

## **ANNOTATION**

on the dissertation for the PhD doctor's degree, in specialty  
«6D110100 - Medicine»

**DMITRIY BABENKO**

### **RESEARCH OF CONCORDANCE AND DISCRIMINATING ABILITY OF MOLECULAR METHODS USED FOR MICROORGANISM TYPING**

#### **Relevance**

Typing methods for discriminating different bacterial isolates of the same species are essential epidemiological tools in infection prevention and control. Microbial typing allows the differentiation of epidemiological related from unrelated isolates of the same bacterial species. It is used to elucidate the source and route of transmission of microorganisms causing outbreaks of infection diseases. This information is important in the source of infection determining and identify ways and mechanisms of pathogen spreading during an infectious disease outbreak (MacCannell 2013). Molecular typing methods of microorganisms provide an opportunity for quantitative assessment of the effectiveness of the infection prevention and control system and as well as an analysis of the development of public health strategies.

There are many typing methods with various principles that are differing both in time per analysis and labor intensive, reliability, reproducibility and cost, and in the ability to differentiate isolates (Foxman, Zhang et al. 2005) and all of them can be divided on phenotypic and genotypic. Due to the large number of weaknesses in the phenotypic approaches, such as labor intensity, the time per analysis, the variability and a relatively low resolution (Keelara, Scott et al. 2014, Duarte, Seliwiorstow et al. 2016), molecular genetic typing methods are currently dominant in solving various problems of microbiology and epidemiology (Shahinian 2000).

Now, many different typing methods have been developed and used. For every typing method that is available in the molecular epidemiology and a newly developed approach, the scientific community evaluates the method based on some criteria, in order to understand how well and in what cases this development may be used. One of the main parameters is the discriminatory power of the typing method, which is can be assessed with Simpson index.

Using various methods for pathogen typing by different research groups has led to the fact that publications and databases have accumulated and continues growing a large number of the results what, in turn, leads to the need for comparative analysis of typing methods based on their resolution power and agreement (concordance) of the results obtained with different typing methods.

### **Aim of the study**

Comparative evaluation of concordance and discriminatory ability of modern molecular genetic methods for microorganism typing.

### **Objectives of the research**

1. To form a representative collection of clinical *S.aureus* isolates, annotate and to determine antibiogram pattern, and as well as receive *S.aureus* genome-wide data, including information about the place and time of isolation, from the NCBI GenBank, NCBI PubMed, Pathosystems Resource Integration Center - PATRIC base data.
2. To carry out subspecies typing of clinical isolates of *S.aureus* with MLVA and Spa methods, as well as *in silico* determination of MLST, Spa, MVLST, MLVA, PFGE, cgMLST, wgMLST and panGenome types using genome-wide data of *S.aureus*.
3. To assess the discriminatory power of typing methods for *S.aureus* by Simpson index.
4. To assess the concordance (agreement) between the results obtained by different typing methods at the type level and closely related groups/complexes/clusters by adjusted Rand coefficient.
5. To investigate the parameters affecting the values of concordance and discriminatory ability of the bacteria typing methods.

### **Scientific novelty**

- For the first time in Kazakhstan the collection of clinical *S.aureus* isolates have been typed by MLVA and Spa method in order to determine the discriminatory power of each method and the concordance between results.
- For the first time the comparative evaluation of 7 modern molecular-genetic typing methods, including approaches based on whole genome sequencing, such as cgMLST, wgMLST and panGenome, was done on a large genome collection (n = 4976), what made it possible to rank the typing method on the discriminatory power and as well as to assess the values of concordance at the type and closely related groups/complexes/clusters levels.
- Comprehensive assessment of the factors influencing the discriminatory power of the typing method and concordance between results have been performed.

### **Practical significance**

Investigations carried out in the thesis extend ideas about the advantages and disadvantages of each typing method of microorganisms and allow to perform an adequate choice of method depends on the problems in public health, including an investigation of both local outbreaks and global epidemics.

### **Practical implementation**

1 Certificate of intellectual property have been received based on the results of the research (Appendix A). Methods of concordance and discriminatory power of the typing methods used in the thesis have been introduced in the education process of the Department of Microbiology and Department of Molecular Biology

and Medical Genetics. Method clustering of Spa types in closely related groups of staphylococci implemented in KSMU Shared Laboratory (Appendix B).

### **The main provisions for the defense**

1. The level of resistance *S.aureus*, isolated in the Central-Kazakhstan region, to macrolides (azithromycin) was 25%. The resistance to cephalosporin (cefazolin, cefoxitin) was observed in 18% of *Staphylococcus* isolates. Resistance to oxacillin was at 18% caused by methicillin-resistant strains of *Staphylococcus* (MRSA) what confirmed with the presence of *mecA* gene.

2. Cluster analysis of Spa and MLVA typing results determined the dominance of the two clonal groups/complexes, the structure of which indicated the regional circulation of the certain intraspecific groups. Discovered t002, t030, t008 Spa types indicated belonging to clonal groups/complexes of *S.aureus* with global epidemic spread.

3. Discriminatory power of MLVA and Spa methods estimated on the collection of clinical *S.aureus* isolates ( $n = 98$ ) was 90-92%. Concordance of the results of the two typing methods was 20.7% at the type level and 86.2% at the closely related groups/complexes level.

4. *In silico* analysis of 4976 *S.aureus* genome-wide data identified 100% discriminatory ability for panGenome, wgMLST and cgMLST, 98% for MLVA, 92% for MVLST, 90% for Spa and 85.5% for MLST typing methods. The concordance of the results has not exceeded 50% at the type level, except cgMLST and wgMLST (85%). At the closely related group/complex/cluster level, the concordance between typing data was in the range of 66% to 95%. An exception was the concordance between MLVA typing and other typing methods (39-61%).

5. The parameters that affect the discriminatory power of scheme typing were: the number of loci in the typing method, values of interlocus concordance and discriminatory power of each locus included in the method. The leading role in discriminatory power belongs to the parameter "number of loci". The collection of microorganisms, which has undergone the typing, also affects the value of Simpson.

6. All stages of cluster analysis of bacteria typing, including the algorithm of the genetic distance matrix calculation, an algorithm clustering of the dendrogram and the rules for the groups' formation or cut-off value effect on the grouping of microorganisms and therefore on the value of concordance between the results.

### **Link of thesis with other research projects**

Thesis performed at Shared Laboratory KSMU in the framework of a research project funded by the MES RK "A multicenter study of antibiotic methicillin-resistant strains of *Staphylococcus aureus* in Central Kazakhstan» (2012-2014) (№ state registration 0112RK00812).

### **Approbation of the work**

The main provisions and the results of the work were presented at: International Conference of Young Scientists "World of Science and youth: Achievements and Prospects", Karaganda (Kazakhstan) 26 February 2015; 16<sup>th</sup> International Congress on Infectious Diseases - ISID, Cape Town (South Africa), 2-5 April 2014; 25<sup>th</sup> European Congress of Clinical Microbiology and Infectious Diseases - ECCMID, Copenhagen (Denmark), 25-28 April 2015; 7 International

Congress of Asian Pacific Society of Infection Control - APSIC, Taipei (Taiwan), 26-29 March 2015; XVII International Congress on Antimicrobial Therapy - IACMAC, Moscow (Russia), May 20-22, 2015; 17<sup>th</sup> International Congress on Infectious Diseases - ISID, Hyderabad (India), 2-5 March 2016; Middle East And South Asia Conference On Epigenetics And Genomics Of Infectious Diseases - Amman (Jordan), 7-8 March 2016; 26<sup>th</sup> European Congress of Clinical Microbiology and Infectious Diseases - ECCMID, Amsterdam (Netherlands), 9-12 April 2016; International Conference of Young Scientists "World of Science and youth: new ways of development», Karaganda (Kazakhstan), April 12, 2016; XIX International biomedical conference of young researchers "Basic Science And Clinical Medicine - Man And His Health", St. Petersburg (Russia), April 26, 2016; XVIII International Congress on Antimicrobial Therapy - IACMAC (IACMAC), Moscow (Russia), 25-27 May 2016

Approbation was held at the enlarged meeting of the Department of Biological chemistry of KSMU, report № 12 dated 17 June 2016.

### **Publications**

Based on thesis results there have been published 42 publications, including chapters in 2 books for laboratory diagnosis (Lambert Academic Publishing). 5 articles have been published in the journals recommended by the Committee for Control of Education and Science. There are 13 publications in the database journals (Scopus and/or Thomson Reuters), including 6 articles and 7 theses.

### **Scope and structure of the dissertation**

Thesis is stated on 117 pages and consists of introduction, main part, conclusion and list of references, including 228 literature sources. The dissertation includes 36 tables, 48 figures, and 2 applications.

### **Materials and methods**

Clinical isolates of *S.aureus* (n = 98) were collected from various hospital departments, including pediatric, surgical, venereal diseases, pulmonology and others of the cities: Karaganda, Astana, Zhezkazgan and Semey, and as well as 116 completed annotated genomes of *S.aureus* and 4976 *S.aureus* genomes sequences as contigs/scaffolds from GenBank NCBI database.

Methods: Evaluation of antimicrobial resistance was performed using serial dilution test. *S.aureus* strain typing was done by Spa and MLVA methods. EBURST/MST and Jaccard/UPGMA algorithms were used for clustering. *In silico* identification of MLST, MVLST, Spa, MLVA, cgMLST, wgMLST, panGenome types was carried out on genome-wide data with using bioinformatics tools. Analysis of the discriminatory ability of typing method was done based on the Simpson index. Agreement (concordance) of the results was assessed using the adjusted Rand coefficient at the type level, and at the closely related groups/complexes level.

### **Conclusions**

1. The level of antimicrobial resistance of the clinical isolates of *S.aureus* (n = 98) from Central Kazakhstan region was 25% to macrolides (azithromycin) and about 18% to cephalosporin (cefazolin, cefoxitin). In 18% cases there were resistance to oxacillin provided MRSA strains, which were confirmed by PCR for the presence of *mecA* gene.

2. Spa and MLVA data clustering of *S.aureus* allowed determining 4 subspecies group/complex with the dominance of the two clonal complexes. The clonal complex structure, including isolates from different cities of Kazakhstan, indicated the circulation of certain intraspecific groups. At the same time, the discovery of t002, t030, t008 Spa type allow to assume about the involvement of isolates of *S.aureus*, isolated in the Republic of Kazakhstan, in global epidemiology.

3. Discriminatory power of MLVA and Spa methods obtained on the typing of the collection of clinical isolates *S.aureus* (n = 98) was 90-92%.

4. The concordance between results of the two methods (Spa and MLVA typing) was 20.7% at the type level and 86.2% at the closely related group level what indicated a good agreement for subspecies groups defined by the typing methods.

5. *In silico* analysis of 4976 genomes of *S.aureus* global collection allowed to rank 7 modern molecular genetic methods for *S.aureus* typing based on their discriminatory ability. Methods based on genome-sequence of microorganisms, such as panGenome, wgMLST, cgMLST, have demonstrated the highest resolution power (100%). Typing method, using tandem repeats, have showed 98% discriminatory power. The resolution power of MVLST and Spa typing was just above 90% (92% and 90%, respectively). The lowest discriminatory power had MLST method (85.5%).

6. Concordance of typing data did not exceed 50% at the type level, except for cgMLST and wgMLST (85%). The agreement between the results, grouped in closely related groups/complexes/clusters, was in the range of 66% to 95%, with the exception of values between MLVA typing and other methods (39-61%).

7. The parameters that affect the discriminatory power of scheme typing were: the number of loci in the typing method, values of interlocus concordance and discriminatory power of each locus included in the method. The leading role in discriminatory power belongs to the parameter "number of loci". The collection of microorganisms, which has undergone the typing, also affects the value of Simpson.

8. All stages of cluster analysis of bacteria typing, including the algorithm of the genetic distance matrix calculation, an algorithm clustering of the dendrogram and the rules for the groups' formation or cut-off value effect on the grouping of microorganisms and therefore on the value of concordance between the results.