Use of environmental DNA in investigating the Arctic terrestrial ecosystem

Sylwia Zielińska

Environmental DNA (eDNA) is a DNA extracted from various environmental samples, such as soil, sediments, water, faeces, without first isolating any target organism. The concept of the isolation of eDNA is based on the fact that all biological components of a given environment contain nucleic acids, and their isolation enables detection of a variety of species regardless of their origin, life stage or gender. Ability to obtain and analyse environmental DNA has led to a new approach in order to monitor a given species' presence in different environments and has an enormous potential for the non-invasive detection, i.e. the identification, spread and abundance, population size estimates or conservation of a highly sensitive target species. eDNA can provide a lot of valuable information, even when it is based only on basic molecular techniques, like PCR amplification, agarose gel electrophoresis or traditional sequencing, which can be still easier and faster to perform than traditional approaches based on observation, behaviour and trapping.

Due to the lack of ability to culture a vast majority of bacteria under laboratory conditions, metagenomics has become a method of choice for analysis of environmental samples. Metagenomics allows to study the genetic material (eDNA) recovered from microorganisms which currently cannot be cultured directly from their natural environment. Although metagenomics is still quite a new approach in microbiological research, it has already delivered plenty of information about microbial communities from various environments, like soil, sediments, water, sludge, faeces, skin, etc. Recent advances in high-throughput sequencing technologies provide a much wider access to the genetic content of uncultured microbial communities and make bacterial communities structure analyses much easier and faster to perform, without the need to clone the DNA sequences into a vector. Metagenomics has become a powerful tool in discovery of new genes or estimating microbial biodiversity, and also in microbial ecology description of the taxonomic distribution of the microbial community members.

The Arctic region comprises about 4.8% of the Earth's land surface and its environment is often considered to be generally simple, with poor species diversity, and characterized by short food chains and a permanent shortage of nutrients. The unique area of the Arctic is receiving much attention mostly because of consequences of the observed and predicted rapid environmental changes associated with the global climate change. Thus, all variations in the Arctic biodiversity and the functioning of the ecosystem at all trophic levels, from microorganisms to vertebrates,

are being carefully observed and reported. This is very important in the case of polar ecosystems, given to the climate changes and the approximately three times faster temperature increase than in other regions of the world. The ongoing development of molecular tools, metagenomics and environmental DNA techniques increases our ability to gain knowledge about microorganisms, plants and animals based only on easily accessible samples, like soil or faeces. Metagenomics, together with the Next Generation Sequencing (NGS) and eDNA techniques, could be used for a broad analysis of the Arctic environment.

The high Arctic archipelago of Svalbard, with Spitsbergen being the largest island, is a breeding ground for many seabirds, including the little auk (*Alle alle*). The little auk population in Hornsund exceeds one million members and is one of the largest colonies in the world. Generally, marine birds that forage at sea and breed on land deposit large amounts of guano, eggshells, feathers, and carcasses near their colonies and this large-scale transport of organic and inorganic matter from the sea to the land is crucial for the local ecosystems. Soil fertilization close to the seabird colonies facilitates the diversity and abundance of terrestrial plant communities, microorganisms, as well as associated invertebrate and vertebrate communities, as rich vegetation attracts populations of herbivores, and subsequently predators, scavengers and decomposers. Their presence can be noticed by numerous faeces and prey remnants, which also enhance the soil nutrient concentrations. Moreover, there are several reports presenting influence of the guano deposition on the physical and chemical soil parameters and in consequence enhancing formation of ornithogenic soils, thus facilitating the development of associated terrestrial plant communities. However, until now the influence of bird-derived soil fertilization on the structure and abundance of microorganisms has not been reported in the Arctic environment. Also, despite the harsh environmental conditions, microbial communities in the Arctic soil are as diverse as those found in other biomes.

Birds are not the only ones to influence the Arctic ecosystem. Large mammals, like the Svalbard reindeer (*Rangifer tarandus platyrhynchus*), which is the most isolated subspecies of Rangifer, endemic to the Svalbard archipelago, also shape this environment by grazing and enhancing the soil nutrient concentrations (by their urine and faeces). In Hornsund, the southernmost fiord of Spitsbergen, the reindeer began occurring regularly only since the 1990s, thus the local population is still small and does not exceed 20 individuals. They live individually or in small groups, and do not perform seasonal migrations. They show seasonal variation in selection of the grazing areas and diet. During the summer, they spend most of their time accumulating fat, particularly feeding on a high quality vascular plant species growing in the wet areas, in the valleys and lowland plains, often in the vicinity of large seabird colonies.

Reindeer are foregut fermenters whose digestion depends on a symbiotic association with the complex microbiota community resident within their rumen. This resulted in studying of the reindeer`s microbial community inhabiting mainly rumen, causing death of the animal or at the very least causing a considerable stress to the animal during sampling. This practice during sampling prevents tracking changes of rumen microbiota composition over time, e.g. according to seasonal and ecological changes, including food availability. In case of studying wild populations with limited or no previous exposure to humans, it seems reasonable to use non-invasive sampling methods, that is to test the faecal samples.

Collected faecal samples can be a valuable source of information about the Svalbard reindeer population. DNA extracted from faeces constitutes a source of information about the presence of Shiga toxin producing *Escherichia coli* (STEC). Ruminants, mainly cattle, which are asymptomatic carriers, are considered as a main reservoir for the Shiga toxin pathogens. Based on amino acid sequence, we can distinguish two major types of this protein: Stx1 and Stx2. The Stx1 protein, produced by STEC, is identical to the Shiga toxin derived from *Shigella dysenteriae* I. The Stx2 protein is homologous to Stx1 at the level of 55% for the A subunit, and 57% for the B subunit. Furthermore, DNA-based approach can deliver information about the animal which delivered the faeces, like the gender of reindeer through the male-specific DNA amplification, in order to facilitate inference for instance in the microbiological context.

Based on the faecal samples, the symbiotic bacterial community structure can be also explored. Previously, microbial community inhabiting the reindeer gut was mainly studied by traditional cultivation methods, which accounted for 10-20% of the total gut microbial community or with the use of classical molecular techniques. Recent advances in high-throughput sequencing technologies provide an opportunity to also analyse the uncultivable bacteria and those with low abundance, often contributing to less than 1% of the total population.

The primary aim of my PhD thesis was to describe the biodiversity of ornithogenic soil ecosystem of the Arctic and to investigate microbiome of the Svalbard reindeer, based on non-invasive sampling, with the use of molecular approaches. In the conducted analysis, I combined the classical biology methods, like PCR reaction, agarose gel electrophoresis and traditional sequencing with modern molecular biology methods: metagenomics and the Next Generation Sequencing. PCR reactions and agarose gel electrophoresis enable fast determination of the presence of the searched DNA sequences and traditional sequencing allows for identification of potential DNA differences. Thus, environmental DNA techniques, based on the collected environmental samples, allow demonstrating the presence of a species without observing or capturing individuals and moreover allow to define its features. Further, metagenomics enable

testing of the environmental samples without the cultivation of microorganisms under the laboratory conditions. This experimental design allowed better access to the information about the group of uncultured microorganism, which can constitute even 99% of the total micro-community [1–4].

The main goal was completed by evaluating the microbial community structure of the soil samples and reindeer faeces, based on the 16S rRNA gene analysis with the use of the NGS. Therefore, comparison of bacterial communities' structure in the soil samples originating from two topographically similar locations: in the vicinity of a large breeding colony of planktivorous seabird, the little auk and control sample, not influenced by seabirds, hence experiencing only a negligible ornithogenic impact. Both soil samples were collected in the north part of the Hornsund fiord at the beginning of August 2013.

Physical and chemical analysis performed on the samples allowed to detect clear differences in the ion concentration and soil conductivity, as well as in the soil dry mass and pH, between samples [2]. DNA from both tested samples was extracted with the use of FastDNA® SPIN Kit for Soil using the FastPrep® Instrument (MP Biomedicals, Santa Ana, CA), suitable for both, the clay and sand consistency of the studied samples. The obtained 3 replicates of the extracted DNA for each soil sample were send for NGS sequencing. V3-V4 hypervariable regions of bacterial 16S rRNA gene were chosen for analysis, as those gene regions have been shown to be the most suitable for Illumina sequencing.

Subsequently, sequencing results were processed and analysed with the use of the Quantitative Insights Into Microbial Ecology (Qiime) pipeline v 1.8.0 software with maintaining appropriate quality standards for the obtained sequences. Moreover, clustering of the operational taxonomic units (OTUs) was established at the 97% similarity level and OTUs were assigned to taxa using GreenGenes v13_5 as the reference. In order to describe the diversity of bacterial structures, Chao1, Shannon, and Simpson indices were estimated, which are proper for ecological analysis. Furthermore, to make the data freely available, they were deposited at ENA – the European Nucleotide Archive [2].

Further, data analysis revealed that 230 253 good quality 16S rRNA gene sequences were obtained, for samples in the vicinity of a little auk colony and 279 122 for control samples. Also, in the sample in the vicinity of a bird colony, 3600 OTUs were found, and 4844 OTUs in control sample; both of the tested samples shared 2822 OTUs. More than 99.97% of all the obtained sequences were classified at the phylum level, in both soil samples.

Taxonomy analysis indicated that both soil samples consisted of 34 phyla, and 32 were common for both samples [2]. The most abundant phyla found in the vicinity of a large breeding colony

were: *Proteobacteria*, *Acidobacteria*, *Actinobacteria* and *Chloroflexi* and these phyla jointly accounted for more than 72.5% of the total bacterial sequences. Furthermore, in control sample, the most abundant phylum was *Actinobacteria*, and then *Chloroflexi*, *Proteobacteria*, *Acidobacteria* and these phyla jointly accounted for more than 79% of the total bacterial sequences [2]. Rarefaction trends analysis indicated that the sampling of bacterial communities were close to complete, which indicates sufficient efficiency of the DNA extraction method.

The received results had shown that, despite striking differences in physicochemical soil characteristics between the two tested soil samples, there are no substantial differences at the phylum level of taxonomic ranks in the microbial communities. The similarity between these two tested bacterial populations was 78% and the chi-square test did not indicate a statistically significant difference between the tested samples. Those results led to conclusion that although vicinity of a seabird colony has an impact on the soil physicochemical features, it has no substantial impact on the high taxonomic ranks of the bacterial community structure, while a limited number of phyla may still be impacted, like *Actinobacteria* and *Proteobacteria*. Moreover, due to a high amount of guano deposition and thus a high nitrogen content, reduction in the number of the *Actinobacteria* phylum, which is able to efficiently fix nitrogen available in the air, seems natural in sample in the vicinity of the bird colony [2].

Following the soil sample analysis, analysis of the bacterial community structure of the ten reindeer faecal samples was performed with the use of a non-invasive sampling method in order to provide a comprehensive view of the faecal microbiota of the Svalbard reindeer from Hornsund [3]. Faecal samples (R1-R10) were collected in the north part of the Hornsund fiord in the late summer of 2013 [3].

Analogically to the soil analysis, DNA from the faecal samples was extracted with the use of the GeneMATRIX Faecal DNA Purification Kit (Eurx Ltd.) using the FastPrep® Instrument (MP Biomedicals, Santa Ana, CA), and was send for NGS sequencing. Similarly, as with soil samples, V3-V4 hypervariable regions of bacterial 16S rRNA gene were selected for evaluation [3].

In accordance with the soil samples, the obtained sequence data were processed and analysed with the use of the Quantitative Insights Into Microbial Ecology (Qiime) pipeline v 1.8.0 software maintaining appropriate quality standards [3]. Also Chao1, Shannon, and Simpson indices were estimated. Moreover, independent analysis with the use of BaseSpace Application 16S Metagenomics v1.0 (Illumina, INC.) was performed. As previously, to make data freely available, they were deposited at ENA – the European Nucleotide Archive [3].

The conducted analysis had shown that, taken together, 380 849 good quality 16S rRNA gene sequences were obtained, from all reindeer faecal samples, in the range of 22 997 – 54 042 for each individual. Rarefaction analysis of the obtained data revealed trends indicating that sampling of the bacterial communities was almost complete, which indicates sufficient efficiency of the DNA extraction method [3]. Additionally, in this case all of the obtained sequences were classified at the phylum level. Moreover, taxonomy-based analysis indicated that the reindeer faecal bacterial communities consisted of 14 different phyla, 30 classes, 46 orders, and 94 families and 141 genera [3]. The most abundant phyla across the population were *Firmicutes* and *Bacteroidetes,* and they jointly accounted for more than 95% of total bacterial sequences. Dominance of these phyla is commonly ascertained in the gut microbiomes of ruminants able to digest and utilize a high variety of vegetation. The remaining 5% of the population reads was comprised of *Tenericutes, Cyanobacteria,* TM7, *Actinobacteria, Proteobacteria, Verrucomicrobia, Elusimicrobia, Planctomycetes, Fibrobacteres, Spirochaetes, Chloroflexi,* and *Deferribacteres* [3].

Lack of statistically significant differences in the bacterial community structure among individuals was expected. The observed similarity for all samples was 92%, most probably due to the small size of the population, relatively small area utilized by the individuals, small travel distances and philopatry of females [3]. Considering harsh and extremely seasonal polar conditions and non-migratory lifestyle of the Svalbard reindeer, high diversity of gut bacterial community may be a natural evolutionary strategy for survival associated with seasonal food changes and enables a rapid response to varying food availability at different sites. Additionally, those results show that sampling of faeces, instead of rumen, can be a good alternative and can reduce interference with a given population's natural environment [3].

The same faeces samples were used for further investigation if the *stx* gene reservoirs are present in such remote and isolated areas, where contact of native fauna with humans and domesticated animals or other wild mammals is minimal, hence probability of transmission of bacteria carrying the *stx* genes from any human-associated or wild mammal-associated reservoir is very limited. To accomplish that, the presence of the *stx1* and *stx2* genes in the reindeer faeces and soil samples was tested. These results may provide useful information about prevalence of the *stx*-producing microbes in such isolated areas characterized by severe weather conditions. It is worth noting that presence of the *stx* genes does not necessarily conclude the presence of the Shiga toxin producing bacteria, however it can indicate a natural source of these genes [4].

Previously obtained genetic material from the reindeer faeces and soil [2–4] was used in PCR and nested PCR amplification in order to amplify the *stx1* and *stx2* genes [4]. Products were visualised by gel electrophoresis in order to confirm their size and sent for sequencing. After that, sequencing results were analysed with the use of the MEGA (version 6) software and aligned using BLASTN (NCBI/BLAST). Additionally, the Stx2 sequences were aligned using BLASTP (NCBI/BLAST) [4]. In order to make inference more complete and to support presented research with additional information about this local reindeer population, molecular sex identification targeting male-specific DNA was performed, as during sample collection sex of only some of the reindeers was established [3, 4]. In order to do so, PCR amplification was performed with the use of primers constructed for the reindeer origin samples [4].

It was shown that the *stx1* gene was present in 9 out of 10 individuals and the *stx2* gene in half of the tested samples. For each *stx2* positive sample, the *stx1* gene was also detected, and in one individual neither the *stx1* nor *stx2* gene was detected [4]. The soil sample from the area of bird colony was negative for both, the *stx1* and *stx2* genes, however, sample not influenced by seabirds was positive only for *stx1*. Furthermore, detailed comparison of *stx1* and *stx2* sequences from the reindeer population and the soil sample revealed no differences in sequences [4]. Additionally, protein alignment of the *stx2* gene showed that the obtained sequences are consistent with the c subtype [4]. Although a limited number of samples was analysed in this study, it was enough to prove the occurrence of the *stx* genes in the wild reindeer population inhabiting an area almost untouched by humans. A high frequency of occurrence of the *stx* genes in the reindeer population might be explained by small pasture area, which may affect transmission of the faecal bacterial flora between the individuals. Moreover, the *stx1* gene was detected more frequently than *stx2,* which may suggest that microbes carrying *stx1* are dominant. Additionally, the presence of the *stx1* gene in the soil sample localized away from the routine flight route of the seabirds may suggest a natural presence of the *stx* genes in the Arctic environment. Furthermore, this study had demonstrated that the DBY7 and DBY8 primers can be used in the future studies as effective molecular markers in the sex identification for the Svalbard reindeer, especially when there is no possibility to determine the sex of individual reindeer by observation [4].

The original reports which have become the basis for my PhD thesis, had broadened the knowledge on the biodiversity of microorganisms connected with the Arctic ornithogenic soil ecosystem, with the use of non-invasive testing. The main implication of this study is that by using only easily accessible samples, like soil or faeces, and molecular biology methods, it is possible to obtain a high quality data about micro- and macro- organisms, with only minimal

interference into local environment. Moreover, despite the fact that this study is limited to the Hornsund area, it demonstrates the occurrence of the *stx* genes and may suggest a natural and very broad distribution of bacteria bearing these genes in the Arctic environment. All of the findings included in my dissertation show that molecular techniques may be potentially valuable for a fast and easy monitoring of the rapid changes taking place in the Arctic ecosystem, especially since this region is characterised by a much faster temperature increase than the other regions of the world.

References

1. Zielińska S, Kidawa D, Stempniewicz L, Łoś M, Łoś JM (2017) Environmental DNA as a valuable and unique source of information about ecological networks in Arctic terrestrial ecosystems. Environmental Reviews er-2016-0060. doi: 10.1139/er-2016-0060

2. Zielińska S, Kidawa D, Stempniewicz L, Łoś M, Łoś JM (2016) The Arctic Soil Bacterial Communities in the Vicinity of a Little Auk Colony. Frontiers in Microbiology 7:1298. doi: 10.3389/fmicb.2016.01298

3. Zielińska S, Kidawa D, Stempniewicz L, Łoś M, Łoś JM (2016) New Insights into the Microbiota of the Svalbard Reindeer Rangifer tarandus platyrhynchus. Frontiers in Microbiology 7:1–9. doi: 10.3389/fmicb.2016.00170

4. Zielińska S, Kidawa D, Stempniewicz L, Łoś M, Łoś JM (2016) DNA extracted from faeces as a source of information about endemic reindeer from the High Arctic: detection of Shiga toxin genes and the analysis of reindeer male-specific DNA. Polar Biology 1–8. doi: 10.1007/s00300-016-1990-2