

### The dependence of discriminatory ability and concordance degree of typing system on clustering methods

D. Babenko<sup>1</sup>, A. Turmuhambetova<sup>1</sup>, I. Azizov<sup>1</sup>, A. Chesca<sup>2</sup>

<sup>1</sup>Karaganda State Medical University, Karaganda, Kazakhstan

<sup>2</sup>Transilvania University, Brasov, Romania

#### Background

A variety of bacterial typing methods based on genetic technology are currently available and successfully used in practice. A large amount of data are obtained now and part of them are stored in Databases which are accessible through web portals ([www.pubmlst.org](http://www.pubmlst.org), [www.pasteur.fr/mlva](http://www.pasteur.fr/mlva), [www.spaserver.ridom.de](http://www.spaserver.ridom.de)). Although results obtained with different typing systems are certainly useful, they contain specific information based on the typing method with its advantages and limitations. There are many attempts to compare the different typing methods in terms of typeability, reproducibility, discriminatory power, type assignment and other parameters suggested by Struelens et al. (1996). The aim of the present study was to show the dependence of discriminatory ability and congruence indices on clustering methods.

#### Methods & Materials.

48 and 27 genomes of *E.coli* and *P.aeruginosa* were used to get PFGE fragments and MLST profiles using SnapGene and Microbial In Silico Typer software, respectively. The PFGE patterns were analyzed as binary data (1 – band, 0 – no band) with using different indices for calculation dissimilarity matrix (ade4 package for R statistics). Clusters formed on 20 % dissimilarity (80% similarity) as a commonly used threshold cutoff. The clustering of MLST data was performed with EMA packages. Discriminatory power and concordance indices were calculated with Ridom EpiCompare software.

#### Results

Based on 48 and 27 PFGE profiles of *E.coli* and *P.aeruginosa* with 80% similarity cut off, the clusters were created using different indices such as Jaccard (1901), Sokal & Michener (1958), Sokal & Sneath(1963), Rogers & Tanimoto (1960), Dice (1945), Hamann (1986) for distance calculation and UPGMA method. PFGE profiles were compared with MLST data. The comparative analysis of PFGE and MLST data is presented in Table1.

Table 1. Insilico comparing the MLST and PFGE methods using different indices for distance calculation.

	Simpson index (CI, 95%)	Rand's coefficient	Adjusted Rand's coefficient	Wallace's coefficient W <sub>1</sub>	Wallace's coefficient W <sub>2</sub>	Reproducibility (with MLST)
MLST ( <i>E.coli</i> )	0.967 [0.95 - 0.985]	-	-	-	-	-
PFGE - Jaccard/UPGMA	0.997 [0.993 - 1.0]	0.968	0.096	0.667	0.054	4.167%
PFGE - Sokal & Michener/UPGMA	0.987 [0.975 - 0.999]	0.97	0.334	0.6	0.243	16.667%
PFGE - Rogers & Tanimoto/UPGMA	0.994 [0.987 - 1.0]	0.972	0.265	0.857	0.162	6.25%
MLST ( <i>P.aeruginosa</i> )	0.92 [0.848 - 0.993]	-	-	-	-	-
PFGE - Jaccard/UPGMA	0.986 [0.967 - 1.0]	0.929	0.224	0.8	0.143	22.222%
PFGE - Sokal & Sneath/UPGMA	0.994 [0.984 - 1.0]	0.926	0.124	1.0	0.071	14.815%
PFGE - Dice/UPGMA	0.977 [0.954 - 1.0]	0.932	0.309	0.75	0.214	18.519%
PFGE - Hamann coefficient	0.926 [0.855 - 0.997]	0.903	0.318	0.385	0.357	66.667%

#### Conclusion

Our study revealed that using different dissimilarity distance calculations the various clusters are formed what is reflected in discriminatory power of method and concordance degree when compared with other typing methods.