbes growth conditions are hard to mimic. Since the development of polymerase chain reaction (PCR, Saiki et al. 1985) to amplify DNA, the molecular methods for studying microbes and the knowledge about microbial diversity have substantially increased.

1.2 Ribosomal RNA genes (rRNA)

The genes coding for ribosomal RNA (rRNA) have been broadly used in studying microbial diversity with molecular methods. Ribosomal RNA molecules are ubiquitous in cellular life forms. The small subunits, which are often utilised, are small enough to be handled in the laboratory, easy to amplify and sequence, and most importantly, they show the evolutionary changes between different organisms (Torsvik & Øvreås 2006). The genes coding for bacterial and archaeal small and large ribosomal subunits, 16S and 23S rRNA genes, respectively, and eukaryal 18S and 28S rRNA genes, contain both conserved and variable regions, and the differences in these genes can be used to infer the relationships between DNA sequences from different species (Van de Peer et al. 1996). Several primers have been developed for the PCR amplification of the ribosomal RNA gene regions from bacteria, fungi and archaea (e.g. Edwards et al. 1989, White et al. 1990, Jürgens et al. 1997). Because of the superior sensitivity of PCR there is no need to cultivate the microbes as the gene of interest can be amplified directly

from the sample. The high conservation of rRNA genes has led to extensive use of them in microbial diversity studies and in June 2008 there were 86 2343 bacterial, 82 771 fungal and 65 295 archaeal ribosomal small subunit RNA gene sequences in the EMBL sequence database (searched with words [bacteria & 16S], [fungi & 18S] and [archaea & 16S], respectively). For fungi the internal transcribed spacer region (ITS), located between the 18S and 28S rRNA genes, is often used because the 18S rRNA sequences may not provide enough sequence resolution for species level identification (Hugenholtz & Pace 1996, Anderson & Cairney 2004, Bruns 2006). The ITS regions are transcribed but processed out of the transcript (Hibbet 1992) and therefore contain more variation than in the rRNA genes from which the ribosomes are formed.

However, there are problems in the use of ribosomal genes for diversity studies. Ribosomal genes are present in operons and the number of rRNA genes can be up to 220 in fungi (Hibbet 1992). In published bacterial genomes, the number of ribosomal operons varies between 1 and 15 (Acinas et al. 2004a). In archaeal genomes usually only one ribosomal operon was found (Acinas et al. 2004a). Results from studies based on rRNA are therefore not quantitative as one copy of an rRNA gene does not correlate to one microbe. There can also be intraspecies differences in rRNA resulting in 16S rRNA sequences from the same species with different nucleotide compositions (Hibbet 1992, Acinas et al. 2004b). The definition of species in microbial diversity studies has been widely discussed and it is a fundamental problem that is not yet resolved. Clonally amplifying organisms like bacteria, archaea and some fungi, cannot be defined as species based on the same criteria as, for example, animals, and basing such definitions on sequence similarity is somewhat arbitrary. This is because the organism is not always extracted and characterised, and in fact the studied unit is a bare gene sequence.

Therefore it is not possible to specifically discuss species but instead sequence diversity. As a result, in sequencing studies the term species is often replaced by the definition of operational taxonomic unit (OTU), ecotype or phylotype. The molecular characterisation of organisms has brought novel information to the taxonomy of microbes, which has been in continuous evolution. Tree of life projects on the phylogeny and biodiversity of bacteria and fungi have been and are constantly under development (Ciccarelli et al. 2006, Lutzoni et al. 2004). These trees are constructed with phylogenetic models from which the organisms can be classified based on similarity, for example in rRNA genes.

1.3 Diversity of Bacteria and Fungi

Bacteria are unicellular organisms without a nucleus that are found in a wide range of environments and environmental conditions. In 1987 Carl Woese published the 12 bacterial divisions for the Bacterial kingdom and ten years later the number of genera had nearly tripled to 33 divisions (Pace 1997, Hugenholtz et al. 1998). An expansion to 54 divisions (Handelsman 2004) was published as sequencing techniques developed and the price of sequencing came down. A recent bacterial taxonomy consists of 105 phyla in the EMBL sequence database. In the past Archaea were considered to be part of the prokaryotes in the classification of Eukaryotes and Prokaryotes, but as the Bacteria and Archaea were found to be systematically distinct, they were separated into kingdoms of their own in the three kingdom classification by Woese (1987, see Fig. 1).

As mentioned above, species definition in microbes is not straightforward. In addition, in bacteria horizontal gene transfer (HGT) and homologous recombination affect the gene content and thus the systematics. A similarity of 96% in 16S rRNA genes has often been considered to distinguish two bacterial species.

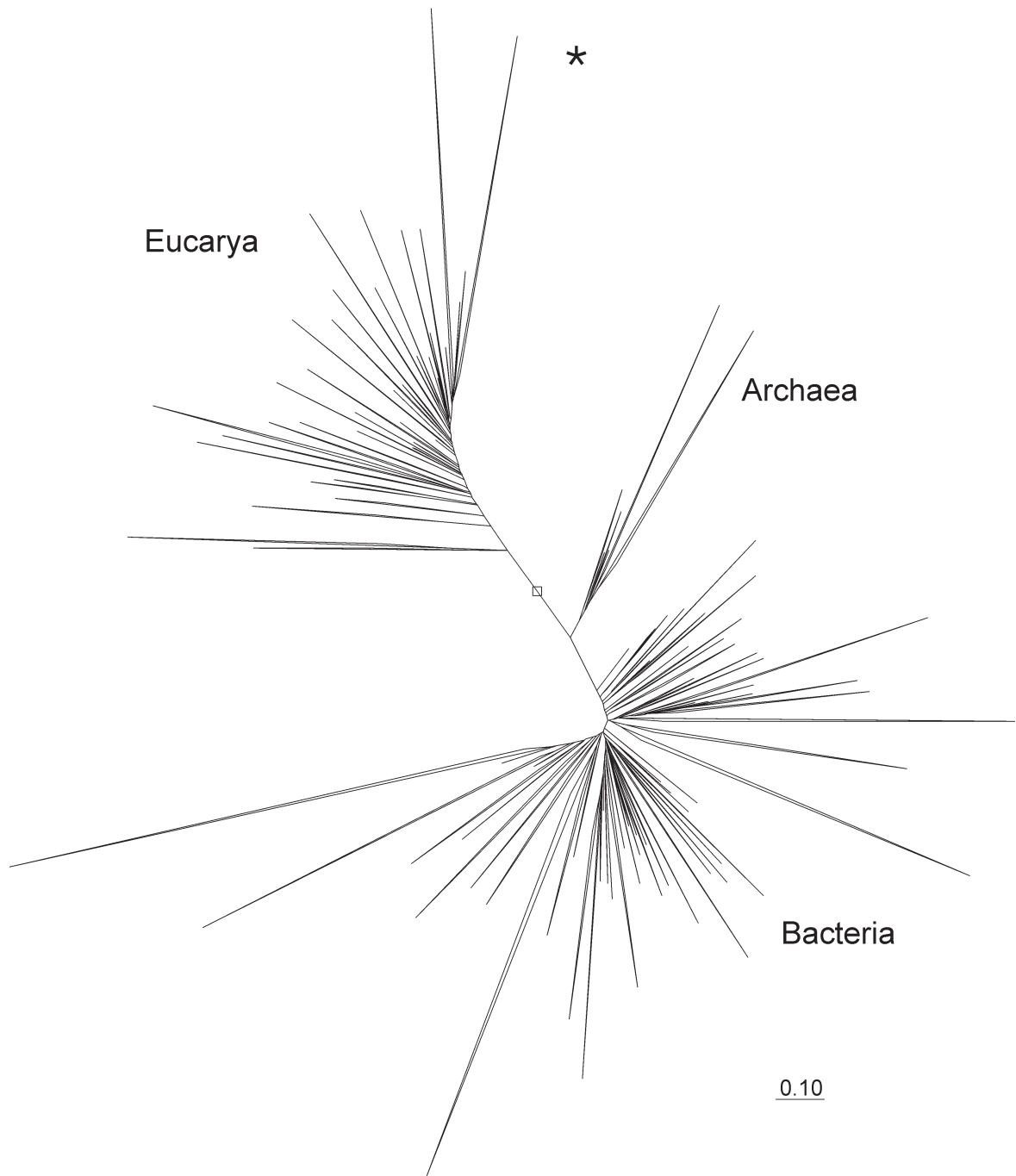


Figure 1. The universal phylogenetic tree built from ribosomal small subunit sequences in the Arbo-Silva 96 database (Pruesse et al. 2007). The branch, in which humans, among other vertebrates and invertebrates, belong to based on this phylogenetic tree, is marked with an asterisk (*). The scale bar indicates 0.10 changes per position. The box in the middle of the tree represents the origin of the tree.

It should be noticed, that the interoperonic differences in 16S rRNA operons within fully sequenced bacterial genomes were found to differ <1% and HGT was assumed to occur between closely related organisms (Acinas et al. 2000a). Ciccarelli et al. (2006) found that HGT had occurred in ribosomal proteins of one class among the 191 species studied. However, when genes of three cultivated *Escherichia coli*-bacteria that were 99% similar by 16S sequencing were compared they had only 39.2% of genes in common and therefore were considered to represent different ecotypes (Welch et al. 2002). Thus it must be kept in mind that similar 16S rRNA does not correlate with similar ecological function as these microbes may have specialised to different environments (Hunt et al. 2008). Despite these limitations, ribosomal genes are valuable tools for the study of microbial ecology as long as these limitations are kept in mind.

Fungi were separated from the plant kingdom into their own when the five kingdom classification by Whittaker (1969) was published. In the three kingdoms classification Fungi belong to Eukarya. In Eukarya the taxonomical resolution was higher because of their greater morphological diversity. Therefore two species that would have been assigned to separate phyla in the Eukarya are assigned to the same phylum in the Bacterial classification (Ciccarelli et al. 2006). The classification of fungi by their phenotypic characteristics is difficult in some phyla as those characteristics do not provide enough resolution. As an example, when the phylogenetics of the fungal class Saccharomycetales were studied, ascomyceteus yeasts from the genus *Pichia* were found to extend across the phylogenetic spectrum of the ascomycetous yeasts. The reason for this resides in the classification of these yeasts by phenotypic characteristics, in this case the shape of the budding cell (Suh et al. 2006). The relationship between two species cannot be therefore determined based on the simple, convergent phenotype (Suh et al. 2006) as the resolution is not high enough.

The number of fungal divisions varies depending on the source and the use of molecular methods in systematics has revolutionised the fungal phylogeny. Besides the four widely accepted phyla Basidiomycota, Ascomycota, Chytridiomycota and Zygomycota, fungal systematics were recently revised by sequencing six gene regions from nearly 200 species (James et al. 2006). The number of phyla as classified by Hibbet et al. (2007) is seven, comprising the Basidiomycota, Ascomycota, Chytridiomycota, Glomeromycota, Microsporidia, Neocallimastigomycota and Blastocladiomycota.

The diversity and activities of the soil fungal community are not as well known as those of the soil bacterial communities as mycologists have often relied on culture based methods in ecological studies on soil fungi (Anderson & Cairney 2004). When molecular methods have been used to study soil fungi, several novel phylotypes have been discovered. Vandenkoornhuysen et al. (2002) found 49 fungal species in plant roots of which 7 had a match (>99% similarity) in the sequence database and the amount of novel species was high. Fungi from tundra soil during different seasons were studied by Schadt et al. (2003) and the results from sequencing 125 fungal ribosomal large subunit genes revealed several novel fungal groups as 40% of the sequences clustered into unique groups. A novel clade was discovered containing many previously unknown fungi. Sequences belonging to this novel group were also found by Porter and colleagues (2008). It has been estimated that the number of fungal species on earth is circa 1.5 million (Hawksworth 1991) of which only a small fraction, around 5%, have been characterised so far.

1.4 Composting

Composting refers to aerobic degradation of organic waste into humus-like material. It is not only a waste treatment technique but also a recycling method as the end product can be used in agriculture as fertiliser, in gardening or in landscaping. The composting process can be divided into four different stages based on the

temperature levels (Gray et al. 1971). In the early, mesophilic stages the temperature is close to the ambient temperature and the pH is low. As the material starts to degrade due to microbial activity the amount of mesophilic microbes, often lactic acid bacteria and yeasts, increases. The increase and activity of these acid-producing bacteria cause a drop in pH to acidic levels. As the temperature starts to increase as a result of the vigorous exothermal microbial activity, it reaches levels above 45°C and thermophilic microbes take over the degradation and the pH rises. A falling temperature indicates the start of the cooling phase, which is followed by the maturation phase. Microbes play key roles in all the phases of the composting process as active decomposers.

Presently composting, besides rotting, is the main method of treating separately collected organic waste in Finland. There are 20 large scale composting facilities in Finland. In 10 of these industrial and household biowaste is treated while the remaining 10 treat wastewater sludge. These facilities are based on drum, tunnel, and tower composting technologies. Although the principle of composting is rather easy there are intrinsic problems in industrial composting. As the scale of the process and the volume of material increases controlling the process becomes more difficult. In order to utilise the composted material as a fertiliser in agriculture or as soil enrichment the compost end product should be free from human, animal and plant pathogens that are often abundant during the early stages of composting. Therefore, quality control parameters have been set by the authorities. These include either proof showing that the temperature in the composted mass reached 70°C for 60 minutes or validation of the process by measuring the reduction of growth of certain indicator organisms. Hygienisation of the compost is achieved by microbial succession and competition where the mesophilic pathogens such as *Enterococcus faecalis*, *Clostridium* sp. and various enteric pathogens are killed by high temperatures and the self heating of the

process. When the composting process does not proceed optimally there is a risk that pathogens originating from plant and food residues, for example, will survive and be present in the composted material. External heating of compost has been utilised but when used too extensively it may also cause the extermination of desired degrading microbes. Problems with low temperatures in the process are often linked to high moisture content and low oxygen levels. Anaerobic conditions cause the decomposition rate to decline (Tiquia et al. 1996) and severe odour problems.

Various methods such as DGGE (denaturing gradient gel electrophoresis, Ishii et al. 2000, Steger et al. 2007), SSCP (single strand conformation polymorphisms, Peters et al. 2000), ARISA (automated ribosomal intergenic spacer analysis, Schloss et al. 2003), cultivation (Beffa et al. 1996, Ryckeboer et al. 2003a), restriction analysis and sequencing (Dees & Ghiorse 2001) and microarrays (Franke-Whittle et al. 2005, Franke-Whittle et al. 2008) have been used to study compost microbial communities to find out which microbes are present in certain stages of the composting process. Compost microbiota have been studied from a wide spectrum of different composting scales, from 500 ml laboratory batch units (Schloss et al. 2003), 30 l synthetic compost made from dog food (Dees & Ghiorse 2001), to actual composting drums (Ishii & Takii 2003, Vuorinen & Saharinen 1997) and composting piles (Steger et al. 2007). Unlike normal laboratory scale studies, the composition of municipal biowaste varies from day to day and thus cannot be controlled. In addition, the microbiota at the composting plant provides a seed inoculum for the composting process.

1.5 Microbes in the composting process

In the early phases of composting the microbial succession is extremely rapid. In studies regarding compost bacteria, the material has included artificial biowaste (eg. Dees and Ghiorse 2001), biowaste composted in laboratory scale batch

units (eg. Strom 1985), pilot-scale reactors (eg. van Heerden et al. 2002) and full-scale facilities (eg. Ishii & Takii 2003). In the mesophilic stages bacteria from the genera *Lactobacillus* and *Bacillus* are often detected (Ryckeboer et al. 2003a, Ryckeboer et al. 2003b). In the thermophilic stages the community becomes more thermophilic and thermotolerant bacteria such as Actinobacteria (Fergus 1964), *Bacillus* (Blanc et al. 1997) and *Thermus* (Beffa et al. 1996) become abundant.

Although it has been found that the composting process is most efficient when both bacteria and fungi are present (Gray et al. 1971), the study of compost microbes has mainly focused on compost bacteria. Fungi have been discovered to have an important role in the composting process as degraders of lignin and cellulose (Tuomela et al. 2000). Eukarya have been found to be numerically abundant in the early stages of the composting process (Schloss et al. 2005) as well. In the mesophilic stages, yeasts (Choi & Park 1998) have been detected, and thermophilic fungi such as representatives from the Pezizomycota (Fergus 1964) and Zygomycota (von Klopotek 1962, Kane & Mullins 1973) have been discovered in the thermophilic stages of the process. Basidiomycota become abundant in the cooling and maturation phase of compost (Von Klopotek 1962). Such fungi have been studied with culture-based methods (Ryckeboer et al. 2003b, Anastasi et al. 2004), by fingerprinting (Hansgate et al. 2005), and by sequencing (Peters et al. 2000), but of these the studies based on molecular methods have been rather limited. With cultivation based methods the microbial diversity cannot be fully determined and thus the microbes present in the composting process require more attention.

1.6 Methods of studying microbial diversity

1.6.1 DNA sequencing

The microbial diversity of many different environments has been studied by amplifying

the gene of interest (often ribosomal genes or genes presenting metabolic functions of interest), cloning the amplified fragments and sequencing them. There are different approaches to sequencing; in addition to ribosomal or metabolic genes the sequenced fragment may also be chosen through functional screening or by sequencing active cells or an interesting part of the genome. In functional screening the metabolic activity of clones in the clone library is studied and no sequence information is required. New classes of genes that encode either known or new functions can be identified (Daniel 2005). Active cells can be sequenced with, for example, use of stable isotope probing (SIP). In SIP, microbes utilise an isotopically enriched substrate and the isotope incorporates into either DNA or RNA. The genomes that have the incorporated isotope in them can then be isolated and further studied (Radajewski et al. 2000). In recent years the study of microbial communities has expanded and microbes have been studied with both phylogenetic and metagenomic approaches (Hugenholtz & Tyson 2008). The microbial diversity found in these studies has been high and the results have advanced the study of microbial ecology with novel findings. For example, the archaea have been found to be important in ammonia oxidation (Leininger et al. 2006), which was previously considered to be a process of bacterial origin. Nonetheless, some problems exist in cloning-based approaches in isolating and sequencing gene fragments from environmental samples. First, it is challenging to get a representative sample to study the diversity and abundance of microbes in the sample (Ranjard et al. 2003). It is also difficult to estimate the number of clones that need to be sequenced to reach the desired coverage in a sample. Microbial groups that are present abundantly in samples overshadow the ones present in lower amounts that might not be sampled and sequenced (Curtis & Sloan 2005). In recent years several next generation sequencing techniques, such as the 454 tag sequencing first used for genome sequencing (Margulies et al.

2005), have been developed, in addition to the broadly used Sanger sequencing (Sanger et al. 1977). The rRNA tag sequences were introduced to microbial ecology as short (from circa 100 to 210 bp) tag-reads that were sequenced from environments such as the deep marine biosphere (Sogin et al. 2006, Huber et al. 2007) and soil (Leininger et al. 2006, Roesch et al. 2007). The amount of data gained is massive in these studies, the number of sequences gathered was up to 906 347 (Huber et al. 2007).

As microbes have been characterised from different environments using sequencing based methods, several separate sequence databases have been developed and maintained (Amaral-Zettler et al. 2008). By aligning the sequences from the studied sample to those available from databases it is possible to determine whether the same sequence has been detected from different environments and the taxonomical annotation can also be compared. There is however much redundancy in these databases, and there are also mis-annotated sequences (Nilsson et al. 2006) or sequences of chimeric origin (Hugenholtz & Huber 2003).

The magnitude of microbial diversity is far too large to be understood and it is demanding to sequence every microbe in a sample. At the moment novel sequencing methods require large amounts of funding although prices are decreasing. There have been several methods developed for characterising microbial communities from different environments in a fingerprint-like manner that are economical and easily accessible.

1.6.2 DNA fingerprinting methods

Fingerprinting methods are commonly used in microbial ecology. For example, restriction fragment length polymorphism (RFLP), amplified ribosomal DNA restriction analysis (ARDRA) and DGGE have been used prior to sequencing in order to lower the number of sequenced clones. Terminal restriction fragment length polymorphisms (t-RFLP) and ARISA have been applied to study the community profile in

the samples of interest. Moreover, these methods can be used before sequencing to determine how much sequencing is required with the sample of interest and also which sequencing technique to use, since the diversity in samples varies.

Although these methods give a profile of the community composition which can be utilised in, for example, studying community changes in different locations (Liu et al. 1997, Osborn et al. 2000, Brodie et al. 2003) they underestimate the true diversity. Microbes from different species or genera can share an identical restriction pattern (Dunbar et al. 2001), a single DGGE band can contain several different ribotypes (Costa et al. 2006) and numerically rare phylotypes are generally not detected (Bent and Forney 2008). These methods are therefore not suitable if the aim is to detect rare phylotypes or certain species instead of community profiles. While bacterial species abundance curves generally have log-normal distribution with very long tails (Martiny et al. 2006) and dominating phylotypes make up the minority of the diversity (Curtis & Sloan 2005), capturing the tail-end of communities is difficult with fingerprinting methods. With cloning they can be captured, but in addition, the abundant types in samples are captured in large numbers.

Dot-plotting methods have been applied to the study of microbes in different environments. In these methods species specific probes are blotted on a nylon membrane and the labelled sample is then hybridised on the membrane (e.g. Valinsky et al. 2002). The species present are detected with either colorimetric or radioactive markers. These methods lack specificity and sensitivity and are therefore not in wider use in microbial diagnostics (Bodrossy & Sessitsch 2004).

1.6.3 Diagnostic microarrays

Microbial diagnostic microarrays, also known as phylochips or phylogenetic oligonucleotide arrays, became a popular high-throughput tool for microbial detection from various environments

a decade ago (Guschin et al. 1997). With these microarrays, parallel detection of hundreds to thousands of microorganisms was achieved and they have been applied in the study of microbes in various environments such as soil (Small et al. 2001), landfills (Bodrossy et al. 2003), water (Rich et al. 2008, Vora et al. 2005), compost (Franke-Whittle et al. 2005, Franke-Whittle et al. 2008) and urban air (Brodie et al. 2007). Often in microarray platforms short (18-28 mer) oligonucleotide probes specific to certain species, genus or strains, are printed on a glass slide. In some applications longer (50-70 mer) probes are used. The short probes have been shown to be specific, but only sequences representing more than 1-5% are detected. Longer probes are more sensitive in reaching lower detection limits but on the other hand they are not as specific as the short probes (reviewed in Bodrossy & Sessitsch 2004). Fluorescently labelled total DNA from the sample or a PCR amplified gene of interest is hybridised on the array and the results, spots to which the labelled sample has hybridised, are read with a scanner.

Although attention is paid in microarray probe design to reaching the highest possible nucleotide specificity for the probes, the thermodynamic properties of the probes need to be verified as well. The microarray probes are printed on the same glass slide and therefore their thermodynamic properties must be identical in order to avoid cross hybridisation. In addition, the sensitivity and specificity parameters should be similar for all probes as well. Besides the probe length and the probe and target sequence similarity, there are other factors affecting the hybridisation specificity. These include the GC-content of the probe, position of matching and mismatching nucleotides in the probe and the secondary structures the probe may form. In addition, hybridisation kinetics influence to the specificity. The amount of non-target molecules, short duration of hybridisation and long posthybridisation washes have been shown to reduce the specificity. The specific hybridisation is thus dependent on

the association or disassociation constant of competing correct and incorrect probe-target molecules (reviewed in Koltai & Weingarten-Baror 2008). Since ribosomal genes do not have sufficiently discrimination power for the use of only one probe to detect one species, microarrays with several probes specific for a single species are widely in use in the nested-probe approach with parallel or hierarchical probe specificity (e.g. Loy et al. 2004, Brodie et al. 2007). These microarray platforms are feasible for the study of the microbiota in the sample, but as the microbes cannot be distinguished at species level and as there are problems with detection limits (Loy et al. 2004, Bodrossy et al. 2003, Franke-Whittle et al. 2005, DeSantis et al. 2007), different platforms have been developed for the study of different microbial communities in a specific and sufficiently sensitive manner.

1.6.4 Ligation detection reaction (LDR)

Enzyme assisted detection methods have been used in combination with microarray hybridisation (Busti et al. 2002, Baner et al. 2003, Castiglioni et al. 2004) as the specificity of the oligonucleotide microarray has not been high enough. A ligation detection reaction (LDR) based method has been applied for detecting single base mutations associated with genetic diseases (Landegren et al. 1988, Barany 1991) and this method has been adapted to characterise microbial communities (Castiglioni et al. 2004, Rantala et al. 2008). In LDR two target specific probes are adjacently ligated together by a thermostable ligase if a perfect complementarity between the two probes and the target DNA is achieved (Fig. 2). The probes are designed so that the nucleotide in the junction point of the probes distinguishes the target from other species. The ligation products are linearly amplified using a thermostable ligase (Barany 1991). One of the two probes is fluorescently labelled (Gerry et al. 1999, Busti et al. 2002) while the other probe contains a 3' tag sequence (Zip sequence, Fig. 2) which directs it to the correct address on

the microarray containing a complementary zip code-sequence (c-zip sequence). These zip sequences have uniform hybridisation properties and the same array platform can be used with multiple ligation probe sets. A complementary control probe (c-control probe) sequence is included in each printed spot to help normalising

oligo hybridisation signals over different spots and arrays, and to set limits for detection. In hybridisation the control probe is hybridised to the microarray with ligation products. It has a different fluorescent label in the 5' end and the thermodynamic melting temperature properties are similar to the zip-sequences.

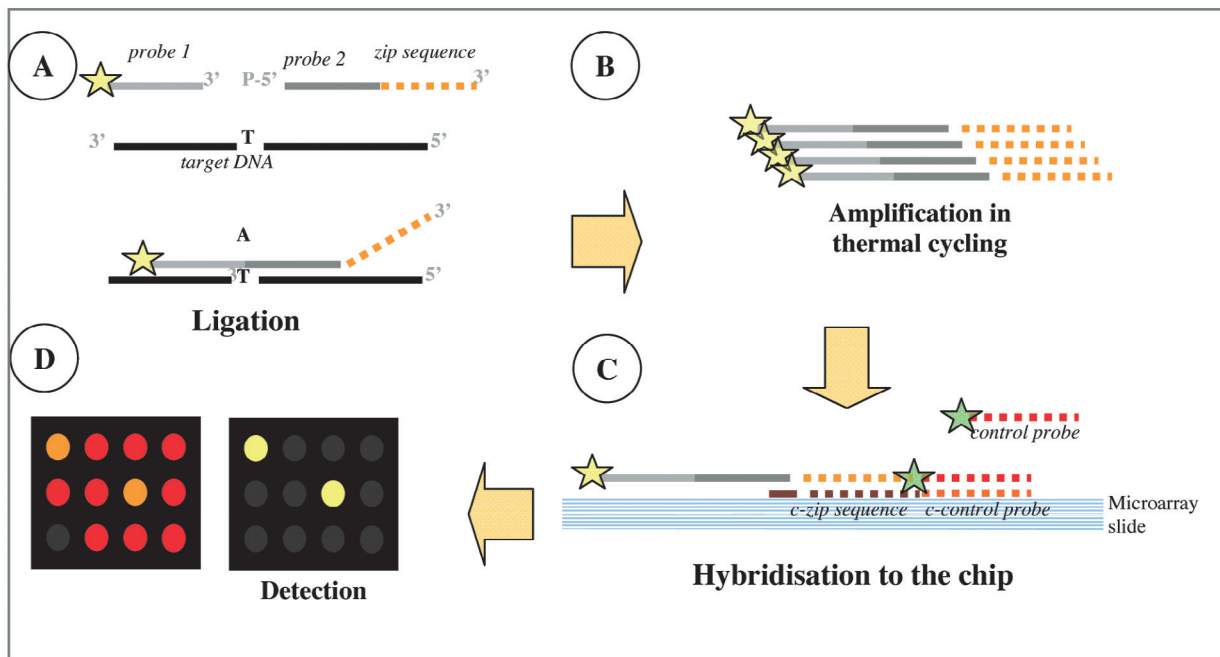


Figure 2. A schematic picture of the ligation detection reaction (LDR) (Barany 1991) and hybridisation to the microarray by the zip-code sequences (Gerry et al. 1999).

- A) Ligation. The ligase enzyme ligates the two probes together if high similarity between target and probe sequences exist.
- B) Linear amplification of the ligation products.
- C) Hybridisation to the microarray with control probe. Ligation products hybridise with the complementary zip sequence (c-zip sequence) and the control probe with the complementary control probe sequence (c-control probe) printed on the microarray.
- D) Detection of results with scanner, different wavelengths are used for the species probe and the control probe.

2. AIMS OF THE STUDY

Microbes degrade the organic waste in compost and the composition of microbes affecting composting was studied. As the process is different in industrial composting facilities where the throughput levels are high and the composted mass differs from day to day, the focus was on studying the microbial communities in these facilities. The general aim of this thesis was to develop a species specific and sensitive detection method for compost microbes. As the microbes in real full-scale composting processes are only partly known, since many of the previous studies have been carried out on a pilot or laboratory reactor scale, the fungi and bacteria in full-scale composting processes had to be characterised before the development of detection methods. The specific objectives were:

- Determination of the microbial community composition in an industrial composting process and to study the differences between compost processes of different scales. In addition, microbes in sub-optimally and optimally operating processes were studied.
- Developing a method by which the communities could be characterised more efficiently.
- To develop a method by which the presence or absence of species of interest could be determined in a specific and sensitive manner. The possibility for multiplex detection of hundreds of different microbes was also one requirement.

3. MATERIALS AND METHODS

3.1 Study sites and sampling

Samples were taken from five different composting plants (Table 1). The composted material was separately collected municipal biowaste, green garden waste and industrial biowaste. The processes were conducted in either drums, drum and tunnel combinations (Fig.1 in paper I) or in windrows. In paper I the effect of wood-ash amendment to the fungal communities in the composting process was also studied.

Table 1. Sampling sites, process scale and the publication the samples were used in.

Composting plant	Scale	Publication
Kiertokapula, Hyvinkää, Finland	full	
feed		I, II
drum (160 m ³) loading-end		I, II
drum (160 m ³), unloading-end		I, II
tunnel (3 pieces, 100 m ³ each)		I, II
drum (160 m ³) loading-end, ash amendment		I, III
drum (160 m ³), unloading-end, ash amendment		I, III
tunnel (3 pieces, 100 m ³ each), ash amendment		I, III
Kujala, Lahti, Finland	pilot	
Drum (5 m ³)		I, II
Drum (5 m ³), ash amendment		I, III
Compost reactor, SLU Sweden ¹	pilot	
Reactor, 200-litre, 55°C		IV
Reactor, 200-litre, 70°C		IV
Ivar, Hogstad, Norway ²	full	
Windrow, 100 m ³ , 55°C		IV
Windrow, 100 m ³ , 65°C		IV
Ämmässuo, Espoo, Finland ³	full	
Windrow, floor area 110 m ³		IV

¹ Samples provided by Cecilia Sundberg, Swedish University of Agricultural Sciences, Uppsala, Sweden

² Samples provided by Erik Norgaard, Norsk Jordförbedring, Norway

³ Samples provided by Christoph Gareis, YTV, Espoo, Finland.

3.2 Experimental methods

Table 2. *Methods used in the publications included in this thesis.*

Method	Publication
DNA extraction from compost samples	I, II, III, IV
Primer design	I
Clone library construction for fungal ITS-region	I, III, IV
Clone library construction for bacterial 16S rRNA gene	II
High throughput sequencing	I, II, III, IV
Phylogenetic analyses	I, II, III, IV
Phospholipid fatty acid analysis (PLFA)	I
Richness and coverage estimators	I, II, III, IV
Ordination analysis	I
LDR probe design	IV
Macroarray probe design	III
Macroarray hybridisation	III
LDR array detection	IV

3.2.1 DNA extraction, PCR and clone library construction

DNA was extracted from compost samples with a bead beating method. For bacterial studies the 16S rRNA gene was amplified with the common bacterial polymerase chain reaction primers pA and pH' (Edwards et al. 1989) located at the 3' and 5' ends of the 16S rRNA gene, creating an amplicon of approximately 1500 base pairs (bp). For the amplification of the fungal internal transcribed spacer (ITS) region, comprising of partial 18S, 5.8S and partial 28S rRNA genes and the flanking ITS1 and ITS2 regions, a new primer was designed (paper I). The amount of knowledge about fungi and the number of fungal sequences in public databases has increased since the development of the often used ITS1 and ITS4 PCR primers (White et al. 1990). Since the ITS1 primer did not amplify all of the fungi found in previous studies, a new universal primer was needed. The new forward primer was situated in the 18S rRNA-gene, approximately 220 base pairs upstream of the beginning of the ITS1 region and used with primer ITS4 (White et al. 1990).

Since the ITS region varies in length between different species the amplicon produced was between 500 and 1200 bp in size. PCR for the samples was done in triplicate and pooled prior to cloning. For some of the samples in paper I semi-nested PCR was used in order to increase the PCR yield sufficiently for cloning. In semi-nested PCR after the first round of PCR the amplified fragments were re-amplified with a pair of primers of which one was used in the first round of PCR and the other was nested.

3.2.2 Sequencing the clones with high-throughput sequencing pipeline

The purified PCR products were cloned into a commercial T/A cloning vector and transformed into competent *Escherichia coli* cells. The plasmids were extracted and the inserts re-amplified with flanking vector specific primers (Fig. 3). Sequencing was conducted with an inner flanking vector primer with fungal clones. For bacteria an inner flanking vector primer and primers situated in the 16S rRNA gene were used because the sequenced region was nearly double in length compared to the fungal clones.

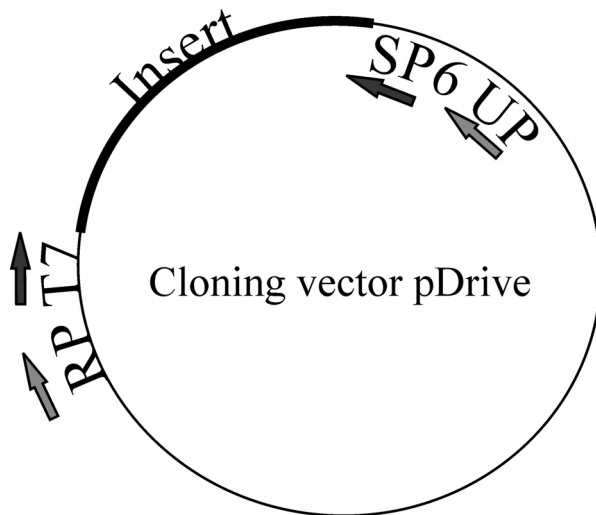


Figure 3. Schematic picture of the PCR and sequencing primer locations in the cloning vector. In the sequencing strategy used the insert was amplified prior to sequencing with vector primers UP and RP. The sequencing was done with inner vector primer T7 and the sequence was verified with inner vector primer SP6.

3.2.3 Sequence and phylogenetic analysis

The sequences were analysed with the Staden package (Staden et al. 2000) programs Pregap, Trev and Gap4 in papers I, II and III and with Phred (Ewing and Green 1998 and Ewing et al. 1998) and Gap4 in paper IV. Sequences with $\geq 99\%$ similarity were joined into contigs. A representative sequence from each contig was sequenced in both directions with vector primers to verify the sequence. The sequences were aligned with either the FASTA (Pearson & Lippman, 1988) or BLAST (Altschul et al. 1990) algorithms to the EMBL and Genebank nucleotide database sequences, respectively, and used in phylogenetic analyses. The sequence alignments were done with either ClustalW (Thompson et al. 1994, I, II, IV) or MUSCLE (Edgar 2004, III). In paper I the phylogenetic analyses were done with the ARB program package (Ludwig et al. 2004) and with the Phylip package (Felsenstein 2006) in papers

II, III and IV. The neighbor-joining distance method (Saitou & Nei 1987) was used in the analyses.

3.2.4 Probe design for macroarray and array hybridisation

The macroarray probes were manually designed for six abundant fungal phylotypes using alignments obtained with ClustalW (Thompson et al. 1994) and Gap4 (Staden package, Staden et al. 2000). The probes were labelled with digoxigenin and hybridised on nylon membranes that had 1536 PCR products from clone libraries gridded on them (Fig. 4). The clones that did not hybridise with the probes were collected for sequencing and sequence analysis was done as described above.

3.2.5 LDR probe design, ligation and hybridisation

The LDR probes were designed using the ARB program package (Ludwig et al. 2004) and a fungal sequence database made from ITS-sequences in public nucleotide databases and in in-house databases (I, III and Pitkäranta et al. 2008). Probes were designed to be species specific so that at least the nucleotide in the junction point of the two probes was unique. A control probe named B3 was designed in addition to the species specific probes and a complementary sequence to this probe was printed on the microarray glass in addition to the complementary zip-code sequence (Fig. 2). Both the ligated probe pairs and the control probe were hybridised on the microarray.

3.2.6 LDR data-analysis

The data were analysed using the R statistical environment (R development core team 2007 and Gentleman et al. 2004). Briefly, the log ratio for the signal from the species probe (LDR) and control probe (B3) was calculated. Missing spots in which printing to the microarray was not

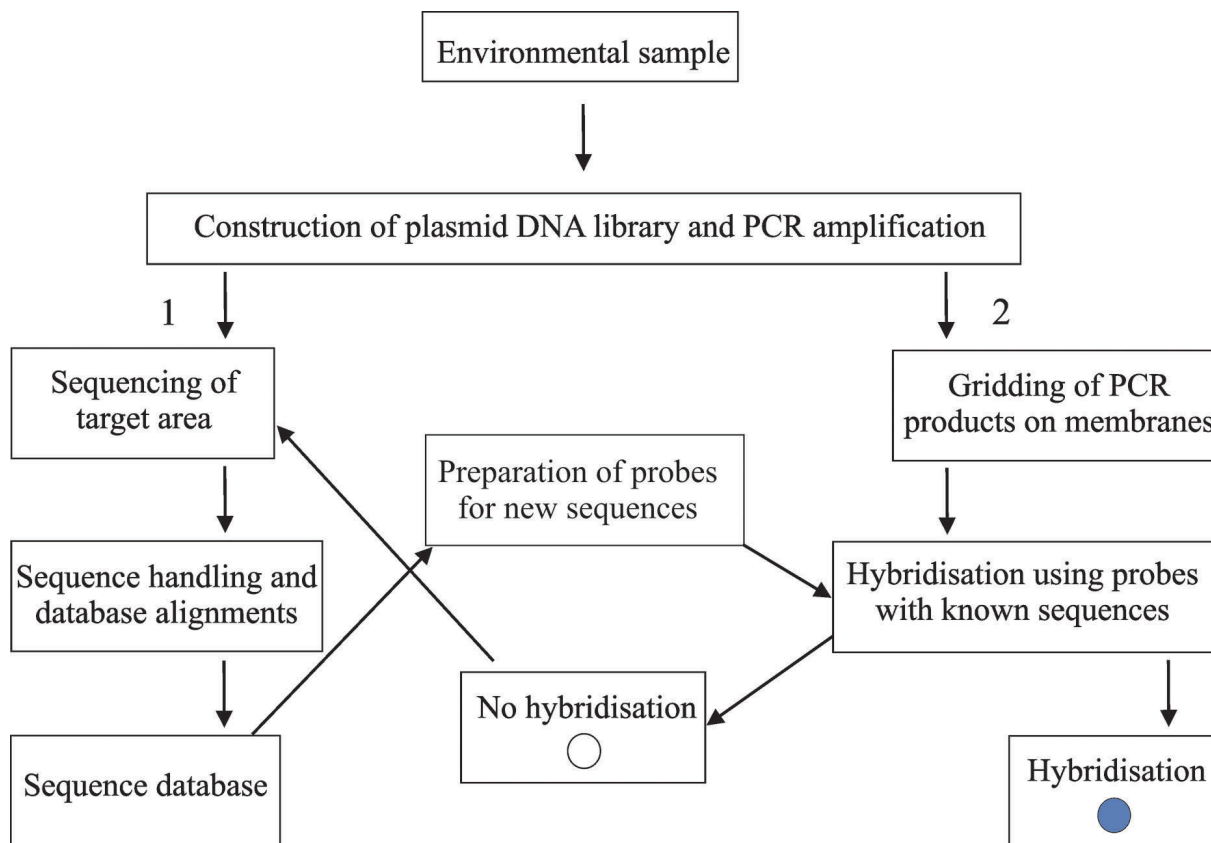


Figure 4. The principle of the macroarray experiment in clone library enrichment by negative selection. The left-hand side (1) describes the clone library construction and sequencing (papers I and II). The right-hand side shows the macroarray protocol. The PCR products were gridded on the membrane and probes specific for the abundant species hybridised on the membrane. The PCR products that did not hybridise with the probes represent the rare phlotypes and were picked for sequencing. As knowledge on the studied environment increases, new probes can be added to the probe pool.

successful were identified as having control probe signal values more than 2 standard deviations below median. Spots with no LDR probe were used as a background set, against which spot replicates of each probe were compared. Spots with ratios over 2.5 SD of background median indicated the detection of the target sequence. Each microarray contained 16 subarrays and thus 16 samples could be tested at a time.

3.2.7 Additional sequence similarity analysis

SONS (Shared OTUs and similarity, Schoss and Handelsman 2006) analysis was conducted

for fungal clone library sequences from study IV. First, a representative sequence from each phylogroup was chosen for alignment in Gap4. The similar phlotypes were joined together and a list consisting of the names of the sequences in each phlotype and library designation files were used in the analysis. The abundance-based Jaccard similarity indices and the nonparametric maximum likelihood estimator (theta, θ) were chosen for the similarity estimations. The abundance-based Jaccard similarity index estimates the fraction of sequences belonging to the same phlotype and community overlap, and the θ index estimates the community structure similarity.

4. RESULTS

4.1 Fungal and bacterial clone libraries and sequencing

The composition of the microbiota in the industrial drum and tunnel composting plants was only partly known. To characterise these microbes clone libraries were constructed from altogether 26 compost samples. The samples were

taken from composting plants located in Lahti, Hyvinkää and Espoo in Finland, from Högstad in Sweden and from a plant in Stavanger, Norway (Table 2). The internal transcribed spacer region (ITS) was the ribosomal region analysed in fungi (papers I, III and IV) and 16S rRNA gene in bacteria (paper II). The total number of good quality sequences was 5 859 (Table 3).

Table 3. The number of good quality sequences from each stage of composting from the four papers (I-IV) included in this thesis.

Sample	Composting plant	Scale	I	II ¹	III	IV
1	feed	full	73	77	-	-
2	drum (160 m ³) feeding-end	full	140	307	-	-
3	drum (160 m ³), unloading-end	full	159	89	-	-
4	tunnel (3 pieces, 100 m ³ each)	full	332	223	-	-
5	drum (160 m ³) feeding-end	full	214	-	-	-
6	tunnel (3 pieces, 100 m ³ each)	full	132	-	-	-
7	drum (160 m ³), unloading-end	full	317	-	-	-
8	tunnel (3 pieces, 100 m ³ each)	full	255	-	-	-
9	drum (5 m ³) feeding-end	pilot	172	571	-	-
10	drum (5 m ³) unloading-end	pilot	93	408	-	-
11	drum (5 m ³) feeding-end	pilot	95	-	-	-
12	drum (5 m ³) unloading-end	pilot	91	-	-	-
13	drum (5 m ³) feeding-end	pilot	118	-	-	-
14	drum (5 m ³) unloading-end	pilot	101	-	-	-
15	drum (160 m ³) feeding-end, ash amendment	full	155	-	144	-
16	drum (160 m ³), unloading-end, ash amendment	full	185	-	230	-
17	tunnel (3 pieces, 100 m ³ each), ash amendment	full	60	-	51	-
18	drum (5 m ³) feeding-end ash amendment	pilot	130	-	114	-
19	drum (5 m ³) unloading-end ash amendment	pilot	162	-	73	-
20	reactor, 200 litre, 55°C	pilot	-	-	-	69
21	reactor, 200-litre, 55°C	pilot	-	-	-	17
22	reactor, 200-litre, 70°C	pilot	-	-	-	91
23	reactor, 200-litre, 70°C	pilot	-	-	-	83
24	windrow, 100 m ³ , 55°C	full	-	-	-	69
25a	windrow, 100 m ³ , 65°C	full	-	-	-	97
25b	windrow, 100 m ³ , 65°C	full	-	-	-	89
26	windrow, floor area 110 m ²	full	-	-	-	73

¹ Samples were taken over two or three days and the number of clones were combined from the different samples.

4.2 Fungal diversity in samples (I, III, IV)

The good quality sequences based on Pregap analyses (I, III) or Phred scores (IV) were joined as phylotypes, and phylogenetic analyses were conducted to these sequences. Based on the clustering in the phylogenetic tree phylotypes were further grouped as phylogroups that were used in studying the fungal succession in the composting process (I, III) and in redundancy analyse (paper I). A total of 280 representative sequences from each fungal phylotype were submitted to the public DNA sequence database (EMBL Nucleotide Sequence Database) with accession numbers AM711356-AM711521, FM173041-FM173102 and FM177644-FM177696. The sequences were clustered based on the distribution in the phylogenetic analyses into 8 different phylogroups named Saccharomycetales 1, Saccharomycetales 2, Dipodascaceae, Saccharomycetales 3, Pezizomycotina, Basidiomycetes, Zygomycetes and Group 1. Hence, fungi were detected from three different phyla: Basidiomycota, Ascomycota and Zygomycota. Group 1 could not be classified into any of the known phyla as reference sequences did not exist in the EMBL database for the ITS-region of these fungi.

The species richness of the sample was estimated with the abundance based coverage estimator (ACE, Chao et al. 1992) and Chao1 Richness estimator (Chao 1984). Basically these two methods differ in the classification of species into rare and abundant. In Chao 1 the species present in singletons and doubletons are considered as rare whereas in ACE the species with fewer than 10 individuals are marked as the rare species. In addition, non-parametric estimators were used to calculate coverage of the sequenced clone libraries and how much of the total richness was covered by the number of clones sequenced. The total estimated 100% coverage was reached in only one of the 26 compost samples studied (Table 4) with both methods, the abundance based coverage estimator (ACE, Chao et al. 1992) and Chao1

Richness estimator (Chao 1984). This indicates high fungal richness and the need to analyse more clones for proper determination of the total microbial diversity in the compost samples studied. In drum composting, especially at the unloading end of the process, the richness was not as high as in the loading end where the highest estimated amount was 180 different phylotypes with Chao1 estimate. The lowest estimate, 4 fungal phylotypes, was in the unloading end of full-scale drum compost (I) and in a compost windrow (IV). The reason for this is presumably that the temperature reached thermophilic levels, in combination with low oxygen levels, rotating movement and a packed compost mass, and thus only provided suitable growth conditions for certain species.

4.3 Fungal community composition and similarity between communities

SONS (Shared OTUs and similarity, Schoss and Handelsman 2006) analysis was run on the fungal clone libraries in paper IV to combine the gathered sequence information. SONS estimates the similarity between communities using non-parametric estimators. The Jaccard similarity index was used to estimate the similarity of the phylotype composition in the samples and theta (θ) indices to analyse the similarity of the communities. With the Jaccard index the phylotype composition is correlated, whereas with theta (θ) indices the abundance of the phylotypes is taken into account. Based on these analyses (Table 5) the difference between the microbiota in reactor experiments and the samples from the full-scale composting plants can be seen in the Nordic composting plants. The Jaccard similarity was from 0.32 to 0.96 between the full-scale plants (1 = full similarity). The indices between the reactor and full-scale samples were from 0 (0 = no similarity) to 0.07. There were differences between the reactor samples as well and the highest similarities were detected between samples taken from different batches of same age (Reactor sample 20 and

Table 4. The non-parametric coverage and richness estimations for the studied fungal samples. With richness estimators the estimated number of different species in the sample was analysed. The coverage estimators were used to estimate the reached coverage of the total diversity in the studied sample.

Paper	Sample number ²	Origin of sample	ACE ¹ estimate	Chao1 ¹ estimate	ACE coverage (%)	Chao1 coverage (%)
I	1	Feed, full-scale	35.84	35.4	58.59	59.32
	2	Drum feeding-end, full-scale	25.61	43.5	74.18	43.5
	3	Drum unloading-end, full-scale	5.6	5.5	89.29	90.91
	4	Drum tunnel, full-scale	8.54	10	93.63	80
	5	Drum feeding-end, full-scale	137.89	180	26.11	20
	6	Drum tunnel, full-scale	14.13	14.25	84.95	84.21
	7	Drum unloading-end, full-scale	4	4	100	100
	8	Drum tunnel, full-scale	18.9	15.5	68.79	83.87
	9	Drum feeding -end, pilot-scale	25.37	24.57	86.7	89.53
	10	Drum unloading-end, pilot-scale	40.91	36.08	63.55	72.06
	11	Drum feeding -end, pilot-scale	55.16	43.13	39.88	51.01
	12	Drum unloading-end, pilot-scale	22.32	20.08	80.65	89.63
	13	Drum feeding -end, pilot-scale	12	10.5	50	57.14
	14	Drum unloading-end, pilot-scale	9	11.25	100	80
	15	Drum feeding -end, full-scale, ash-amendment	27.24	27.25	55.07	55.05
	16	Drum unloading-end, full-scale, ash-amendment	4	5.25	100	76.19
	17	Drum tunnel, full-scale, ash-amendment	10.99	9.67	81.9	93.1
	18	Drum feeding -end, pilot-scale, ash-amendment	72.99	53.67	28.77	39.13
	19	Drum unloading-end, pilot-scale, ash-amendment	29.67	19.5	23.6	35.9
III	16	Drum unloading-end, full-scale, ash-amendment	8.88	8.25	67.61	72.73
	18	Drum feeding end, pilot-scale, ash-amendment	26.16	26.67	61.15	60.00
	18	Drum feeding -end, pilot-scale, ash-amendment	12.15	12.12	90.51	90.72
	19	Drum unloading-end, pilot-scale, ash-amendment	9.63	8.00	62.34	75.00
IV	20	Reactor 200 l, day 8	37.0	30.0	37.9	46.7
	21	Reactor 200 l, day 16	6.4	6.0	78.1	83.3
	22	Reactor 200 l, day 3	17.1	24.3	70.0	49.5
	23	Reactor 200 l, day 8	14.2	11.0	78.4	100.0
	24	Windrow, IVAR	40.6	34.7	46.9	54.8
	25a	Windrow, IVAR	24.3	29.0	45.3	37.9
	25b	Windrow, IVAR	27.5	9.0	31.6	100.0
	26	Windrow, YTV	4.0	3.5	75.0	85.7

¹ ACE= abundance based coverage estimator (Chao et al. 1992)

Chao1= Chao1 Richness estimator (Chao 1984)

These indices estimate the number of different phylotypes or species in a sample.

² Samples 25a and 25b were replicates from a same sample

Reactor sample 23, Jaccard 0.92, theta 0.51). In successive samples of different ages from the same composting batches the communities differed (Reactor sample 20 - Reactor sample 21, Jaccard 0.16, Reactor sample 22 - Reactor sample 23, Jaccard 0.04). In samples 25a (Windrow, IVAR) and 25b (Windrow, IVAR), which were replicates of the same sample but for which the DNA extraction and clone library construction were conducted separately, the indices were 0.96 (Jaccard, standard error 0.18) and 0.92 (theta, standard error 0.04), respectively, showing high similarity.

4.4 Fungal succession in the composting processes

The size of the composting plants was found to affect the microbiota in samples greatly. The difference in community composition between different process scales can be seen from SONS analysis on the phylotype level for the samples in paper IV (Table 5), and on the phylogroup level for all of the fungal clone libraries (Fig. 5).

The phylotype composition in the pilot and full-scale plants (I, III) was not similar, and the phylotype composition in the research reactor compared to the pile composting plant (IV) was not similar either. In the full-scale drum composting process yeasts were abundant in all stages of composting process, also in the thermophilic stages in the drum unloading end and in the tunnel. In the early stages of composting process the community in the normal state process comprised of yeasts belonging to the Dipodascaceae-group with sequences sharing similarity with *Geotrichum candidum* and *Dipodascus australiensis*. In the unloading end of drum fungi from group Saccharomycetales 1 became abundant. This group comprised of sequences with similarity to for example *Candida ethanolica* and *Issatchenkia orientalis* yeasts. In tunnel stages the detected fungi clustered to groups Saccharomycetales 1, Pezizomycotina, Basidiomycetes and Dipodascaceae. After

wood-ash amendment (samples 15-17) the community in the drum feeding end was similar to the normal state drum unloading end. The community composed mainly of fungi from group Saccharomycetales 1. The same group was detected abundantly in the drum unloading end and tunnel stages of composting. Basidiomycetes were observed in the tunnel stage as well.

In IVAR pile composting plants (Fig. 5d) there was no rotating movement and the community composed of yeasts from the group Saccharomycetales 1, of Pezizomycotina such as *Thermomyces lanuginosus* and Zygomycetes such as fungi from genus *Absidia* and *Rhizomucor*. The samples 25a and 25b were replicate clone libraries constructed from the same sample of which two independent DNA extractions were conducted. They showed to share a similar community composition based on the sequencing results. Phylotypes detected in sample 26 (YTV) clustered to Pezizomycotina group and were mostly similar to *T. lanuginosus*.

In normal state pilot-scale drum (samples 9-14 Fig. 5a) the diversity was higher than in the full-scale. Sequences from a novel group were detected in the pilot-scale drum and the community composition was thus different than in the full-scale process. The detected fungi in the feeding end of the drum clustered to phylogroups Pezizomycotina, Saccharomycetales 2, Zygomycetes, Basidiomycetes and the novel Group 1. In the unloading end the diversity decreased and fungi such as *Rhizomucor miehei*, clustering to Zygomycetes and representatives from the Group 1 were observed. Ash-amendment changed the phylotype composition: in the feeding end yeasts such as *I. orientalis* and *Kluyveromyces marxianus* were present. The Zygomycota group fungi that were detected in the normal-state pilot drum were not observed in the ash amended pilot drum. In the unloading end the fungal community comprised of Pezizomycotina such as *Scytalidium thermophilum* and Saccharomycetales 1 fungi.

Communities in the reactor experiments (Fig. 5b), which were laboratory scale, shared

phylotypes from the same novel phylogroup Group1 that were observed in the pilot-scale drum. In addition Saccharomycetales 1 fungi, Pezizomycotina and Saccharomycetales 2 fungi were sequenced.

4.5 Bacterial diversity and succession (II)

Altogether 1678 good quality sequences were gathered from the bacterial 16S rRNA clone library sequencing. The sequences clustered into

Table 5. Results from the SONS analysis. Reactor was a laboratory scale process, IVAR and YTV were full-scale processes. A and B denote the two samples in which the phylotype composition was correlated. Index is the fraction of sequences found in shared phylotypes. The Jaccard similarity index estimates the similarity of the phylotype composition between the samples. The theta (Θ) index analyses the similarity of the communities taking the phylotype abundance into account. S error indicates the standard error calculated by the SONS program with 1000 iterations. Samples “Windrow, IVAR – 25a” and “Windrow, IVAR – 25b” were replicates from the same sample.

				Jaccard similarity		Theta (Θ) similarity	
		A	B	index	s error	index	s error
Reactor vs reactor	Reactor -20	Reactor -21	0.16	0.15	0.15	0.05	
	Reactor -20	Reactor -22	0.01	0.03	0.00	0.00	
	Reactor -20	Reactor -23	0.92	0.14	0.51	0.11	
	Reactor -21	Reactor -22	0.01	0.04	0.00	0.00	
	Reactor -21	Reactor -23	0.13	0.15	0.05	0.02	
	Reactor -22	Reactor -23	0.04	0.07	0.00	0.00	
IVAR vs reactor and YTV	Reactor -20	Windrow, IVAR - 24	0.01	0.02	0.00	0.00	
	Reactor -20	Windrow, IVAR – 25a	0.01	0.02	0.00	0.00	
	Reactor -20	Windrow, IVAR – 25b	0.01	0.02	0.00	0.00	
	Reactor -21	Windrow, IVAR - 24	0.06	0.07	0.02	0.02	
	Reactor -21	Windrow, IVAR – 25a	0.01	0.02	0.00	0.00	
	Reactor -21	Windrow, IVAR – 25b	0.07	0.10	0.02	0.02	
	Reactor -22	Windrow, IVAR - 24	0.04	0.07	0.00	0.00	
	Reactor -22	Windrow, IVAR – 25a	0.01	0.04	0.00	0.00	
	Reactor -22	Windrow, IVAR – 25b	0.01	0.03	0.00	0.00	
	Reactor -23	Windrow, IVAR - 24	0.03	0.04	0.01	0.01	
	Reactor -23	Windrow, IVAR – 25a	0.01	0.02	0.00	0.00	
	Reactor -23	Windrow, IVAR – 25b	0.03	0.05	0.01	0.00	
	Windrow, IVAR - 24	Windrow, IVAR – 25a	0.93	0.20	0.67	0.11	
	Windrow, IVAR - 24	Windrow, IVAR – 25b	0.80	0.23	0.67	0.10	
Windrow, IVAR – 25b	Windrow, IVAR – 25a	0.96	0.18	0.92	0.04		
YTV vs reactor and IVAR	Reactor -20	Windrow, YTV – 26	0.00	0.00	0.00	0.00	
	Reactor -21	Windrow, YTV – 26	0.00	0.00	0.00	0.00	
	Reactor -22	Windrow, YTV – 26	0.00	0.00	0.00	0.00	
	Reactor -23	Windrow, YTV – 26	0.00	0.00	0.00	0.00	
	Windrow, IVAR - 24	Windrow, YTV – 26	0.36	0.24	0.45	0.09	
	Windrow, IVAR – 25a	Windrow, YTV – 26	0.62	0.35	0.80	0.06	
	Windrow, IVAR – 25b	Windrow, YTV – 26	0.57	0.34	0.73	0.07	

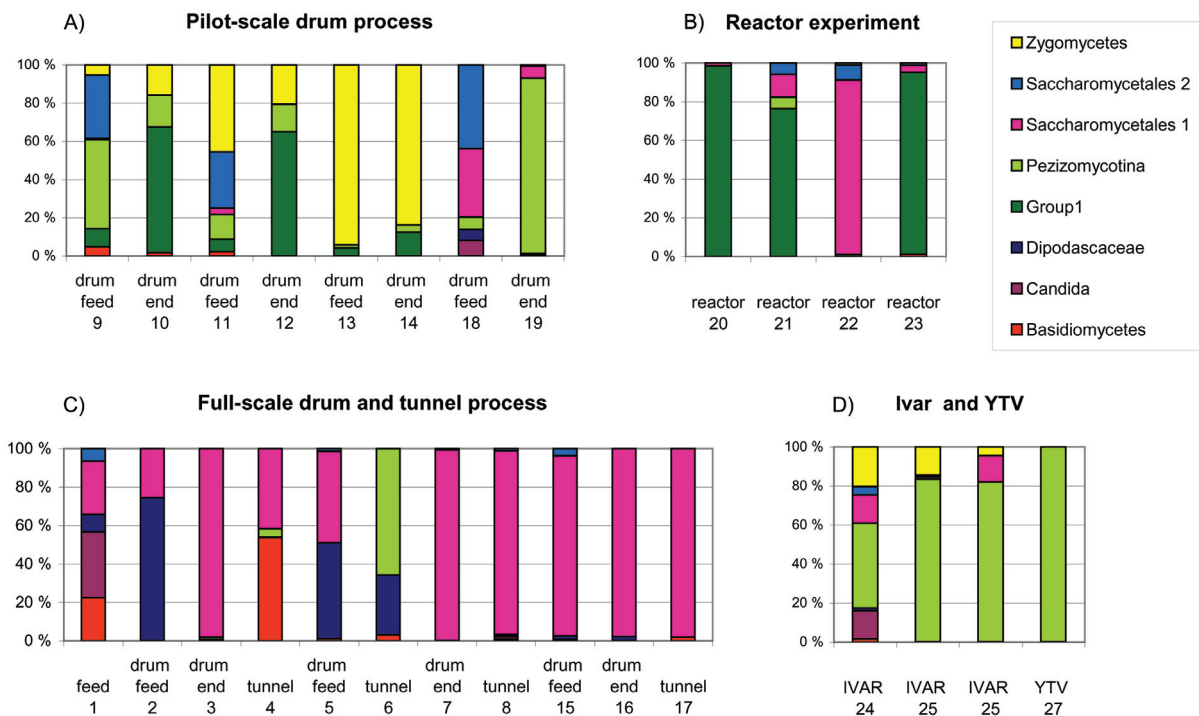


Figure 5. The proportion of clones belonging to different fungal phylogroups in the studied samples. A and B show the phylogroup composition in the pilot-scale drum (Kujala) and reactor experiments. The phylogroup compositions in the full-scale drum and tunnel facility (Hyvinkää) and the pile composting unit (IVAR and YTV) are presented in C and D. The sample numbers corresponding to numbering in Table 3 are marked after the stage of composting. Group 1 contained fungal sequenced which in the phylogenetic analyses did not cluster into any of the sequenced fungal groups. drum feed = feeding end of the compost drum
drum end = unloading end of the compost drum.

566 phylotypes of which 255 were found in samples from the full-scale composting plants and 286 were present in the pilot-scale compost. The phylotype composition differed between the two process scales, full- and pilot-scale. At phylotype level there were only 25 phylotypes (4.4%) shared between the processes, at species and division level the similarity was higher (7.3% and 23%, respectively). The sequences were clustered into nine groups by the phylogenetic analysis. The groups Actinobacteria, *Bacillus*, *Clostridium*, *Lactobacillus*, *Acetobacter* and *Thermoactinomyces* were named by sequences clustering in the phylogenetic tree. The group “other bacteria” included bacterial sequences annotated to other bacterial classes or genera than the six bacterial groups used in this classification.

The “unidentified”-group represents bacterial sequences which were not similar to the sequences in the sequence databases. The “uncultivated”-group included sequences with similarity to bacteria that were reported in the EMBL database as uncultivated.

In the feeding end of both process scales sequences from the *Lactobacillus*-group were abundant as well as sequences with similarity to the Acetobacteria (Fig. 6). On the pilot-scale *Bacillus* and Actinobacteria were frequently detected. Sequences from the *Bacillus*-group were abundant in the unloading end of the drums and Actinobacteria were detected abundantly in the pilot-scale processes. The amount of unidentified bacteria was high as well. In the full-scale processes *Lactobacillus* were still observed in the

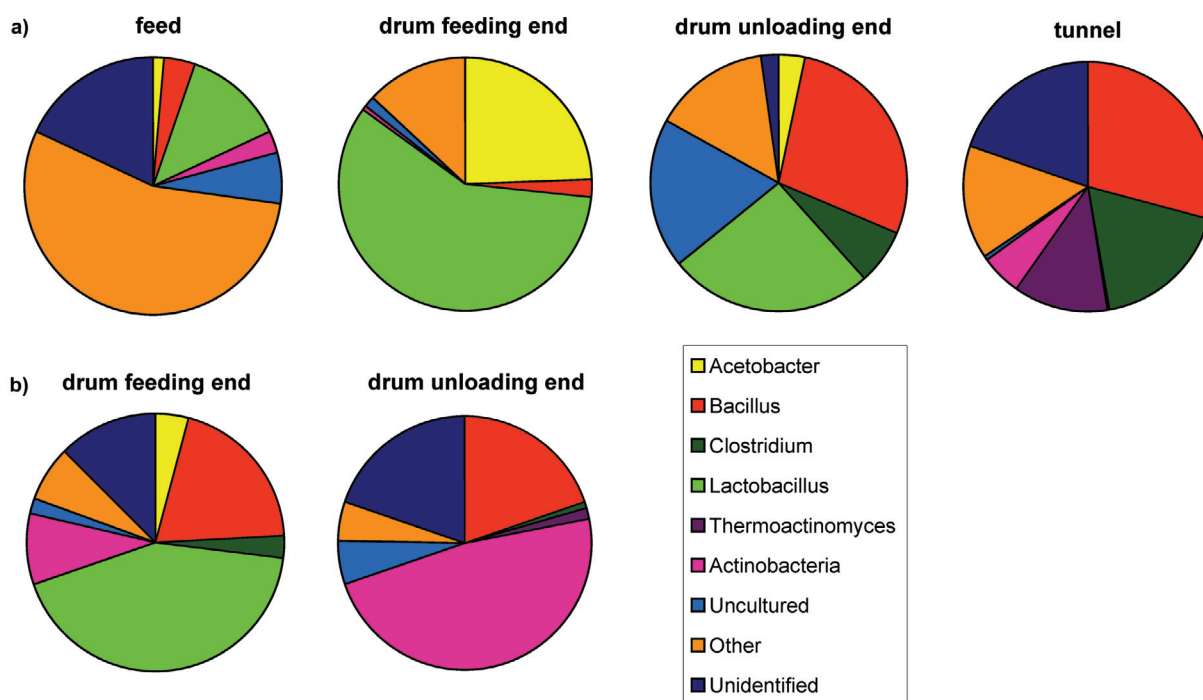


Figure 6. The bacterial 16S rRNA sequence group composition in a full-scale (a) and a pilot-scale (b) composting facility. The composting processes correspond to the plants in Figure 5a and 5c.

“other bacteria”	bacterial sequences similar to other than the six bacterial classes or genera used in classification.
“uncultivated”-group	sequences similar to bacteria that were reported in the EMBL database as uncultivated bacteria.
“unidentified”-group	sequences with no close similarity to sequences in the nucleotide database.

unloading end of the drum as well as sequences similar to *Clostridium* species, which indicates anoxic conditions. In the tunnel stage of the full-scale composting the diversity increased and the bacterial sequence groups present comprised of *Bacillus*, *Thermoactinomyces*, Actinobacteria and *Lactobacillus*. The *Clostridia* were also found.

The estimated bacterial diversity was evidently higher than the fungal diversity (Table 5 and Table 6). The highest estimation of the number of bacterial phylotypes was 961 and 905 from the feeding end of the pilot-scale drum, with ACE and Chao1 models respectively (Table 6). The lowest estimated abundance was 188 from the full-scale feed with the ACE model and 142 from the

unloading end of the drum with the Chao1 model. As with fungi, the bacterial diversity decreased in the drum unloading end before increasing again in the tunnel stages of the full-scale composting.

4.6 Macroarray (III)

The macroarray test was developed to find the sequences present as a minor fraction of the total population. The method was used for 1536 clones that were gridded with a robot on a nylon membrane (paper III). With hybridisation using six probes corresponding to abundant phylotypes in paper I, 900 of the 1536 clones were recognised and were not further sequenced. The remaining

41% of the clones were sequenced as they did not belong to the six common phylotypes (Fig. 4). In addition, 384 clones were sequenced to verify the hybridisation results and the specificity and sensitivity of the method and the probes used. The rate for the false negatives was 5.2%. In false negatives, the probe had not hybridised to the spot containing the corresponding phylotype and a sequence representing one of the common

phylotypes was sequenced. They were mainly from species that were present in high numbers in these samples. False positives were found in 15 cases (n=384, 3.9%). The false positives gave a hybridisation signal but the clone did not contain a phylotype similar to the probed phylotypes. Some of the false positives appeared to be caused by impure PCR-products that contained PCR-products from two clones.

Table 6. The non-parametric coverage and richness estimations for the studied bacterial samples. With richness estimators the estimated number of different species in the sample was analysed. The coverage estimators were used to estimate the reached coverage of the total diversity in the studied sample.

Paper	Origin of sample	ACE ¹ estimate	Chao1 ¹ estimate	ACE coverage (%)	Chao1 coverage (%)
II	Feed, full-scale	188.75	193.08	28.08	27.45
	Drum feeding-end, full-scale	540.19	209.06	17.77	45.92
	Drum unloading-end, full-scale	188.89	142.57	26.47	35.07
	Drum tunnel, full-scale	426.76	190.56	22.26	49.85
	Drum feeding-end, pilot-scale	961.31	905.62	20.7	21.97
	Drum unloading-end, pilot-scale	954.1	514.69	16.77	31.09

¹ ACE= abundance based coverage estimator (Chao et al. 1992)

Chao1= Chao1 Richness estimator (Chao 1984)

These indices estimate the number of different phylotypes or species in a sample.

4.7 Microarray (IV)

4.7.1 Specificity

In order to find the specificity of the 17 probes designed for specific species or phylotypes, all probe pairs were used in a ligation and a microarray hybridisation with one template at a time. Prior to ligation, the template DNA was PCR amplified with universal fungal primers. Despite the close relatedness of the tested species and phylotypes, the probes proved to be specific for the target nucleic acids. In the microarray assay a positive signal was detected only from the corresponding zip-code spot and it was straightforward to differentiate between the specific probe and other probes used for ligation and hybridisation.

4.7.2 Sensitivity

In diagnostic microarrays the sensitivity is an important issue. The sensitivity of the LDR-array was tested by titrating different amounts of target PCR-product for the ligation experiments. Genomic DNA from a *Penicillium* strain was used as a background DNA in ligation to perform the test in a spiked manner. The tested template amounts were 10 fmol, 1 fmol and 0.1 fmol, corresponding to 4%, 0.4% and 0.04% of target PCR-product versus the background DNA. The lowest amount, 0.04%, was detected with the microarray (Fig. 7) showing the sensitivity of the method.

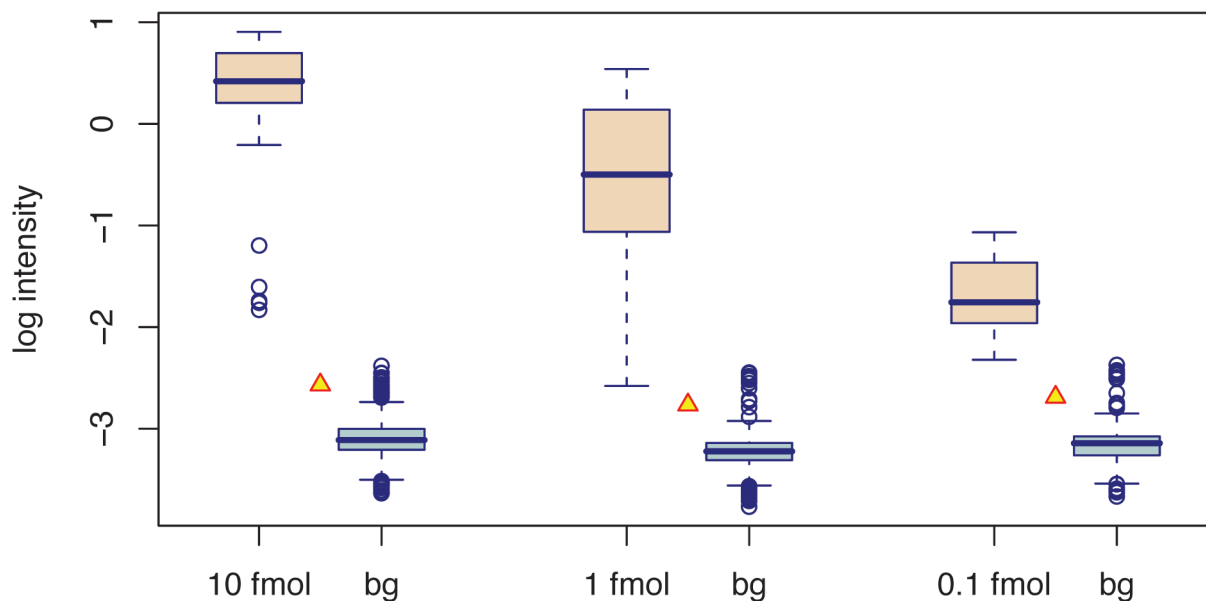


Figure 7. A boxplot showing the signal distribution of probes with different concentrations of template used in the ligation reactions. “bg” is the background distribution from the corresponding subarray. Yellow triangles denote the detection limit (2.5 SD above the background median). The false positives above the detection limit are from cZip number 17 which was not used in any of the probes. Even with 0.1 fmol of template PCR-product used, the signal can be clearly differentiated from the background signal.

4.7.3 Application of the microarray to environmental samples

After testing and optimising the microarray with pure cultures and clones, environmental samples were used as templates for the microarray test. Ten samples were taken from composting plants in Finland and Norway and from a research reactor located in Sweden. Extracted genomic DNA and PCR amplification products were used as a template. The extracted DNA from an amount of 10 ng was not sufficient for use as a template. However, 20 ng of fungal PCR amplified DNA gave good fluorescent signals. Probes also functioned in a multiplex manner and they were all used in the same ligation reaction.

As these samples were uncharacterised, clone libraries were constructed from eight

samples and 96 clones were sequenced per library, except for one library in which there were problems with cloning. The sequencing results were mainly in concordance with the microarray results (paper IV, table 3). There was a problem with one probe, specific for *K. marxianus*, which did not give a signal although the phylotype was found with cloning and sequencing. Clones similar to *G. candidum*, *Aspergillus fumigatus* and *Candida rugosa* were not among the sequenced clones, however they were detected with the microarray, showing the high sensitivity of the microarray platform. With the use of a control probe (Fig. 2) the differentiation between positive and negative spots was straightforward and the signals could be normalised within each microarray and between different microarray hybridisations.

5. DISCUSSION

Most of the genetic diversity on earth is from microbial life forms (Hugelholz & Pace 1996). While the number of insects has been assumed to be extremely high, the number of microbial species exceeds this (Pace 1997). In recent years the study of microbial communities has expanded as the tools for studying them have developed. Microbes have been studied using both phylogenetic and metagenomic approaches. Examples of some of the target environments include soil (Tringe et al. 2005, O'Brien et al. 2005), acid mine drainage (Tyson et al. 2004), oceans (Venter et al. 2004, DeLong et al. 2006), marine sediments (Inagaki et al. 2006) and human intestines (Eckburg et al. 2005). These studies have focused mainly on bacterial communities, whereas the fungal communities have not been studied to the same extent. Although fungi are important components of soil ecosystems (Daniel 2005), the fungal diversity in soil and many other natural habitats is only partly characterised (Vandenkoornhuysen et al. 2002).

The widespread use of modern DNA-based techniques for determining biodiversity and the presence of specific microbes has also opened up new possibilities for the determination of microbial diversity and succession during composting. Composting differs from microbial succession in natural environments such as soil, because composting, as described in this thesis work, is an enclosed system where the substrate level is kept close to maximum. As a result of the microbial decomposition activity the temperature rises and the pH changes radically within a relatively short period of time, causing the decline of certain species while the amount of others increases (Epstein 1997). The methods used here were large scale clone library sequencing, selection of rare phylotypes from clone libraries, and diagnostic microarrays, which were used independently or in combination. As a result, a large dataset of information on the microbes present in different

stages of composting processes of different scales was gathered and a method for monitoring these microbes was developed.

In this thesis fungi were studied broadly as the information on the fungi involved in composting, obtained using molecular methods, is limited (Hansgate et al. 2005, Peters et al. 2000). It was found that the fungal communities in the industrial composting plants and the pilot and laboratory scale reactors comprised of microbes that were different from the microbes previously detected in compost environments (reviewed by Ryckeboer et al. 2003b, Anastasi et al. 2004, Hansgate et al. 2005). Ascomycetous yeasts were frequently found in the studied composting processes at different time points and seasons (I, III, IV). In previous studies the fungi detected have mainly included Pezizomycotina such as *Aspergillus* or *Penicillium* species (e.g. Anastasi et al. 2004), fungi from the genus *Mucor* or *Rhizomucor* (Von Klopotek 1962, Fergus 1964, Anastasi et al. 2004), or basidiomycetes such as *Coprinus cinereus* (Von Klopotek 1962). Yeasts have been rarely detected (von Klopotek 1962, Peters et al. 2000) in these studies although they are known to be important in the first phases of the composting process for overcoming the acidophilic early stages (Choi & Park 1998). In Nordic composting facilities, in particular, a low pH in the initial stages of composting and its inhibitory effect on the composting process has been observed (Sundberg et al. 2004), which may explain the fungal community composition found in this thesis. With phospholipid fatty acid analysis (PLFA, paper I) fungi were found to represent up to 20% of the biomass in samples. As composting is a dynamic process, the fungal biomass detected with PLFA can be considered to be from active, living cells. PLFA results cannot be correlated with the sequencing results as the resolution of PLFA is low and different fungal groups cannot be differentiated.

Yeasts were observed from the full-scale drum and tunnel facility (Fig. 5c) from

all stages, whereas in the two other full-scale facilities the community composed of moulds (Pezizomycotina-group) and Zygomycetes in addition to the yeast-groups. The abundance of yeasts might be due to the rotating movement of the drum compost, which prevents the growth of the filamentous fungi such as the Pezizomycotina-fungi. The fact that yeasts were not detected as often in the composting batches studied in paper IV (IVAR and YTV in Fig. 5d) supports this hypothesis. Yeasts can be considered to be important for process operation, especially the yeasts clustering to the Saccharomycetales 1b-group (paper I), which were present in every stage of each process, including one that was treated with wood ash and was functioning efficiently. It was assumed that the yeasts would be associated with the first stages of composting where *Lactobacillus* are abundantly present. This was observed in paper I and II, but yeasts were also found in the tunnel stages where the mass had been composting for 7 to 21 days and where the bacterial community consisted of *Bacillus*, *Thermoactinomyces* and Actinobacteria.

Fungal taxonomy is based on morphological properties. The sequencing of fungal cultures and the use of molecular methods has caused fungal taxonomical classification to change. For example, the existence of sexual and asexual types of some fungal species, which have been categorised into different taxa as a result of their different morphology, is one reason that some fungal sequences, mainly yeasts, in the sequence databases have been wrongly annotated. This has caused problems in this study as the amount of reference sequences has been low and sequences with annotation to certain genera have clustered differently in phylogenetic analyses. It was found (papers I, III, IV) based on the phylogenetic analyses that yeasts from the genus *Pichia* and *Candida* were observed in different branches of the phylogenetic tree. The almost certain polyphyly of *Pichia* and the misidentification of yeasts into the genus *Candida* was seen to cause the same phenomenon with the use of sequences from

five *loci* (Suh et al., 2006). The names for the phylogroups used in papers I, III, and IV were therefore not from the genus in certain branches but those of higher taxa. For example, sequences from the order Saccharomycetales were clustered into 4 different branches. This makes the comparison with previous studies where the fungi have been characterised from composts based on morphology (Fergus 1964, Kane & Mullins 1973, Ryckeboer et al. 2003a, Anastasi et al. 2004) or restriction patterns (Peters et al. 2000, Hansgate et al. 2005) difficult. In this study, only one area, the ITS-region of the rRNA gene cluster, was examined. The use of two or more regions (*loci*) would have made the assumptions more accurate, but the databases still contain relatively few other sequences, meaning a low number of reference sequences.

A wide picture of the bacterial communities in composting processes was gathered in paper II as over 1500 bacterial 16S rRNA clones were sequenced. After comparing the results between the suboptimally functioning full-scale facility and an optimised pilot-scale facility, it was found that there were no single species that could be an indicator of the different conditions. However, based on the data, a bacterial family or genus may be used as an indicator as these were found to differ between certain phases or conditions in the processes. Nonetheless, the fact that bacteria from the genus *Clostridia*, which indicated anaerobic conditions, were detected in both processes shows that a wide spectrum of different microbes should be studied to get a good overview of the state of a composting process. Even in optimally functioning compost processes there can be anaerobic pockets containing anaerobic microbes (Ryckeboer et al. 2003b).

Based on the theoretical non-parametric coverage and abundance estimations the coverage and abundance of phylotypes in the studied samples varied according to the stage of the composting process. Microbes are too diverse to count exhaustively and methods to estimate

the diversity are needed. These methods have been adapted to microbial ecology and are based on the capture-recapture methodology (Hughes et al. 2001). The estimated number of bacteria and fungi in these compost samples was lower than in farm soil (3000 phylotypes, Tringe et al. 2005) or urban aerosols (1500-1800 phylotypes, Brodie et al. 2007). The reason for this could be that the composting process is a closed system. The conditions during the process can be very extreme: the pH can drop to acidic levels close to pH 4 and the temperature can range from below zero up to plus 80°C. This can reduce the number of different bacteria, fungi and archaea in these processes as the harsh conditions favour growth of only certain organisms. The same kind of situation has been found for example in acid mine drainage (Tyson et al. 2004). The community composition has been extensively studied in these low pH conditions with a metagenomic approach and only a few bacterial phylotypes were found to be present.

In these estimations there were a lower number of fungal phylotypes than bacterial ones. This may be because the calculations were based on phylotypes that can be comprised of several species, depending on the differences in the ITS-region between two fungal species. There could also be bias caused by the method used to detect the diversity, cloning and sequencing. Several novel bacterial species were found with pyrosequencing from previously characterised samples (Sogin et al. 2006) showing that cloning and sequencing selected certain bacteria. The reason for this could be in the nucleotide composition. For example, sequences making secondary structures are difficult to clone. In addition, the sequence number was much higher in the pyrosequencing based study than with cloning and sequencing. Pyrosequencing should not be the sole method of identification from now on, as only short tag-sequences are produced instead of the full-length ribosomal RNA coding gene sequences that are necessary for phylogenetic alignments and which provide more phylogenetic resolution.

The distinctions between the different scales of process were clearly seen as the phylotype composition and also the succession of microbial communities varied between the two processes. Therefore assumptions cannot be made based on laboratory experiments regarding compost studies although there have been attempts to analyse the microbial succession at different stages of composting on the laboratory scale (Alfreider et al. 2002, Hansgate et al. 2005, Peters et al. 2000, Schloss et al. 2003). The problems with microcosm experiments are that the process is closed, the material used is usually known and that the scale is smaller than in the industrial composting process. In the composting process new material of varying composition is fed into process and the microbes already present in the composting facility function as a seed for the process. When the process progresses, the microbial activity alters the conditions (Epstein 1997) from acid to neutral and from mesophilic to thermophilic, before returning to the mesophilic stage. The results of this thesis showed that the microbial community composition in the industrial, full-scale composting process had similarities to the communities studied in laboratory-scale experiments (Schloss et al. 2003, Dees & Ghiorse 2001, Hansgate et al. 2005) but that there were clear differences as well. It was found that in pilot-scale composting the processes proceeded faster as the community composition was similar in the full-scale tunnel and pilot-scale drum unloading stages. This was seen with both bacteria and fungi.

With the method developed in paper III the rare phylotypes in the studied sample were detected in an effective manner. Microbial communities are often rank-abundance distributed (Curtis & Sloan 2005) and capturing the tail-end of those communities, i.e. the species present in small numbers, is difficult with fingerprinting methods (Bent & Forney 2008). With cloning they can be captured, but in addition, the abundant types in samples are captured in large amounts. With the membrane

hybridisation method the common phylotypes are marked and only the rare ones, that there is no probe for, are further sequenced. In samples from paper I it was found that certain phylotypes were found repeatedly and other phylotypes were rare in clone numbers. By using phylotype specific probes these abundant phylotypes were selected out from the sequenced clones and only the rare ones sequenced. New probes can be added to the probe pool when the studied sample is further characterised and more of the abundant phylotypes have been detected.

With the use of LDR-microarray technology the presence or absence of the species being studied can be determined in 1.5 days, whereas with clone library construction and sequencing this would take at least 3 days. The diversity of natural microbial communities can cause problems with cloning as it has been estimated for a soil sample that over 40 000 sequencing reactions are required to reach 50% of the diversity (Dunbar et al. 2002). Cloning and sequencing of microbes is not the best method to choose if the aim is to detect certain species in a sample, as it is laborious and sequencing costs are still relatively high. With the new generation of sequencing technologies it is easy to gain sequence information on hundreds or thousands of tag-fragments, but the technology is not always available and it is rather expensive. The different microarray platforms provide tools for microbial diagnostics (Bodrossy & Sessitch 2004), diversity studies (Brodie et al. 2007, DeSantis et al. 2007) and community functional potential studies (GeoChip, He et al. 2007). Recently, a genome proxy microarray was published in which fragments of genes from ecologically relevant uncultivated microbes were detected in a sensitive (0.1-1% of total DNA sufficient for detection) manner, though there were problems with the specificity (Rich et al. 2008). Depending on the aim of the research, one of these different platforms can be chosen.

The LDR based microarray (Busti et al. 2002) was further developed (paper IV) and the sensitivity of the microarray proved to be very

high. The detection limit was lowered from the previously published limits (Rantala et al. 2008) and 0.04% of target DNA in total DNA was adequate for detection. The sensitivity reached the level of quantitative-PCR, which is considered extremely sensitive. With qPCR the presence of one species or phylotype can be analysed at a time. At the moment, with the LDR-array developed in this thesis, up to 120 species can be studied with one microarray from 16 different samples (paper IV). In many oligonucleotide studies detection limits of 1-5% have been calculated (e.g. Loy et al. 2004, Franke-Whittle et al. 2005) and the increase in sensitivity of the microarray method is considerable. Furthermore, the microarray was species or phylotype specific despite the close relatedness of the target microbes. Species that are highly similar based on their rRNA small subunit sequences can have different niche spaces (Jaspers and Overmann 2004) and with this method they can be distinguished. The use of control probes increased the functionality of the array and results within and between arrays could be normalised.

With the development of faster, cheaper and more accessible of sequencing techniques the amount of sequence data is growing. This brings great challenges to the maintenance of the public sequence databases. These challenges include the minimisation of redundancy, the accuracy of the annotations, and the computational challenges involved in keeping the search and aligning processes efficient, even though the number of sequences is growing nearly exponentially. The taxonomy should be correct in these databases as well. As whole genomes have been sequenced novel information based on microbial phylogeny and ecology has been gathered (Hugenholtz & Tyson 2008). Still, genome and gene sequencing mainly focuses on certain divisions that are often biotechnically or clinically interesting. The fungal genomes that have been sequenced have been pathogens or model genetic systems (Bruns 2006): of the 82 finished fungal genomes (In NCBI database 13th of October 2008) many

were from pathogens (e.g. *Aspergillus nidulans*, Galagan et al. 2005) or industrially interesting fungi (e.g. *Trichoderma reesei*, Martinez et al. 2008). To understand microbial diversity, and more importantly, microbial ecology as whole, genomes from the whole spectrum of bacteria, archaea, protists and fungi should be studied. As knowledge about the vast diversity of microbes is provided in many studies, the focus of research has changed towards studying the ecology and evolution of organisms instead of descriptive studies on their diversity.

Methods which are based on PCR amplification, such as the ones used in this thesis, face certain problems. PCR does not provide amplification products in the same ratios as the genomic templates that are added to the reaction (Poltz & Cavanaugh 1998) and problems with primer selection and the formation of PCR artefacts have been reported (reviewed in von Wintzingerode et al. 1997). The possible bias that has emerged with the use of PCR based methods could be avoided by using whole genome amplification instead of PCR prior to ligation detection reaction and microarray hybridisation. This amplification has been successfully applied to a soil microbiome study in which community composition was further studied with cloning and sequencing (Abulencia et al. 2006) and to a microarray study (Wu et al. 2006). The development of the LDR microarray test into a quantitative method is of interest. In the LDR-microarray method (paper IV) the control probe can be important when quantifying the results since the data can be normalised between different subarrays and microarray slides, and the uneven printing of DNA onto the microarray glass is not a problem. With methods such as the microarray, the composition of the microbial community, based on the species of interest, can be studied efficiently in a specific and sensitive manner.

6. CONCLUSIONS

In this thesis compost microbiota were studied in a comprehensive way. Fungal and bacterial communities from different phases of composting on two scales were characterised with cloning and sequencing. The stage and scale of the composting was found to shape the communities. The full-scale industrial drum composting process was found to contain a diverse yeast community in both suboptimally and optimally functioning processes. In industrial batch composts the filamentous fungi such as Pezizomycota and fungi from Zygomycota were present in low numbers among the sequenced clones. Both fungal and bacterial communities were dissimilar between the different scales showing that full-scale composting processes should be studied to understand compost microbiology. The bacterial diversity was higher than the fungal diversity showing the wider diversity of bacteria and their better adaptation to partly extreme conditions with high temperatures, low pH and low oxygen levels.

Nowadays, there are many methods for studying microbial diversity. In this thesis two methods were developed for diversity studies. It was demonstrated that with the macroarray method the clones that are present in high numbers in environmental clone libraries can be selected out from the sequenced clones. This method can be used in diversity studies to target the rare phylotypes that may be of importance in microbial communities. In the studied samples 41% of the clones were sequenced after the negative selection and so the method was cost-effective.

In this thesis the LDR microarray method was further developed and shown to be capable of sensitive and specific detection of species of interest in compost samples. With the multiplex detection method just 0.04% of the studied organism's target gene was sufficient for detection from the sample's total DNA. This detection limit is lower than in previously published studies and species level discrimination was

achieved as well. The microarray was tested with compost samples. To verify the gathered results of the diversity in the samples, clone libraries were constructed and sequenced. The results were in concordance and the LDR microarray can be further developed as a tool for compost process monitoring in industrial composting facilities.

With cloning, sequencing and microarrays microbial community composition can be determined, but the next goal is to utilise the data. This can be achieved by additionally studying the physical and chemical properties of samples which was already done to some extent in this thesis. By studying samples from different depths and from different timescales it is possible to get more information about the community and the ecology of the microorganisms detected.

7. ACKNOWLEDGEMENTS

This study was carried out in a collaborative project between the Institute of Biotechnology and the Department of Ecological and Environmental Sciences, University of Helsinki. Viikki Graduate School in Biosciences (VGSB), Tekes and the Research Foundation of the University of Helsinki are acknowledged for the financial support.

I wish to express my gratitude to my supervisors Professor Martin Romantschuk and docent Petri Auvinen. Mara, I started in your group as a MSc student and I thank you for the opportunity to continue towards PhD under your supervision. The relaxed atmosphere in your group, your patient guidance and trust in me has made this process to be successful. Petri I thank you for always having time for guidance and comments. I really appreciate the scientific and non-scientific discussions we had and especially making me to think life after PhD is acknowledged. I'm also forever thankful for the ancient molecular biology lessons you have given to the kit-generation student.

I am deeply grateful to Academy professors Kaarina Sivonen and Jari Valkonen for careful reviewing this thesis and for the constructive comments for improving it.

My co-authors Miia Pitkäranta, Lars Paulin, Tuija Vasara, Jarmo Ritari, Jukka Kurola, Merja Kontro and Pasi Partanen are kindly acknowledged for their expertise and smooth collaboration. Miia, your peer-support has meant me a lot and I've been lucky to have a fungal orientated molecular microbiologist as a colleague and as a friend.

This thesis was conducted in the DNA sequencing and genomics lab at the Institute of Biotechnology. I thank the director of BI Mart Saarma for providing the great state-of-art facilities to carry out this study. I wish to express my gratitude to Lars Paulin for building this high-throughput facility in Finland and for the endless reservoir of ideas. Paula Collin-Olkkonen, Päivi Laamanen, Ritva Rajala,

Kirsi Lipponen and Lea Merviä, thank you for the excellent technical assistance and for being so flexible. The former and present staff at the sequencing lab: Noora, Sini, Markku, Matias, Hannu, Mari, Tanja, Tuula, Riikka, Eetu, Anna-Liisa, Eeva-Marja, Panu, Dario, Tuomas, Ari-Matti, Juha and Olli, thanks for all the help with robotics, machines, computers, programming, cars... Anu warm thanks on grammatical issues. Pia and Kaisa, the cheerful company and support on both good and bad (food)days, especially during the final stages of getting this thesis done is acknowledged.

I wish to acknowledge Timo Kairesalo, Olli-Pekka Penttinen, Taru Nordman and Hanna Mikkonen for the administrative help regarding finishing this thesis and conducting research between to different labs. Sari Kauppi is thanked for always being friendly and helpful, and having time to assist with sampling and DNA-extractions.

I express my gratitude to the former dean of VGSB Marja Makarow for accepting me as a student in Viikki Graduate School in Biosciences. The acceptance to grad school pushed my studies into a new level, all the fantastic courses, the people I met through school and the peer support from other students has been of value. The coordinators Eeva Sievi and Sandra Falk and the current dean Dennis Bamford are acknowledged for help and guidance. Repe, thanks for providing the right kind (and colour) of gloves, pens, folders and lab stuff.

I've been privileged to have many dear friends around me, both old and new. Tiina Mehto-Vaittinen, Elina Stenvall, Jonna Ketola and Tarja Ikonen AKA ruokakerho, thank you for being there for me! Your presence in the final stages of getting this book done was invaluable. Also friends including but not limited to Elina, Melli, Essi, Suvi, Muikku and Hanna, thank you and your families for all the good times during these studies of mine. Essi also did the layout of this thesis for which I'm deeply grateful.

Mum, thank you for encouragement in all stages of my studies and for your love and support. Dad, you will always be in my heart. Antti, it's good to know that when in doubt I have a brother to lean on. Your help in arranging karonkka is also acknowledged. Jussi, my dear husband, I thank you for your love and for staying by my side all these years.

Viikki, December 2008

A handwritten signature in black ink, appearing to read 'Jussi', with a small dot at the end.

8. REFERENCES

- Abulencia, C. B., Wyborski, D. L., Garcia, J. A., Podar, M., Chen, W., Chang, S. H. et al. (2006)** Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* 72: 3291-3301.
- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., & Polz, M. F. (2004)** Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J Bacteriol* 186: 2629-2635.
- Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L., & Polz, M. F. (2004)** Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430: 551-554.
- Alfreider, A., Peters, S., Tebbe, C. C., Ranggner, A., & Insam, H. (2002)** Microbial community dynamics during composting of organic matter determined by 16S ribosomal analysis. *Compost Sci Util* 10: 303-312.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990)** Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995)** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.
- Amaral-Zettler, L., Peplies, J., Ramette, A., Fuchs, B., Ludwig, W., & Glockner, F. O. (2008)** Proceedings of the international workshop on Ribosomal RNA technology, April 7-9, 2008, Bremen, Germany. *Syst Appl Microbiol* 31: 258-268.
- Anastasi, A., Varese, G., Voyron, S., Scannerini, S., & Morchisio, V. F. (2004)** Characterization of fungal biodiversity in compost and vermicompost. *Compost Sci Util* 12: 185-191.
- Anderson, I. C., & Cairney, J. W. (2004)** Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. *Environ Microbiol* 6: 769-779.
- Banér, J., Isaksson, A., Waldenström, E., Jarvius, J., Landegren, U., & Nilsson, M. (2003)** Parallel gene analysis with allele-specific padlock probes and tag microarrays. *Nucleic Acids Res* 31: e103.
- Barany, F. (1991)** Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc Natl Acad Sci U S A* 88: 189-193.
- Beffa, T., Blanc, M., Lyon, P. F., Vogt, G., Marchiani, M., Fischer, J. L., & Aragno, M. (1996)** Isolation of *Thermus* strains from hot composts (60 to 80 degrees C). *Appl Environ Microbiol* 62: 1723-1727.
- Bent, S. J., & Forney, L. J. (2008)** The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J* 2: 689-695.
- Blanc, M., Marilley, L., Beffa, T., & Aragno, M. (1997)** Rapid identification of heterotrophic, thermophilic, spore-forming bacteria isolated from hot compost. *Int J syst Bacteriol* 47: 1246-1248.
- Bodrossy, L., & Sessitsch, A. (2004)** Oligonucleotide microarrays in microbial diagnostics. *Curr Opin Microbiol* 7: 245-254.
- Bodrossy, L., Stralis-Pavese, N., Murrell, J. C., Radajewski, S., Weilharter, A., & Sessitsch, A. (2003)** Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* 5: 566-582.
- Brodie, E. L., Edwards, S., & Clipson, N. (2003)** Soil fungal community structure in a temperate upland grassland soil. *FEMS Microbiol Ecol* 45: 105-114.
- Brodie, E. L., DeSantis, T. Z., Parker, J. P., Zubieta, I. X., Piceno, Y. M., & Andersen, G. L. (2007)** Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci U S A* 104: 299-304.
- Bruns, T. (2006)** Evolutionary biology: a kingdom revised. *Nature* 443: 758-761.

- Busti, E., Bordoni, R., Castiglioni, B., Monciardini, P., Sosio, M., Donadio, S. et al. (2002)** Bacterial discrimination by means of a universal array approach mediated by LDR (ligase detection reaction). *BMC Microbiol* 2: 27.
- Castiglioni, B., Rizzi, E., Frosini, A., Sivonen, K., Rajaniemi, P., Rantala, A. et al. (2004)** Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. *Appl Environ Microbiol* 70: 7161-7172.
- Chao, A. (1984)** Nonparametric estimation of the number of classes in a population. *Scand J Statist* 11: 265-270.
- Chao, A., & Lee, S. (1992)** Estimating the number of classes via sample coverage. *J Am Stat Assoc* 87: 210-217.
- Choi, M. H., & Park, Y. H. (1998)** The influence of yeast on thermophilic composting of food waste. *Lett Appl Microbiol* 26: 175-178.
- Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., & Bork, P. (2006)** Toward automatic reconstruction of a highly resolved tree of life. *Science* 311: 1283-1287.
- Costa, R., Gotz, M., Mrotzek, N., Lottmann, J., Berg, G., & Smalla, K. (2006)** Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *FEMS Microbiol Ecol* 56: 236-249.
- Curtis, T. P., & Sloan, W. T. (2005)** Microbiology. Exploring microbial diversity--a vast below. *Science* 309: 1331-1333.
- Daniel, R. (2005)** The metagenomics of soil. *Nat Rev Microbiol* 3: 470-478.
- Dees, P. M., & Ghiorse, W. C. (2001)** Microbial diversity in hot synthetic compost as revealed by PCR-amplified rRNA sequences from cultivated isolates and extracted DNA. *FEMS Microbiol Ecol* 35: 207-216.
- DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U. et al. (2006)** Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311: 496-503.
- DeSantis, T. Z., Brodie, E. L., Moberg, J. P., Zubietta, I. X., Piceno, Y. M., & Andersen, G. L. (2007)** High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* 53: 371-383.
- Dunbar, J., Ticknor, L. O., & Kuske, C. R. (2001)** Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Appl Environ Microbiol* 67: 190-197.
- Dunbar, J., Barns, S. M., Ticknor, L. O., & Kuske, C. R. (2002)** Empirical and theoretical bacterial diversity in four Arizona soils. *Appl Environ Microbiol* 68: 3035-3045.
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M. et al. (2005)** Diversity of the human intestinal microbial flora. *Science* 308: 1635-1638.
- Edgar, R. C. (2004)** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-1797.
- Edwards, U., Rogall, T., Blocker, H., Emde, M., & Bottger, E. C. (1989)** Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* 17: 7843-7853.
- Epstein, E. (1997)** The science of composting. Lancaster, Technomic Publishing Company, 487 pp.
- Ewing, B., & Green, P. (1998)** Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8: 186-194.
- Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998)** Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185.
- Felsenstein, J. (2006)** PHYLIP (Phylogeny inference package) version 3.66. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Fergus, C. L. (1964)** Thermophilic and thermotolerant molds and actinomycetes of

mushroom compost during peak heating. *Mycologia* 56: 267-284.

Franke-Whittle, I. H., Klammer, S. H., & Insam, H. (2005) Design and application of an oligonucleotide microarray for the investigation of compost microbial communities. *J Microbiol Methods* 62: 37-56.

Franke-Whittle, I. H., Knapp, B. A., Fuchs, J., Kaufmann, R., & Insam, H. (2008) Application of COMPOCHIP Microarray to Investigate the Bacterial Communities of Different Composts. *Microb Ecol* in press DOI: 10.1007/s00248-008-9435-2

Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S. et al. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438: 1105-1115.

Gans, J., Wolinsky, M., & Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* 309: 1387-1390.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.

Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., & Barany, F. (1999) Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol Biol* 292: 251-262.

Gray, K. R., Sherman, K., & Biddlestone, A. J. (1971) A Review of composting – Part 1. *Process Biochem* 6: 32-36.

Guschin, D. Y., Mobarry, B. K., Proudnikov, D., Stahl, D. A., Rittmann, B. E., & Mirzabekov, A. D. (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl Environ Microbiol* 63: 2397-2402.

Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68: 669-685.

Hansgate, A. M., Schloss, P. D., Hay, A. G., & Walker, L. P. (2005) Molecular

characterization of fungal community dynamics in the initial stages of composting. *FEMS Microbiol Ecol* 51: 209-214.

Hawksworth, D. L. (1991) The fungal dimension of biodiversity: Magnitude, significance, and conservation. *Mycol Res* 95: 641-655.

He, Z., Gentry, T. J., Schadt, C. W., Wu, L., Liebich, J., Chong, S. C. et al. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 1: 67-77.

Hibbet, D. S. (1992) Ribosomal RNA and fungal systematics. *Trans Mycol Soc Japan* 33: 533-556.

Hibbet, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E. et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycol Res* 111: 509-547.

Huber, J. A., Mark Welch, D. B., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., & Sogin, M. L. (2007) Microbial population structures in the deep marine biosphere. *Science* 318: 97-100.

Hugenholtz, P., & Tyson, G. W. (2008) Microbiology: metagenomics. *Nature* 455: 481-483.

Hugenholtz, P., & Huber, T. (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int J Syst Evol Micr* 53: 289-293.

Hugenholtz, P., & Pace, N. R. (1996) Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol* 14: 190-197.

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

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

- Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., & Polz, M. F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320: 1081-1085.
- Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever, M., Lauer, A. et al. (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc Natl Acad Sci U S A* 103: 2815-2820.
- Ishii, K., & Takii, S. (2003) Comparison of microbial communities in four different composting processes as evaluated by denaturing gradient gel electrophoresis analysis. *J Appl Microbiol* 95: 109-119.
- Ishii, K., Fukui, M., & Takii, S. (2000) Microbial succession during a composting process as evaluated by denaturing gradient gel electrophoresis analysis. *J Appl Microbiol* 89: 768-777.
- James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J. et al. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443: 818-822.
- Jaspers, E., & Overmann, J. (2004) Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologicals. *Appl Environ Microbiol* 70: 4831-4839.
- Jürgens, G., Lindström, K., & Saano, A. (1997) Novel group within the kingdom Crenarchaeota from boreal forest soil. *Appl Environ Microbiol* 63: 803-805.
- Kane, B. E., & Mullins, J. T. (1973) Thermophilic fungi in a municipal waste compost system. *Mycologia* 65: 1087-1100.
- Koltai, H., & Weingarten-Baror, C. (2008) Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction. *Nucleic Acids Res* 36: 2395-2405.
- Landegren, U., Kaiser, R., Sanders, J., & Hood, L. (1988) A ligase-mediated gene detection technique. *Science* 241: 1077-1080.
- Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W. et al. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442: 806-809.
- Liu, W. T., Marsh, T. L., Cheng, H., & Forney, L. J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* 63: 4516-4522.
- Loy, A., Kusel, K., Lehner, A., Drake, H. L., & Wagner, M. (2004) Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal cooccurrence of recognized genera and novel lineages. *Appl Environ Microbiol* 70: 6998-7009.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32: 1363-1371.
- Lutzoni, F., Kauf, F., Cox, C. J., McLaughlin, D., Celio, G., Dentinger, B. et al. (2004) Assembling the fungal tree of life: progress, classification and evolution of subcellular traits. *Am J Bot* 91: 1446-1480.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
- Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E. et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol* 26: 553-560.
- Martiny, J. B., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L. et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 4: 102-112.
- Nilsson, R. H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K. H., & Koljalg,

- U. (2006)** Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE* 1: e59.
- O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J. M., & Vilgalys, R. (2005)** Fungal community analysis by large-scale sequencing of environmental samples. *Appl Environ Microbiol* 71: 5544-5550.
- Osborn, A. M., Moore, E. R., & Timmis, K. N. (2000)** An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* 2: 39-50.
- Pace, N. R. (1997)** A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.
- Pearson, W. R., & Lipman, D. J. (1988)** Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444-2448.
- Peters, S., Koschinsky, S., Schwieger, F., & Tebbe, C. C. (2000)** Succession of microbial communities during hot composting as detected by PCR-single-strand-conformation polymorphism-based genetic profiles of small-subunit rRNA genes. *Appl Environ Microbiol* 66: 930-936.
- Pitkäranta, M., Meklin, T., Hyvärinen, A., Paulin, L., Auvinen, P., Nevalainen, A., & Rintala, H. (2008)** Analysis of fungal flora in indoor dust by ribosomal DNA sequence analysis, quantitative PCR, and culture. *Appl Environ Microbiol* 74: 233-244.
- Polz, M. F., & Cavanaugh, C. M. (1998)** Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* 64: 3724-3730.
- Porter, T. M., Schadt, C. W., Rizvi, L., Martin, A. P., Schmidt, S. K., Scott-Denton, L. et al. (2008)** Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. *Mol Phylogenet Evol* 46: 635-644.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007)** SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188-7196.
- R Development core team. (2007)** R: A language and environment for statistical computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Radajewski, S., Ineson, P., Parekh, N. R., & Murrell, J. C. (2000)** Stable-isotope probing as a tool in microbial ecology. *Nature* 403: 646-649.
- Ranjard, L., Lejon, D. P., Mougel, C., Schehrer, L., Merdinoglu, D., & Chaussod, R. (2003)** Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environ Microbiol* 5: 1111-1120.
- Rantala, A., Rizzi, E., Castiglioni, B., de Bellis, G., & Sivonen, K. (2008)** Identification of hepatotoxin-producing cyanobacteria by DNA-chip. *Environ Microbiol* 10: 653-664.
- Rich, V. I., Konstantinidis, K., & DeLong, E. F. (2008)** Design and testing of 'genome-proxy' microarrays to profile marine microbial communities. *Environ Microbiol* 10: 506-521.
- Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D. et al. (2007)** Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 1: 283-290.
- Ryckeboer, J., Mergaert, J., Coosemans, J., Deprins, K., & Swings, J. (2003a)** Microbiological aspects of biowaste during composting in a monitored compost bin. *J Appl Microbiol* 94: 127-137.
- Ryckeboer, J., Mergaert, J., Vaes, K., Klammer, S., De Clercq, D., Coosemans, J. et al. (2003b)** A survey of bacteria and fungi occurring during composting and self-heating processes. *Ann Microbiol* 53: 349-410.
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985)** Enzymatic amplification

of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.

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

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.

Santelli, C. M., Orcutt, B. N., Banning, E., Bach, W., Moyer, C. L., Sogin, M. L. et al. (2008) Abundance and diversity of microbial life in ocean crust. *Nature* 453: 653-656.

Schadt, C. W., Martin, A. P., Lipson, D. A., & Schmidt, S. K. (2003) Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* 301: 1359-1361.

Schloss, P. D., & Handelsman, J. (2006) Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microbiol* 72: 6773-6779.

Schloss, P. D., Hay, A. G., Wilson, D. B., Gossett, J. M., & Walker, L. P. (2005) Quantifying bacterial population dynamics in compost using 16S rRNA gene probes. *Appl Microbiol Biotechnol* 66: 457-463.

Schloss, P. D., Hay, A. G., Wilson, D. B., & Walker, L. P. (2003) Tracking temporal changes of bacterial community fingerprints during initial stages of composting. *FEMS Microbiol Ecol* 46: 1-9.

Small, J., Call, D. R., Brockman, F. J., Straub, T. M., & Chandler, D. P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol* 67: 4708-4716.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R. et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* 103: 12115-12120.

Staden, R., Beal, K. F., & Bonfield, J. K. (2000) The Staden package, 1998. *Methods*

Mol Biol 132: 115-130.

Steger, K., Sjogren, A. M., Jarvis, A., Jansson, J. K., & Sundh, I. (2007) Development of compost maturity and Actinobacteria populations during full-scale composting of organic household waste. *J Appl Microbiol* 103: 487-498.

Strom, P. F. (1985) Identification of thermophilic bacteria in solid-waste composting. *Appl Environ Microbiol* 50: 906-913.

Suh, S. O., Blackwell, M., Kurtzman, C. P., & Lachance, M. A. (2006) Phylogenetics of Saccharomycetales, the ascomycete yeasts. *Mycologia* 98: 1006-1017.

Sundberg, C., Smårs, S., & Jönsson, H. (2004) Low pH as an inhibiting factor in the transition from mesophilic to thermophilic phase in composting. *Bioresource Technology* 95: 145-150.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.

Tiquia, S. M., Tam, N. F., & Hodgkiss, I. J. (1996) Effects of composting on phytotoxicity of spent pig-manure sawdust litter. *Environ Pollut* 93: 249-256.

Torsvik, V., & Øvreås, L. (2006) Microbial phylogeny and diversity in soil. In *Modern soil microbiology*. J. D. van Elsas, J. K. Jansson & J. T. Trevors (eds). FL, USA, CRC press, pp. 23-54.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W. et al. (2005) Comparative metagenomics of microbial communities. *Science* 308: 554-557.

Tuomela, M., Vikman, M., Hatakka, A., & Itävaara, M. (2000) Biodegradation of lignin in a compost environment: a review. *Bioresource Technol* 72: 169-183.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M. et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37-43.

- Valinsky, L., Della Vedova, G., Jiang, T., & Borneman, J. (2002)** Oligonucleotide fingerprinting of rRNA genes for analysis of fungal community composition. *Appl Environ Microbiol* 68: 5999-6004.
- Van de Peer, Y., Chappelle, S., & De Wachter, R. (1996)** A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res* 24: 3381-3391.
- Van Heerden, I., Cronje, C., Swart, S. H., & Kotze, J. M. (2002)** Microbial, chemical and physical aspects of citrus waste composting. *Bioresour Technol* 81: 71-76.
- Vandenkoornhuysse, P., Baldauf, S. L., Leyval, C., Straczek, J., & Young, J. P. (2002)** Extensive fungal diversity in plant roots. *Science* 295: 2051.
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A. et al. (2004)** Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66-74.
- Von Klopotek A. (1962)** Über das vorkommen und verhalten von Schimmelpilzen bei der Kompstierung Städtischer Abfallstoffe. *Antonie van Leeuwenhoek* 28: 141-160.
- Von Wintzingerode, F., Gobel, U. B., & Stackebrandt, E. (1997)** Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21: 213-229.
- Vora, G. J., Meador, C. E., Bird, M. M., Bopp, C. A., Andreadis, J. D., & Stenger, D. A. (2005)** Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic *Vibrio* spp. *Proc Natl Acad Sci U S A* 102: 19109-19114.
- Vuorinen, A. H., & Saharinen, M. H. (1997)** Evolution of microbial and chemical parameters during manure and straw co-composting in a drum composting system. *Agr Ecosyst Environ* 66: 19-29.
- Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D. et al. (2002)** Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 99: 17020-17024.
- White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990)** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. San Diego, Academic Press, Inc, 482 pp.
- Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998)** Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95: 6578-6583.
- Whittaker, R. H. (1969)** New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science* 163: 150-160.
- Woese, C. R. (1987)** Bacterial evolution. *Microbiol Rev* 51: 221-271.
- Wu, L., Liu, X., Schadt, C. W., & Zhou, J. (2006)** Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Appl Environ Microbiol* 72: 4931-4941.