

IN SILICO COMPARISON OF DIFFERENT TYPES OF MLVA WITH PFGE BASED ON PSEUDOMONAS AERUGINOSA GENOMES

DMITRIY BABENKO¹, ANAR TURMUHAMBETOVA¹, TIM SANDLE², SORINA ANAMARIA PESTREA³, DAN MORARU⁴, ANTONELLA CHEȘCĂ^{4,5}

¹Karaganda State Medical University, Kazakhstan - ²University of Manchester, United Kingdom - ³Clinical Hospital of Psychiatry and Neurology Brașov - ⁴Clinical Hospital of Pneumophysiology Brașov - ⁵Transilvania University Brașov, Romania

ABSTRACT

Introduction: *Pseudomonas aeruginosa* is an important pathogen in nosocomial infections and developed typing techniques are essential allowing researchers to understand hospital epidemiology. Monitoring the emergence and transmission of *Pseudomonas aeruginosa* strains permits the elucidation of the source of infection and routes of bacterial transmission. The aim of the present study was an *in silico* comparison of Pulsed-field gel electrophoresis with different schemes of Multiple Locus Variable-number Tandem Repeat Analysis in terms of discriminatory power and concordance.

Materials and methods: 58 *P.aeruginosa* whole genomes have been analyzed *in silico* to determine *SpeI*-digested PFGE type and subspecies types using different MLVA methods. Resolution power, strength and direction of the concordance between typing methods have been estimated by calculation of the Simpson's index, the adjusted Rand and the adjusted Wallace coefficients.

Results: The Simpson's indices of diversity were 1.0 for PFGE and from 0.995 to 0.999 for MLVA schemes with 6-19 markers. The congruence between PFGE and different MLVA methods measured by the adjusted Rand index were from 0.306 to 0.665 on cluster level for PFGE and type level for MLVA. The congruence was slightly higher at the clonal cluster level - from 0.46 to 0.694.

Conclusion: Our *in silico* study for comparing different MLVA schemes with PFGE, based on *Pseudomonas aeruginosa* genomes showed, on the one hand, the same high level of discriminatory power of PFGE and MLVA even with 6 tandem markers; nonetheless, on the other hand, there was moderate/poor congruence (no more 70%) between PFGE and MLVA schemes on cluster level.

Key words: MLVA, PFGE, *Pseudomonas aeruginosa*, *In Silico*, Simpson index, Adjusted Rand index, Adjusted Wallace coefficient.

DOI: 10.19193/0393-6384_2017_4_090

Received November 30, 2016; Accepted February 02, 2017

Introduction

Identifying different types of very diverse organisms within a species is critical for outbreak investigations and for the control of infectious diseases⁽¹⁾. A large variety of approaches for typing have been developed. These can be divided into two groups - phenotypic and genotypic methods. Importantly, such typing methods must have the ability to distinguish epidemic from endemic or sporadic isolates.

At the same time, methods must be low-cost, rapid, highly reproducible and easy to carry out and to interpret^(1,2).

The results obtained with phenotypic techniques have a tendency to vary due to the influence of growth conditions and/or growth phase on the products of gene expression. Furthermore, some phenotypic approaches are time consuming, tend to be more manual, and require greater skill to run⁽¹⁾. Importantly, as it has been shown in other studies⁽³⁾ the discriminatory power of phenotypic methods is lower in comparison with genotypic variants. Each of these reasons has led to genotypic techniques gradually replacing phenotypic approaches over the past two decades.

There are a large number of genetic typing methods with a high degree of discriminatory

power. One of the most frequently used molecular-genetic methods is pulsed-field gel electrophoresis (PFGE). PFGE is considered as the “gold standard” typing method, due to its excellent discriminatory power and high epidemiological concordance⁽⁴⁾; however, the method remains relatively expensive and time-consuming^(5,6). Furthermore, this technique tends to suffer from a lack of interlaboratory reproducibility⁽⁶⁾.

Another relatively new genetic approach is Multilocus Variable Number Tandem Repeat Analysis (MLVA), which is based on a set of polymorphic tandem repeat loci of bacteria and this method has been successfully used for epidemiological investigations. MLVA is considered as a perspective method with “ideal” criteria for typing, such as being easy to perform, rapid, possessing a high discriminatory power, and having good reproducibility. There are several MLVA schemes for typing of the pathogen *Pseudomonas aeruginosa* currently available. This range from having six to sixteen variable number tandem repeat (VNTR) markers⁽⁷⁻¹⁰⁾.

In the present study, we compared different approaches of MLVA with PFGE *in silico* in terms of discriminatory power and coefficients congruence, against the clinical pathogen *Pseudomonas aeruginosa*.

Materials and methods

The study examined 58 *Pseudomonas aeruginosa* genomes (that have been completed and published). These were downloaded from the National Center for Biotechnology Information (NCBI) GenBank⁽¹¹⁾.

In silico simulation of SpeI enzyme digestion PFGE was performed on *Pseudomonas aeruginosa* genomes transformed into circular structure with

Geneious software (Biomatters Lmd.)⁽¹²⁾. PFGE profiles, ranging from 10 to 950 kb, were saved as a TIFF file for further analysis in TotalLab™ 1D software (Nonlinear Dynamics, Newcastle upon Tyne, UK).

Eight different MLVA schemes (reported in table 1) were analyzed by using primer pairs (presented in table 2) and SnapGene (GSL Biotech) and Geneious (Biomatters Lmd.) tools in order to produce amplification products^(12, 13). Based on PCR products, the number of repeats were calculated and rounded up to the nearest integer to obtain the numeric digit code of the isolates. These codes, referred to MLVA profiles, were subsequently used for clustering.

To evaluate genetic variability and relatedness of *Pseudomonas aeruginosa* isolates based on a comparison of bands, the genetic distance was calculated using the Sørensen-Dice coefficient for similarity⁽¹⁶⁾. From this, a dendrogram was built using unweight-pair group method with arithmetic means via the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method. Categorical data (MLVA digit profile) were analyzed using Nei’s standard genetic distance⁽¹⁷⁾ and UPGMA algorithm clustering with the “EMA” package for R statistics⁽¹⁸⁾. Isolates that clustered together, with more than 80% similarity, were considered as genetically related strains^(5, 19). A dendextend R package was used for visualization and dendrogram comparison⁽²⁰⁾. Diversity (from Simpson’s index) and cluster concordance (Adjusted Rand, Adjusted Wallace) were evaluated via a web source⁽²¹⁾.

Results

Fifty-eight genomes of *Pseudomonas aeruginosa* were downloaded and analyzed *in silico*. Prior to converting into circular shape, the genomes were

| | | | | | | | | | | | | | | | | | | | |
|----------------------------|------|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| MLVA19 Tandem [This study] | ms10 | ms061 | ms77 | ms127 | ms142 | ms172 | ms173 | ms194 | ms207 | ms209 | ms211 | ms212 | ms213 | ms214 | ms215 | ms216 | ms217 | ms222 | ms223 |
| MLVA16 Orsay (10) | | ms061 | ms77 | ms127 | ms142 | ms172 | | | ms207 | ms209 | ms211 | ms212 | ms213 | ms214 | ms215 | ms216 | ms217 | ms222 | ms223 |
| MLVA15 Orsay (7) | | | ms77 | ms127 | ms142 | ms172 | | | ms207 | ms209 | ms211 | ms212 | ms213 | ms214 | ms215 | ms216 | ms217 | ms222 | ms223 |
| MLVA12 Tandem | | ms061 | ms77 | | ms142 | ms172 | | | ms207 | ms209 | ms211 | ms212 | ms213 | ms214 | | | ms217 | ms222 | |
| [this study] | | | | | | | | | | | | | | | | | | | |
| MLVA9 Utercht (14) | | | ms77 | ms127 | ms142 | | | | | | ms211 | | ms213 | | ms215 | ms216 | ms217 | | ms223 |
| MLVA9 London (9) | | ms061 | | | | ms172 | | | ms207 | ms209 | ms211 | | ms213 | ms214 | | | ms217 | ms222 | |
| MLVA7 Tandem (8) | ms10 | ms061 | ms77 | ms127 | ms142 | ms172 | ms173 | | | | | | | | | | | | |
| MLVA6 Tandem (15) | ms10 | ms061 | ms77 | ms127 | ms142 | ms172 | | | | | | | | | | | | | |

∖: MLVA schemes with different number of loci (from 6 to 19) for *Pseudomonas aeruginosa* typing.

| Typing method | group number | Simpson's index [95% CI] | Adjusted Rand coefficient (PFGE(80%)–MLVA) | Adjusted Wallace PFGE (80%) -> MLVA | Adjusted Wallace MLVA -> PFGE (80%) | *Entanglement in comparison with PFGE |
|-----------------------------|--------------|--------------------------|--|-------------------------------------|-------------------------------------|---------------------------------------|
| PFGE | 58 | 1 [1.000-1.000] | - | 0.000 | 1.000 | - |
| PFGE (80% cut-off) | 49 | 0.994 [0.989-0.999] | - | - | - | - |
| MLVA19 Tandem | 56 | 0.999 [0.996-1.000] | 0.332 | 0.199 | 1.000 | 0.2756 |
| MLVA19 Tandem (80% cut-off) | 48 | 0.993 [0.987-0.999] | 0.634 | 0.698 | 0.581 | - |
| MLVA16 Orsay | 55 | 0.998 [0.995-1.000] | 0.46 | 0.299 | 1.000 | 0.3211 |
| MLVA16 Orsay (80% cut-off) | 46 | 0.992 [0.985-0.998] | 0.664 | 0.798 | 0.569 | - |
| MLVA15 Orsay | 53 | 0.997 [0.994-1.000] | 0.665 | 0.498 | 1.000 | 0.3162 |
| MLVA15 Orsay (80% cut-off) | 45 | 0.991 [0.985-0.997] | 0.637 | 0.798 | 0.530 | - |
| MLVA12 Tandem | 54 | 0.998 [0.994-1.000] | 0.57 | 0.399 | 1.000 | 0.1521 |
| MLVA12 Tandem (80% cut-off) | 46 | 0.992 [0.985-0.998] | 0.664 | 0.798 | 0.569 | - |
| MLVA9 Utercht | 51 | 0.995 [0.990-1.000] | 0.553 | 0.498 | 0.623 | 0.2353 |
| MLVA9 Utercht (80% cut-off) | 42 | 0.988 [0.981-0.995] | 0.462 | 0.696 | 0.346 | - |
| MLVA9 London | 53 | 0.997 [0.994-1.000] | 0.665 | 0.498 | 1.000 | 0.2269 |
| MLVA9 London (80% cut-off) | 47 | 0.992 [0.986-0.998] | 0.694 | 0.798 | 0.613 | - |
| MLVA7 Tandem | 55 | 0.998 [0.995-1.000] | 0.306 | 0.199 | 0.665 | 0.2045 |
| MLVA7 Tandem (80% cut-off) | 49 | 0.994 [0.989-0.999] | 0.497 | 0.497 | 0.497 | - |
| MLVA6 Tandem | 55 | 0.998 [0.995-1.000] | 0.46 | 0.299 | 1.000 | - |
| MLVA6 Tandem (80% cut-off) | 48 | 0.993 [0.988-0.999] | 0.665 | 0.698 | 0.634 | - |

Table 3: Number of *Pseudomonas aeruginosa* types and groups determined with 80% cut-off, Simpson's index of diversity, and adjusted Rand indices, adjusted Wallace coefficients for PFGE and different MLVA schemes and entanglement of PFGE and MVLA tree comparison.

This allows the user to compare PFGE patterns within ranges from around 10 to 800 kbp.^(6, 24) Different studies have shown the discriminant power of the PFGE-Spe I approach, which is between 0.98 and 0.998^(25, 26). Our *in silico* PFGE simulation generated band numbers from 24 to 48 (mean = 34; SD = + 5) in the range 10-950 kbp what allowed to achieve the ultimate (100%) Simpson's index.

MLVA utilizes PCR to amplify the region containing short, repetitive tandem sequences. Amplification products were separated by electrophoresis to determine the size and the number of repeats in the locus. Different MLVA schemes, involving from 6 to 16 loci with repetitive tandem sequences, have been developed for *Pseudomonas aeruginosa* typing⁽⁷⁻¹⁰⁾.

In our study, all eight MLVA schemes provided discriminatory power between 0.988 and 0.998, with overlapping 95% confidence intervals. Makaoui Maa'tallah et al. in their study on typing of *Pseudomonas aeruginosa* (n=85) received the same Simpson's index of diversity (0.980 (0.970-0.991) using MLVA scheme with 15 markers⁽²⁷⁾.

One of the possible reason why MLVA schemes with 6 or 7 loci produced the same discriminatory ability as MLVA with more loci (e.g. 15 or 16 loci) was because the markers included in each MLVA method had a different discriminatory index (as shown in table 4).

The adjusted Rand indices have a degree of equivalence between the type assignments of two distinct typing methods (PFGE (80% cut-off) and MLVA). These were 0.501 for type level and 0.614 for MLVA cluster level at 80% (p = 0.0493) and these ranged from 0.332 to 0.694. This indicates a poor/moderate congruence between PFGE clusters and MLVA clusters/types approaches (as per table 3).

The adjusted Wallace coefficient indicates the probability that pairs of isolates that are assigned to the same type by one typing method are also typed as identical by the other. PFGE with an 80% cut-off were not a particularly good method for predicting MLVA type (the mean of adj. Wallace was 0.361 (from 0.199 to 0.498)).

By contrast, MLVA clusters (more than 80% similarity) were better predicted by PFGE clusters (here the mean of adj. Wallace was 0.911 (from

0.623 to 1.0)). In general, clusters of MLVA typing provided a good prediction for PFGE clusters (the mean of adj. Wallace was 0.723; from 0.497 to 0.798), but due to the higher resolution of PFGE the MLVA method performed poorly in terms of predicting PFGE (adj. Wallace MLVA clusters -> PFGE clusters were from 0.346 to 0.634)).

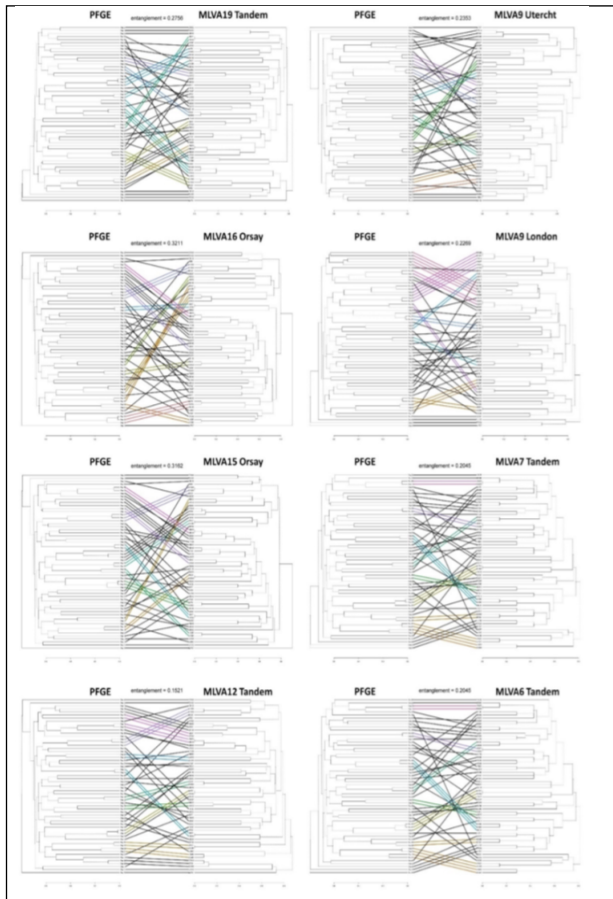


Fig. 2: Tree comparison constructed based on PFGE and different MLVA data of *Pseudomonas aeruginosa* typing with connected lines showing degree of disorder between two dendrograms (entanglement). These pictures and indices were calculated with Dendextend package (R statistics).

In systematic biology, visualization of similarities and differences of phylogenetic trees, in particular in multi-gene analysis, can be provided by drawing two dendrograms side by side with connector lines between species that correspond to each other in the two trees. The resultant tanglegram plot, as presented in this paper, demonstrated discrepancies between the topology structure of MLVA methods and PFGE (figure 2); whereas entanglement outcomes showed approximately the same alignment parameters between different MLVA schemes and PFGE (as shown in table 3).

| MLVA marker | # Samples | # Different types | Discriminatory index | Confidence interval (95% CI) |
|-------------|-----------|-------------------|----------------------|------------------------------|
| ms10 | 58 | 16 | 0.912 | [0.875 - 0.949] |
| ms061 | 58 | 10 | 0.882 | [0.854 - 0.91] |
| ms77 | 58 | 5 | 0.338 | [0.183 - 0.493] |
| ms127 | 58 | 3 | 0.424 | [0.283 - 0.565] |
| ms142 | 58 | 8 | 0.826 | [0.774 - 0.879] |
| ms172 | 58 | 9 | 0.734 | [0.646 - 0.822] |
| ms173 | 58 | 15 | 0.895 | [0.854 - 0.935] |
| ms194 | 58 | 11 | 0.782 | [0.699 - 0.866] |
| ms207 | 58 | 12 | 0.89 | [0.864 - 0.915] |
| ms209 | 58 | 9 | 0.793 | [0.738 - 0.849] |
| ms211 | 58 | 7 | 0.828 | [0.792 - 0.864] |
| ms212 | 58 | 8 | 0.764 | [0.691 - 0.838] |
| ms213 | 58 | 7 | 0.779 | [0.708 - 0.851] |
| ms214 | 58 | 9 | 0.8 | [0.75 - 0.849] |
| ms215 | 58 | 8 | 0.797 | [0.744 - 0.851] |
| ms216 | 58 | 4 | 0.604 | [0.502 - 0.707] |
| ms217 | 58 | 7 | 0.797 | [0.746 - 0.848] |
| ms222 | 58 | 5 | 0.724 | [0.68 - 0.767] |
| ms223 | 58 | 6 | 0.724 | [0.653 - 0.794] |

Table 4: Resolution power and 95% CI of each VNTR marker included in MLVA scheme for *Pseudomonas aeruginosa* typing.

In summary, our in silico study for comparing different MLVA schemes with PFGE, based on *Pseudomonas aeruginosa* genomes showed, on the one hand, the same high level of discriminatory power of PFGE and MLVA even with 6 tandems markers; nonetheless, on the other hand, there was poor congruence between PFGE and some MLVA schemes, and moderate agreement in cases with other MLVA approaches. In general, concordance between PFGE clusters and MLVA clusters was higher than between PFGE clusters and MLVA types, although this was never more than 69.4%. Based on different principles of PFGE and MLVA typing methods, there were understandable different outcomes in terms of the topology of hierarchical clustering.

Interestingly, PFGE and MLVA methods could independently determine the *Pseudomonas aeruginosa* PA7 strains as taxonomic outliers⁽²⁸⁾.

Acknowledgements

We are grateful to Peter Kruczkiewicz for constructive comments and discussion in using Microbial In Silico Typer (MIST).

References

- 1) van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect* 2007; 13 Suppl 3: 1-46.
- 2) Struelens MJ. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin Microbiol Infect* 1996; 2: 2-11.
- 3) Di Cagno R, Minervini G, Sgarbi E, Lazzi C, Bernini V, et al. Comparison of phenotypic (Biolog System) and genotypic (random amplified polymorphic DNA-polymerase chain reaction, RAPD-PCR, and amplified fragment length polymorphism, AFLP) methods for typing *Lactobacillus plantarum* isolates from raw vegetables and fruits. *Int J Food Microbiol* 2010; 143: 246-53.
- 4) Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijk J, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill* 2013; 18: 20380.
- 5) Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 1995; 33: 2233-9.
- 6) Grundmann H, Schneider C, Hartung D, Daschner FD, Pitt TL. Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa*. *J Clin Microbiol* 1995; 33: 528-34.
- 7) Vu-Thien H, Corbinau G, Hormigos K, Fauroux B, Corvol H, et al. Multiple-locus variable-number tandem-repeat analysis for longitudinal survey of sources of *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *J Clin Microbiol* 2007; 45: 3175-83.
- 8) Onteniente L, Brisse S, Tassios PT, Vergnaud G. Evaluation of the polymorphisms associated with tandem repeats for *Pseudomonas aeruginosa* strain typing. *J Clin Microbiol* 2003; 41: 4991-7.
- 9) Turton JF, Turton SE, Yearwood L, Yarde S, Kaufmann ME, et al. Evaluation of a nine-locus variable-number tandem-repeat scheme for typing of *Pseudomonas aeruginosa*. *Clin Microbiol Infect* 2010; 16: 1111-6.
- 10) Sobral D, Mariani-Kurkdjian P, Bingen E, Vu-Thien H, Hormigos K, et al. A new highly discriminatory multiplex capillary-based MLVA assay as a tool for the epidemiological survey of *Pseudomonas aeruginosa* in cystic fibrosis patients. *Eur J Clin Microbiol Infect Dis* 2012; 31: 2247-56.
- 11) Sayers EW, Karsch-Mizrachi I. Using GenBank. *Methods Mol Biol* 2016; 1374: 1-22 [http://www.ncbi.nlm.nih.gov/genome/browse].
- 12) Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012; 28: 1647-9 [http://www.geneious.com].
- 13) SnapGene software (from GSL Biotech; available at snapgene.com).
- 14) Revets H, Vandamme P, Van Zeebroeck A, De Boeck K, Struelens MJ, et al. Burkholderia (*Pseudomonas*) cepacia and cystic fibrosis: the epidemiology in Belgium. *Acta Clin Belg* 1996; 51: 222-30.
- 15) Edelstein MV, Sklelenova EN, Shevchenko OV, D'Souza J W, Tapalski DV, et al. Spread of extensively resistant VIM-2-positive ST235 *Pseudomonas aeruginosa* in Belarus, Kazakhstan, and Russia: a longitudinal epidemiological and clinical study. *Lancet Infect Dis* 2013; 13: 867-76.
- 16) Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology* 1945; 26: 297-302.
- 17) Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* 1983; 19: 153-70.
- 18) Nicolas Servant, Eleonore Gravier, Pierre Gestraud, Cecile Laurent, Caroline Paccard, et al. EMA: Easy Microarray data Analysis. R package version 1.4.4. <https://CRAN.R-project.org/package=EMA> 2014;
- 19) Speijer H, Savelkoul PH, Bonten MJ, Stobberingh EE, Tjhie JH. Application of different genotyping methods for *Pseudomonas aeruginosa* in a setting of endemicity in an intensive care unit. *J Clin Microbiol* 1999; 37: 3654-61.
- 20) Galili T. dendextend: Extending R's dendrogram functionality. R package version 0.17.5. <http://CRAN.R-project.org/package=dendextend> 2014;
- 21) Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, et al. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J Clin Microbiol* 2006; 44: 2524-32 (<http://www.comparingpartitions.info/index.php?link=Tool>).
- 22) Morales G, Wiehlmann L, Gudowius P, van Delden C, Tummler B, et al. Structure of *Pseudomonas aeruginosa* populations analyzed by single nucleotide polymorphism and pulsed-field gel electrophoresis genotyping. *J Bacteriol* 2004; 186: 4228-37.
- 23) Spencker FB, Haupt S, Claros MC, Walter S, Lietz T, et al. Epidemiologic characterization of *Pseudomonas aeruginosa* in patients with cystic fibrosis. *Clin Microbiol Infect* 2000; 6: 600-7.
- 24) Doleans-Jordheim A, Cournoyer B, Bergeron E, Croize J, Salord H, et al. Reliability of *Pseudomonas aeruginosa* semi-automated rep-PCR genotyping in various epidemiological situations. *Eur J Clin Microbiol Infect Dis* 2009; 28: 1105-11.
- 25) Silbert S, Pfaller MA, Hollis RJ, Barth AL, Sader HS. Evaluation of three molecular typing techniques for nonfermentative Gram-negative bacilli. *Infect Control Hosp Epidemiol* 2004; 25: 847-51.
- 26) Talon D, Cailleaux V, Thouverez M, Michel-Briand Y. Discriminatory power and usefulness of pulsed-field gel electrophoresis in epidemiological studies of *Pseudomonas aeruginosa*. *J Hosp Infect* 1996; 32: 135-45.
- 27) Maatallah M, Bakhrouf A, Habeeb MA, Turlej-Rogacka A, Iversen A, et al. Four genotyping schemes for phylogenetic analysis of *Pseudomonas aeruginosa*: comparison of their congruence with multi-locus sequence typing. *PLoS One* 2013; 8: e82069.
- 28) Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, et al. Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* 2010; 5: e8842.

Corresponding author

ANTONELLA CHEȘCĂ

Clinical Hospital of Pneumophysiology Brașov
(Romania)