Introduction

From the human perspective the value of soil is mainly defined by the role it plays in food production processes. However soil also has several other functions which are crucial to the life of humans, animals and plants, although its contribution to the functioning of ecosystems is often non-direct and remains unseen (Brady and Weil, 1999). Soil allows to limit the results of climate changes and acts as a regulator of water drainage, it decreases the temperature in urban areas during hot periods and, at the same time, increases the humidity of air, it also helps to preserve the biological diversity in urban areas, allowing the survival of several species of plants, animals and microorganisms (Lal, 2004; Seneviratne, 2010). One of the fundamental differences between urban soil and soil originating from rural areas is the fact that the first is subjected to strong changes due to anthropogenic activity (de Kimpe and Morel, 2000). The first negative phenomenon associated with the sealing of soil is increasing its density and agglomeration, which results in issues in terms e.g. of proper water intake. The second notable threat to soils in urban areas is associated with their local contamination. Degradation of soils in UE-28 proceeds at a rate of approx. 2.46 t/ha/year (Panagos et al., 2015), and some researchers claim that the rate may be as high as 2.76 t/ha/year (Bosco et al., 2015). In Europe, an average of 0.3–1.4 tons of soil per ha is created annually (Verheijen et al., 2009). Due to this reason several scientists treat soil as a non-renewable resource. Aside from heavy metals and liquid ionic, hydrocarbons are one of the most dangerous types of contaminants. Pollution with hydrocarbons often occurs during accidents and leakage of diesel oil during its transport or storage (Piotrowska-Cyplik and Czarnecki, 2003; Sydow et al., 2015; Ławniczak et al., 2016; Sydow et al., 2016). From human perspective, the environmental contamination
with petroleum hydrocarbons is a notable environmental and health problem. These compounds, due to their hydrophobic nature, may penetrate through cell membranes and damage the cells, which are not adopted to their presence. Such compounds are often toxic, mutagenic and carcinogenic (Marcic et al., 2012).

A detailed analysis of physicochemical parameters, physical phenomena and determination of the contamination rate allows to select a proper remediation technique. The UE Directive.

2004/35/CE indicates that natural bioremediation methods are a solution with limited human intervention in the processes of returning the environment to its primary state. The use of bioremediation aims to clean-up the polluted areas, it does not require the application of complex apparatuses and it can be used in situ. It employs biological systems to decrease the concentrations and (or) completely remove several chemical compounds, such as petroleum and its processing products, aliphatic and aromatic hydrocarbons (with the exception of polycyclic aromatic hydrocarbons and polychlorinated biphenyls), organic solvents and others (Lisiecki et al., 2014). The increasing use of bioremediation for the biodegradation of hydrocarbons results from its ecological and economic merits, as it does not interfere with the natural clean-up processes and soil subjected to bioremediation may be reused (Marquez-Rocha et al., 2001). It is general accepted that in order to achieve complete degradation of a mixture of petroleum hydrocarbons, the contribution of numerous microorganisms which cooperate with each other is necessary. Several authors emphasize the high potential of soil for conducting natural attenuation processes during a long time period, which is of particular importance, since this approach guarantees that its structure remains intact and limits the costs of bioremediation (Szczepaniak et al., 2016). However, in order to allow the natural attenuation processes to proceed, the bacterial microflora must be sufficiently rich in species capable of degrading petroleum hydrocarbons. Evaluation of the soils biodegradation potential has a notable practical implication, as it allows to indicate the need to introduce external microbial cultures with a high biodegradation potential, to establish the range of such operations and to monitor the biodegradation process not only based on the changes of pollutant concentrations, but also by determining the abundance of the introduced bacterial consortia and evaluation of their functioning by measuring the expression of genes crucial for the biodegradation process (Cébron et al., 2008).

The aim of this study was to compare the composition of the bacterial metabiome and the presence of selected genes encoding crucial enzymes responsible for the decomposition of aromatic hydrocarbons in soils originating from urban areas and areas contaminated with hydrocarbons in relation to agricultural soil and soil originating from a protected wild-life area (forest soil). In order to fully identify all the microorganisms present in the soils and evaluate their genetic pool, the sequencing of the soil metabiome using Next Generation Sequencing with the use of MiSeq Illumina was assumed. Changes in the number of enzymes initiating the decomposition of polycyclic aromatic hydrocarbons (dioxygenases (PAHRHDaGN, PAHRHDaGP)), monoaromatic hydrocarbons (catechol 2,3-dioxygenase (xylE, Cat 2,3) and naphtalene 1,2-dioxygenase (ndoB) was analysed in the analysed soil samples with respect to the gene encoding 16S rRNA using the RT-PCR method.

**Experimental**

**Materials and Methods**

The origin of soil samples. The soil used in the studies originated from the Wielkopolska region in Poland and was collected from four sites with different usage. Soil A originated from a forest in the area of the landscape park (Poland) (N 52.534032; E 17.047875); Soil B originated from an agricultural field used to grow wheat (N 52.551539, E 16.99974); Soil C originated from the roadside vicinity of a road near Poznan (Poland) (N 52.524980, E 16.976538); Soil D originated from a petrol oil station (N 52.560036, E 16.999887). The surface layer of soil from each area was collected from 10 different spots, each sample contained 100 g of soil. The samples were collected from a depth of 10–20 cm, then they were combined and mixed. The soil near the road was collected in the direct vicinity of the drive way. In the case of the remaining soils a regular method of sampling was applied, since a uniform distribution of contaminants was assumed. The total mass of the combined soil obtained upon combination of all individual samples was 1 kg.

Identification of microorganisms

DNA extraction. Total DNA was extracted from soil sample using Genomic Mini AX Soil kit (A&A Biotechnology) according to manufacturer’s instruction. The extracted DNA was quantified using Quant-iT HS ds.-DNA assay kit (Invitrogen) on Qubit2 fluorometer (Lawniczak et al., 2016).

PCR amplification. Region IV of bacterial 16S RNA gene was amplified using universal primers 515F and 806R: containing reverse complement of 3’ Illumina adapter, golay barcode, reverse primer pad, reverse primer linker and reverse primer (Table 1). Products
were purified in Clean-Up columns (A&A Biotechnology) according to manufacturer’s protocol. The libraries were constructed from amplicons using NEBNext® DNA Library Prep Master Mix Set for Illumina (New England Biolabs UK). Then the libraries were pooled at equimolar concentration. Sequencing was conducted on an Illumina MiSeq (Illumina, USA) using paired-end (2 × 250) MiSeq Reagent Kits v2 (Illumina, USA). Sequencing primers were based on Caporaso et al. 2012 (Table I). The sequencing reaction was performed with Illumina instrument and MiSeq Reagent Kit v2 (2 × 250bp) (Lawniczak et al., 2016).  

Bioinformatic analysis. The sequencing data was processed using CLC Genomic Workbench 8.5 and CLC Microbial Genomics Module 1.2. (Qiagen, USA). The total number of reads ranged from 438 216 to 461123. After sequencing, the reads were demultiplexed to the probes and the overlapping paired-end reads were merged (70% of total reads) and trimmed to yield fragments of 289 nt. Just fragments which passed the merging were retained for downstream processing. Chimeric reads (from 27121 to 29785) were filtered and remaining sequences were assigned to operational taxonomic units (OTUs). Number of reads which passed merging and trimming ranged from 136576 to 154294. Reads were clustered against the SILVA v119 99% 16S rRNA gene database (Quast et al. 2013).  

RT-PCR. Genes level was analysed using a Power SYBR Green PCR Master Mix (Life Technologies) on ABI 7500 SDS (Applied Biosystems). Primers used for real-time PCR are listed in Table I. Total bacterial RNA was quantitated by real-time PCR amplification of fragment of bacterial 16S ribosomal RNA with universal bacterial primers and TaqMan MGB probe using TaqMan Universal Master Mix II (Life Technologies) on ABI 7500 SDS (Applied Biosystems). Sequences of primers and probe used are listed in Table I. All analysis was done in triplicates. In order to compare the gene expression in each sample, the mean expression index was calculated according to formula: C_{T} target/C_{T} 16S using data from 3 analyses. This parameter reflects the expression level of a specific gene compared to the expression level of the universal gene (16S RNA) in the whole metabiome (Szczepaniak et al., 2016).  

### Results and Discussion  

Analysis of the soil metabiome is a challenging scientific problem due to the properties of soil, methods and efficiency of isolation of genetic material and the use of different molecular biology methods in order to identify the soil microbial species (Kozdrój, 2013; Szczepaniak et al., 2015). The use of metagenomic 16S
analysis method based on the MiSeq platform allows to notably expand the knowledge regarding the composition of the soil microbiome. This tool is particularly useful during comparison studies of genomes originating from different environments (Szulc et al., 2014). In order to identify the bacterial consortium (M), a metagenomes analysis of the gene encoding 16S rRNA was conducted on the basis of V4 hypervariable region of the 16S rRNA gene (Fig. 1) (Caparoso et al., 2012).

Metagenomic 16S analysis of the soil metabiome indicated that soil originating from the forest (A) and agricultural soil (B) were the most diverse in terms of microbial species. In the bacterial metabiome the identified species belonged to 64 classes, including the following dominant classes: Alphaproteobacteria (22.83%) and Gammaproteobacteria (22.88%) as well as Clostridia (11.11%), Bacilli (8.40), Betaproteobacteria (7.55%) and Planctomycetia (5.21%). The growth of crops and application of agricultural processes resulted in the dominance of Gammaproteobacteria in the agricultural soil (soil B), which accounted for 23.06% of the total population. The remaining dominant classes included Alphaproteobacteria (21.92%), Actinobacteria (11.94%) and Clostridia (8.05%). The ratio of the remaining groups of microorganisms did not exceed 5%.

As a result of applying different agricultural processes, including different fertilization methods and use of plant protection agents, the metabiome of agricultural soils was characterized by a lower population variety, which was reflected by a lower value of alpha-diversity indices. In the studies carried out by Newman et al. (2016) regarding the influence of glyphosate on the metabiome of rhizosphere soil in soybean and corn, it was established that three phyla dominated in control soil: Proteobacteria, Acidobacteria, and Actinobacteria. The use of glyphosate significantly increased the ratio of Proteobacteria (particularly Gammaproteobacteria) in the soil metabiome, whereas the ratio of Acidobacteria notably decreased. Furthermore, in the study of Zhou et al. (2015), which focused on the influence of long-term fertilization on the changes of the soil metabiome, a decrease of the total number of 16S rRNA gene and decrease of bacterial diversity was observed in the case of soils treated with mineral fertilizers. Dominant phyla, which increased their abundance as a result of fertilization, included: Proteobacteria, Acidobacteria, Actinobacteria, Actinobacteria and Proteobacteria. On the other hand, the ratio of Acidobacteria and Nitrospirae in the population was notably decreased. It should therefore be assumed that other factors may also lead to changes in the composition of soil bacteria populations. The presence of hydrocarbon contaminants in soil may be one of such factors.

Petroleum hydrocarbons are compounds which undergo decomposition in soil due to activity of several groups of microorganisms. Several different microbial species participate in the biodegradation of hydrocarbons, ranging from strictly aerobic to strict anaerobic bacteria. Several Gram-positive (Rhodococcus or Bacillus) as well as Gram-negative (Alcaligenes, Acinetobacter, Pseudomonas) species are also characterized by relatively broad substrate spectrum (Szczepaniak et al., 2015).

The progress of hydrocarbon biodegradation processes carried out in situ depends on the conditions in a given area and the present microflora. During the early stages n-alkanes are biodegraded by the bacteria and when only persistent compounds are left (branched alkanes, mono- and polycyclic aromatic hydrocarbons), the biodegradation process is conducted by bacteria.
characterized by a lower growth rate and unique metabolic abilities (Mrozik and Piotrowska-Seget, 2010).

In the soil samples collected from the roadside (soil C) and soil collected from a former fuel station (soil D), bacteria belonging to the Gammaproteobacteria class were clearly the dominant group. Their ratio in the soil samples collected from the roadside (soil C) was at 28.4%, whereas in the case of soil collected from a former fuel station (soil D) the value reached 36.63%. The ratio of Alphaproteobacteria was similar to the remaining soils and ranged from 17.86 to 24.78%. A notable ratio of bacteria belonging to the Sphingobacteria (11–12%) and Actinobacteria (8.0–8.4%) class was also observed. The ratio of the remaining identified classes of microorganisms did not exceed 6%.

The Principal Component Analysis (PCA) was used in order to evaluate the changes in the bacterial soil metabiome. Figure 2 shows how the analysed bacterial metabiomes were grouped. Two first main components are decisive and describe the variability of the primary data in 84.0%. The first main component carried approx. 67.23% of the data regarding microbial populations contained in the input variables. It includes the following positively correlated variables: Gammaproteobacteria, Alphaproteobacteria, Betaproteobacteria and Bacilli; as well as the negatively correlated: Clostridia and Flavobacterium. The second main component is mainly responsible for the presence of Actinobacteria, Planctomycelia and Others, and described the variability of the analysed data in 20.84%. The high variability of the soil metabiome depending on its place of origin was presented on Fig. 2. The presence of anthropogenic contaminants in soil (soil C and D) caused notable changes in the soil metabiome, which are visible in the lower-left part of the figure. The remaining metabiomes of soils A and B are present in separate quarters of the figure.

The determined alpha diversity (Table II) of soil originating from the forest area (soil A) and agricultural soil (soil B) were significantly higher compared to soil originating from urban and contaminated areas (soil C and D) (p < 0.05).

A rational approach to the bioremediation process should limit the activities which may change the

<table>
<thead>
<tr>
<th></th>
<th>Soil A</th>
<th>Soil B</th>
<th>Soil C</th>
<th>Soil D</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTU observed</td>
<td>1893 ± 51</td>
<td>1944 ± 71</td>
<td>1426 ± 29</td>
<td>1312 ± 44</td>
</tr>
<tr>
<td>Shannon's index</td>
<td>5.8 ± 0.2</td>
<td>5.7 ± 0.3</td>
<td>4.8 ± 0.2</td>
<td>4.1 ± 0.1</td>
</tr>
<tr>
<td>Chao 1 bias-corrected</td>
<td>1358 ± 187</td>
<td>1398 ± 154</td>
<td>1089 ± 122</td>
<td>1012 ± 115</td>
</tr>
<tr>
<td>Phylogenetic diversity</td>
<td>4.98 ± 0.31</td>
<td>4.86 ± 0.22</td>
<td>3.95 ± 0.34</td>
<td>3.68 ± 0.21</td>
</tr>
</tbody>
</table>
composition and structure of soil and therefore disrupt the biological equilibrium by introducing new groups of microorganisms. To date, the studies regarding the bioremediation of natural environments rarely include analyses of the environmental genetic pool, which would contribute to valuable data regarding its natural attenuation potential (Afzal et al., 2011). Selected microbial communities characterized by a high hydrocarbon-degradation potential are often prepared for the purpose of bioremediation technologies (Cerqueira et al., 2011). However, it is known that introduction of external organisms (consortia of isolated microorganisms) is not always possible – e.g. in the case of protected areas, which become contaminated. The same applies to physical removal of polluted soil, since the removal of the contaminated layer may also eliminate species of insects, molluscs and representatives of other organisms, which are under protection.

Furthermore, the proper functioning of a microbial consortium is based on its composition as well as the combinations of specific functions (defined by genes) which are provided by specific microbial species (Panicker et al., 2010). Selection of specific traits (gene sequences) allows to monitor the changes in the structure of the consortium and it the general genetic poll of the remediated environment. Such studies may be the foundation for determining the natural attenuation potential of a given environment and to decide whether bioaugmentation or use of additional agricultural agents is necessary (Szczepaniak et al., 2016). Usually, prokaryotic organisms oxidise PAHs using dioxygenases by introducing two oxygen atoms into the substrate and formation of cis-hydrodiols, which are then transformed into dihydroxyl compounds. This reaction is crucial for initiating the conversion of aromatic hydrocarbons into environmentally sage forms (Cébron et al., 2008; Fuentes et al., 2014).

Prior to the determination of the presence and abundance of genes responsible for the biodegradation of aromatic hydrocarbons, an indirect method was used based on the measurement of the number of cycles for the analysed gene (Ct gene) in comparison to an internal standard – the number of cycles encoding the 16S rDNA subunit (Ct16S). The employed method allowed to determine the relative changes in the expression of genes compared to the number of bacteria present in soil. The measured Ct16S value for the 16S RNA gene, which was used as an internal standard, was in the range of 6.2 to 7.3 for a given sample. (see Fig. 3). The presence of the gene encoding dioxygenases in Gram-negative bacteria (PAH-RHDα GN) was highest in soil originating from the fuel station (soil D) and from soil originating from the roadside (soil C). The relative number of these genes reached 0.54 ± 0.21 and 0.26 ± 0.14, accordingly, and these values were significantly different (p < 0.05). Whereas in soil originating from the forest area (soil A) and agricultural soil (soil B) the content of these genes in the soil metabiome gene pool was lowest and ranged from 0.08 to 0.14. The differences between these values were not statistically significant (p > 0.05).

A different result was obtained in the case of studies regarding the presence of the gene encoding the dioxygenase in Gram-positive bacteria (PAH-RHDα GP), as its number in the microbial gene pool was significantly lower in samples collected from the roadside (soil C) and soil collected from the fuel station (soil D) and ranged from 0.15 to 0.19. On the other hand, in soil originating from the forest area and agricultural soil its abundance did not exceed 0.08.
The relative number of genes encoding enzymes crucial for the degradation of aromatic hydrocarbons (xyLE, Cat 2,3 and ndoB) was highest in case of soil originating from anthropogenic areas (soils C and D) and ranged from 0.3 to 0.45. In case of agricultural soil and soil originating from the forest area (soils A and B) these values did not exceed 0.10. The observations of several authors confirm that the origin of soils and their contact with contaminants results in the increase of genes responsible for the biodegradation of hydrocarbons in the gene pools. The obtained results confirm the observations of other authors regarding the uneven distribution of genes encoding enzymes decomposing hydrocarbons in different terrestrial environments (Cébron et al., 2008). This distribution is mainly based on the area of origin. In soils in which microorganisms had contact with contaminants (soils C and D) a higher ratio of genes associated with biodegradation processes was observed in comparison to non-contaminated soil. However non-contaminated soils were characterized by the presence of microflora capable of biodegradation processes and therefore the level of genes associated with biodegradation processes may increase after contact with contaminants, especially after a prolonged period, which confirms the high potential of soils to conduct natural attenuation during a long term period.

Conclusions

The selection of specific genes, which are relevant to the decomposition of hydrocarbons, allows to evaluate the biodegradation potential and monitor the changes in the microbial community structure and the overall gene pool in the remediated environment. Such monitoring is particularly justified in the case of freshly contaminated areas, providing solid foundation for additional, controlled bioaugmentation. In the case of permanent or long-term contaminations the detailed analysis of gene pool of the soil metabiome allows to determine the natural attenuation potential and decide whether bioaugmentation or application of agricultural agents is necessary. Such actions would facilitate the remediation of areas subjected to long-term anthropogenic activity. A precise diagnostic carried out in order to determine the genetic deficiencies or elimination of some consortium members may indicate that aside from the composition of a consortium, the combination of specific functions (defined by genes carried by those organisms) is a crucial factor.

Acknowledgements

Research work funded by National Science Centre in Poland in the years 2014–2017 as a research project Opus no 2013/11/B/ NZ9/01908.

Literature


