Institute of Medical Biology

Polish Academy of Sciences

Anna Bogumila Kubiak

Genetic epidemiological analysis of pathogenic *Escherichia coli* strains

- abstract

doctoral supervisor

dr hab. Paweł Parniewski, prof. IBM PAN

Lodz 2014

ABSTRACT

Gram-negative bacterial foodborne pathogens are one of the main causes of morbidity and mortality worldwide. What is more, hospital infections are also a significant issue as far as public health is concerned. In both cases *Escherichia coli* is one of the most frequently isolated gram-negative pathogen. Therefore, there is a still need to develop new useful methods in this field.

Escherichia coli is highly diverse bacterial species, including intestinal (IPEC) and extraintestinal (ExPEC) pathogens. Both groups are subdivided into pathovars. Being a membership of a given pathovar can be stated based on many features, such as presence of specific virulence factors, inducing various clinical symptoms, O-serotype andd phylogenetic group.

There are several techniques for microorganisms' fingerprinting, based on repetitive-elements polymerase chain reaction (rep-PCR). Trinucleotide repeat sequences (TRS) were detected in many bacterial genomes and were effectively used in our laboratory for discrimination uropathogenic *Escherichia coli*, *Mycobacterium gordonae*, *Mycobacterium kansasii*, *Mycobacterium avium* strains and *Salmonella enterica ssp. enterica* serovars. There are three main strong points of this method: it's simple and easy to perform, it's highly reproducible and what is the most important, it has a great discriminative power.

There are many VNTR loci in *E. coli* genome, which are excellent estimators of the genetic relationship. In most studies researches use the same VNTR loci, their modifications or add new ones to the set to make the test more universal to wider range of *E. coli* types.
Near the *purU* gene in *E. coli* genome, the *tyrT* genes are surrounded by the repeated elements of different length: motif B and motif A. The number of copies and complexity of the region differs among 40 analyzed *E. coli* genomes. It is thought, by our team, to be interesting from the viewpoint of genotyping and phylogenetic investigations.

The main objectives of the study were:

- determination of the diversity indices for TRS-PCR in study of intestinal pathogenic *E. coli*,
- development and evaluation of the differentiation effectiveness for the intestinal pathogenic *E. coli* with the use of a new VNTR-COLI-175 element,

- implementation of similar study for uropathogenic *E. coli* and its comparative analysis with the intestinal strains.

We tested the collection of 101 genomic DNA from diarrheal IPEC strains, two IPEC strains form another source and the collection of 50 UPEC strains. Each strain was characterized in terms of O-serotype, assignment to the phylogenetic group, assignment to the pathovar (for IPEC strains) and the presence of virulence factors specific for IPEC pathotype and UPEC pathovar. TRS-PCR tests were performed with the use of primers containing $N_6(CGG)_4$ and $N_6(GTG)_4$ motifs separately. The tests conditions differ only in annealing temperature. For $N_6(CGG)_4$-PCR it is very high and lower for $N_6(GTG)_4$-PCR.
Analyses of reproducibility gave the best results for CGG-PCR, however all results were at very similar levels. The similarity comparisons of the $N_6(CGG)_4$- and $N_6(GTG)_4$-band pattern profiles for IPEC collection revealed that $N_6(CGG)_4$-PCR and $N_6(GTG)_4$-PCR tests were able to distinguish strains with very good efficacy. It was confirmed by high values of discriminatory indices for each analyses. What is more evaluated values of diversity indices were almost at the same high level.

As similar analyses were applied to uropathogenic *E. coli* collection, we compared them with the results for the IPEC collection. The composite analysis of $N_6(CGG)_4$ and $N_6(GTG)_4$-PCR for the IPEC collection with combination with the UPEC collection gave results which are promising. Composite data analyses of $N_6(CGG)_4$ and $N_6(GTG)_4$-PCR assays had the highest discriminative power (DI= 0,996) for this collection of pathogenic *E. coli* strains. The main conclusion is that intestinal pathogenic strains clustered separately from uropathogenic strains, which is significant from epidemiological point of view.

Investigation of the potential utility of the VNTR-COLI-175 element in genotyping *E. coli* strains was another stage of this study. It was concluded that analysis of the presence of VNTR-COLI-175 element could be an additional differentiating factor in combination with TRS-PCR. The main remark is that strains, which were almost identical in TRS-PCR, are in many cases different as far as analysis of this VNTR region is concerned. Based on this result, it may be stated the analysis of the VNTR-COLI-175 element is helpful in further differentiation between similar strains.

To sum up we can conclude that:

1. $N_6(CGG)_4$ and $N_6(GTG)_4$-PCR tests have very good power of discrimination at the strain level in the case of IPEC collection, therefore, they may be useful in

epidemiology of the human intestinal pathogenic *E. coli* strains. However, they are not useful in cross-pathovar discrimination.

2. Some of the IPEC strains, which are almost identical in TRS-PCR analyses are different as far as analysis of the VNTR-COLI-175 region is concerned. It allows for their deeper diversification. It is similar in the case of the UPEC strains.
3. Diversification of the IPEC and UPEC strains in TRS-PCR analysis shows that the UPEC strains belonging to the ExPEC pathotype grouped separately from the IPEC strains. What is more, this differentiation correlates with such factors as phylogeny, virulence factors, O-serotype in same cases and some other factors which are hard to define at this stage of the knowledge.

These factors need to be farther examined and identified.