Sampling, metadata and DNA extraction — important steps in metagenomic studies

Agnieszka Felczykowska†*, Anna Krajewska†, Sylwia Zielińska† and Joanna M. Łoś

Department of Molecular Biology, University of Gdansk, Gdansk, Poland

Metagenomic studies have become increasingly popular. They allow for the estimation of biodiversity in complex populations. This diversity presents an enormous but largely unexpected genetic and biological pool and can be exploited for the recovery of novel genes, entire metabolic pathways and their products. Generally, metagenomic studies are as follows: majority of the microorganisms present in the environment cannot be cultivated by standard techniques, DNA extraction methods are not very effective, isolated DNA is contaminated with various compounds, a choice for a screening method is not obvious.

Key words: sampling, DNA extraction, DNA purification, soil, water, wastewater, metadata

Received: 10 October, 2014; revised: 30 October, 2014; accepted: 12 November, 2014; available on-line: 13 February, 2015

INTRODUCTION

Due to the lack of ability to culture almost 99% of bacteria living in the natural environment with traditional microbiology methods, isolation of bacterial DNA directly from environmental samples has become a useful tool in molecular biology and biotechnology. Metagenomics is the study of genetic material recovered from microorganisms that cannot be cultured using already known methods and takes place directly in their natural environment. Isolation of bacterial DNA from natural environments has become a useful tool in detection of bacteria that cannot be cultivated in a traditional way, to determine the fates of selected bacteria or recombinant genes under natural conditions and to reveal genotypic diversity and its change in the microbial ecosystem (Zhou, 1996). Metagenomics, a term first coined by Handelsman in 1998, is a habitat based investigation of mixed microbial populations at the DNA level. The idea of cloning DNA directly from environmental samples was first propose by Pace, and in 1991 when the first such clone in phage vector was reported. Meagenomics is a tool used in genomics analysis of a population of microorganisms. Metagenomics combines many molecular techniques developed in the last century, enabling researchers to further study the diversity of microorganisms, their dependencies, and unlock the potential of biotechnology.

Sampling is one of the crucial steps in a metagenomic analysis. The way of collection and then storage of the gathered material affects the quality and quantity of the results. Metagenomics analysis involves extracting DNA from an environmental sample, cloning DNA into a suitable vector (cosmid, fosmid or BAC) producing large insert libraries, transformation of the host bacteria with the DNA obtained and screening of the resulting transformants.

Theoretically, a metagenomic library should contain clones representing the entire genetic complement of a single habitat, although this is dependent on the efficiency of DNA extraction and cloning methods. The information held within a metagenomic library can be used to determine community diversity and activity, presence of specific microorganisms or biosynthetic pathways as well as simply for searching for the presence of individual genes.

Construction of libraries with DNA extracted from different environmental samples lagged due to difficulties associated with maintaining the integrity of DNA during its extraction and purification.

This review highlights the most common problems in the early steps of collecting samples and metadata with DNA extraction from different environments in metagenomics research. In Fig. 1 we present the main pattern of the DNA extraction procedure from any environmental sample.

Problems associated with screening data, analysis and storage, are discussed in more details in “The most widespread problems in the function-based microbial metagenomics” review.

DNA ISOLATION

The first step in the isolation of nucleic acids from environmental samples is to choose the appropriate method of isolation. In the case of microbiological tests, inappropriate decision at this stage can result in erroneous identification of microorganisms and the absence of reliable knowledge about their function in the environment. The degree of our knowledge on biodiversity

†These authors contributed equally to this work.

Abbreviations: BAC, bacterial artificial chromosome; Bermuda Atlantic Time-series Study site; CAMERA, The Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis; CTAB, cetyltrimethyl-ammonium bromide; DOC, dissolved organic carbon; eDNA, environmental DNA; DGGE, denaturant gradient gel electrophoresis; ENS, European Nucleotide Archive; FISH, fluorescent in situ hybridization; GOLD,IMG, Integrated Microbial Genomes; MIGS, Minimum Information about a Genome Sequence; MIMS, Minimum Information about a Metagenome Sequence; POC, particulate organic carbon; PVPP, Polyvinylpyrrolidone; RBS, ribosomal binding site; VOCs, volatile organic compounds.

e-mail: agnieszka.felczykowska@biol.ug.edu.pl
in the environment to a great extent depends on what methods are used in the study, whether they were based on traditional microbial cultivation or on the analysis of the isolated genetic material (Kozdrój, 2010). Every method has its limitations. The type of environment (e.g. homogenous vs. heterologous), method of sampling and transport of the samples to the laboratory, are the main factors which affect the efficiency of the research methods used in metagenomics.

Methods which are based on the cultivation of microorganisms in specific media are simple, convenient and allow for simultaneous and rapid comparison of multiple samples using simple laboratory equipment. The disadvantage of these methods is significant underestimation of the actual variety of microorganisms, because it is confined to extract only organisms able to grow in the media (Welington et al., 1997). The techniques based on direct detection of microbial cells by analysis of nucleic acids are an alternative for traditional microbial cultivation.

Differences between nucleotide sequences determine the functional and structural differences between organisms. When one tries to assess the diversity of organisms in the environment, it is very important to develop efficient methods for extracting nucleic acids either from microbial cells or directly from the environment, and skillfully read the information embedded in their structure by billions of years of evolution (Milling et al., 2005).

When selecting methods of isolation, one must remember that DNA should be isolated from the whole spectrum of microorganisms present in the biotope, and most importantly, isolation should not physically disrupt the genetic material. Contamination with proteins, humic acids and metals should be kept to a minimum. Different groups of microorganisms (bacteria, fungi, protozoa) have different susceptibility to lytic reagents due to the differences in the structures of their cells. Majority of the microorganisms which are analyzed are present in the environment in the form of spores, which are metabolically dormant and show high resistance to lytic agents. Poor yields and small sample size of the DNA isolated from these organisms make them inapplicable for further metagenome analysis (Steele & Streit, 2006).
Many studies of DNA extracted from environmental samples are focusing on 16S rRNA sequence data analysis, obtained by PCR amplification (Bernhard & Field, 1999; Peters et al., 2000; Cai et al., 2003; Verhelst et al., 2004; Yu et al., 2008). It has been shown that due to differences in the cell wall and membrane structure of bacteria, effectiveness of DNA extraction can depend on the extraction protocol used (Wellington et al., 1997; Krsek & Wellington, 1999; Carrigg et al., 2007). It is highly important to choose and optimize a proper DNA extraction protocol for the target group in a given study, not only due to DNA extraction process per se, but also because of further analysis challenges. DNA isolation can be crucial when environmental DNA extraction is involved. It is important to obtain a good quality DNA with high purity and low degree of fragmentation. Mechanical methods used for bacterial cell lysis, like sonication, bead-beating homogenization or freeze-thaw cycles can increase efficiency of cell lysis, but can also shear DNA, and such DNA is often not suitable for further molecular use. In order to receive reliable PCR amplification or enzyme digestion, it is also required to obtain nucleic acids that are free from enzymatic inhibitors such as humic acids, heavy metals or proteins.

Morgan and coworkers (2010) suggested to use multiple DNA extraction procedures for a single environmental sample to increase the likelihood of including every organism in the tested sample. With a simple test, by in vitro-simulated microbial community, they demonstrated that with the use of two different DNA extraction protocols, two libraries can be created from a single mixture of organisms and thus data suggesting two various communities can be obtained. Therefore, it has to be considered at an early stage of a given study which extraction protocols to choose in order to obtain DNA from the target group of organisms.

### Sampling and Metadata

Sampling can be of premier importance for the quality of data obtained as well as the interpretation of results. It is considered as a crucial step when using metagenomic approaches, since the possessed sample may not be of a representative size (Thomas et al., 2012). Especially when describing biodiversity samples should represent the whole population from which they are taken (Wooley et al., 2010) and when describing a habitat, samples must be representative of the habitat (Handelsman et al., 2007). It is also important to include the whole target group of microorganisms and prepare samples by selected methods. Paul and Clark (1989) indicated how important is the time in which the soil sample is transported and what are the conditions and length of storage. The preservation time should be limited to the minimum and biological analyses should be performed as soon as possible after sampling, to minimize the effect of storage on bacterial communities. It can be particularly important for minimizing the risk of contamination and for obtaining reliable results.

Habitat changes over time in response to changing conditions. This is a central way to understand the community structure and function (Handelsman et al., 2007). Thus, there are plenty of questions to ask and answer before the sampling step. When collect the samples? One needs to consider the time of the year and the time of the day as well. How many samples and what the sample volume are needed to represent various environmental conditions? What are the specific features of the environment and can they be misleading in the interpretation of data obtained?

A detailed description of the environmental context and the methods used appears to be necessary to compare studies and results. It is becoming increasingly important to organize diverse and complex data that the users can locate freely, easily understand and analyse according to their interests (Barret et al., 2012). Handelsman and coworkers (Committee on Metagenomics: Challenges and Functional Applications, National Research Council, 2007) recommended to carefully reflect the strategy of the sampling and the variability of the experimental methods. They created a list of questions worth to be considered before sampling begins. Collected metadata provide information about a source of the sample, and when and under which conditions it was sampled. In microbial ecology, it can refer to physical, chemical or other environmental features of the sample. Barrett and coworkers (2012) presented BioProject databases at NCBI in order to facilitate access and organization of metadata. It is important that collected data is adequately organized and description can also provide appropriate annotations and context. It is greatly appreciated that for complete understanding of experimental results it is necessary to obtain metadata in addition to the actual data.

For better understanding, it is highly important to provide good metadata. Unfortunately, it is not easy to get information on all aspects of the environment tested, especially from marine waters. In this case, a good idea for complex interpretation is to use a well-known habitat. Venter et al. (2004) sampled in the Sargasso Sea, a nutrient-limited, open ocean environment. The aim was to test whether whole-genome shotgun sequencing can be effectively applied to gene and species discovery, as well as for overall environmental characterization. The choice of this location was intentional, as it is a well-known and characterized region of the global ocean, especially the Bermuda Atlantic Time-series Study site (BATS). Due to intensive physical and biogeochemical studies, it provided a great opportunity for interpretation of environmental genomic data in an oceanographic context (Venter et al., 2004).

Field et al. (2008) proposed to use MIGS (the minimum information about a genome sequence) as the formal way of describing genomes and metagenomes in a more detailed way. MIGS allows the use of a comparative genomic analysis to provide a better understanding of the source of each genome, and enables to locate genomes and metagenomes in their geospatial and temporal context (when relevant) through specification of geographical location and sampling data. In addition, the authors prepared a MIGS checklist containing useful information about sampling steps like geographic location, habitat, time of sampling and MIMS specification (Minimum Information about a Metagenomic Sequence) — describing habitat parameters like temperature, pH or salinity. The provision of the large amount of various information (metadata) can simplify data analysis and allow for a better interpretation of results. However, in order for this to happen, there is a need to provide metadata in a standard, simple and unequivocal form.

### The soil habitat

Soil represents the most challenging environmental niche for microorganisms. It harbors enormously diverse microbial communities, and it is a major reservoir of microbial genomic and taxonomic diversity. Before we proceed to the isolation of DNA from soil, we should mind...
that the total number of bacterial cells living on Earth is close to 4–6×10^29 from which about 2.6×10^29 cells exist in the soil (Torsvik & Ovreas, 2002). There is about 10^9 prokaryotic organisms in one gram of soil, and more than 2 thousand of different types of genomes. Without taking into account genomes of rare species and microorganisms whose DNA was not recovered during isolation (Torsvik et al., 1996; Torsvik & Ovreas, 2002; Daniel, 2005), average representation of one type of genome is therefore less than 0.05% (Stein et al., 1996; Rosellomora & Amann, 2001). Studies on diversity of prokaryotes in soil showed that only 0.1–1.0% of the bacteria may be obtained from the environment by means of microbiological methods, and then cultivated in the laboratory (Amann et al., 1995; Hugenholtz et al., 1998; Torsvik & Ovreas, 2002; Steele & Streit, 2006). The remaining 99% of bacterial soil population remains unexplored and can be a source of unknown genes.

In metagenomic studies, a vast number of methods is employed for isolation of nucleic acids directly from the soil (Handelsman et al., 1998). Although many of the methods to isolate soil DNA have been described, none of them is universally applicable in soil metagenomics (Zhou et al., 1996; Harry et al., 1999; Lakay et al., 2007). The origins of bacterial DNA isolation from different types of soil date back to the 80s of 20th century, when Vigdis Torsvik from the University of Bergen has published the first extraction procedure (Torsvik, 1980), which involved the separation of bacterial cells from soil particles, followed by lysis of the cells, and separating DNA and RNA from organic matter by a series of chromatographic separations. Unfortunately, the procedure was time-consuming, required large amounts of soil, and was not very effective. Since then, the extraction of nucleic acids has been simplified. Now, it requires smaller volume of samples, which increases the number of samples that can be analyzed simultaneously.

There are two ways for nucleic acids isolation from soil. The direct extraction of nucleic acids in situ, after lysis of bacterial cells present in this natural environment (e.g. soil matrix) (Ogarm et al., 1987), and the indirect method which requires separation of bacterial cells from soil particles, followed by lysis and final step of nucleic acids purification (Holben et al., 1988; Courtois et al., 2001; Robe et al., 2003). Both approaches have their advantages and disadvantages associated with DNA yield, purity and representation of microorganism diversity (Tsai et al., 1991; Courtois et al., 2001).

Before choosing the appropriate isolation technique, we must consider a number of factors such as: type of environment, size of DNA, and the purpose of its subsequent use. An optimal method should: avoid excessive fragmentation of genetic material by physical factors, prevent the degradation of the DNA by nucleases, and ensure to obtain genetic material of high quality and low contamination with substances that inhibit later analysis.

Direct Methods. The direct in situ lysis extraction method has been widely used during the last decade. This method, which involves complete in situ lysis of all microorganisms, generally provides the highest DNA yields within acceptable processing time. The disruption of the microbial cell wall is the first step, and leads to the release of all nucleic acids from bacteria to the extraction buffer. In the second step, which is preceded by separation of the extraction buffer from soil particles, nucleic acids are isolated from the extraction buffer. This is the most challenging step, because a lot of contaminants such as humic acids, heavy metal ions, and proteins are extracted along with DNA. The choice of the extraction buffer is a compromise between the expected DNA quantity and the required DNA purity (Robe et al., 2003).

Microbial cell disruption is usually a combination of physical, thermal, chemical and enzymatic lysis. Physical treatments such as bead-beating homogenization, sonification, vortexing (Steffan et al., 1988; Miller et al., 1999; Maarit Niemi et al., 2001; Miller, 2001), and thermal shock (thermal treatments: freezing-thawing, freezing-boiling (Tsai et al., 1991; More et al., 1994; Porteous et al., 1997; Orsini & Romano-Spica, 2001) destroy soil structure, and tend to yield the greatest access to the whole bacterial community, including bacteria hidden deep within soil microaggregates. They have also shown efficiency for disruption of vegetative forms, small cells and spores, but they often result in significant DNA shearing (More et al., 1994). The average size of the DNA fragments varies from 600 bp to 25 kbp, when using physical lysis. This allows using them for plasmid, phage or cosmid library construction, as well as for preforming a PCR reaction. However, too intense lysis may cause excessive DNA fragmentation. Chemical lysis either alone or in association with physical methods, has been also used extensively. It requires preliminary grinding of the material which allows the extraction lysis buffer to access the cells imbedded in soil aggregates. Probably the most common chemical used is sodium dodecyl sulfate (SDS) which dissolves the hydrophobic part of cell membranes. Detergents have often been used in combination with heat-treatment and with chelating agents such as EDTA, Chelex 100 (Robe et al., 2003) and various Tris and sodium phosphate buffers (Krsek & Wellington, 1999). Increasing the EDTA concentration results in higher yields, but lowers purity of the isolated nucleic acids. Other chemical reagents that are used, such as cetyltrimethyl-ammonium bromide (CTAB), can partially remove humic acids (Zhou et al., 1996), and form insoluble complexes with denatured proteins, polysaccharides and cell debris (Saano et al., 1995). Polyvinyl-polypyrrolidone (PVPP) can also help to remove humic acids during lysis, but it lowers the DNA yield, thus it is recommended to use PVPP only for the nucleic acids purification step (Krsek & Wellington, 1999).

Enzymatic methods are based on the sample digestion by various enzymes. They affect DNA in the mildest way and are particularly useful in the case of Gram positive bacteria, which are resistant to physical and chemical methods, and when the size of isolated DNA is of high importance (e.g. BAC libraries). Enzymes can also be used for destroying DNA nucleases, and for removal of RNA. The most commonly used enzymes are: lysozyme, proteinase K, RNase A (Tsai et al., 1991; Tebbe & Vahjen, 1993; Zhou et al., 1996; Maarit Niemi et al., 2001), or achorromopeptidase (effective on lysozyme resistant bacteria) (Simonet et al., 1984).

Indirect methods. The first, and the most important step in indirect methods, is to disperse the soil matrix in order to isolate as many intact bacterial cells (for high quality DNA), representing full diversity of microbial life, as possible. The next step is the cell lysis followed by isolation and purification of DNA. To disperse the soil, one can use both physical and chemical methods. The most common physical method employs homogenization, sonication, shaking or a rotating pestle procedure (Robe et al., 2003). Chemical methods are mostly used in combination with physical ones. The majority of chemical compounds used in these methods are detergents like: SDS, PEG (Steffan et al., 1988), sodium deoxycholate (McDonald, 1986), sodium chloride, and PVPP, which
can lower the humic acids level (Steffan et al., 1988). Cation exchange resin also proved to be effective (McDonald, 1986). However, chemicals can also cause negative effects, such as fragmentation of DNA after disruption of the cell wall, therefore, it is very important to maintain the integrity of the cell during this step.

Another method of separating the bacterial cells from the soil matrix is centrifugation based on differences in sedimentation between the individual components of the sample (Robe et al., 2003). The method consists of two subsequent centrifugations. The first one, performed at low acceleration, serves to remove large pieces of soil and fungal thalli. The second one, performed at high speed, employs supernatant obtained in the first centrifugation to collect the bacterial sediment. After one cycle it is possible to separate about 10% of bacteria present in the soil sample, and according to the authors, this represents the whole biological diversity of the sample (Holben et al., 1988). Subsequent cycles of centrifugation will increase the amount of material obtained.

An alternative method is density gradient centrifugation. For the gradient medium one can use Percoll, metrizamide, Nycodenz (Robe et al., 2003), or sucrose (Pillai et al., 1991). The efficiency of this method of separation varies from 6 to 50% of the total number of bacterial cells contained in the soil sample. The efficacy depends mainly on the composition of the soil. Processing of soil with a high clay content is very challenging. In comparison to the sedimentation method, gradient centrifugation allows to obtain bacterial cells that are less contaminated (Robe et al., 2003). Separation of the cells is followed by a step of isolation and purification of DNA. Isolation procedures are similar to those described in Chapter 3: "The soil habitat".

The choice of the method depends on the outcome we want to gain. Direct methods will give us a relatively large amount of DNA, with a broad spectrum of representativeness of microorganisms present in the soil sample. When high purity of DNA is needed, which is crucial for later molecular analysis, indirect methods are recommended.

Regardless of whether we choose a direct or indirect method, we will always obtain nucleic acids which are contaminated to a different degree with proteins, humic acids, polysaccharides, lipids, minerals, as well as active enzyme inhibitors (e.g., metal ions) used in molecular reactions (Milling et al., 2005). Most of DNA purification methods are based on precipitation with: potassium acetate, PEG, ethanol or isopropanol used alone or in combination. Sephadex gel filtration, ion exchange chromatography column, agarose or PVPP/PVP gel electrophoresis exhibit similar effects of the selective binding and precipitation of proteins as do humic substances present in the crude extract of the DNA (Cullen & Hirsch, 1998). Cesium chloride gradient centrifugation is often used to purify high quality DNA, with sizes up to 100 kb (Robe et al., 2003). This method is time consuming, and faster alternatives (yet with lower yield) are available on the market. By using "ready to use" DNA extraction and purification kits, we can process different types of soil samples and get a relatively pure DNA in a short time.

Nowadays we can witness ongoing efforts for improvement of methods of DNA purification after environmental sampling. Still, there are up to 50% losses in the isolated DNA at this stage (Carrigg et al., 2007). Therefore it is very important to choose an appropriate lysis method and a suitable extraction buffer to provide a lot of DNA with contamination kept at a minimum level. This is why each step of isolation must be carefully planned and considered.

The water habitat

Water covers around 71% of the Earth’s surface. The vast majority of water is found in seas and oceans, just a few percent in groundwater, rivers, and lakes. Freshwater accounts for only about 2.5% of the total volume of water available on our planet and much of it is stored in the form of ice (Debros et al., 2009). There can be plenty of microscopic life forms in the ocean water. Nominal cell counts of >10^9 cells per ml in surface sea water were found, while there are predictions that the oceans harbour 3.6×10^3 microbial cells (Sogin 2006). Groundwater, unlike surface water, is often inhabited by sparse microorganisms with low species diversity due to its oligotrophic character.

Metagenomic approaches have already been applied to many water environments, but oceans and seas are the main part of our planet and many metagenomic projects are focused on DNA extracted from marine microorganisms (Venter et al., 2004; Sogin et al., 2006; Mohamed et al., 2013; Ferreira et al., 2014), including coastal lagoons (Rivera et al., 2003; Ghai et al., 2012). Most of them are focused on exploring biodiversity and genome analysis of unknown taxa, as well as expression of novel and useful genes, or detection of pathogenetic bacteria (Rivera et al., 2003).
Presently, more and more projects concentrate on other than marine water environments, like hot springs (Tckere et al., 2011; Jiménez et al., 2012), lakes (Oh et al., 2011), rivers (Ghai et al., 2011; Amos et al., 2014) or small water ponds (Ranjan et al., 2005; Kapardar et al., 2010) which can also offer discovery of unknown genes as well as documentation of unexpected species in a particular environment. In 1997, phylogenetic analysis of bacterial communities in the Columbia River, its estuary and adjacent coastal ocean, demonstrated a wide diversity of species. DNA sequences found in the river samples were remarkably similar to those found in lakes from the Netherlands, Alaska or the Adirondack Mountains, which confirmed existence of cosmopolitan fresh water bacteria. This research also revealed, in all tested environments, that the clones isolated belonged to clades of common soil bacteria. Probably, due to the interactions between soil and water environments, there is a close relationship and overlap in bacterial communities (Crumп et al., 1999). In another case, also by using metagenomic approaches based on 16S rRNA analysis, actinobacteria (considered as a typical soil inhabitant) were found in the ocean, and later also in lakes, which led to a better understanding of the species and to reconstruction of the genomes of uncultivable marine actinobacteria (Ghai et al., 2012).

Furthermore, it is worth mentioning that drinking water is also explored by metagenomic approaches (Bai et al., 2013; Chao et al., 2013). It can be used to find pathogenic bacteria (Ocepek et al., 2011), which are hard to detect or in the case when it takes a long time to cultivate them with traditional methods. Shi et al. (2013) investigated the chlorination effects on microbial antibiotic resistance in a drinking water treatment plant and also presented phylogenetic analysis based on bacterial DNA extracted from concentrated water samples. In 2012, the Gomez-Alvarez group applied next-generation sequencing techniques to characterize the composition and functional diversity of bacterial populations extracted from drinking water treated with various disinfection strategies. In other studies, metagenomic approaches were used to characterize the viral community found in reclaimed water, and to compare it with viruses in potable water to clarify concerns about reclaimed water used as an alternative water supply (Rosario et al., 2009). When dealing with water samples of different origin, various problems associated with typical features of the environment can occur and expected solutions need to be adapted to the test sample. Drinking water, due to disinfection treatment, has low biomass, while pond water is often more polluted, as is the river water which is also very variable. When sampling and searching for extraction methods, all those aspects need to be considered.

Water is a very difficult environment to describe because of its size, salinity, and variability which stems from exchange of water due to currents and waving, periodical changes of water levels, and anthropological factors. Furthermore, these characteristics can be connected to geolocation, insolation or precipitation. Regardless of the research assumptions and methods used, exploring the genomic diversity of microbial communities by metagenomic approaches begins with sample collection.

Through functional metagenomics, many novel antibiotics were identified as were proteins involved in antibiotic resistance, vitamin production, and pollutant degradation (Handelsman et al., 2007). It is of crucial importance to gain high-quality DNA from a sample to obtain a representative metagenomic data. However, this step could be very challenging. The physical and chemical structure of microbial community affects quality of DNA, as do size, amount and purity.

Many water extraction protocols have been published (Fuhrma et al., 1988; Somerville et al., 1989; Schmidt et al., 1991; Boccuzzi et al., 1998; Crump et al., 1999; Rivera et al., 2003; Ocepek et al., 2011) as well as commercial kits. Regardless of the method chosen, the first step of any water extraction protocol is concentration of the sample, and it can be achieved by centrifugation, filtration or combination of both. Water sample can be filtered with the use of various flow filter systems using different pore sizes which depend on the target group of microorganisms. On the other hand, the real challenge can be the amount of concentrated water. Sabree et al. (2009) draw attention to an important issue when preparing metagenomic DNA from a water sample. To obtain enough DNA to build libraries in order to access planktonic communities, it is required to prepare the equipment that is capable of handling large volumes of water to concentrate sufficient microbial biomass. Due to low biomass of drinking water, Shi et al. (2013) filtered about 2000 L of water in 48h, to concentrate the bacterial cells. Jiménez et al. (2012) concentrated 10 L of water from acidic hot springs from the Columbian Andes and due to the low amount of recovered DNA (about 116 ng per liter) they decided to perform amplification of the DNA with the use of Φ29 polymerase prior to 454 pyrosequencing.

Viral genomes are smaller and comparatively shorter than those of bacteria. The amount of recovered DNA from environmental samples is often insufficient for further analyses such as the construction of cloning libraries or 454 pyrosequencing. Therefore, Kim and Bae (2011), after concentration of 16 L of seawater, decided to amplify viral DNA with the use of the linker amplified shotgun library (LASYL) and multiple displacement amplification (MDA) with the use of random hexamers and Φ29 DNA polymerase.

After concentration, to achieve cell lysis combination of enzymatic treatment, high temperature, detergent treatment, and mechanical disruption often has been used. Authors often introduced slight improvements to adapted protocols (also to those from soil extraction), in order to optimize them for a given test sample. When dealing with water samples it can be relatively easy, in comparison to soil, to remove chemical and enzymatic inhibitors. As contaminations inhibit enzymatic reactions and influence cloning efficiency, it is highly important to remove all inhibitors from a DNA sample. Thus, additional purification steps may be required, especially when dealing with polluted environments.

Contaminated and unique environments are often associated with obtaining exceptional features. The knowledge about special properties of sampled water can be used to reach DNA from uncultivated bacteria with unique physiological mechanisms. Microbial communities from pond water can withstand fluctuations of salinity due to evaporation and dilution, therefore, bacteria living in such an environment are expected to possess unique stress tolerance mechanisms. In 2010, Kapardar et al. (2010) identified and characterized two novel salt tolerance genes from pond water, and in 2005, Ranjan et al. isolated twelve unique genes encoding enzymes with lipolytic activity with low similarity to already known lipolytic proteins. However, when extracting DNA from pond water described as greenish-brown in colour, Ranjan et al. (2005) mentioned that after initial isolation procedure, the metagenomic DNA obtained was resistant to restriction enzyme digestion. It
was further purified, with good effect, by CTAB in order to obtain digestible pure DNA. This indicated that finding proper methods, extraction protocols, and solutions suitable for the environment being tested can be very challenging. It is important to be aware of steps that are required before extraction, as well as any difficulties that can appear during DNA extraction, before it actually begins.

THE SLUDGE HABITAT

Wastewater microbiology is recognized as a mature and dynamic discipline, which offers much towards a deeper understanding of life in complex microbial communities (Daims et al., 2006). Microbial communities inhabiting wastewater environment are of significant interest for applied as well as basic microbiology. This population has been extensively studied for a number of years. However, only with the development of molecular and metagenomic approaches it has become possible to assess the true diversity of wastewater communities (Snaidr et al., 1997). The microbial community of water and wastewater treatment systems has been examined for many years. First, from an ecological point of view, researchers have been interested in determining the ultimate diversity of the system and function such as bulkling, foamation, nitrification, etc. Second, from a health perspective, researchers were interested in determining the identity and level of pathogens (Gilbride et al., 2006). Initial investigation into the composition of wastewater microbial communities was based on traditional microscopy observations (Cruds, 1975; Eikelboom, 1975) or culture-dependent techniques (Ueda & Earle, 1972). But these culture-dependent isolated bacteria do not accurately represent the composition and diversity of natural microbial communities (Ward et al., 1990).

Fresh drinking water is constantly required and wastewater is constantly produced. The provision of drinking water and the management of wastewater have thus been crucial to the success of human civilization. When it comes to the purification of sewage water, microorganisms are superior to humans, their abilities to degrade the most diverse of organic substance and to recycle elements such as nitrogen, phosphorous and carbon, are unmatched in nature (Daims et al., 2006). Successful exploitation of these features has been achieved for almost a century in biological wastewater treatment plants (WWTPs). These facilities are among the most important biotechnological applications, preventing the pollution of natural ecosystems and the spread of sewage-borne diseases.

Wastewater is a mixture of different pollutants, which can be characterized in several groups according to the type of contamination or origin. This variety of pollutants in wastewater causes great difficulties in planning metagenomic experiments. The degree of contamination can be determined by many factors such as: the size of the agglomeration, degree of urbanization, seasons. Generally, nontoxic wastes are contributed mainly by the food industry and by domestic sewage, whereas toxic wastes are contributed by coal processing (phenolic compounds, ammonia, cyanide), petrochemical (oil, petrochemicals, surfactants), pesticide, pharmaceutical, and electroplating (toxic metals such as cadmium, copper, nickel, zinc) industries (Kumaran & Shivaraman, 1988). Major contaminants found in wastewater are biodegradable organic compounds, volatile organic compounds (VOCs), recalcitrant xenobiotics, toxic metals, suspend-
The broadest techniques used in the studies of wastewater are: denaturant gradient gel electrophoresis (DGGE), fluorescent in situ hybridization (FISH), and cloning of 16S rRNA (Sanz & Köchling, 2007). Cloning and sequencing of the gene that codes for 16S rRNA was the most widely used in the field of microbial ecology and sequencing of the gene that codes for 16S rRNA (Sanz & Köchling, 2007). Cloning (DGGE), fluorescent in situ hybridization (FISH), and wastewater are: denaturant gradient gel electrophoresis (Wagner & Loete, 2002; Błaszczyk et al., 2011). The analysis of 16S rRNA genes, aided by using PCR to amplify target sequences in environmental samples, has enabled microbial ecologists to identify and characterize microorganisms in a natural community. The 16S rRNA gene contains both, highly variable and highly conserved fragments that enable us to analyze all organisms in the community. One of the methods employing these properties is the amplified ribosomal DNA restriction analysis (ARDRA) (Błaszczyk et al., 2011).

DNA and RNA are usually extracted from active sludge samples using two separate methods developed for soil and sediment samples. However, active sludge differs from soil and sediment in the last three aspects: high biomass density, low humic acid content, and the presence of bacterial aggregate flocks (Zhongtang & Mohn, 1999). In this case, problems of extraction are similar as in the case of soil and sediment samples (see Chapter: “The water habitat”). Cloning was employed to establish with precision the phylogenetic position of filamentous bacteria in granular sludge that were previously affiliated, by in situ hybridization, to the division of green-sulfur bacteria (Sekiguchi et al., 2001); or to determine the prevalent sulfate reducing bacteria in a biofilm (Ito et al., 2002). Another technique takes advantage of the fact that DNA fragments of the same size band with different nucleic acid sequence differ in mobility when run on a gel under denaturing conditions, thus generating band patterns that directly reflect the genetic biodiversity of a given sample. The number of bands corresponds to the number of dominant species. Coupled with sequencing and phylogenetic analysis of the bands, this method can give an overview of the composition of a given microbial community. DGGE has been used for the evaluation of the granular sludge’s microbial diversity from UASB reactors treating brewery wastewater (Chan et al., 2001), alcohol distillery (Akarsubasi et al., 2001), and unbleached pulp plant wastewaters (Buzzi et al., 2006). This technique is not used alone but rather as a part of a combined approach with other methods, for example with in situ hybridization (Santegoeds et al., 1998; Onda et al., 2002). The most important application of DGGE is monitoring dynamic changes in microbial communities, especially when many samples have to be processed. An excellent way to overcome some of the problems of studying microbial populations of a microcosm, without resorting to traditional methodology, is to use fluorescent probes. These are short sequences of DNA (16–20 nucleotides) labeled with a fluorescent dye. These sequences recognize 16S rRNA sequences in fixed cells and hybridize with them in situ (DNA–RNA matching). Microorganisms can be identified, localized and quantified in almost every ecosystem with hybridization (Amann et al., 1990).

Wastewater, like the soil, is an environment in which multiple factors determine the choice of methods for extracting DNA/RNA and later analysis. The techniques used are most often a combination of several methods to achieve the desired results.

CONCLUSIONS

Due to habitat changes over time, it is important to properly plan experiments in order to obtain data of good quality. Obtained results should have detailed description of the environmental context (metadata) to compare studies and results. Sampling and DNA extraction can be considered as crucial steps when using metagenomic approaches. During sampling, it is important to remember that samples should represent the population from which they are taken, storage time should be also limited to the minimum and biological analyses should be performed as soon as possible after sampling, as it can affect results.

Isolation of bacterial DNA directly from environmental samples has become a useful tool in molecular biology and biotechnology. It can lead to discovery of new genes, ways of resistance to antibiotics or describe microbial biodiversity in a specific environment. In order to effectively deal with these issues, researchers should have access to efficient DNA extraction methods together with appropriate techniques for their further analysis. The best solution would be a universal method of DNA isolation from various environments that would lead to extracting a relatively high amount of high quality DNA. Unfortunately, all efforts to find all-purpose extraction method are insufficient because of too many different variables affecting this process. As a result, an appropriate selection and optimization of extraction methods must be performed for each habitat individually.

Acknowledgements

This work was supported by European Commission (grant no. 245137, project acronym: MAREX, large scale integrating project within KBBE-2009-3-2-01) and Ministry of Science and Higher Education, Poland (grant no. 1544/7.PR UE/2010/7). This work was also supported by National Science Center, Poland (grant no. 2011/01/D/NZ2/04817).

REFERENCES


160 A. Felczyska and others


