PHYLOGENETIC RELATIONSHIP CLUSTERING THRESHOLDS OF POTENTIALLY BACTERIOCINOGENIC CLINICAL AND DAIRY ENTEROCOCCUS SPP. STRAINS WITH RESPECT TO THEIR GEOGRAPHICAL ORIGINS IN BULGARIA

Svetoslav G. Dimov1, Tanya Strateva2, Veronika Petkova3, Maria Dimitrova4, Daniela Atanasova5, Ivan Mitov2

Address(es): Assoc. Prof. Svetoslav G. Dimov, PhD, Sofia University „St. Kliment Ohridski”, Faculty of Biology, Dept. of Genetics, 8, Dragan Tzankov Blvd., 1164 – Sofia, Bulgaria, phone: +359 2 8167 342.
1 Sofia University „St. Kliment Ohridski”, Faculty of Biology, Dept. of Genetics, 8, Dragan Tzankov Blvd., 1164 – Sofia, Bulgaria.
2 Medical University of Sofia, Faculty of Medicine, Dept. of Medical Microbiology, 2, Zdrave Str., 1431 – Sofia, Bulgaria.

*Corresponding author: svetoslav@biofac.uni-sofia.bg
doi: 10.15414/jmbfs.2015/16.5.3.286-289

ARTICLE INFO
Received 13. 5. 2015
Revised 12. 6. 2015
Accepted 15. 10. 2015
Published 1. 12. 2015

ABSTRACT
The aim of the study is to examine different thresholds of similarity coefficients for reliable clustering of clinical and dairy enterococci originating from various geographical regions in Bulgaria, based on randomly amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) unweighted pair group method with arithmetic mean (UPGMA) analysis. For creating subsets of different strains on a relatively independent basis, was created a selection based on the presence of some bacteriocin genes. A collection of 110 Enterococcus faecalis and 51 Enterococcus faecium nosocomial strains from in-patients in five hospitals, covering the major cities of Bulgaria, and additional 11 Enterococcus faecalis and 15 Enterococcus faecium strains isolated from home-made and artisanal cheeses in different mountain regions of the country was screened for the presence of five commonly found bacteriocin genetic determinants – enterolysin A (EntA) and bacteriocin AS-48 for E. faecalis and enterocin A (EntA), enterocin B (EntB) and enterocin P (EntP) for E. faecium. Strains carrying the same bacteriocin gene were subjected separately to RAPD-PCR, followed by UPGMA analysis. The obtained results showed that clustering should be performed with similarity coefficients above 95% in order to separate the strains based on cheeses’ or clinical origin and according to their geographical origins. Our findings indicate that the largely accepted similarity coefficient of 70% for clonal relatedness should be reconsidered and the threshold raised up.

Keywords: Dairy and clinical Enterococcus spp., RAPD-PCR, UPGMA, clustering thresholds

INTRODUCTION
Randomly amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) has become one of the most routinely used methods for phylogenetic studies in microbiology which is still performed for several major reasons: it is fast, it does not require expensive equipment, it is species- and physiological state independent, it is not time consuming, and none the last it possesses good enough reproducibility when it is performed repeatedly by experienced researchers (Caetano-Anolles, 1991, Bowditch et al., 1993, Abed et al., 1995, Al-Badah et al., 2015). The obtained polymorphic electrophoretic patterns are subjected to bioinformatics analyses using two major methods – unweighted pair group method with arithmetic mean (UPGMA) (Legendre et al., 1998) and neighbor-joining (NJ) (Saitou et al., 1997), the latter taking into account also the evolution rates. Both methods are based on similarity matrices filled with the similarity coefficients for each possible pair of two of the analyzed strains. The results are represented in the form of dendrograms which are visually assimilated as phylogenetic trees in which the strains are clustered in groups of more closely related ones. However, the correct interpretation of the dendrograms requires to put in place a similarity percentage threshold to form a cluster of clonally related strains. Among the scientific community such threshold of 70% to 75% similarity has been widely accepted for decades (Webster et al., 1996, Grundmann et al., 1997).

In our study, in order to assess the reliability of the thresholds of 70% similarity, 90% similarity and 95% similarity, we subjected to RAPD-PCR analyses two collections of clinical and dairy products isolates of E. faecalis and E. faecium coming from different and relatively distant and isolated cities in Bulgaria (Fig 1). They were subdivided to separate groups for each analysis according to their potential to be bacteriocinogenic for different bacteriocins as a relatively independent from the source (clinical or non-pathogenic cheese-borne) common feature (Grousu-Tudor et al., 2014). All strains showing the presence of a given bacteriocin gene were subjected separately to an UPGMA analyses, and clustering was performed using the three mentioned thresholds.

MATERIAL AND METHODS
Bacterial strains, media and cultivation conditions
All strains and their origins are listed in Table 1. The clinical Enterococcus spp. isolates from the different hospitals were provided by the Department of Medical Microbiology, Faculty of Medicine, Medical University of Sofia, Bulgaria, where they were identified at species level by polyphasic approach including molecular, microbiological and physiological methods (Strateva et al., 2014). The cheese-borne non-pathogenic Enterococcus spp. were isolated in the Faculty of Biology,
Sofia University “St. Kliment Ohridski”, Bulgaria, where they were identified at species level by the same methods. All strains were cultivated on M-17 broth, liquid or 1.5% agar media at 37 °C (for the clinical isolates) or 30 °C (for the dairy isolates) when used, or conserved as 20% glycerol stocks at -80 °C.

### Table 1 List of the strains presenting bacteriocin genes, and their origins

<table>
<thead>
<tr>
<th>Strain origin</th>
<th>Strain №</th>
<th>Strains presenting the entA gene</th>
<th>Strains presenting the entP gene</th>
<th>Strains presenting the entB gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiprofile Hospital for Active Treatment “Dr. Tota Fenkova” – Gabrov*</td>
<td>83, 84, 87, 91, 139, 141, 143, 219, 221, 236, 238</td>
<td>83, 84, 87, 91, 139, 141, 143, 219, 221</td>
<td>236, 238, 280</td>
<td>122, 123, 124, 125, 126, 126, 130, 131, 134, 135, 136, 137, 138, 135, 136, 137, 138</td>
</tr>
<tr>
<td>University Multiprofile Hospital for Active Treatment “St. Marina” – Varma*</td>
<td>122, 123, 124, 125, 126, 130, 131, 134, 135, 136, 137, 138</td>
<td>137, 138</td>
<td>135, 136, 137, 138</td>
<td>133, 135, 136, 137, 138</td>
</tr>
<tr>
<td>Cheese-borne strains</td>
<td>SD1, SD2, SD4, SD5, SD6, SD14-B, 17, 1L, 53, 122, 126, 137, 138, 139, SH9.1</td>
<td>SD2, SD14-B</td>
<td>SH9.1</td>
<td>SH9.1</td>
</tr>
</tbody>
</table>

* Further within the text the hospitals are designed only by the names of the cities or their proper names for the two hospitals in Sofia.

### DNA isolation

Total DNA from all used strains was isolated with GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich), according to the manufacturer’s instructions, from 3 ml overnight cultures inoculated with a single colony. Unless it was used immediately, the DNA was stored for up to several weeks at -20 °C.

### PCR screening for the presence of bacteriocin genetic determinants

All PCR primers and their properties are listed in Table 2. PCR amplification was carried out in a volume of 20 µl containing 0.4 U Taq polymerase in a KCl buffer system (Thermo Scientific), containing 1.5 mM MgCl₂, 0.2 mM dNTPs and about 50 ng of the template DNA. Primers concentrations were 0.20 µM. After an initial denaturation step for 5 min at 94 °C, 40 amplification cycles were performed. They consisted of incubation for 30 s at 94 °C for denaturation, 30 s at the annealing temperature of each primer, and 1 min at 72 °C for synthesis. A final extension step for 5 min at 72 °C completed the amplification. 8 µl of the PCR products were analyzed by electrophoresis on a 1.5 % agarose gel in TBE buffer system, followed by ethidium-bromide staining.

### Table 2 PCR primers used in this study and their characteristics

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'→3')</th>
<th>Target / Application</th>
<th>Tₐnait (°C)</th>
<th>Product size (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>As48-1</td>
<td>5'-AATAAAACTCATGGGT-3'</td>
<td>E. faecalis enterocin A5-48 gene</td>
<td>50°C</td>
<td>377 bp</td>
<td>Fernández et al., 2007</td>
</tr>
<tr>
<td>As48-5</td>
<td>5'-CCAAGAATACTGCTCCTTT-3'</td>
<td>E. faecalis enterolysin A gene</td>
<td>59°C</td>
<td>1007 bp</td>
<td>Nigutova et al., 2006</td>
</tr>
<tr>
<td>EntL3</td>
<td>5'-GGACAAACAATCCTGGGAGACACT-3'</td>
<td>E. faecium enterocin B gene</td>
<td>55°C</td>
<td>123 bp</td>
<td>du Toit et al., 2001</td>
</tr>
<tr>
<td>EntL9</td>
<td>5'-GGCAGTAAAGATTAGAATAAA-3'</td>
<td>E. faecium enterocin A gene</td>
<td>55°C</td>
<td>123 bp</td>
<td>du Toit et al., 2001</td>
</tr>
<tr>
<td>EntAF</td>
<td>5'-AAATTTATGGAAATGGGTGTTAT-3'</td>
<td>E. faecium enterocin B gene</td>
<td>60°C</td>
<td>160 bp</td>
<td></td>
</tr>
<tr>
<td>EntAR</td>
<td>5'-GCACCTCCCTGGGAATTGCTC-3'</td>
<td>E. faecium enterocin B gene</td>
<td>60°C</td>
<td>160 bp</td>
<td></td>
</tr>
<tr>
<td>EntBF</td>
<td>5'-GAAAATGATCACAAGATGCTTA-3'</td>
<td>E. faecium enterocin P gene</td>
<td>60°C</td>
<td>215 bp</td>
<td>Gutiérrez J. et al., 2005</td>
</tr>
<tr>
<td>EntBR</td>
<td>5'-GGTGCAATTAGATACATCATTG-3'</td>
<td>E. faecium enterocin P gene</td>
<td>60°C</td>
<td>215 bp</td>
<td>Gutiérrez J. et al., 2005</td>
</tr>
<tr>
<td>EntPl</td>
<td>5'-ATGAGAAGGAAAATTATTTAGTTT-3'</td>
<td>RAPD - PCR</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Fitzsimons et al. 1999</td>
</tr>
<tr>
<td>EntP2</td>
<td>5'-TGAATGGTCGCCCATTGCAACAAC-3'</td>
<td>RAPD - PCR</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Fitzsimons et al. 1999</td>
</tr>
</tbody>
</table>

### RAPD – PCR

RAPD was performed by using an already established within the research team protocol (Kirilov et al, 2011), in which the time of synthesis is gradually prolonged and the annealing temperatures raised in order to favor the formation of more diverse patterns. PCR amplification was carried out in a volume of 20 µl containing 0.4 U Taq polymerase in a KCl buffer system (Thermo Scientific), containing 1.5 mM MgCl₂, 0.2 mM dNTPs and about 50 ng of the template DNA. RAPD P2 primer concentrations was 0.40 µM. After an initial denaturation step for 5 min at 94 °C, 4 amplification cycles were performed consisting of denaturation step at 94 °C for 45 s, annealing step for 120 s at 30 °C and synthesis for 30 s at 72 °C, followed by 10 amplification cycles consisting of denaturation step at 94 °C for 5 s, annealing step for 30 s at 36 °C and synthesis for 30 s at 72 °C,10 amplification cycles consisting of denaturation step at 94 °C for 5 s, annealing step for 30 s at 36 °C and synthesis for 50 s at 72 °C, and finally 10 more cycles consisting of denaturation step at 94 °C for 5 s, annealing step for 30 s at 36 °C and synthesis for 60 s at 72 °C. A final extension step for 5 min at 72 °C completed the amplification. 8 µl of the PCR products were analyzed by electrophoresis on a 1.5 % agarose gel in TBE buffer system, followed by ethidium-bromide staining.

### UPGMA analyses

The software GeneTools v. 4.1 (Syngen) was used for the similarity matrices calculations on the basis of the electrophoretic gels images and the construction of the UPGMA dendrograms using the “profile” and “band position” options. Calculations were performed by means of the Jaccard algorithm (Jaccard, 1901) accounting not only for their positions but the peaks surfaces of the fluorescent bands as well.
RESULTS AND DISCUSSION

Screening for the presence of bacteriocin genetic determinants

The predominant thresholds for UPGMA clustering of clonally related bacterial strains described in the scientific publications usually ranges between 70% and 75% similarity (Webster et al., 1996; Grundmann et al., 1997). However, our previous experience shows that this threshold results in at least inconclusive results. This is why our aim was to check the reliability of the clustering using thresholds of 70%, 90% and 95% similarity. E. faecalis and E. faecium were chosen as test organisms because of their dualistic nature – GRAS (generally recognized as safe) status can be attributed to some strains of both species (Giraffa, 2003), while many strains are responsible for nosocomial infections (Sood et al., 2008). Purposefully for the experiments were selected non-pathogenic strains isolated from cheeses and clinical ones originating from different cities in Bulgaria. As an additional and independent criterion for forming subset groups for RAPD-PCR UPGMA analyses was used their potential to produce different bacteriocins accordingly to the presence of the respective genes.

The results from the screening for presence of bacteriocin genes are listed in Table 1. Among the collection of E. faecalis strains 24 strains revealed to be EntA positive but only one gave positive result for enterocin AS-48, so it was impossible to use this bacteriocin in our further analyses. Among the isolates from cheese only one possessed entA gene. More diverse results were obtained within the E. faecium collection. Among them all the three bacteriocin genetic determinants were found – 58 strains were entA+, 7 – entB+ and 3 – entP+. Among the strains isolated from cheese 14 were potential producers of EntA, 4 of EntB and 2 of EntP. An interesting detail is that all the three entP+ strains were positive also for the entA gene while this was valid for 5 of the 7 entB+ strains. No combined entP+ and entB+ strains were found.

Phylogenetic relatedness of the different potentially bacteriocinogenic groups of strains

Bacteriocin AS-48 was produced by only one E. faecalis clinical strain, so it was excluded as there were no counterparts for comparisons, and the UPGMA analysis of E. faecalis was performed only on the basis of the entA+ genotype. The UPGMA dendrogram of the entA+ E. faecalis strains is shown on Fig. 2. If 70% similarity threshold is applied, one small cluster comprising three strains from “St. Ivan Rilski” and two major clusters comprising strains from different cities are formed. An interesting element is that one of the major clusters comprises strains from the two hospitals in Sofia and the hospital in Gabrovo in the central part of Bulgaria, while the other one from the southern cities Blagoevgrad and Kurdzhali, as well the only cheese strain. One strain from the hospital in Gabrovo remains unclustered. If a threshold of 90% similarity is applied, one cluster of five strains and 7 small clusters of two strains are formed. In this case two more strains from Blagoevgrad and Sofia remain unclustered. The bigger cluster groups are composed only of strains from Sofia, while concerning the smaller ones, only two of them comprise strains from the same hospitals. Increasing the similarity threshold to 95% only slightly underline the tendency to form clusters from the same hospitals, while significantly increases the number of the unclustered strains up to 13. An intriguing observation is that the only cheese strain in this group shows more than 95% similarity to one of the Blagoevgrad’s strains.

The UPGMA analysis of the E. faecalis collection was with further limited applications concerning the study of the segregation of the dairy and the clinical isolates because it comprised only one cheese isolate therefore making impossible to speculate about a threshold dividing the clinical and the non-pathogenic strains of E. faecalis. However interesting results concerning the geographical origins of the clinical strains were obtained. It was found that good clustering corresponding to the geographic origins of the strains is achieved at a threshold of 90% similarity. Lowering the threshold down to 70% clustered together strains from different hospitals but an interesting tendency emerged – one of the clusters comprised strains from the hospital in cities in the central part of Bulgaria, while the other one from the southern part. This observation indicate that threshold of 70% to 75% indeed should be considered as indicating some clonal relatedness but with a limited extend. Leveling up the threshold of similarity up to 95% resulted in partial hospital-specific clustering, however the number of the unclustered strains increased significantly.

Because of the presence of enough strains of both types – clinical and dairy, the most integral results were obtained from the three UPGMA analyses performed within the E. faecium isolates. The dendrograms representing the UPGMA analyses of the entP+ and entB+ E. faecium strains are shown on Fig. 3 and Fig. 4. Both groups comprise only cheese-borne strains and strains from the Hospital “St. Ivan Rilski”, so conclusions concerning geographical origins cannot be made.

The UPGMA analysis of the entP+ strains can be examined only with a great caution because these strains are only three – one clinical and two from cheeses. However, the only clinical isolate segregated from the non-pathogenic ones, which formed a cluster, in a threshold of 70% similarity. Increasing the threshold up to 90% resulted of unclustering of all strains. More intriguing results were obtained from the UPGMA analysis of the second collection of entB+ E. faecium strains including 4 dairy and 3 clinical strains originating uniquely from only one hospital. Still at 70% similarity threshold two clusters formed – one comprising the three clinical strains, and another one comprising 3 of the 4 dairy strains (one dairy strain remained unclustered). Increasing the threshold up to 90% resulted in unclustering of the clinical strains while the cheeses’ cluster was conserved even at 95%. This observation was the first indicator that a higher threshold for clustering the food-borne strains should be applied in comparison to the clinical strains.

Logically the most informative results were obtained from the larger collection – the collection of entA+ E. faecium strains comprising 44 clinical and 14 dairy isolates where a tendency that a higher threshold for clustering the food-borne strains should be applied in comparison to the clinical strains was observed inside the collection of the entA+ E. faecium strains. The dendrogram obtained from the UPGMA analysis (Fig. 5) shows that if 70% similarity threshold is applied, two major clusters of 22 and 36 strains are formed, comprising strains from different sources – different cheeses and hospitals in different geographic locations. From the total of 58 strains, only two of them, one clinical and one cheese-borne, remain unclustered. The heterogeneity observed within the two clusters does not
allow segregation of the strains on the basis of their nature or geographic origins. Therefore, the use of 70% similarity threshold is obviously inappropriate when clonal relatedness strains of different types and originating from different locations is studied. At a threshold of 90% similarity the number of the clusters increases up to 8, comprising between 2 and 12 strains, and number of the unclustered strains increases to 4. At this level of discrimination slight tendency of conformity between the clusters and the geographic origins emerges, however the dairy strains segregate with the clinical ones. Therefore increasing the threshold up to 90% similarity also did not result in segregation on the basis of the type of the strains. However, a tendency within the clinical strains emerged in some of the clusters for segregation on the basis the geographical origins. Complete discrimination of the dairy strains and the clinical ones was achieved only when 95% similarity threshold was used. It led to the formation of three “cheese” clusters. At this level the tendency for geographical segregation of the clinical isolates were further pronounced leading to the formation four “pure” hospital clusters – those of Gabrovo hospital, Military Medical Academy, “St. Ivan Rilski” hospital, and Varna hospital. Three additional predominantly geographical clusters were also formed – one dominated by strains from the two hospitals in Sofia and two dominated by isolates from the Varna hospital. Only one “mixed” cluster of clinical strains comprising different cities was formed at this threshold. These observations led to the conclusion that within the entA E. faecium strains 95% similarity threshold was more appropriate for the clustering of potentially clonally related strains, as well for the segregation of the dairy isolates from the clinical ones.

**CONCLUSION**

The combined investigations of the four RAPD-PCR UPGMA analyses of both Enterococcus species used in our work led us to several conclusions: 1. Using 70% similarity threshold is both inappropriate for strains clustering and discrimination on the basis of the type (clinical or dairy), as well upon the geographical origins; 2. In many cases a 90% similarity threshold resulted at least in geographical if not clinical segregation of the nosocomial enterococci but was not enough for separation of the dairy and clinical isolates; 3. Complete segregation of the isolates from cheeses and the clinical ones was achieved only 95% similarity threshold was applied, however some of the hospitals’ or cities’ clusters were cracked; 4. The higher threshold of 95% needed for clustering the dairy isolates should probably be an indication for the separate evolutionary lineage of the non-pathogenic enterococci, showing greater relatedness among them.

**Acknowledgments:** This work was supported by a grant from the Medical University of Sofia (Council of Medical Science, project no. 6/2014, grant no. 11/2014). The authors thank Prof. E. Savov (Military Medical Academy, Sofia), Assoc. Prof. K. Bojkova (UMHAT “St. Marina”; Varna), Dr. A. Katrandzhieva (MHT “Dr. Tota Venkova”, Gabrovo), Dr. S. Raycheva (PMDL “Mikro-2”, Kurdzhalii) and Dr. I. Ivanova (MDL “Medilab”, Blagoevgrad) for their assistance in collection of clinical Enterococcus spp. isolates.

**REFERENCES**


Figure 5 UPGMA dendrogram of the entA E. faecium strains. The different clusters formed at the different similarity thresholds are indicated with brackets.