Chapter I
Introduction
1.1 Microbial diversity in soil.

Soil is fundamental and irreplaceable entity that governs plant productivity of terrestrial ecosystems and it maintains biogeochemical cycles because microorganisms in the soil degrade, sooner or later, virtually all organic compounds including persistent xenobiotics and naturally occurring polyphenolic compounds. The living population inhabiting soil includes macrofauna, mesofauna, microfauna and microflora. It has been estimated that 80–90% of the processes in soil are reactions mediated by microbes (Coleman and Crossley, 1996; Nannipieri and Badalucco, 2003). Indeed, bacteria and fungi are highly versatile; they can carry out almost all known biological reactions. To provide a comprehensive view of the complex relations between microbial diversity and soil functionality people consider:

a) complexity of soil as a biological system;

b) microbial diversity and their functions in soil;

c) the link between microbial diversity and soil functions;

d) time points that represent instances when measurements of microbial diversity are unnecessary for a better understanding of soil functionality;

and

e) methods for better evaluation and manipulation of microbial diversity and soil functionality.

What Microbes Can Do?

There are countless ways in which microbes influence daily life. Earth is a biological entity as much as it is a physical one, and most of the vital biology, on which all life depends, is microbiology. But because microbes are individually invisible, we (even microbiologists) need to be reminded of our debt to them. Here are some of the reasons.

Microbes Modulate and Maintain the Atmosphere.

Carbon is the most abundant chemical element in all living things, including humans (excluding the hydrogen and oxygen in the water, which makes up the bulk of our weight). Carbon dioxide (CO₂) in the atmosphere is the most abundant source of carbon on Earth, but in this form it is inaccessible to animals
and most bacteria. Plants and some bacteria "fix" carbon through photosynthesis, a light-driven conversion of CO₂ to sugars that generates the oxygen that fuels all aerobic forms of life. Although plants tend to get most of the credit, bacteria are responsible for about half of the photosynthesis on Earth (Pedros-Alio 2006). Ocean microbes, collectively present at billions of cells per liter, grow at rates of about one doubling per day in surface waters and are consumed at about the same rate (Whitman et al. 1998). The organisms that carry out photosynthesis turn over rapidly in the ocean as well, on the average about once per week. Net primary productivity in the global ocean is estimated to fix 45-50 billion tons of CO₂ per year (Falkowski et al. 1998). Chemical transformations mediated by marine microbes play a critical role in global biogeochemical cycles.

![The global carbon cycle](http://www.bigelow.org/foodweb/carbon_cycle.jpg)

**Fig. 1 The global carbon cycle (Source:http://www.bigelow.org/foodweb/carbon_cycle.jpg)**

The collective metabolism of marine microbial communities has global effects on fluxes of energy and matter in the sea, on the composition of Earth's
atmosphere, and on global climate. In essence, the combined activities of 
microbial communities affect the chemistry of the entire ocean and maintain the 
habitability of the entire planet. Hidden within the population dynamics of these 
complex communities are fundamental lessons of environmental response and 
sensing, species and community interactions, gene regulation, and genomic 
plasticity and evolution. Microbes are the stewards of Earth's biosphere and are 
Nature's biosensors par excellence. Perhaps most obviously today, the living 
oceans play a critical role in the global carbon cycle (Falkowski et al. 1998). The 
coupling of the upper ocean and the atmosphere results in higher concentrations 
of dissolved CO$_2$ in surface seawater than in the rest of the ocean. Much of the 
elevated carbon input can move through the action of the ocean's "biological 
pump," which depends on microbial communities in the surface water that 
transform inorganic CO$_2$ into organic carbon. The organic carbon can either be 
respired and recycled back to the upper ocean-atmosphere system or sink out of 
the surface water and be sequestered in the deep ocean. Complex microbial 
community interactions help to regulate the proportion of recycled versus 
sequestered carbon. The structure of the phytoplankton community, the rates at 
which phytoplankton are attacked and destroyed by viruses, and the capacity of 
other microbes to turn organic carbon back into CO$_2$ all influence the fate of 
carbon, and the ability of the ocean to act as a source of, or a sink for, CO$_2$. CO$_2$ 
is a very important greenhouse gas, so photosynthetic bacteria serve the planet 
in two ways: they convert carbon into biologically accessible forms and they 
remove CO$_2$ from the atmosphere, thereby mitigating some of the anthropogenic 
release of CO$_2$ and other greenhouse gases.

**Microbes Keep Us Healthy.**

It should come as no surprise that in the microbe-dominated biosphere, close 
relationships between microbes and animals are an ancient theme. Humans are 
no exception. The numbers are staggering. The microbes that reside on the 
surface of the human body alone outnumber human cells by about a factor of 10. 
The genomes of members of our indigenous microbial communities (the human 
metagenome) contain thousands of times more genes than the human genome
Microbial communities also inhabit the human mouth, skin, and respiratory and female reproductive tracts. The compositions of these communities change over time and, for some body sites, like the oral cavity, there is already evidence that certain community compositions are associated with periodontal disease. Understanding how microbial community structure affects health and disease may contribute to better diagnosis, prevention, and treatment of disease. The vast majority of these microbial partners live in the intestine, where a diverse community of microbes, 10 to 100 trillion in number, perform functions that humans have not had to evolve, including the extraction of calories from otherwise indigestible components of our diet and the synthesis of essential vitamins and amino acids. The complex communities of microbes that dwell in the human gut shape key aspects of postnatal life, such as the development of the immune system, and influence important aspects of adult physiology, including energy balance. Gut microbes serve their host by functioning as a key interface with the environment; for example, they defend us from encroachment by pathogens that cause infectious diarrhea, and they detoxify potentially harmful chemicals that we ingest (intentionally or unintentionally). In light of the crisis in management of infectious pathogens due to emergence of antibiotic resistance, we would be well served to understand the role of microbial communities in protecting us from infectious agents. Our microbes are master physiological chemists: identifying the chemical entities that they have learned to manufacture and characterizing the functions of human genes and gene products that they manipulate should lead to valuable additions to our 21st-century medicine cabinet (pharmacopeia).

**Microbes Support Plant Growth and Suppress Plant Disease.**

The microbial communities on and around plants play a central role in the health and productivity of crops. The most complex of these communities reside in the soil, which is a composite of mineral and organic materials teeming with bacteria and archaea. Some functions of these microbes are well known. Some bacteria fix atmospheric nitrogen, converting it from dinitrogen gas, a form unusable by plants and animals to ammonia, which is readily used. Other soil microbes
recycle nutrients from decaying plants and animals, and others convert elements, such as iron and manganese to forms that can be used for plant nutrition. Soil microbial communities determine whether plants will become infected by pathogens. A lingering mystery is the "suppressive soil" phenomenon (Mazzola 2004). In some soils, plants stay healthy even when pathogens are present at high density; when the soil is sterilized, the disease suppression disappears, suggesting a biological basis of the phenomenon. After decades of wrestling with the enigma of suppressive soils, plant pathologists have concluded that in many cases a complex community is responsible for the suppressive activity, which is hugely beneficial to agriculture. No organism has been found to provide the same effect in isolation, because the community members modify each other's behavior.

**Microbes Clean Up Fuel Leaks.**

There are hundreds of thousands of underground storage tanks around the globe, most of which are used for storing gasoline. In fact, almost every part, the gasoline station in the United States uses three or more of these tanks to dispense regular, premium, and super-premium versions of gasoline. The sad truth about these underground tanks is that the vast majorities of them are already leaking or will leak and send gasoline into the subsurface, where it has the potential to contaminate the groundwater. Given the ubiquity and magnitude of the gasoline leaks and the fact that 50% of the US population relies on groundwater as a drinking-water source, one must wonder how it is that we are not all drinking water contaminated with gasoline! The answer is that we are being protected by the omnipresent and vastly adaptable subsurface microbial community (Mazzola 2004). As gasoline is released into the subsurface, relatively dormant members of the microbial communities are triggered to become active and biodegrade the gasoline constituents. Gasoline is composed of thousands of organic chemicals and a variety of microbes containing complementary metabolic systems are required to degrade them all. Furthermore, because there is too little of any single electron acceptor in the subsurface to react with all the electron donors of gasoline, different bacteria with
different respiratory capabilities are required to complete the gasoline remediation. For example, when oxygen is depleted in the groundwater in the vicinity of a gasoline spill, bacteria that can respire nitrate take over followed by bacteria that respire iron, manganese, sulfate, and eventually CO$_2$. This complicated community of microbes works together in a self-organized pattern triggered by the movement of the leaking gasoline until the contaminants have been transformed into harmless CO$_2$ and water. The microbial community then becomes dormant again awaiting the next influx of substrate (either natural or anthropogenic) to return to activity.

**Microbes and soil health.**

The presence of an active and diverse microbial community in soil can be considered a good indicator of a healthy and functioning soil ecosystem. In contemporary ecology, diversity of higher organisms is commonly used as an indicator of ecosystem health because this property of communities is important for the maintenance and stability of ecosystem processes. However, for soil microbial communities, it has not been possible to measure species diversity accurately (Griller et al., 1997; Trevors, 1998). The relative diversity of functions performed by microbial communities is possibly more ecologically relevant to soil processes than species diversity (Zak et al., 1994; Trevors, 1998). Furthermore, in practical terms, it is the diversity of soil functions that is of interest to farmers, agronomists and policy makers. It is unlikely that the functional diversity of soil microbial communities can be determined from species diversity (Giller et al., 1997; Degens, 1999).

Microbial characteristics of soils are being evaluated increasingly as sensitive indicators of soil health because of the clear relationships between microbial diversity, soil and plant quality, and ecosystem sustainability (Doran et al., 1994). While the understanding of microbial properties such as biomass, activity, and diversity are important to scientists in increasing knowledge of the factors contributing to soil health, results of such analyses may also be useful to extension personnel and farmers in devising practical measures of soil quality. Community-level microbial interactions are complex with individual species
relying on the presence, function, and interaction of many other species. Therefore, quantitative and qualitative changes in the composition of soil microbial communities may serve as important and sensitive indicators of both short and long-term changes in soil health. The analysis of soil microbial communities should involve not only the determination of microbial biomass and diversity, but also determination of microbial growth, distribution, function, and, if possible, the nature of interactions among species. Over the past 10 years, the approach to analyzing soil microbial communities has changed dramatically. Many new methods and approaches are now available, allowing soil microbiologists to gain access to more of the microorganisms residing in soil and allowing for better assessments of microbial diversity.

**Identity of soil microbes.**

In soil, 80 to 99% of the microorganisms remain unidentified [Bakken and Olsen, 1987; Leisack et al., 1997]. The use of the polymerase chain reaction (PCR) technique to amplify a gene common to all organisms now allows the identification of these previously unknown organisms [Billi and Potts, 2002; Boettger, 1989; Edwards et al., 1989]. The gene commonly amplified for this purpose codes for the 16S and 23S rRNA sequences of the small subunit (SSU) and the large subunit (LSU) of the ribosome respectively [Bintrim et al., 1997; Boettger, 1989] (Fig. 2). It has been found that the misidentification of bacteria is far less common with the gene coding for SSU rRNA (SSU rDNA) sequence [Head et al, 1998] than with more traditional methods of microbial identification such as morphology, Gram stain, enzyme activities, and the utilization of several substrates as sole carbon and energy sources. The challenge for the soil microbial ecologist is to identify the populations and assemblage of microorganisms that have key functional roles in specific soil processes. PCR amplification of 16S rRNA genes (16S rDNA) using consensus bacterial primers and separation of the resultant PCR amplicons by Agarose gel electrophoresis (AGE) constitutes one of the popular techniques used to describe soil bacterial ecology. Bands on the gel can be sequenced and the resultant information used to infer something about the diversity of the original sample.
Introduction

[Chandler et al., 1997; Chowdhury et al., 2004; Whiteley et al., 1996]. Over the last few years we have seen a progressive development of these studies applied to soils, as molecular techniques have been systematically applied to many diverse environments [Brintrim et al., 1997; Chin et al., 1999; Hershberger et al., 1996; Macrae, 2000]. We are just beginning to expand our understanding of important soil processes, their regulation and how they might be manipulated. Thus, the main aim of studying genetic diversity in soil using 16S rDNA methods should be to facilitate the isolation of novel organisms for the biotechnology industry [Felske and Akkermans, 1998] and identification and characterization of novel biotechnologically important microbial enzymes and metabolites. Both culture dependent and culture independent techniques have contributed in our understanding in microbial diversity in soil.

Fig. 2 Domain structure of *Escherichia coli* 16S rRNA gene (SSU). V1 to V10 represent the variable domains present in the 16S rRNA gene sequence
1.1.1 Culture-dependent methods of community analysis.

1.1.1.a Dilution plating and culturing methods.
Traditionally, the analysis of soil microbial communities has relied on culturing techniques using a variety of culture media designed to maximize the recovery of different microbial species. This is particularly the case for soil health studies. There are numerous examples where these techniques have revealed a diversity of microorganisms associated with various soil quality parameters such as disease suppression and organic matter decomposition (Tunlid et al., 1989; Boehm et al., 1993, 1997; de Leij et al., 1993; Workneh et al., 1993; Alvarez et al., 1995; Hu and van Bruggen, 1997; Maloney et al., 1997). Although there have been recent attempts to devise suites of culture media to maximize the recovery of diverse microbial groups from soils (Balestra and Misaghi, 1997; Mitsui et al., 1997), it has been estimated that less than 1% of the microorganisms found in typical agricultural soils are culturable using current culture media formulations (Torsvik et al., 1990; Atlas and Bartha, 1998). This is based on comparisons between direct microscopic counts of microbes in soil samples and recoverable colony forming units.

1.1.1.b Community-level physiological profiles.
One of the more widely used culture-dependent methods for analyzing soil microbial communities has been that of community-level physiological profiles (Garland and Mills, 1991; Winding, 1994; Zak et al., 1994; Konopka et al., 1998). This technique takes advantage of the traditional methods of bacterial taxonomy in which bacterial species are identified based on their utilization of different carbon sources. Community-level physiological profiles have been facilitated by the use of a commercial taxonomic system, known as the BIOLOG® system, which is currently available and has been used extensively for the analysis of soil microbial communities (Winding, 1994; Lehman et al., 1995; Garland, 1996). This BIOLOG® system is based on the utilization of a suite of 95 different carbon sources that have been described previously (Garland and Mills, 1991). Utilization of each substrate is detected by the reduction of a tetrazolium dye, which results in a color change that can be quantified spectrophotometrically.
The pattern of substrates that are oxidized can be compared among different soil samples from a series of times or locations as an indication of differences in the physiological functions of microbial communities. Most commonly, multivariate statistical techniques are necessary to analyze the substrate utilization profile data (Hackett and Griffiths, 1997; Hitzl et al., 1997). For example, in this analysis, communities are considered to be functionally similar if the utilization profile of the 95 different carbon sources from one community clusters with that from another community. If the profiles segregate, communities would be considered functionally different. As such, community-level physiological profiles can be useful in assessing gross functional diversity (Zak et al., 1994; Garland, 1996; Campbell et al., 1997). There are a number of important considerations in the use of this method for community analysis. First, the density of the initial inoculums must be standardized because it affects the rate at which color develops in the wells and thus the time at which color development should be measured (Garland and Mills, 1991; Haack et al., 1995). Visible color will not develop within a well until the total number of cells able to utilize that substrate reaches approximately $10^8$ cells/ml (Haack et al., 1995). Because the number of cells directly inoculated into the wells may be well below $10^8$ cells/ml, there can be a substantial lag phase while the cells grow within the well. This may lead to false negatives if wells are read too soon. Inaccurate physiological profiles may also result if samples are dominated by only a few species capable of growing on particular substrates. Furthermore, the period of microbial growth within the well may also lead to competition effects which again may bias the substrate utilization profile (Haack et al., 1995). Perhaps the best way to standardize inoculums levels is to employ vital stains combined with epifluorescence microscopy as a means to quantify actively-respiring cells (Garland, 1996). This way, a standard population of metabolically active cells can be introduced into each well. A second methodological consideration is that an analysis of functional diversity is based on the assumption that color development in each well is solely a function of the proportion of organisms present in the sample which are able to utilize a particular substrate (Garland, 1997). However, this
may not be valid given that some strains may utilize certain substrates more efficiently than others in the guild, predominating in the well and resulting in proportions of strains that differ from the original sample (Smalla et al., 1998). Furthermore, the ability of different taxa in the sample to utilize the same carbon sources is generally unknown. A third problem is that the substrates found in commercially available BIOLOG® plates are not necessarily ecologically relevant and most likely do not reflect the diversity of substrates found in the environment (Konopka et al., 1998). This is supported by the recent study of Campbell et al. (1997) in which plant root exudate compounds were included as carbon sources in a functional analysis of nine upland grassland sites. The carbon sources most useful in differentiating the different sites were predominantly these plant root exudates. All of these compounds had particularly low utilization rates suggesting they were utilized by organisms that were present in the soil in low numbers. Campbell et al. (1997) hypothesized that these compounds have greater differentiating ability both because they are biochemically more diverse and because they select for the more slow growing organisms that are usually present in the sample in smaller numbers. While community level physiological profiles may provide information useful for assessments of soil microbial community diversity, the method still suffers from the same bias problems encountered with culture plating methods, making data interpretation problematic. Future work with ecologically meaningful substrates (i.e. those that are likely to be found in soil habitats in nature) should make the method more appropriate for use with soil microbial communities. Despite the fact that culture-dependent techniques are not ideal for studies of the composition of natural microbial communities when used alone, they provide one of the most useful means of understanding the growth habit, development, and potential function of microorganisms from soil habitats. A combination of culture-based and culture-independent approaches is likely to reveal more complete information regarding the composition of soil microbial communities (Liesack et al., 1997).
1.1.2 Functional screening based community profiles.
Although widely used in recent years, interpretation of carbon source utilization patterns by microbial communities and their ecological relevance is still in a developmental phase. However, this approach may help to explain the evolution of functional relationships among communities, and may contribute to our current understanding of the complex interactions that form the basis of coexistence in a communal context. The objective of these studies (Ghosh et al. 2007) was to determine the functional and structural responses of a degradative microbial community to substrates of varying complexity, as well as to evaluate methods for the characterization of microbial communities. The functional evolution of the community over time was also investigated.

"Functional Screening" based method facilitates the isolation of organisms with specific sets of characteristics. Functional screening could be an useful method to isolate organisms with specific characters i.e. production of extracellular enzymes like protease, cellulase, amylase, xylanase and pectinase, nitrogen fixation, phosphate solublization and insecticidal toxin production. Recent studies on functional screening based methods results in isolation of different biocatalyst synthesizing microorganisms. Analyzing isolates cultivated on plates has traditionally compared bacterial communities. Both cultivable and non-cultivable organism populations in any soil system could be exploited by the method of "Functional screening". Although this method has been successfully applied to cultivable part of the soil microbiota (only 1%) [Tapp and Stotzky, 1995; Chowdhury et al., 2004; Ghosh et al., 2007], still people are working to exploit the potential in uncultivable microbial community by functional Metagenomics.

1.1.3 Culture-independent methods of community analysis (Metagenomic studies).
Because of the inherent limitations of culture-based methods, soil microbial ecologists are turning increasingly to culture-independent (Metagenomic) methods of community analysis. Using culture-independent methods, the composition of communities can be inferred based on:
(1) the extraction, quantification, and identification of molecules from soil that are specific to certain microorganisms or microbial groups; or

(2) advanced fluorescence microscopic techniques. Useful molecules for such studies include phospholipids, fatty acids and nucleic acids (Morgan and Winstanley, 1997) whereas the microscopic techniques involve either the hybridization of fluorescent-labeled nucleic acid probes with total RNA extracted from soils or hybridizations with cells in situ.

**What Is Metagenomics?**

Like genomics, metagenomics is both a set of research techniques, comprising many related approaches and methods, and a research field. In Greek, meta means "transcendent." In its approach and methods, metagenomics overcomes the twin problems of the unculturability and genomic diversity of most microbes, the biggest roadblocks to advancement in clinical and environmental microbiology. Meta in the first sense means that this new science seeks to understand biology at the aggregate level, transcending the individual organism to focus on the genes in the community and how genes might influence each other's activities in serving collective functions. In the second sense, meta also recognizes the need to develop computational methods that maximize understanding of the genetic composition and activities of communities so complex that they can only be sampled, never completely characterized. Metagenomics, still a very new science, has already produced a wealth of knowledge about the uncultured microbial world because of its radically new ways of doing microbiology. All metagenomics studies take the same first step: DNA is extracted directly from all the microbes living in a particular environment. The mixed sample of DNA can then be analyzed directly, or cloned into a form maintainable in laboratory bacteria, creating a library that contains the genomes of all the microbes found in that environment. The library can then be studied in several ways, based primarily either on analyzing the nucleotide sequence of the cloned DNA or on determining what the cloned genes can do when they are expressed as proteins. It is important to recognize that the library is not organized into neat volumes, each containing the genome of one community.
member. Instead, it consists of millions of clones, each holding a random fragment of DNA.

A metagenomics library is like thousands of jigsaw puzzles jumbled into a single box—putting the puzzles together again is one of this new science's great challenges. The metagenomics approach is now possible because of the availability of inexpensive, high-throughput DNA sequencing and the advanced computing capabilities needed to make sense of the millions of random sequences contained in the libraries. Sequence-based metagenomics captures a massive amount of information on the microbial community under study. A study of the metagenome of the microbial inhabitants of the Sargasso Sea, for example, generated sequences of about a million genes and revealed whole classes of genes that were more diverse than could ever have been anticipated on the basis of studies of cultured organisms. At the other end of the spectrum, studies of a simple microbial community that lives in the extremely acidic water draining from metal mines demonstrated the potential of metagenomics to dissect detailed interactions among microbial-community members. Metagenomics, thus, is more than just large-scale sequencing. In function-based metagenomics, millions of random DNA fragments in a library are translated into proteins by bacteria that grow in the laboratory. Clones producing "foreign" proteins are then screened for various capabilities, such as vitamin production or antibiotic resistance. This enables researchers to access the tremendous genetic diversity in a microbial community without knowing anything about the underlying gene sequence, the structure of the desired protein, or the microbe of origin. New antibiotics and resistance mechanisms have already been discovered using function based metagenomics.

Until recently, the complex microbial communities inhabiting nearly every environment and organism on Earth have essentially been invisible. With metagenomics, the astonishing genetic and metabolic diversity of the microbial world will be increasingly revealed. The practical applications of knowledge of these previously unseen realms of nature will be only part of the result. It is likely that as new biological strategies are brought to light, fundamental biological
Introduction

concepts will be affected. Basic ideas that organize biologists’ understanding of the living world may need refinement in the face of greater understanding of how microbial communities function. New concepts of genomes, species, evolution, and ecosystem robustness will have effects beyond the specific field of microbiology. The questions that must be asked are “deep” ones, but answers will in all cases inform and guide the work of putting increased knowledge of microbial communities to practical use.

The study of microbes has focused on single species in pure culture, so understanding of these complex communities lags behind understanding of their individual members. We know enough, however, to confirm that microbes, as communities, are key players in maintaining environmental stability.

What Metagenomics Offers Future?

The pure culture paradigm has not only limited what microbiologists have studied; it has also limited how they think about microbes. Microbes have been studied as sovereign entities and examined only for their responses to the simple chemicals that can be added to their media. We know little about their behavior as partners in the strategic alliances that are metabolic consortia, such as the consortia that decontaminate drinking water or that make up the complex structured biofilms that keep dental hygienists busy. The invisible members of a microbial community can differ vastly in their biochemical activities and interactions, not only between species but also within species. Such understanding can be achieved only with methods that go beyond the pure-culture and single-whole-genome approaches that have dominated microbial genomics. We must move directly to the genes, to defining environments by the potential and realized biochemical and geochemical activities of the genes that are there, and the complex patterns of interactions with in and between cells that regulate their responses to changes in their physical and biological surroundings.

We must do this while recognizing that except in restricted environments and specialized consortia with limited numbers of genetically homogeneous constituents—we will be dealing with enormous amounts of data that will represent an incomplete sampling of the genetic diversity present. In short, we
must adopt the methods of metagenomics. Pioneering steps in this direction, which illustrate the character and range of such methods, are described later in this report; but in metagenomics, necessity not only is the mother of invention but will be the grandmother of a paradigm shift. It will refocus us one level higher in the biological hierarchy (molecules, cells, organisms, species, populations, communities, the biosphere). It will shift the emphasis from individuals to interactions, from parts to processes—a change that would be timely and highly desirable even if it were not also technologically necessary. Not coincidentally, this shift will parallel the new focus of organismal genomics on interactions between cellular components and how they are coordinated within the complex systems called organisms. This new focus is called systems biology. Metagenomics will be the systems biology of the biosphere.

Metagenomics provides a means for studying microbial communities on their own "turf." Complex ecological interactions—including lateral gene transfer, phage-host dynamics, and metabolic complementation—can now be studied with the lens of metagenomics. Community composition, function, and dynamics can now be measured and modeled in the environment with universal microbial-community genomic approaches. These approaches have the potential to provide new insights into the environmentally relevant microbial communities and activities that control matter and energy flux on Earth. With such information in hand, it will become possible to interpret the interplay between natural cycles and human activities that together shape the future of the planet (Fig. 3).
Fig. 3 Historical and step-by-step evolution of microbial ecology

Metagenomics Can Contribute to Advances in Many Fields.
Metagenomics (Fig. 4) offers a means of solving practical problems facing humanity. Cracking the secrets of some of Earth's countless microbial communities will reveal ways to meet myriad challenges in biomedicine, agriculture, and environmental stewardship. These are among the most important potential contributions:

- **Earth Sciences**: the development of genome-based microbial ecosystem models to describe and predict global environmental processes, change, and sustainability.
- **Life Sciences**: the advancement of new theory and predictive capabilities in community-based microbial biology, ecology, and evolution.
- **Biomedical Sciences**: the description, on a global scale, of the role of the human microbiome (the collective genome of our symbionts) in health and
disease in individuals and populations, and the development of novel diagnostic and treatment strategies based on this knowledge.

- **Bioenergy**: the development of microbial systems and processes for new bioenergy resources that will be more economical and environmentally sustainable and less vulnerable to disruption by world politics.
- **Bioremediation**: the development of tools for monitoring environmental damage at all levels (from climate change to leaking gas storage tanks) and microbe-based (green) methods for restoring healthy ecosystems.
- **Biotechnology**: the identification and exploitation of the remarkably versatile and diverse biosynthetic capacities of microbial communities to generate beneficial industrial, food, and health products.
- **Agriculture**: the development of more effective and comprehensive methods for early detection of threats to food production (crop and animal diseases) and food safety (monitoring and early detection of dangerous microbial contaminants) and the development of management practices that maximize the beneficial attributes of microbial communities in and around domestic plants and animals.
- **Biodefense and Microbial Forensics**: the development of more effective vaccines and therapeutics against potential bioterror agents, the deployment of genomic biosensors to monitor microbial ecosystems for known and potential pathogens, and the ability to precisely identify and characterize microbes that have played a role in war, terrorism, and crime events, thus contributing to discovering the source of the microbes and the party responsible for their use.
1.2 East-Calcutta Wetland soil and the microbial diversity.

In the metropolitan cities of India, about 0.1 million tonnes of municipal solid waste is generated every day, of which Calcutta accounts for 2920 tones per day. The solid wastes generated are dumped in some areas of the country old landfill site at the eastern lowland fringes known as “Dhapa”, located at the eastern fringe of Kolkata since 1868 (Fig. 5). The municipal solid wastes of the landfill site have transformed into stabilized compost under natural conditions over time. The landfill area in Calcutta has now become an environmental hazard due to lack of space for accommodating further volume of wastes generated by the ever increasing population. Expansion and shifting of the landfill area is not feasible. For deriving the benefits of effective disposal and resource recovery, Calcutta Municipal Coorporation Development Authority (CMDA) has commissioned a plant to prepare municipal solid waste compost for use in agriculture production practice. Short-term field study with municipal solid waste compost both in India (Bhattacharyya, 2002) and abroad (Pascual et.al., 1999) revealed no detrimental influence on microbial biomass and enzyme activities of soil. Long term effect of the application of municipal solid wastes to soil is still largely unknown. The assessment of the ecological scenario of the Calcutta landfill soils might reflect on the potential implications of using the compost on a
long-term basis in agricultural land. Growing vegetables with sewage water either fully or partly, in an area of 320 hectares, at the Calcutta landfill site is still continuing, since about a century (Kundu, 1994). Luxurious growth and considerable yield of the crops indicate normal edaphological condition. Soils of Dhapa landfill site are basically composed of thoroughly weathered Kolkata's solid waste. Sanitary land filling has never been practiced here. Toxic materials have not also been segregated out of the wastes before disposal. In the preliminary investigation, Olaniya et al. (Olaniya et al., 1998) indicated a considerable amount of heavy metals in the landfill soils (LFS) and in the crops grown there. Soils of Dhapa, therefore, could be looked upon as something different from natural soils that are subjected to tremendous anthropogenic stress. The Municipal solid waste (MSW) has been found to be compostable and contains plant nutrients (Gallardo-Lara and Nogales, 1987). There are very few available literatures indicating microbial counterpart of this Dhapa landfill area. Healthy growth of vegetative crops in this area indicates normal soil processes are continuing (Perucci, 1990; Perucci, 1992). It is the soil's microbial populations that are responsible for the recycling of carbon and other nutrients and thereby continue the soil process. These soil organisms occupy endless foam of tiny soil particles; they purify water, detoxify harmful substances, and recycle waste products. They restore carbon dioxide to the air and make the atmosphere's nitrogen available to the plant. Making of an inventory of such microbial consortia may provide useful information regarding the microbial diversity of the soil system like that of the Dhapa landfill. The assessment of the microbial community diversity of the East Kolkata Landfill soils (Dhapa soils) might reflect on the potential implications of using the compost on long-term basis in agricultural land. Moreover, the microbial diversity of Dhapa landfill can be exploited for the biotechnological purposes, like isolation of different biotechnologically relevant enzymes and proteins.
1.3 Enzymes from soil.

Enzymes are produced from animals, plants, and microorganisms, but isolation of enzymes from the first two groups is limited due to several reasons. In contrast microbial enzymes can be produced in amounts meeting all demands of the market. Seasonal fluctuations of raw materials do not count and there are possibilities for genetic and environmental manipulation of microorganisms (bacteria and fungi) to give increased yields of desired enzymes in a way not possible with higher organisms. Moreover, the diversity of enzymes available from microorganism is very great. Lastly, microbial enzymes present a wide spectrum of characteristics makes them useful for industrial applications. Soil could be a useful source of microorganisms that are capable of synthesizing a variety of biotechnologically important biocatalysts. Enzymes like protease, cellulase, pectinase, xylanase, lipase are proved to be useful in different biotechnological applications. Most of these enzymes, used industrially, are from either bacteria or fungal origin. Moreover, recent advancement in the field of recombinant DNA technology further facilitates large scale production of these enzymes in biotechnology industries. One of the advantages in working with microbial enzymes is that the genetic manipulation of bacteria and fungus is well understood and comparatively easier.
1.3.a Proteolytic enzymes.
Proteases are the single class of enzymes that occupy a pivotal position with respect to their applications in both physiological and commercial fields. Proteolytic enzymes catalyze the cleavage of peptide bonds in other proteins. The term protease appeared in the German literature of physiological chemistry in the latter part of the nineteenth century in reference to proteolytic enzymes and was used as a general term embracing all the hydrolases that act on proteins, or further degrade the fragments of them. Around 1930, the need to distinguish different kinds of protease activity began to be recognized and two sets of terms were independently proposed, one in Germany and one in the USA. In Germany, Grassmann and Dyckerhoff (Grassmann and Dyckerhoff, 1928) proposed that the enzymes that act on proteins are ‘proteinases’, and those that act preferentially on oligopeptides are ‘peptidases’. The present concept on the nomenclature of proteolytic enzymes is still under debate but the most accepted thought was proposed by Bergmann and Ross (Bergmann and Ross, 1936).

According to them, the term peptidases represent all the peptide bond hydrolases. They also suggested the terminology of ‘endopeptidase’ and ‘exopeptidase’ (Fig. 6).

Fig. 6 The synonymous terms for proteolytic enzymes
Advances in analytical techniques have demonstrated that proteases conduct highly specific and selective modifications of proteins such as activation of zymogenic forms of enzymes by limited proteolysis, blood clotting and lysis of fibrin clots and processing and transport of secretory proteins across the membrane. Recent advances in genomic and proteomin tools further increases our understanding on protease action. The estimated value of the worldwide sales of industrial enzymes is $1 billion (Godfrey and West, 1996). Of the industrial enzymes, 75% are hydrolytic. Proteases represent one of the three largest groups of industrial enzymes and account for about 60% of the total worldwide sale of enzymes. Proteases have a long history of application in the food and detergent industries. Their application in the leather industry for dehairing and bating of hides to substitute currently used toxic chemicals is a relatively new development and has conferred added biotechnological importance (Zen-Yoji et.al., 1961). The vast diversity of proteases, in contrast to the specificity of their action, has attracted worldwide attention in attempts to exploit their physiological and biotechnological applications (Fox et.al.,1991; Yang et.al., 1996).

Proteases are grossly subdivided into two major groups, i.e., exopeptidases and endopeptidases, depending on their site of action. Exopeptidases cleave the peptide bond proximal to the amino or carboxy termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate. Based on the functional group present at the active site, proteases are further classified into four prominent groups, i.e., serine proteases, aspartic proteases, cysteine proteases, and metalloproteases (Rao et.al.; 1998). There are a few miscellaneous proteases which do not precisely fit into the standard classification, e.g., ATP-dependent proteases which require ATP for activity (Menon and Goldberg;1987). Based on their amino acid sequences, proteases are classified into different families (Argos; 1987) and further subdivided into “clans” to accommodate sets of peptidases that have diverged from a common ancestor (Zwickl et.al.; 1992). Each family of peptidases has
been assigned a code letter denoting the type of catalysis, i.e., S, C, A, M, or U for serine, cysteine, aspartic, metallo-, or unknown type, respectively.

**Exopeptidases.**
The exopeptidases act only near the ends of polypeptide chains. Based on their site of action at the N or C terminus, they are classified as amino- and carboxyopeptidases, respectively.

**Aminopeptidases.**
Aminopeptidases act at a free N terminus of the polypeptide chain and liberate a single amino acid residue, a dipeptide, or a tripeptide. They are known to remove the N-terminal Met that may be found in heterologously expressed proteins but not in many naturally occurring mature proteins. Aminopeptidases occur in a wide variety of microbial species including bacteria and fungi (Watson; 1976). In general, aminopeptidases are intracellular enzymes, but there has been a single report on an extracellular aminopeptidase produced by A. oryzae (Labbe et.al., 1974). The substrate specificities of the enzymes from bacteria and fungi are distinctly different in that the organisms can be differentiated on the basis of the profiles of the products of hydrolysis (Cerny; 1978).

**Carboxyopeptidases.**
The carboxyopeptidases act at C terminals of the polypeptide chain and liberate a single amino acid or a dipeptide. Carboxyopeptidases can be divided into three major groups, serine carboxyopeptidases, metallocarboxyopeptidases, and cysteine carboxyopeptidases, based on the nature of the amino acid residues at the active site of the enzymes.

**Endopeptidases.**
Endopeptidases are characterized by their preferential action at the peptide bonds in the inner regions of the polypeptide chain away from the N and C termini. The presence of the free amino or carboxyl group has a negative influence on enzyme activity. The endopeptidases are divided into four subgroups based on their catalytic mechanism, (i) serine proteases, (ii) aspartic proteases, (iii) cysteine proteases, and (iv) metalloproteases.
Serine proteases.
Serine proteases are characterized by the presence of a serine group in their active site. They are numerous and widespread among viruses, bacteria, and eukaryotes, suggesting that they are vital to the organisms. Serine proteases are found in the exopeptidase, endopeptidase, oligopeptidase, and omega peptidase groups. Based on their structural similarities, serine proteases have been grouped into 20 families, which have been further subdivided into about six clans with common ancestors (Barett, 1994). The primary structures of the members of four clans, chymotrypsin (SA), subtilisin (SB), carboxypeptidase C (SC), and Escherichia D-Ala-D-Ala peptidase A (SE) are totally unrelated, suggesting that there are at least four separate evolutionary origins for serine proteases. Clans SA, SB, and SC have a common reaction mechanism consisting of a common catalytic triad of the three amino acids, serine (nucleophile), aspartate (electrophile), and histidine (base). Although the geometric orientations of these residues are similar, the protein folds are quite different, forming a typical example of a convergent evolution. The catalytic mechanisms of clans SE and SF (repressor LexA) are distinctly different from those of clans SA, SB, and SC, since they lack the classical Ser-His-Asp triad. Another interesting feature of the serine proteases is the conservation of glycine residues in the vicinity of the catalytic serine residue to form the motif Gly-Xaa-Ser-Yaa-Gly (Brenner, 1988). Serine proteases are recognized by their irreversible inhibition by 3,4-dichloroisocoumarin (3,4-DCI), L-3-carboxytrans 2,3-epoxypropyl-leucylamido (4-guanidine) butane (E.64), di-isopropylfluorophosphate (DFP), phenylmethylsulfonyl fluoride (PMSF) and tosyl-L-lysine chloromethyl ketone (TLCK). Some of the serine proteases are inhibited by thiol reagents such as p-chloromercuribenzoate (PCMB) due to the presence of a cysteine residue near the active site. Serine proteases are generally active at neutral and alkaline pH, with an optimum between pH 7 and 11. They have broad substrate specificities including esterolytic and amidase activity.
(i) Serine alkaline proteases.
Serine alkaline proteases are produced by several bacteria, molds, yeasts, and fungi. They are inhibited by DFP or a potato protease inhibitor but not by tosyl-L-phenylalanine chloromethyl ketone (TPCK) or TLCK. They hydrolyze a peptide bond which has tyrosine, phenylalanine, or leucine at the carboxyl side of the splitting bond. The optimal pH of alkaline proteases is around pH 10, and their isoelectric point is around pH 9. Although alkaline serine proteases are produced by several bacteria such as Arthrobacter, Streptomyces, and Flavobacterium spp. (Boguslawski et.al., 1983), subtilisins produced by Bacillus spp. are the best known.

(ii) Subtilisins.
Subtilisins of Bacillus origin represent the second largest family of serine proteases. Two different types of alkaline proteases, subtilisin Carlsberg and subtilisin Novo or bacterial protease Nagase (BPN9), have been identified. Subtilisin Carlsberg produced by Bacillus licheniformis was discovered in 1947 by Linderstrom, Lang, and Ottesen at the Carlsberg laboratory. Subtilisin Novo or BPN9 is produced by Bacillus amyloliquefaciens. Subtilisin Carlsberg is widely used in detergents. Both enzymes exhibit a broad substrate specificity and have an active-site triad made up of Ser221, His64 and Asp32.

Aspartic acid proteases.
Aspartic acid proteases, commonly known as acidic proteases, are the endopeptidases that depend on aspartic acid residues for their catalytic activity. The active-site aspartic acid residue is situated within the motif Asp-Xaa-Gly, in which Xaa can be Ser or Thr. The aspartic proteases are inhibited by pepstatin (Fitzgerald et.al., 1990). They are also sensitive to diazoketone compounds such as diazoacetyl-DL-norleucine methyl ester (DAN) and 1,2-epoxy-3-(p-nitrophenoxo)propane (EPNP) in the presence of copper ions. Microbial acid proteases exhibit specificity against aromatic or bulky amino acid residues on both sides of the peptide bond, which is similar to pepsin, but their action is less stringent than that of pepsin. Microbial aspartic proteases can be broadly divided into two groups, (i) pepsin-like enzymes produced by Aspergillus, Penicillium,
Rhizopus, and Neurospora and (ii) rennin-like enzymes produced by Endothia and Mucor spp.

**Cysteine/thiol proteases.**

Cysteine proteases occur in both prokaryotes and eukaryotes. About 20 families of cysteine proteases have been recognized. The activity of all cysteine proteases depends on a catalytic dyad consisting of cysteine and histidine. The order of Cys and His (Cys-His or His-Cys) residues differ among the families (Barett, 1994). Generally, cysteine proteases are active only in the presence of reducing agents such as HCN or cysteine. Based on their side chain specificity, they are broadly divided into four groups: (i) papain-like, (ii) trypsin-like with preference for cleavage at the arginine residue, (iii) specific to glutamic acid, and (iv) others. Papain is the best-known cysteine protease.

**Metalloproteases.**

Metalloproteases are the most diverse of the catalytic types of proteases (Barett, 1995). They are characterized by the requirement for a divalent metal ion for their activity. They include enzymes from a variety of origins such as collagenses from higher organisms, hemorrhagic toxins from snake venoms, and thermolysin from bacteria (Hibbs et.al., 1985; Okada et.al., 1986; Shannon et.al.,1989; Weaver et.al., 1977; Wilhelm et.al., 1987). Based on the specificity of their action, metalloproteases can be divided into four groups, (i) neutral, (ii) alkaline, (iii) Myxobacter I, and (iv) Myxobacter II. The neutral proteases show specificity for hydrophobic amino acids, while the alkaline proteases possess a very broad specificity. Myxobacter protease I is specific for small amino acid residues on either side of the cleavage bond, whereas protease II is specific for lysine residue on the amino side of the peptide bond. All of them are inhibited by chelating agents such as EDTA but not by sulphydryl agents or DFP.

1.3.b Keratinolytic enzymes (proteases).

Shortage of protein in food and feed necessitates the search for a food source which is both economic as well as high in protein content. Waste management seems to be the only possible option answering both the needs. Recently lot of emphasis has been given by different research groups around the world by...
Introduction

finding ways to recover proteins out of waste products in Poultry industry. The major waste produced in poultry industry is feather that contains mostly keratin (80-90% on dry mass basis). But the utilization of feather waste as a dietary protein supplement for animal feedstuffs has been carried out only on a limited basis due to its poor digestibility. The poor digestibility of native feather protein is due to the constituent amino acid composition and configuration that provides structural rigidity. The mechanical stability of keratin and its resistance to microbial degradation depend on tight packing of the protein chains in α-helix (α-keratin) and β-sheet (β-keratin) structures (Parry and North 1998). Moreover, cross-linking of protein chains by cysteine bridges confers high mechanical stability and resistance to proteolytic degradation. Previously, steam pressure-cooked treatment was used to make the feather more digestible (Dalev 1994), although that process required significant energy and resulted in the deleterious destruction of some essential amino acids (Papadoulos and Ketelaars 1986). Therefore biological treatment is another choice to improve the nutritional value of feather waste (William et al. 1990).

The major difference between keratin and other proteins is the presence of a higher level of disulfide bonds in keratin. Reports on feather degradation propose that keratinases act on disulfide bonds, which are responsible for the mechanical stability of keratin (Williams, 1990); thus making it easier for proteolytic enzymes to act. Other reports suggest that the reduction of disulfide bonds by disulfide reductases (Bockle & Mueller, 1997) or the production of sulfite and thiosulphate (Kunert & Stransky, 1988) to be involved in keratin degradation. Involvement of cell-bound redox system for disulfide bond reduction has also been suggested in prokaryotes (Bockle & Muller, 1997). Therefore it has been well documented that keratinolysis is often assisted by disulfide reductase activity in extracellular milieu. In most cases of keratinolytic proteases, it was found that the purified enzyme can't degrade keratin (Ignatova, 1999); and only in case of keratinase from B.licheniformis PWB-1, it has been shown that the purified enzyme was capable of degrading feather keratin (Lin et.al.,1992).
Therefore, it can be said that for feather degradation, reduction in sulfydryl groups is the crucial step.

Keratinolytic enzymes have important usages in environment friendly biotechnological processes that utilize keratin containing wastes from poultry industry (Shih 1993; Onifade et al. 1998) as a protein source. Moreover, keratinolytic enzymes could be interesting for pharmaceutical and cosmetic industries. These enzymes are produced by some species of saprophytic and parasitic fungi (Bahuguna et al. 1989; Rajak et al. 1992; Safranek et al. 1982), a few actinomycetes (Mukhopadhyay and Chandra 1990; Noval and Assen 1974), some Bacillus licheniformis strain (Lin et al. 1992) and the thermophilic Fervidobacterium pennavorans (Friedrich 1994). The use of keratinolytic protease to nutritionally upgrade feather meal has been demonstrated. A comparable growth rate was observed between chicken fed isolated soybean and a methionine supplemented feather meal fermented with Streptomyces fradiae (Elmayergi et al. 1971). The utilization of a Bacillus licheniformis feather lysate with amino acid supplementation produced a growth curve identical to that of soybean meal (William et al. 1991). The use of crude keratinolytic protease significantly increased the amino acid digestibility of raw feathers and commercial feather meal (Lee et al. 1991).