Polish Journal of Environmental Studies Vol. 14, No 5 (2005), 551-557

Original Research

Specific Genomic Fingerprints of *Escherichia coli* Strains with Repetitive Sequences and PCR as an Effective Tool For Monitoring Freshwater Environments

K. Baldy-Chudzik*, M. Stosik

Department of Microbiology and Genetics, Institute of Biotechnology and Environmental Science, University of Zielona Góra, ul. Monte Cassino 21B, 65-561 Zielona Góra, Poland

Received: December 29, 2003
Accepted: January 14, 2005

Abstract

The Rep-PCR fingerprinting method has been applied to identify genomic diversity of 252 *E. coli* strains derived in the area of a flowing-water basin. The received results show that applying UPGMA and Nearest Neighbour-Joining clustering methods to statistical analysis of rep-PCR fingerprints has made it possible to discriminate and group the strains, revealing a characteristic structure of *E. coli* population for particular stands of sample drawing. The proposed procedure of the analysis may be useful for routinely monitoring water quality.

Keywords: freshwater environment, *E. coli* diversity, rep-PCR fingerprinting

Introduction

Water, both stagnant and flowing, is routinely monitored for public health. Faecal contamination of water is traditionally proved with the occurrence of *Escherichia coli* strains [1]. Although the identified *E. coli* are basically nonpathogenic (commensal), their occurrence warns against a simultaneous occurrence of the pathogenic ones. The number of data proving a considerable genomic diversity among *E. coli* strains increases steadily, and the clonal nature of the species has been definitely confirmed [2]. Hence, there is a need to supplement the quantity research of *E. coli* strains of water environment with quality research, demonstrating the genomic diversity of the strains and providing information important in terms of ecology and epidemiology. Nowadays, there are a few kinds of DNA analyses in use which identify and characterize the interstrain diversity of bacteria, including *E. coli*. They are: pulsed–field electrophoresis (PFGE) [3], ribotyping [4, 5], ribosomal DNA heterogeneity [6], and repetitive extragenic palindromic-PCR (rep-PCR) [7, 8]. The analyses, depending on the kind, apply digestion with restriction enzyme or PCR amplification with deliberately selected reaction primers. The reaction products are analyzed using electrophoresis. It results in obtaining band patterns which are specific and characteristic for particular bacterial genomes. The band patterns are then compared with each other, enabling the analysis of homology/similarity degree of the identified strains [9]. Laboratories monitoring water environment all over the world have been more and more frequently using rep-PCR fingerprinting methods in recent years [7]. This is due to the fact that the method is cheaper and faster than PFGE, which is of fundamental importance for water environment screening. Moreover, its considerable discrimination power makes it better than ribotyp-

*Corresponding author; e-mail: K.Baldy-Chudzik@ibos.uz.zgora.pl

ing-based methods for determination of DNA polymorphism within a species. Primers of rep-PCR fingerprinting reaction are homologic to the repetitive sequences, belonging to the families of the following types: REP, ERIC, BOX [10,11]. The characteristic feature of the sequences is the fact that both the number of copies per genome and their distribution in a genome are unique and characteristic for an individual strain/clone [12,13]. Hence, it is assumed that the detection of rep-PCR fingerprinting band patterns of 100% homology/similarity for more than one isolates allows asserting, with a considerable possibility, that the same clone/strain is dealt with.

The presented research has included the analysis of *E. coli* strains identified in the area of the Wojnowskie East and West Lakes. The rep-PCR fingerprinting method with the use of BOX-type primer has been applied to illustrate the genomic diversity of the examined strains. Next, the obtained fingerprinting patterns have been analyzed to determine homolgy/similarity relation between *E. coli* identified in the particular stands of water sampling.

Materials and Methods

Object of the Study

The Wojnowskie (East and West) Lakes are situated in the west of Poland, in Lubuskie Province. Wojnowskie East is situated in the drainage-basin of the following rivers: Obra Leniwa, Obrzyca, and Odra.

The Obra Leniwa River flows into the lake from the north and flows out of the lake southwards into the Wojnowskie West Lake.

So, the Obra Leniwa River flows into Wojnowskie West Lake from the north (out of the Wojnowskie East) and flows out southwards, as the Obrzyca River. The Wojnowskie Lakes, situated along the rivers, are of an open, flowing type, with an eight-day period of water interchange. The lakes are situated in the first zone of the rigorous protection area of a water intake for the town of Zielona Góra. On the basis of the faecal-type coli indicator, the waters of the lakes are rated as the first (the Wojnowskie West Lake) and the second (the Wojnowskie East Lake) purity classes [14].

Water Sampling

The water sampling stands have been situated in five places (Fig. 1).

Stand I – north of Wojnowskie East Lake, where the Obra Leniwa River flows into the reservoir. Stand II – in the north-central part of Wojnowskie West Lake. Stand III – in the south-central part of the lake, not far from the place where the Obra Leniwa River flows into the reservoir (from Wojnowskie East Lake).

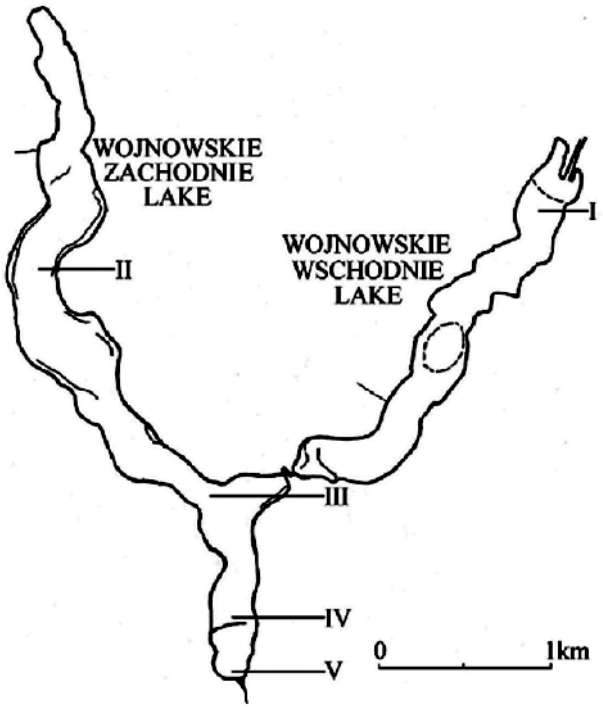

Fig. 1. Location of stands of water sampling in the area of the Wojnowskie Lakes.

Stand IV – situated in the southern part of Wojnowskie West Lake. Stand V – on the periphery of the southern part of the lake, where the Obrzyca River flows out of the lake.

In each stand water samples were drawn at two different depth-zones: a subsurface zone, i. e. 50 cm below the water surface, and a bottom zone, i. e. 50 cm above the bottom of the reservoir. Stand V was the only exception, where the samples were drawn only in the subsurface zone, where the lake is 2m deep. The samples were drawn within a period of 9 months, from March to November.

Isolation of *Escherichia coli*

E. coli bacteria have been isolated from water samples according to Polish standards PN-77/C-04615/07. The determination of the faecal coliform has been carried out with the use of the fermentative test-tube method [15]. 252 *E. coli* strains have been identified altogether. The numbers of the identified strains in particular stands and zones are presented in Table 1.

DNA Isolation and rep-PCR (BOX) Fingerprinting Condition

18-hour *E. coli* cultures, growing in Luria-Bertani bullion to the optical density $OD_{\lambda 600nm}$ = 0.6-0.9 have composed the initial material for DNA isolation. The genom-

Table 1. Number of *E. coli* strains identified in stands and zones of water sampling.

Stand	Number of *E.coli* in the sampling zone				Number of strains in both zones of a stand	
	surface		bottom			
I	16	6.35%	26	10.32%	42	16.67%
II	20	7.94%	22	8.73%	42	16.67%
III	27	10.71%	20	7.94%	47	18.64%
IV	43	17.06%	45	17.86%	88	34.92%
V	33	13.10%	0	0.0%	33	13.10%
Total	139	55.16%	113	44.84%	252	100.00%

ic DNA has been isolated with the use of Wizard Genomic DNA Purification kit (Promega), according to the procedure given by the manufacturer. The PCR reaction has been carried out basically in a way suggested by Rademaker and de Bruijn [16] with some modifications. Briefly, rep-PCR fingerprints have been obtained with the use of the following primer: BOX A1R (5'-CTA CGG CAA GGC GAC GCT GAC G –3'). The reaction mixture of the final volume 25µl consisted of: 19µl of 10x reaction buffer supplemented with 15mM $MgCl_2$ (Finnzymes) and 2 mg/ml BSA albumine (Sigma); 1µl 10mM mix nukleotydes (dNTP, Promega); 1µl 30 pmol primer BOX A1R; 1µl polymerase DNA 2.5 U/µl (Finnzymes), and 3µl of genomic DNA. The amplification reaction has been carried out in PTC 200 thermalcycler (MJ Reserch Inc. Watertowa) applying the following amplification protocol: the initial denaturation 95°C, 2min, and next 30 cycles consisting of a denaturation step 94°C, 30s; annealing 50°C, 30s; extension 72°C 1min as well as a final extension step 72°C for 4min and storage at 4°C.

The reaction products have been analyzed electrophoretically in 1.2% agarose gel (Serva), dyed with ethidium bromide (Sigma), some photographs have been taken. To prove the credibility of the repetitiveness of the obtained fingerprinting patterns, the PCR reaction was repeated three times for each strain.

Computer-Assisted rep-PCR DNA Fingerprint Analysis

Gel images were analyzed using BIO-GENE analysis software (Version 99.03, Vilber-Lourmat). The positions of the bands on each lane and each gel were normalized using the 1kb DNA ladder (Fermentas) as an external reference standard. Fingerprint images were added to a database and compared in a statistical analysis. The similarity between the strains was determined on the basis of the Jaccard similarity coefficient with a 5-percent confidence interval. A database has been built in

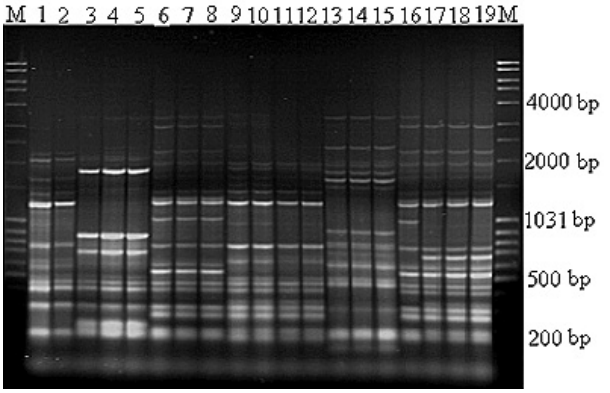

Fig. 2. Representative picture of rep-PCR (BOX) fingerprints of *E. coli* strains of stand IV.
M – 1kb DNA Ladder, size marker (Fermentas); Lane1-2 – strains of similarity group A; 3-5 – strains of similarity group B; 6-8 similarity group D; 9 and 10 – similarity group E; 13-15 –group G; 16- group I; 17-19 –group J.

the form of a matrix expressing (in percentage) all the homologies/similarities of rep-PCR fingerprints resulting from the comparison of the electrophoretical paths with each other. The dendrogram was constructed on the basis of the averaged similarity of the matrix with the use of the algorithm of the Unweighted Pair-Group Method (UPGMA). The data of the matrixes obtained from BioGene sofware have been further used in Statgraphics (version Plus 5). Categorical crosstabulation analysis (row by column frequency test) has been used to show the frequency of occurrence of strains of particular similarity groups in particular stands of water sampling. The nearest neighbour-joining clustering method has been used to show relations between similar groups of the dendrogram and the stand of water sampling, on one hand, and the date of water sampling and the zone on the other hand.

Results and Discussion

The genomic diversity analysis of 252 strains of *E. coli* has been carried out with the use of the rep-PCR fingerprinting method with BOX-type primer.

Complex patterns of fingerprints have been obtained for all the examined strains. Generally, the electrophoretic analysis of the PCR reaction products has revealed that the number of bands in particular electrophoretic paths ranged from 8-18. The sizes of the PCR products ranged from slightly less than 250 bp to about 4,000 bp. Products ranging from 350-400bp have commonly occurred. The greatest diversity of DNA fragments have been observed within the range of 400bp–1,500bp (Fig. 2).

UPGMA statistical analysis has been used to determine mutual homology/similarity relations among the strains. The received dendrogram has grouped the 252

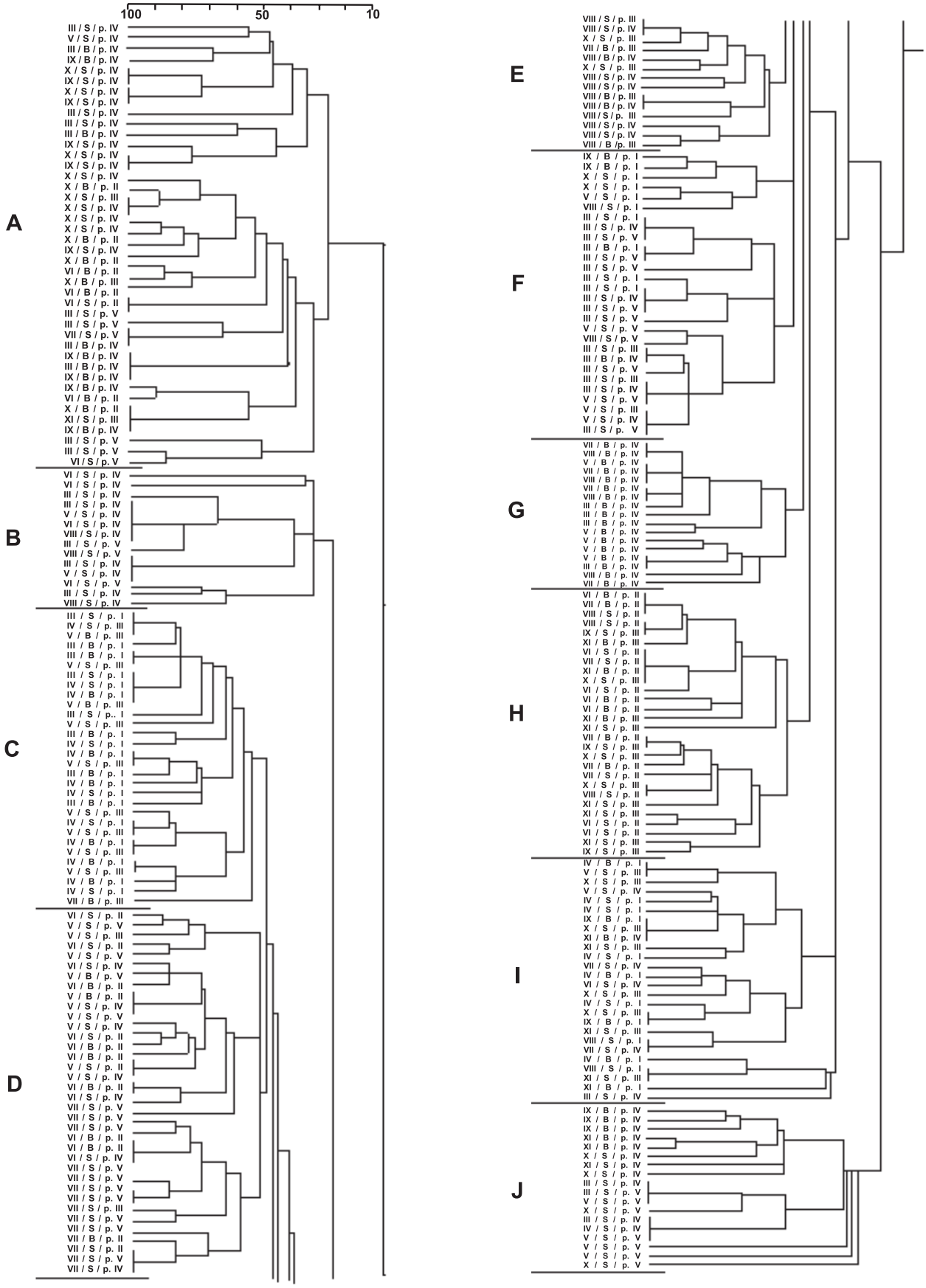

Fig. 3. Homology/similarity dendrogram of 252 *E. coli* strains.
A-J similarity group denotations; X/Y/Z stand for month of water sampling (roman number) /sampling zone marked S (subsurface) or B (bottom) /stand of water sampling (roman number), respectively.

strains of *E. coli* into ten similarity groups (Fig. 3). Each group has demonstrated a characteristic number of strains as well as separate inter- and intra-group similarity relations. Groups C, E, D and H reveal the highest interstrain similarity of features (100->50%). Groups B (100-35.4%), I (100-33%), A (100-27.7%) and J (100-23.3%) reveal the highest heterogeneity, and so the span of intergroup similarity. The homology relations between the groups are diverse – from 53.5% to 7.7%. Groups C, D and E remain in the highest mutual relations (>50%). Groups F, G and H reveal 50-40% similarity in decreasing order. Groups with a clear inner heterogeneity such as A, B, I and J remain also in low intergroup similarity relations (7.7%, 24%, 33%, 16.2%, respectively, Fig. 3.) The detailed analysis of similarity relations with respect to the identification of the source of the strains was not possible because of a nonpoint character of the origin of the strains. Categorical crosstabulation analysis has been applied to show the relation between the stand of water sampling, where the strains have been identified, and their affiliation to the presented similarity groups. The flowing character of the lakes makes the relations between the affiliation of the strains to a similarity group and the suitable stands of water sampling differ considerably (Table 2). Strains that have been classified to a similarity group appearing for the first time in a given stand along the way of the water flowing through the basin, have been assumed as characteristic/specific. Strains identified in stand I belong to three similarity groups: C, F and I. Strains derived from stand II have also distributed into three, but different similarity groups: A, D and H.

Stand III is represented by seven similarity groups. Strains occurring in this stand, apart from the ones that are characteristic exclusively for stand III (group E), affiliate with groups specific both for stand I (groups C, F and I) as well as stand II (groups A, D and H). *E. coli* identified in stand IV distribute to as many as eight similarity groups. The strains discovered here affiliate with groups characteristic for stands situated higher: I, III (similarity groups F, I) and stands: II, III (similarity groups A, D and E) as well as three other similarity groups. Since groups B, G and J have not occurred in stands situated higher, but have occurred in stand IV, it has been assumed that they characterize specific strains of the stand. Strains of stand V have been classified to similarity groups B and J, also characteristic for strains of stand IV as well as groups D and F, which comprise strains from more than a single stand out of the ones situated higher, such as stands II and IV (group D) and stands I and IV (group F). Strains of stand V, situated in the place where the Obrzyca River flows out of the Wojnowskie Lakes, reveal lower diversity when compared to *E. coli* of stand IV. In order to confirm the fact that the received picture of distribution of the strains is not the effect of random events but a certain regularity, the analysis has been improved by including the date of water sampling. Although the water samples were not always drawn in the same period of time (month) in all the stands, the date of water sampling has been assumed to be the factor proving the repetitiveness of the observed relations.

The nearest neighbour-joining clustering method has been used to illustrate relations considering the dates and the zones of water sampling (Fig. 4). The plot presents the distribution of strains with regard to the stand of water sampling, the similarity group and the months of sampling. The picture presents 2-5 clusters composed of strains from 2-5 months of drawing for particular similarity groups in subsequent stands of water sampling. Despite the fact that the regularity of water sampling was not maintained, the particular clusters demonstrate distinct regularity of occurrence of strains of particular similarity groups in particular stands. It should be emphasised that each of the similarity group has revealed a considerable percentage of strains of 100% – pattern simi-

Table 2. Frequency of occurrence of *E. coli* strains of particular stands in relation to homology/similarity groups of dendrogram.

Stand	Dendrogram similarity groups:										Number of strains in a stand
	A	B	C	D	E	F	G	H	I	J	
I	0 0%	0 0%	20 7.94%	0 0%	0 0%	10 3.97%	0 0%	0 0%	12 4.76%	0 0%	42 16.67%
II	13 5.16%	0 0%	0 0%	13 5.16%	0 0%	0 0%	0 0%	16 6.35%	0 0%	0 0%	42 16.67%
III	5 1.98%	0 0%	10 3.97%	2 0.79%	7 2.78%	3 1.19%	0 0%	12 4.76%	8 3.17%	0 0%	47 18.64%
IV	23 9.13%	11 4.37%	0 0%	9 3.57%	7 2.78%	5 1.98%	16 6.35%	0 0%	6 2.38%	11 4.37%	88 34.92%
V	0 0%	4 1.59%	0 0%	12 4.76%	0 0%	10 3.97%	0 0%	0 0%	0 0%	7 2.78%	33 13.10%
Total in a group	41 16.27%	15 5.95%	30 11.90%	36 14.29%	14 5.56%	28 11.11%	16 6.35%	28 11.11%	26 10.325	18 7.14%	252 100.00%

row-by-column frequency test

Similarity groups represented by strains of particular stands have been typed on a grey background.

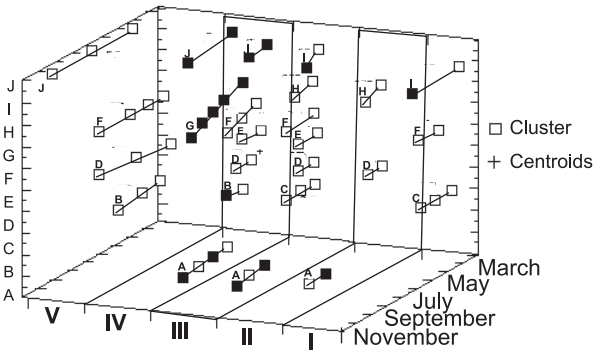

Fig. 4. Plot Nearest Neighbour – Joining Method – relations between: a stand (X axis, marked as I – IV) – a similarity group (Y axis, A-J) – a month of water sampling (Z axis).
Inside the plot, particular clusters have been marked with a letter character (a similarity group). Clusters of the bottom zone have been typed on a black background.

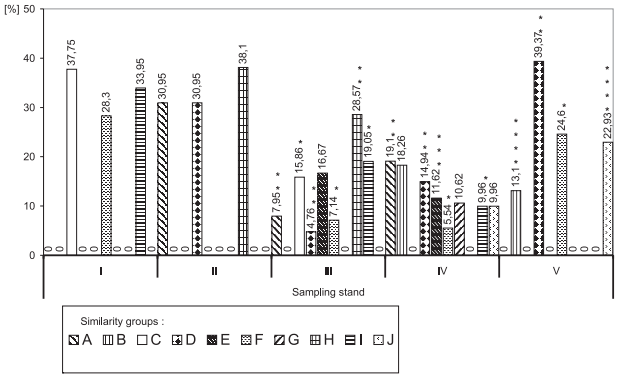

Fig. 5. The structure of *E. coli* population in particular stands of water sampling.
Roman number under the horizontal axis – a stand of water sampling; Asterisk above the bar – a stand of a similarity group's origin (*, **, ***, **** – stand I, II, III, IV respectively).

larity, although they were identified in different stands and in different time periods (Fig. 3). The full homology of patterns among such strains proves their mobility from one stand to another. The probability of such interpretation increases the fact that the water interchange in the examined basin takes place every 8 days. Additional evidence confirming the possibility are the bibliography reports on quite long periods of *E. coli* surveillance outside its natural habitat (30-120 days) [17].

Taking the zones of sampling into consideration has revealed that the strains are conveyed in the subsurface zone in the vast majority of cases. The largest number of strains of 100% pattern homology within the similarity groups has been found in this zone. Generally, groups with higher inter-diversity of similarity relations reveal a higher percentage of strains of the bottom zone (e. g. group A). Only group G, characteristic for strains of stand IV, is different.

Despite the compact structure of the inter-group similarity (50-40%), it consists exclusively of the bottom zone strains. The observation proves that the picture of micro-variability expressed in diversification both within and between the similarity groups, is also reflected by the ability of strains to live temporarily in different zones of a water reservoir.

The calculated frequency of occurrence of strains of particular similarity groups in each stand (Table 2), as well as the known frequency of water sampling (Fig. 4) enabled (in a form of mean values expressed in percentage), estimation of the *E. coli* population structure in a given stand with regard to the contingent of strains derived from stands situated higher (Fig. 5).

Stands I and II are located the furthest distance from the other ones. Waters within these stands do not unite and so their characteristic/specific strains reveal an individual character of genomic polymorphism expressed in separate similarity groups in the dendrogram. The structure of *E. coli* population within the area of stand III consists of: specific strains – in 16.67%, strains of stand I – 42.05% and strains of stand II – 41.28%. In stand IV the specific strains constitute 38.84% of the population (groups B, G, J together, Fig. 5); 15.5% of the population flows in from stand I and 34.04% from stand II; 11.62% of the population consists of stand III's contingent. The *E. coli* population of stand V comes from strains of stand I – 24.6%, stand II – 39.37% and 36.03% of the population consists of specific strains of stand IV. No characteristic/specific strains have been found for stand V, nor for stand III. The structure of population of the identified strains in particular stands of water sampling reveals that the migration ability of the examined *E. coli* through the subsequent stands varies. It is expressed in a gradual decrease in number and/or lack of strains of particular similarity groups in subsequent stands (Fig. 5). The observation may prove the fact that micro-variability of strains expressed in the affiliation to certain similarity groups is demonstrated by a higher rate of *E. coli* survival in water, which is a secondary environment.

In summary, the received results indicate that applying UPGMA and nearest neighbour-joining clustering method to the statistical analysis of rep-PCR (BOX) fingerprinting made it possible to differentiate and group the strains effectively, revealing a characteristic structure of *E. coli* population for particular stands of water sampling. The basic advantages of BOX primer, such as: the repetitiveness of the received patterns and the optimal degree of complexity with relation to primers of the remaining families of repetitive sequences, i. e. ERIC and REP, have already been reported [16, 18, 13]. Rep-PCR fingerprinting method has been successfully applied to differentiate and group *E. coli* isolates depending on the known source of their origin [3, 5, 8]. The method has been applied to the presented research to characterize strains/clones derived from nonpoint sources of contamination. The proposed interpretation is simplified, since it does not use a

database with a whole gamut of rep-PCR fingerprinting patterns of various defined source organisms, enabling, on the basis of the further comparison, identification of presumed source of strains.

The presented approach made it possible to rate the influence of the neighbouring stands on the structure of the population of *E. coli* identified in particular stands of water sampling and to what degree it results from the characteristic/specific contingents of the particular stand. In the face of reports on the possibility of changes in rep-PCR fingerprint patterns for *E. coli* strains connected with the change of its habitat from primary (homeothermal organism) into secondary (water, soil), the presented approach seems to be justified [7, 19, 20].

Moreover, the presented method of the analysis simplifies the results interpretation, particularly for flowing-water reservoirs. Additionally, it is economical with both time and money, which is important for routine examinations of water quality [9].

References

1. NIEMIR. M., HEIKKILA M. P., LAHTI K., KALSO S., NIEMELÄ S. I. Comparison of methods for determining the numbers and species distribution of coliform bacteria in well water samples. J. Appl. Microbiol. **90**, 850, **2001**.
2. GORDON D., M. Geographical structure and host specificity in bacteria and the implication for tracing the source of coliform contamination. Microbiol. **147**, 1079, **2001.**
3. KARIUKI S., GILKS C., KIMARI J., OLANDA A., MUYODI F., WAIYAKI P., HART C., A. Genotype analysis of *Escherichia coli* strains isolated from children and chicken living in close contact. Appl. Environ. Microbiol. **65**, 472, **1999.**
4. CARSON C. A., SHEAR B. L., ELLERSIECK M. R., ASFAW A. Identification of fecal *Escherichia coli* from human and animals by ribotyping. Appl. Environ. Microbiol. **67**, 1503, **2001.**
5. PARVEEN S., PORTER K. M., ROBINSON K., EDMINSTON L., TAMPLIN M. L. Discrimination analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. **65**, 3142, **1999.**
6. BERNHARD A. E., FIELD K. G. A PCR assay to discriminate human and animal feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. **66**, 4571, **2000.**
7. CARSON C. A., SHEAR B. L., ELLERSIECK M. R., SCHNELL J. D. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from human and animals. Appl. Environ. Microbiol. **69**, 1836, **2003.**
8. DOMBEK P. E., JOHNSON L. K., ZIMMERLEY S. J., SADOWSKY M. J. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66**, 2572, **2000**.
9. OLIVE D. M., BEAN P. Principles and applications of methods for DNA-based typing of microbial organisms. J. Clin. Microbiol. **37**, 1661, **1999**.
10. HULTON C. S. J., HIGGINS C. F., SHARP P. M. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. **5**, 825, **1991.**
11. VERSALOVIC J., KOEUTH T., LUPSKI J., R. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. **19**, 6823, **1991**.
12. GILSON E., CLEMENT J. M., BRUTLAG D., HOFNUNG M. A. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. **3**, 1417, **1984**
13. VERSALOVIC J., SCHNEIDER M., DE BRUIJN F. J., LUPSKI J. R. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. **5**, 25, **1994.**
14. Evaluation of the environmental in Lubuskie province in 2000. Provincial Inspectorate of Environmental Protection in Zielona Góra, Library of environmental monitoring Zielona Góra – Gorzów Wielkopolski. (in Polish) **2001**.
15. Polish standards PN-77 (C-04615 (07, Determination of the fecal coliform by the fermentative test-tube method. **1990.**
16. RADEMAKER J. L., DE BRUIJN F. J. Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer-assisted pattern analysis. IN: DNA markers: protocols, applications and overviews. Willey-Liss, INC., New York, pp151-171, **1997.**
17. LECLERC H., MOSSEL D. A. A., EDBERG S. C., STRUIJK C. B. Advances in the bacteriology of the coliform group: Their suitability as markers of microbial water safety. Annu. Rev. Microbiol. **55**, 201, **2001.**
18. BALDY-CHUDZUK K., NIEDBACH J., STOSIK M. Heterogeneity of *Escherichia coli* derived from artiodactyla animals analyzed with the use of rep-PCR fingerprinting, Folia Microbiol. **48**, 162, **2003.**
19. GORDON D. M., FITZ-GIBBON F. The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiol **145**, 2663, **1999.**
20. HARTEL P. G., SUMMER J. D., HILL J. L., COLLINS J. V., ENTRY J. A., SEGERS W. I. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31**, 1273, **2002.**