Comparison of MALDI-TOF mass spectrometry and 16S rDNA sequencing for identification of environmental bacteria: a case study of cave mussel-associated culturable microorganisms

Ana Bielen[1,*], Ivana Babić[2], Marija Vuk Surjan[3], Snježana Kazazić[2], Ana Šimatović[4], Jasna Lajtner[3], Nikolina Udiković-Kolić[2], Zrinka Mesić[5], Sandra Hudina[3]

[1]Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, Pierottijeva 6, 10000 Zagreb, Croatia

[2]Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia

[3]Department of Biology, Faculty of Science, University of Zagreb, Horvatovac 102a, Zagreb, Croatia

[4]Laboratory of Microbial Biotechnology, Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy

[5]Oikon Ltd., Zagreb, Trg Senjskih uskoka 1-2, 10020 Zagreb, Croatia

*Corresponding author: Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, Pierottijeva 6, 10000 Zagreb, Croatia. Tel: +385 1 4836 013; Fax: +385 1 4836 016; E-mail: abielen@pbf.hr

Abstract

Matrix Assisted Laser Desorption Ionization - Time of Flight Mass Spectrometry (MALDI-TOF MS) is routinely used as a rapid and cost-effective method for pathogen identification in clinical settings. In comparison, its performance in other microbiological fields, such as environmental microbiology, is still being tested, although isolates of environmental microbes are essential for in-depth in vivo studies of their biology, including biotechnological applications. We investigated the applicability of MALDI-TOF MS for the identification of bacterial isolates from a highly oligotrophic environment - Dinaric Karst caves, which likely harbor specific microorganisms. We cultured bacteria from the shell surface of the endemic mussel *Congeria jalzici*, one of the three known cave mussel in the world that lives in the Dinaric karst underground. The bacterial isolates were obtained by swabbing the shell surface of mussels living in microhabitats with different amounts of water: 10 air-exposed mussels, 10 submerged mussels and 10 mussels in the hygropetric zone. A collection of 87 pure culture isolates was obtained, mostly belonging to the phylum Bacillota (72%), followed by Pseudomonadota (16%), Actinomycetota (11%) and Bacteroidota (1%). We compared the results of MALDI-TOF MS identification (Bruker databases DB-5989 and version 11, v11) with the results of 16S rDNA-based phylogenetic analysis, a standard procedure for bacterial identification. Identification to the genus level based on 16S rDNA was possible for all isolates and clearly outperformed the results from MALDI-TOF MS, although the updated MALDI-TOF MS database v11 gave better results than the DB-5989 version (85% versus 62%). However, identification to the species-level identification by 16S rDNA sequencing was achieved for only 17% of isolates, compared with 14% and 40% for the MALDI-TOF MS databases DB-5989 and v11 database, respectively. In conclusion, our results suggest that continued enrichment of MALDI-TOF MS libraries will result with this method soon becoming a rapid, accurate, and efficient tool for assessing the diversity of culturable bacteria from different environmental niches.

Keywords: bacterial isolates; biofilm; *Congeria jalzici*; Dinaric karst underground; MALDI-TOF MS; oligotrophic environment

Introduction

Matrix-Assisted Laser Desorption Ionisation - Time of Flight Mass Spectrometry (MALDI-TOF MS) is a method for rapid identification of bacterial isolates by analysing the mass spectra of proteins from whole bacterial cells. Such protein fingerprints are species-specific and allow identification of bacterial species by comparison with reference spectra from commercially available databases (Lavigne et al. 2013; Singhal et al. 2015). It has many advantages, such as high accuracy of identification, low cost, rapid results and no need for highly skilled laboratory personnel (Angeletti 2017; Torres-Sangiao et al. 2021). The entire MALDI-TOF MS protocol takes about 30 minutes, which is significantly shorter than 2-3 days required for traditional phenotypic testing or 16S rDNA sequencing combined with phylogenetic analysis (He et al. 2010). In the clinical setting, MALDI-TOF MS has repeatedly been shown to provide identification results comparable to or even more accurate than other identification methods (Mellmann et al. 2008; He et al. 2010; Zhu et al. 2011; Barberis et al. 2014; Rossel et al. 2019). This is understandable as the reference databases contain the vast majority of clinically important species (Hou et al. 2019). Therefore, MALDI-TOF MS is currently being widely used to identify human pathogenic bacteria and fungi and is gradually replacing traditional methods, as it can accurately identify 90% of relevant bacterial and yeast isolates, including antibiotic-resistant bacteria and fungi that pose a global public health threat (Wilson et al. 2017; Elbehiry et al. 2022).

However, the use of MALDI-TOF MS to identify bacteria in other areas of microbiology, such as environmental microbiology, lags behind and is hampered by inadequate spectral databases with respect to a variety of culturable environmental microbes (Santos et al. 2016; Ashfaq et al. 2022). Here, culture-independent methods have been routinely and widely used for decades for microbial community analysis (Su et al. 2012), but pure cultures are still needed as environmental microorganisms are an immense resource for many biotechnological applications and are essential for in-depth *in vivo* studies of their biology, including epizootiology in the case of pathogens (Austin 2017; Bodor et al. 2020). As the rapid identification of new isolates is very important in this context, MALDI-TOF MS should be considered as a promising tool for the routine identification of bacteria (Ashfaq et al. 2022). To date, there are a number of studies using MALDI-TOF MS for the analysis of environmental isolates from different habitats (e.g. Dieckmann et al. 2005; Uhlik et al. 2011; Dybwad et al. 2012; Koubek et al. 2012; Viver et al. 2015; Emami et al. 2016, Brauge et al. 2021), including the identification of culturable microbes from acidic wastewater (Kopcakova et al. 2014), Arctic seawater (Timperio et al. 2017), copper mines (Avanzi et al. 2017), soil samples from the high altitude Indian Himalayan region (Pandey et al. 2019) and caves (Mulec et al. 2015; Mudgil et al. 2022). However, a significant proportion of the collected isolates could not be identified or were misidentified by MALDI-TOF MS due to limited coverage of environmental bacteria in the MALDI-TOF MS spectral database (Emami et al. 2016; Timperio et al. 2017; Brauge et al. 2021). Several studies conducted over the past decade have

compared MALDI-TOF MS identification of environmental isolates from a variety of environmental niches with 16S rDNA-based analysis (Emami et al. 2012; Böhme et al. 2013; Fykse et al. 2015; Assis et al. 2017; Avanzi et al. 2017; Timperio et al. 2017; Strejcek et al. 2018; Kačániová et al. 2019; Pandey et al. 2019; Pomastowski et al. 2019; Brauge et al. 2021). The reported concordances between the two identification approaches varied widely and were often low. In addition, the comparative studies mentioned above have analysed bacteria isolated from a limited number of environmental niches, mostly from nutrient-rich habitats such as seawater (Emami et al. 2012; Timperio et al. 2017; Brauge et al. 2021), soil (Strejcek et al. 2018; Pandey et al. 2019) and host tissues that are important as food sources (Böhme et al. 2013; Assis et al. 2017; Kačániová et al. 2019; Brauge et al. 2021). In addition, only one version of the MALDI-TOF MS spectral database was used per study, making it impossible to quantify the putative improvements in MALDI-TOF MS identification of environmental isolates over time. All this suggests that further analyses of MALDI-TOF MS as a tool for rapid identification of environmental isolates are needed, especially those covering previously unanalysed habitats.

The aim of this work was therefore to compare the performance of MALDI-TOF MS with standard DNA-based bacterial identification, i.e. phylogenetic analysis of the 16S rRNA gene, in a specific, previously unexplored, environmental setting. We focused on the Dinaric karst underground, a highly specific environment which will likely harbour unique environmental microorganisms. Dinaric caves are characterized by stable microclimatic conditions. Regardless of the season, relative humidity is high, usually between 85 and 100%, while air temperature corresponds to the surface mean annual air temperature of the area (Matočec et al. 2002; Surić et al. 2020). Furthermore, caves are challenging and extreme environments, mainly because they are highly oligotrophic, lack primary production and depend on limited nutrient input brought in with percolating water (Kováč 2018). Finally, the Dinaric karst underground is considered a hotspot of invertebrate biodiversity with one of the most complex and diverse subterranean faunas in the world, including many endemic and relict species (Culver and Sket 2000; Sket 2005; Ozimec et al. 2009). While most biospeleological studies of this hotspot focus on invertebrates (Bilandžija et al. 2013; Deharveng and Bedos 2018; Bedek et al. 2019; Kozel et al. 2020), studies on the microbial communities of the Dinaric karst are scarce and mainly describe limestone-associated biofilms (Mulec 2008; Pašić et al. 2010; Velikonja et al. 2014; Mulec et al. 2015). Animal-associated microbial communities of the Dinaric karst are even less studied. There is only one recent study describing the bacterial communities on the skin of *Proteus anguinus*, an endemic cave salamander (Kostanjšek et al. 2019).

In this study, we provide the first insight into the culturable bacteria from the shell surface of *Congeria jalzici*, a unique endemic and critically endangered cave-dwelling bivalve (Bilandžija et al. 2013). Unlike other cave invertebrates, species of the genus Congeria form dense colonies (Jovanović Glavaš et al. 2017), that make a very specific microenvironment in karst cave ecosystems. We used the collected Congeria-associated bacterial isolates to analyze the applicability of MALDI-TOF MS

as a rapid and accurate method for identifying bacteria from a highly specific environment. In particular, we compared the success of MALDI-TOF MS identification using two versions of the Bruker database with the results of 16S rDNA-based phylogenetic analysis, a standard approach for bacterial identification.

Materials and methods

Sampling and cultivation

Swab samples of biofilm communities from the surface of *C. jalzici* shells were taken in Markov Ponor (Figure 1) in October 2016. Markov Ponor is one of the few caves with living colonies of *C. jalzic*i due to which it was included in Natura 2000 network of protected areas. During sampling, the water parameters were: temperature = 18.3 °C, pH = 7.78, conductivity = 351 µS/cm, biological oxygen demand measured after 5 days at 20 °C (BOD_5) < 1 mg O_2/L, ortophosphates < 0.03 mg P/L, nitrites < 0.02 mg N/L, nitrates < 1.13 mg N/L, total dissolved solids (TDS) = 168 mg/L, hardness = 235 mg $CaCO_3$/L, dissolved oxygen = 5.03 mg/L, oxygen saturation = 55.1%, dissolved organic carbon (DOC) < 1 mg/L, total organic carbon (TOC) < 1 mg/L.

Figure 1. The location of the sampling site Markov Ponor (red dot) within the map of the Dinaric karst (in grey). Adapted from Bilandžija *et al.* (2013).

Biofilm samples were collected from surfaces (shells) of mussels living in three microhabitats within Markov ponor with different amount of water: (i) 10 air-exposed mussels, found *in situ* above the water level and abbreviated throughout the manuscript as (**A**); (ii) 10 submerged mussels (**S**), and (iii) 10 mussels in hygropetric area (**H**), characterized by a thin layer of water flowing over the rock surface. Such sampling procedure was performed in order to collect greater variety of isolates, since each microhabitat might be dominated by different bacterial taxa (Zhu et al., 2019). One swab sample was taken per animal using sterile swab sticks. Swabs

were put on ice and transported to the laboratory where microorganisms from each swab were immediately inoculated on Tryptic Soy Agar (TSA, Biolife, Italy), a non-selective culture medium. Following incubation at 21 °C for seven days, up to 10 morphologically distinct single bacterial colonies per individual swab/mussel were selected, purified by repeated streaking to obtain individual pure culture isolates, and then stored in 20% glycerol (GRAM-MOL, Croatia) at -80 °C until further use.

Identification of isolates by sequencing and phylogenetic analysis of 16S rRNA gene fragments

Genomic DNA was extracted from isolates using a NucleoSpin® Microbial DNA kit (Macherey Nagel, Germany) following the manufacturer´s instructions. Almost full-length 16S rDNA sequence (1465 bp, variable regions V1 to V9) of each isolate was amplified by PCR using universal bacterial primers 27F and 1492R (Lane 1991). Reaction mixture contained 1 µL of DNA solution, 12.5 µL of EmeraldAmp® PCR 2X Master Mix (TAKARA), 0.5 µL of 10 µM forward primer 27F, 0.5 µL of 10 µM reverse primer 1492R and 9 µL of dH_2O. Thermal cycling was performed in Alpha Cycler 1 (PCRmax, United Kingdom) as follows: 5 min at 95 °C, followed by 28 cycles of 1 min at 94 °C, 1 min at 55 °C and 1.5 min at 72 °C, and 10 min at 72 °C. PCR products were analyzed by gel electrophoresis to confirm the appropriate PCR product size and then purified and Sanger sequenced using the same primers by a commercial service (Macrogen, Inc., Amsterdam, Netherlands). The resulting 16S rDNA sequences were assembled and analyzed using GeneStudio (GeneStudio, Inc.). 16S rRNA gene sequences were deposited in GenBank and accession numbers are listed in Supplementary table S1.

The taxonomic identity of the cave isolates was determined by phylogenetic analyses of 16S rRNA gene fragments. Firstly, Blastn (megablast algorithm) was used to compare the sequences of the isolates to NCBI 16S ribosomal RNA sequences database (Bacteria and Archea) and preliminarly determine the genus of each isolate. Then, all 16S sequences from the respective genera were downloaded from https://lpsn.dsmz.de/, an expert-curated database for prokaryotic nomenclature (Parte et al. 2020; Meier-Kolthoff et al. 2022), to serve as a complete set of referent sequences in the subsequent phylogenetic analyses. Next, multiple sequence alignments (MSA) of 16S rDNA sequences of the cave isolates and referent sequences were constructed in MAFFT (Katoh and Standley 2013). MSAs were constructed separately for phyla Actinomycetota, Bacillota, Bacteroidota and Pseudomonadota, and edited in SeaView (Gouy et al. 2010) and BioEdit (Hall 2011). Maximum likelihood phylogenetic trees based on the Kimura 2-parameter model (Kimura 1980) were constructed using MEGA v.7 software (Kumar et al. 2016). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, +G). The rate variation model allowed for some sites to be evolutionarily invariable (+I). Positions containing gaps and missing data were eliminated. Branch lengths correspond to the number of substitutions per site, and the reliability of the internal branches was assessed by bootstrapping with 1000 replicates. Phylogenetic trees were edited in iTOL web app (Letunic and Bork 2019). Finally, isolates were identified to species level if they were clustered with only one

reference sequence. If they were grouped with more than one reference sequence, they were identified to the genus level, as it was not possible to distinguish between two or more closely related species.

Identification of isolates by matrix assisted laser desorption ionisation – time of flight mass spectrometry (MALDI-TOF MS)

Fresh single colonies of each isolate were cultured on TSA plates at 22 °C for one to three days, depending on the time needed to obtain visible colonies. Next, the collected biomass was analyzed by MALDI-TOF MS. Sample preparation was performed according to the Bruker protocol using ethanol/formic acid extraction (Topić Popović et al. 2015).

Briefly, fresh colony was suspended in 300 µL of deionized water followed by addition of 900 µL of absolute ethanol (Kemika, Croatia), centrifuged at 13,000 rpm for 2 min, and the supernatant was discarded. The centrifugation was repeated, and the residual ethanol was discarded. The pellet was air dried and dissolved in 20 µL each of 70% formic acid (Fisher Scientific, Spain) and acetonitrile (Fisher Scientific, UK). After centrifugation at 13,000 rpm for 2 min, 1 µL of the supernatant was transferred to a polished steel MSP 96 target (Bruker Daltonik, Germany) and allowed to air dry at room temperature. Matrix solution of 10 mg/mL α-cyano-4-hydroxy-cinnamic acid (HCCA, Bruker Daltonics, Germany) in 50% acetonitrile and 2.5% trifluoroacetic acid (Sigma-Aldrich, Germany) was prepared fresh. The sample spot on the steel target was then overlaid with 1 µL matrix solution and dried again. Measurements were performed using a Microflex LT mass spectrometer (Bruker Daltonics, Germany), and spectra were recorded in positive linear mode within a mass range of 2000 to 20000 Da. External calibration was performed using the Bacterial Test Standard (Bruker Daltonics, Germany).

Recorded mass spectra were processed using two versions of the MALDI Biotyper software package and database, and the standard settings: (i) MALDI Biotyper 3.0 software package and Bruker database DB-5989, (ii) MALDI Biotyper Compass Explorer 4.1 software package and Bruker database version 11 (v11). Each isolate obtained a logarithm score between 0 and 3. According to the manufacturer's recommendations, scores ≥ 2.300 are recognized as highly probable species identification, scores ranging from 2.000 to 2.299 as a secure genus identification and probable species identification, scores ranging from 1.700 to 1.999 as a probable genus identification, and scores < 1.700 as not reliable identification. Here, we assigned a species-level identification to all isolates with a score ≥ 2.000 and a genus-level identification to all isolates with a score ≥ 1.700.

Results

Taxonomic identification of isolates was firstly performed by phylogenetic analyses where 16S rDNA sequences of the isolates were clustered with referent sequences collected from an expert-curated LPSN database for prokaryotic nomenclature (Parte et al. 2020; Meier-Kolthoff et al. 2022) (Supplementary Figures S1 to S4). We collected 87 isolates, represented with 47 unique 16S rDNA sequences. 16S rDNA analyses showed that the majority of isolates belonged to the phylum Bacillota (72%), followed by Pseudomonadota (16%), Actinomycetota (11%) and Bacteroidota (1%) (Supplementary Tables S1 and S2, Supplementary Figures S1 to S4). The most abundant genera were gram-positive *Bacillus* (Supplementary Figure S5i, j, k and l) with 39 isolates and gram-negative *Pseudomonas* (Supplementary Figure S5b) with 13 isolates. Some of the genera, i.e. *Bacillus*, *Paenibacillus*, *Rhodococcus*, were found in all three analyzed microhabitats (i.e. air exposed, hygropetric, submerged), but the majority were more or less microhabitat-specific (Supplementary Table S2).

In overall, based on the clustering of 16S rDNA sequences of the cave isolates with a complete set of referent sequences from the respective genera (Supplementary Figures S1 to S4), all 87 isolates were identified to the genus level and 17% (15/87) to the species level (Figure 2, Supplementary Table S1). Further, using the best blast hit searches we obtained sequences from the database with high, 99 - 100% identity to our queries. However, we often obtained multiple best blast hits to the same query (32/87, 37%), making it hard to accurately assign the species using only best blast hit results (see Supplementary Table S1). In fact, even for the 55 queries that yielded a single best blast hit, the phylogenetic analysis could not confirm it as the correct species identification (48/55, 87%).

Next, we performed MALDI-TOF MS identification of the isolates and compared them with 16S rDNA-based identification Figure 2, Supplementary Table S1). The recorded MALDI-TOF protein spectra were matched against two different versions of the Bruker databases (DB-5989 and v11) to determine if the identification accuracy of MALDI-TOF increased with updates to the database.

MALDI-TOF MS identification results obtained with the database DB-5989 somewhat underperformed 16S rDNA-based identification: 17% of isolates were identified to species level by 16S rDNA sequencing, whereas in the case of MALDI-TOF MS species identification was successful for 14% of isolates (Figure 2). Next, 100% of isolates were identified to the genus level by 16S rDNA analysis (correct identification to species + to genus level), compared to 62% for MALDI-TOF MS identification. Finally, 38% (33/87) of isolates could not be reliably identified to either the genus or species level by MALDI-TOF MS. This was not the case with 16S rDNA sequencing, where all isolates were successfully identified at least to the genus level. Although 15% of the isolates could still not be assigned to either a genus or a species using the updated version of the MALDI-TOF MS database (v11), identification results were significantly improved (Figure 2). Namely, 40% of the isolates could be identified to

species level (compared to 19.5% for 16S rDNA analysis) and 85% to genus level (100% for 16S rDNA analysis).

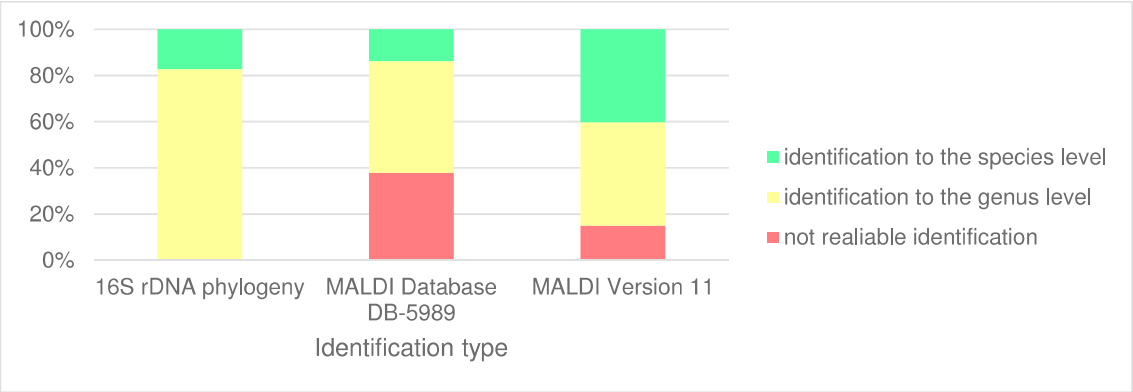
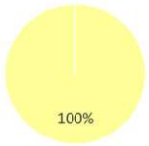
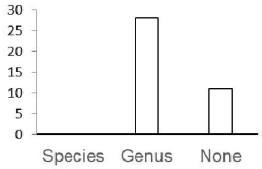
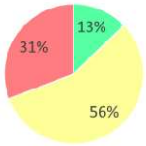
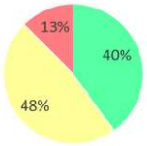
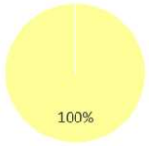
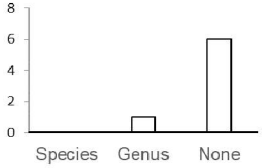
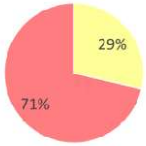
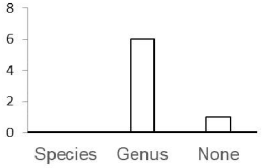
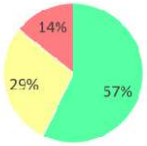
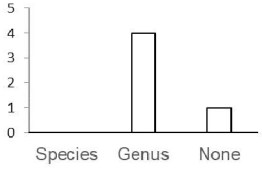
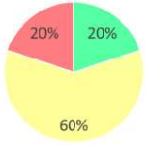
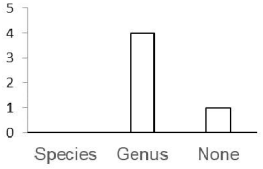
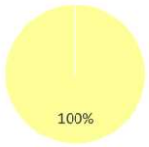
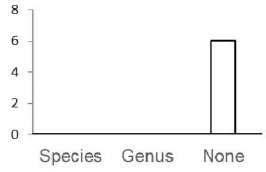
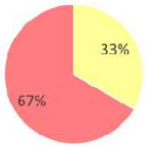
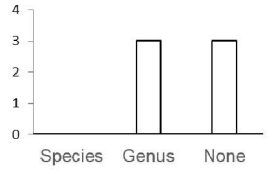
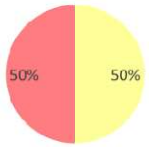
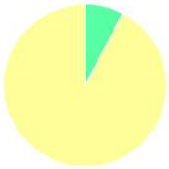
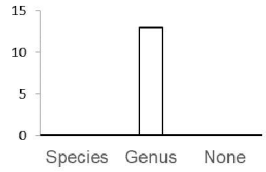
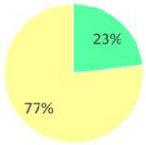
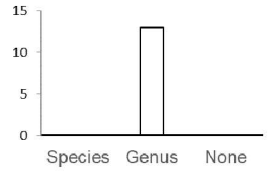
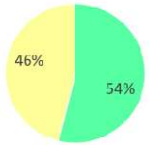


Figure 2. Overview of MALDI-TOF MS identification – Database DB-5989 and Database V11 of *Congeria*-associated bacterial isolates in comparison with 16S rDNA-based identification.

We also examined in detail the identification success of the most common genera in our collection (Table 1), in particular *Pseudomonas* and *Bacillus*. For the *Pseudomonas* isolates, phylogenetic analysis of 16S rDNA resulted in 100% genus identification, but only one isolate (1/13, 8%) could be identified to species (Table 1, Supplementary Figure S4), *Pseudomonas capeferrum* isolate H.8. MALDI-TOF MS genus identification was also 100% accurate, and identification to species level outperformed the 16S rDNA phylogenetic analysis: 3/13 (23%) for Database DB-5989 and 7/13 (54%) for Database V11. Further, not a single *Bacillus* isolate could be identified to the species by 16S rDNA phylogenetic analysis. In comparison, MALDI-TOF MS Database DB-5989, successfully identified 5/39 (13%) *Bacillus* isolates to species, 23/39 (59%) to genus, and 11/39 (28%) were not identified. With the new database v11, 15/39 *Bacillus* isolates (38%) could be identified to species level and 18/39 (46%) to genus level, while others could not be correctly classified (6/39, 15%).

Table 1. Identification of *Congeria*-associated bacterial isolates from different genera by two methods: 16S rDNA phylogenetic analyses and MALDI-TOF MS (Database DB-5989 and Database version 11). Only the genera containing five or more isolates in our collection are presented. Identification success of different methods is color-coded: green – percentage of isolated identified to the species level; yellow – percentage of isolates identified to the genus level; red – percentage of isolates that could not be reliably identified. Concordance between methods at species or genus level is also presented.

Genus	No. of isolates	16S rDNA identification	Concordance 16S vs DB-5989	MALDI-TOF MS identification DB-5989	Concordance 16S vs v11	MALDI-TOF MS identification v11
Bacillus	39					
Lysinibacillus	7					
Paenibacillus	5					

Peribacillus	6					
Pseudomonas	13					

314

Finally, we analyzed the level of corcodance between the results of the two identification methods (Figure 3, Tables 1 and 2). The results of the 16S rDNA-based identification compared to the MALDI-TOF MS database DB-5989 were in agreement for more than half of the cave isolates (59%), i.e., for 4% of the isolates at the species-level and for 55% of the isolates at the genus-level level. The MALDI-TOF MS database v11 results matched the 16S rDNA results for 79% of the isolates, i.e., 6% at the species level and 73% at the genus level. In addition, we compared our results to a number of other studies that have used both MALDI-TOF MS and 16S rDNA-based identification in parallel to identify isolates from a variety of environmental niches ranging from aquatic to terrestrial to aerial environments, including some extreme environments such as copper mines. Reported genus level concordance percentages ranged from 57% to 100%, depending on the study, and were mostly higher than the concordance obtained here (Table 2). Species-level concordance could not be compared because the different studies used different 16S rDNA-based identification approaches, ranging from best blast hits to phylogeny-based analyzes with different sets of reference sequences.

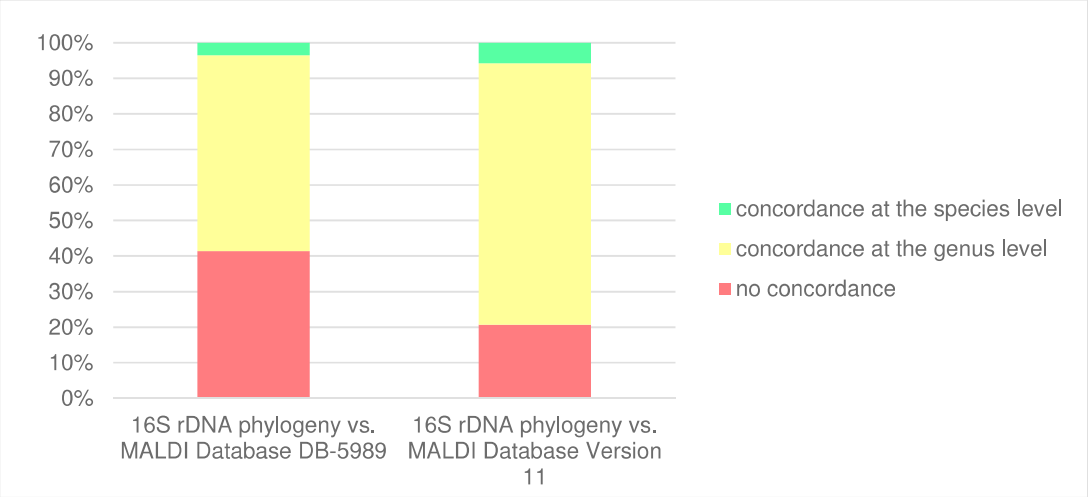


Figure 3. Concordance between MALDI-TOF MS identification (Database DB-5989 and Database V11) and 16S rDNA-based identification of *Congeria*-associated bacterial isolates.

336 Table 2. Comparison of 16S rDNA sequencing and MALDI-TOF MS identification
337 success of environmental bacteria from different habitats.

Source	**Number of isolates**	**Dominant genera**	**Concordance between 16S rDNA sequencing and MALDI-TOF MS identification on the genus level (%)**	**Study**
Ballast waters (marine)	39	*Bacillus*, *Pseudomonas*	92	Emami et al. 2012
Seafood	50	*Pseudomonas*, *Bacillus*	96	Böhme et al. 2013
Air	103	*Bacillus*, *Staphylococcus*	87	Fykse et al. 2015
Freshwater fish farms (nile tilapia)	131	*Streptococcus*, *Lactococcus*	83	Assis et al. 2017
Copper mine	72	*Pseudomonas*, *Enterobacter*	82	Avanzi et al. 2017
Arctic sea water	45	*Pseudomonas*, *Serratia*	100	Timperio et al. 2017
Soil, sediment	86	*Pseudomonas*	95	Strejcek et al. 2018
High altitude soil	61	*Bacillus*, *Pseudomonas*	57	Pandey et al. 2019
Freshwater fish	15	*Pseudomonas*	100	Kačániová et al. 2019
Honey	38	*Bacillus*	92	Pomastowski et al. 2019
Seafood, sea water	713	*Bacillus*, *Psychrobacter*, *Staphylococcus*	86	Brauge et al. 2021
Karst cave	87	*Bacillus*, *Pseudomonas*	59	This study, MALDI-TOF MS Database DB-5989
			79	This study, MALDI-TOF MS Database V11

338

339

Discussion

We present the first insight into the invertebrate-associated bacterial isolates in Dinaric karst as well as the first report on the culturable bacteria associated with the shells of endemic and endangered *Congeria* mussels. We have used this collection to analyze the applicability of MALDI-TOF MS as an inexpensive and fast tool for identification of environmental bacteria. To our knowledge, this is the first study that identified the environmental bacterial isolates using two different updates to the Bruker libraries. Our results suggest that the continued expansion of MALDI-TOF MS databases will soon make the method an effective mean of identifying bacterial isolates from various environments outside the usual medical setting. Rapid and reliable identification of bacterial isolates from the environment is a prerequisite for their in-depth analysis, ranging from pathogenicity studies to biotechnological applications.

The bank of isolates presented here is representative of a very specific environment: on the one hand, a very stable habitat characterised by the absence of light, relatively constant temperature and high humidity, and on the other, drastic annual fluctuations in the water level (Bonacci et al. 2009; Jovanović Glavaš et al. 2017). Moreover, *Congeria* colonies, often comprising hundreds of individuals per m^2, represent a specific microhabitat in the Dinaric karst underground, i.e. a peak of relatively dense biomass (Jovanović Glavaš et al. 2017), that represents a potential hotspot of bacterial diversity in this extremely oligotrophic environment. All this confirms the ecological and microbiological uniqueness of the habitat we have selected for testing the MALDI-TOF MS performance.

All of our isolates were reliably identified to the genus level by 16S rDNA. However, resolution to the species level was poor, i.e. 83% of isolates could not be undoubtedly identified to the species level by phylogenetic analysis of nearly complete 16S rRNA gene fragments since they clustered with more than one reference sequence from the expert-curated LPSN database (Meier-Kolthoff et al. 2022). In general, the advantages of 16S rDNA analysis that make it a widely used tool for prokaryote identification are (i) 16S rRNA gene is universally found in all bacteria and archaea, (ii) it contains highly conserved regions that can be used as primer targets for amplification and subsequent sequencing, (iii) it contains hypervariable regions that can be analyzed to distinguish and identify different taxa, and (iv) the enormous and ever-growing number of 16S rDNA sequences in databases (GenBank at NCBI, Ribosomal Database Project, RDP, GreenGenes, SILVA, etc.) is available and can be used as a reference (Federhen 2012; McDonald et al. 2012; Cole et al. 2014; Yilmaz et al. 2014; Balvočiūtė et al. 2017). For example, the “Bacteria and Archaea 16S ribosomal RNA database” at NCBI currently contains over 20,000 curated entries corresponding to bacterial and archaeal type strains. However, 16S rDNA-based identification approach has serious disadvantages. It is time-consuming, especially when sequencing is followed by computationally intensive phylogenetic analyses. Also, it often does not provide resolution down to species

level, as visible also from the phylogenetic analyses conducted in the scope of this study. Closely related prokaryotic species often share high, sometimes even 100% identity, in 16S rRNA sequence (Vos et al. 2012; Seuylemezian et al. 2018). Further, it has been shown that sequence databases (such as the widely used NCBI) contain a (not negligible) portion of misidentified sequences and even genomes (Beaz-Hidalgo et al. 2015; Jin et al. 2020; Morimoto et al. 2020), that can lead to further misclassifications. However, due to the omnipresence of 16S rRNA analysis, attempts are often made to use it for species identification. Firstly, best blast hits (with sequence identities ≥ 97 or 99%) are often reported and mistakenly considered as unambiguous species identifications (e.g. Emami et al. 2012; Böhme et al. 2013; Assis et al. 2017; Avanzi et al. 2017; Pandey et al. 2019). Alternatively, if 16S rDNA phylogenetic analyses are performed, an incomplete set of referent species is often used, which can again lead to an inaccurate assignment of species (e.g. Timperio et al. 2017).

Regarding the identification of isolates in our collection by the MALDI-TOF MS approach, we tested two different updates of the database and compared their identification success with 16S rDNA-based analysis. Our results suggest that with the expansion of the Bruker database, there is a trend toward improved identification of environmental bacteria by MALDI-TOF MS, although it still lags behind 16S rDNA analysis in the percentage of isolates for which even the genus cannot be accurately determined. This is due to the fact that, in comparison to genetic procedures for identification of bacteria, MALDI-TOF MS relies on much smaller referent databases that lack the bacterial diversity found in nature (Mulec et al. 2015; Topić Popović et al. 2016). The latest Bruker database v11 has 10,833 entries in total and covers 3,893 species of 664 microorganism genera. Nevertheless, MALDI-TOF MS has been increasingly used in environmental microbiology, due to its attractive properties such as high speed and low cost (see for instance Mulec et al. 2015; Kittinger et al. 2016; Topić Popović et al. 2016; Avanzi et al. 2017; Timperio et al. 2017; Seuylemezian et al. 2018; Strejcek et al. 2018; Pomastowski et al. 2019; Ashfaq et al. 2022).

When inspecting the identification of environmental isolates to the genus level, 16S rDNA analysis is highly reliable, while MALDI-TOF MS identification gives variable results (Table 2 and references therein). Expectedly, the overall MALDI-TOF MS identification performance for environmental isolates (i.e. between 57% to 100% at a genus level) is poorer than previously reported for clinically important bacteria where correct genus identification was mostly above 90% (Mellmann et al. 2008; Bizzini et al. 2011; Schmitt et al. 2013; Rodríguez-Sánchez et al. 2016; Schulthess et al. 2016; Alcalá et al. 2021). This reflects the underrepresentation of environmental bacteria in MALDI-TOF MS libraries (Popović et al. 2017). In addition, we have observed variable success in identifying genera in different environmental studies. This is most likely due to a combination of factors, such as different environmental niches (inhabited by bacteria not consistently represented in MALDI-TOF MS databases), different versions of the Bruker database used, and different colony preparation

protocols. Compared with other studies (Table 2), MALDI-TOF MS database DB-5989 performed the worst in identifying *Congeria*-associated isolates, i.e., genus-level concordance with 16S rDNA-based identification was the lowest of all studies examined, indicating the poor representation of our isolates in the Bruker library. This also indicates the uniqueness of the studied habitat and its microbiota with a high representation of rare and specific bacterial species. It has been previously shown that different niches in karst cave systems contain unique niche-specific bacterial assemblages (Zhu et al. 2019). However, when the MALDI-TOF MS database was updated to v11, the identification results improved significantly, and the concordance level with 16S rDNA-based identification was now comparable with other studies (Table 2). The observed improvement in the identification capabilities of MALDI-TOF MS with v11 of the database suggests that this method is gaining importance for routine identification of environmental bacteria (Ashfaq et al. 2022).

The improvement of the Bruker libraries was further confirmed by the observed increase in the number of isolates that obtained high identification scores and could be identified to the species level (probable species identification) (Figure 2). If a matching bacterial species is present in the database and is well annotated, MALDI-TOF MS has the potential for more accurate species-level identification than the 16S rRNA gene, since MALDI-TOF MS approach is based on whole proteome analysis. The environmental isolates from the genera *Pseudomonas* and *Bacillus*, abundant in our collection and elsewhere (Böhme et al. 2013; Avanzi et al. 2017; Timperio et al. 2017; Strejcek et al. 2018; Pomastowski et al. 2019; Samad et al. 2020; Brauge et al. 2021), were often reported to be difficult for identification to the species level by MALDI-TOF MS, partly because of their difficult phylogeny (Gomila et al. 2015; Wang and Ash 2015; Lalucat et al. 2022) and partly because of the limitations of the Bruker library. In this study, the species-level identification of the isolates from these genera was improved when the newer version of the database was used, and the identities obtained were in line with 16S rDNA results despite the fact that none of our isolates obtained a score > 2.300 (highly probable species identification).

Ultimately, accurate identification using both MALDI-TOF MS and 16S rDNA depends upon the precise classification of reference taxa. Nonetheless, the taxonomy of many bacterial genera undergoes constant refinement and correction due to the ongoing discovery of novel species and the increasing number of whole genome sequences (see for instance Huang et al. 2019). Consequently, databases, despite undergoing continuous updates and curation, inevitably harbor inaccuracies and inconsistencies, particularly within large and complicated genera such as *Pseudomonas* and *Bacillus*. In our study, this caused some discrepancies in identifications by two approaches. For instance, isolate H.8 was identified as *Pseudomonas putida* by MALDI-TOF MS Database V11, and as *P. capefferum* by 16S rDNA sequencing. Its 16S rRNA sequence was highly similar to that of *P. capeferrum* WCS358 isolate that belongs to the *P. putida* species group (Berendsen et al. 2015) and was even classified as *P. putida* for decades (see for instance Marugg et al., 1985; Lemanceau et al., 1992; Duijff et al., 1999, Meziane et al. 2005), before being reclassified as *P. capeferrum*

based on a whole genome analyses (Berendsen et al. 2015). In addition, three isolates (S.10, S.12 and S.13) identified as *Bacillus* sp. by 16S rDNA were classified by MALDI-TOF MS as members of *Niallia* sp., a new genus that until recently was classified as *Bacillus* (Gupta et al. 2020). In line with this, both MALDI-TOF MS and 16S rDNA-based identification should be used only as a first step in the identification of environmental isolates, while in-depth phylogenetic analysis is needed for precise species identification. For example, it has been proposed to use a combination of housekeeping genes, often based on whole genome sequences, to distinguish between related bacterial isolates of *Pseudomonas*, *Bacillus* and other bacterial genera (Gomila et al. 2015; Wang and Ash 2015; Lalucat et al. 2020). 16S rRNA gene, *gyrB*, *rpoB*, *rpoD* and 92 housekeeping genes have recently been used to investigate the correct taxonomic classification of *Pseudomonas* species (Saati-Santamar et al., 2021). For the genus *Bacillus*, it has been proposed to use gyrA instead of the 16S rRNA gene sequence, as it provides better phylogenetic resolution (Liu et al., 2022), or to move from single gene to whole genome-based species identification (Khurana et al., 2020).

In summary, our results demonstrate the progress of MALDI-TOF MS-based identification of environmental bacteria using a collection of isolates from a highly specific environmental compartment: a biofilm from the shell surface of the cave-dwelling bivalve *Congeria jalzici*. However, the identification scores obtained were still < 2.300, suggesting that further enrichment of commercial MALDI-TOF MS libraries is required before this method can be used as a reliable tool for assessing the diversity of culturable bacteria from such habitats. The construction of specific in-house MALDI-TOF databases for particular genera/environments could also be beneficial (e.g. Pinar-Méndez et al. 2021).

References

Alcalá L, Marín M, Ruiz A, Quiroga L, Zamora-Cintas M, Fernández-Chico MA, Muñoz P, Rodríguez-Sánchez B (2021) Identifying anaerobic bacteria using MALDI-TOF mass spectrometry: a four-year experience. Front Cell Infect Microbiol 11:521014. https://doi.org/10.3389/fcimb.2021.521014

Angeletti S (2017) Matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) in clinical microbiology. J Microbiol Methods 138:20-29. doi: 10.1016/j.mimet.2016.09.003

Assis GBN, Pereira FL, Zegarra AU, Tavares GC, Leal CA, Figueiredo HCP (2017) Use of MALDI-TOF mass spectrometry for the fast identification of gram-positive fish pathogens. Front Microbiol 8:1492. https://doi.org/10.3389/fmicb.2017.01492

Ashfaq MY, Da'na DA, Al-Ghouti MA (2022) Application of MALDI-TOF MS for identification of environmental bacteria: A review. Journal of Environmental Management, 305, 114359. doi: 10.1016/j.jenvman.2021.114359

Austin B (2017) The value of cultures to modern microbiology. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol 110:1247-1256. doi: https://doi.org/10.1007/s10482-017-0840-8

Avanzi IR, Gracioso LH, Baltazar M dos PG, et al (2017) Rapid bacteria identification from environmental mining samples using MALDI-TOF MS analysis. Environ Sci Pollut Res 24:3717-3726. https://doi.org/10.1007/s11356-016-8125-8

Balvočiūtė M, Huson DH (2017). SILVA, RDP, Greengenes, NCBI and OTT—how do these taxonomies compare? BMC genomics 18(2):1-8. https://doi.org/10.1186/s12864-017-3501-4

Barberis C, Almuzara M, Join-Lambert O, Ramírez MS, Famiglietto A, Vay C (2014) Comparison of the Bruker MALDI-TOF mass spectrometry system and conventional phenotypic methods for identification of Gram-positive rods. PLoS One 9:e106303. https://doi.org/10.1371/journal.pone.0106303

Beaz-Hidalgo R, Hossain MJ, Liles MR, Figueras MJ (2015) Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for *Aeromonas* genomes in the GenBank database. PloS one 10(1):e0115813. https://doi.org/10.1371/journal.pone.0115813

Bedek J, Gottstein S, Taiti S (2019) A new species of *Alpioniscus (Illyrionethes)* from the Dinaric Karst (Isopoda, Oniscidea, Trichoniscidae). Subterr Biol 32:33-43. https://doi.org/10.3897/subtbiol.32.37509

Berendsen RL, van Verk MC, Stringlis IA, Zamioudis C, Tommassen J, Pieterse CM, Bakker PA (2015). Unearthing the genomes of plant-beneficial *Pseudomonas* model strains WCS358, WCS374 and WCS417. BMC genomics 16(1):1-23. DOI 10.1186/s12864-015-1632-z

Bizzini A, Jaton K, Romo D, Bille J, Prod'hom G, Greub G (2011) Matrix-assisted laser desorption ionization–time of flight mass spectrometry as an alternative to 16S

rRNA gene sequencing for identification of difficult-to-identify bacterial strains. J Clin Microbiol 49(2):693-696. https://doi.org/10.1128/jcm.01463-10

Bilandžija H, Morton B, Podnar M, Cetković H (2013) Evolutionary history of relict *Congeria* (Bivalvia: Dreissenidae): unearthing the subterranean biodiversity of the Dinaric Karst. Front Zool 10:5. https://doi.org/10.1186/1742-9994-10-5

Bodor A, Bounedjoum N, Vincze GE, Kis ÁE, Laczi K, Bende G, Szilágyi Á, Kovács T, Perei K, Rákhely G (2020) Challenges of unculturable bacteria: environmental perspectives. Rev. Environ. Sci. Biotechnol. 19:1-22. https://doi.org/10.1007/s11157-020-09522-4

Böhme K, Fernández-No IC, Pazos M, Gallardo JM, Barros-Velázquez J, Cañas B, Calo-Mata P (2013) Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS fingerprinting. Electrophoresis 34:877-887. https://doi.org/10.1002/elps.201200532

Bonacci O, Pipan T, Culver DC (2009) A framework for karst ecohydrology. Environ Geol 56:891-900. https://doi.org/10.1007/s00254-008-1189-0

Brauge T, Trigueros S, Briet A, Debuiche S, Leleu G, Gassilloud B, Wilhelm A, Py J-S, Midelet G (2021) MALDI-TOF mass spectrometry fingerprinting performance versus 16S rDNA sequencing to identify bacterial microflora from seafood products and sea water samples. Front Mar Sci 8:650116. https://doi.org/10.3389/fmars.2021.650116

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM (2014) Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42:D633-42. https://doi.org/10.1093/nar/gkt1244

Culver DC, Sket B (2000) Hotspots of subterranean biodiversity in caves and wells. J Cave Karst Stud 62:11-17. https://digitalcommons.usf.edu/kip_articles/2593

Deharveng L, Bedos A (2018) Diversity of Terrestrial Invertebrates in Subterranean Habitats. Cave Ecology 107-172. https://doi.org/10.1007/978-3-319-98852-8_7

Dieckmann R, Graeber I, Kaesler I, Szewzyk U, von Döhren H (2005) Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by Intact-Cell-MALDI-TOF mass spectrometry (ICM-MS). Appl Microbiol Biotechnol 67:539-548. https://doi.org/10.1007/s00253-004-1812-2

Duijff BJ, Recorbet G, Bakker PA, Loper JE, Lemanceau P (1999). Microbial antagonism at the root level is involved in the suppression of Fusarium wilt by the combination of nonpathogenic *Fusarium oxysporum* Fo47 and *Pseudomonas putida* WCS358. Phytopathology 89(11):1073-1079.

Dybwad M, Granum PE, Bruheim PP, Blatnya JM (2012) Characterization of airborne bacteria at an underground subway station. Appl Environ Microbiol 78:1917-1929. https://doi.org/10.1128/AEM.07212-11

573 Emami K, Askari V, Ullrich M, Mohinudeen K, Anil AC, Khandeparker L, Burgess JG,
574 Mesbahi E (2012) Characterization of bacteria in Ballast water using MALDI-TOF
575 mass spectrometry. PLoS One 7 (6):e38515.
576 https://doi.org/10.1371/journal.pone.0038515

577 Emami K, Nelson A, Hack E, Zhang J, Green DH, Caldwell GS, Mesbahi E (2016).
578 MALDI-TOF mass spectrometry discriminates known species and marine
579 environmental isolates of *Pseudoalteromonas*. Front Microbiol 7:104. doi:
580 10.3389/fmicb.2016.00104

581 Elbehiry A, Aldubaib M, Abalkhail A, Marzouk E, ALbeloushi A, Moussa I, Ibrahem M,
582 Albazie H, Alqarni A, Anagreyyah S, Alghamdi S, Rawway M (2022) How MALDI-
583 TOF Mass Spectrometry Technology Contributes to Microbial Infection Control in
584 Healthcare Settings. Vaccines 10(11):1881. doi: 10.3390/vaccines10111881

585 Federhen S (2012) The NCBI Taxonomy database. Nucleic Acids Res 40:136–143.
586 https://doi.org/10.1093/nar/gkr1178

587 Fykse EM, Tjärnhage T, Humppi T, et al (2015) Identification of airborne bacteria by
588 16S rDNA sequencing, MALDI-TOF MS and the MIDI microbial identification system.
589 Aerobiologia (Bologna) 31:271-281. https://doi.org/10.1007/s10453-015-9363-9

590 Girard L, Lood C, Höfte M, Vandamme P, Rokni-Zadeh H, van Noort V, Lavigne R,
591 De Mot R (2021) The ever-expanding *Pseudomonas* genus: description of 43 new
592 species and partition of the *Pseudomonas putida* group. Microorganisms 9(8):1766.
593 doi: 10.3390/microorganisms9081766

594 Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E (2015) Phylogenomics and
595 systematics in *Pseudomonas*. Front Microbiol 6:214.
596 https://doi.org/10.3389/fmicb.2015.00214

597 Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical
598 user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol
599 27:221-224. https://doi.org/10.1093/molbev/msp259

600 Gupta RS, Patel S, Saini N, Chen S (2020) Robust demarcation of 17 distinct
601 *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics
602 and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp.
603 nov. and proposal for an emended genus *Bacillus* limiting it only to the members of
604 the Subtilis and Cereus clades of species. Int J Syst Evol Microbiol 70(11):5753-
605 5798. doi: 10.1099/ijsem.0.004475

606 Hall T (2011) BioEdit: An important software for molecular biology. GERF Bull Biosci
607 2:60-61. https://doi.org/10.1002/prot.24632

608 He Y, Li H, Lu X, Stratton CW, Tang YW (2010) Mass spectrometry biotyper system
609 identifies enteric bacterial pathogens directly from colonies grown on selective stool
610 culture media. J Clin Microbiol 48:3888-3892. https://doi.org/10.1128/JCM.01290-10

611 Huang CH, Wang CL, Liou JS, Lee AY, Blom J, Huang L, Watanabe K (2019)
612 Reclassification of *Micrococcus aloeverae* and *Micrococcus yunnanensis* as later

heterotypic synonyms of *Micrococcus luteus*. Int J Syst Evol Microbiol, 69(11):3512-3518. https://doi.org/10.1099/ijsem.0.003654

Hou TY, Chiang-Ni C, Teng SH (2019) Current status of MALDI-TOF mass spectrometry in clinical microbiology. J Food Drug Anal 27:404-414. doi: 10.1016/j.jfda.2019.01.001

Jin S, Kim KY, Kim MS, Park C (2020) An assessment of the taxonomic reliability of DNA barcode sequences in publicly available databases. Algae 35(3):293-301. https://doi.org/10.4490/algae.2020.35.9.4

Jovanović Glavaš O, Jalžić B, Bilandžija H (2017) Population density, habitat dynamic and aerial survival of relict cave bivalves from genus *Congeria* in the Dinaric karst. Int J Speleol 46:13-22. https://doi.org/10.5038/1827-806X.46.1.2020

Kačániová M, Klūga A, Kántor A, Medo J, Žiarovská J, Puchalski C, Terentjeva M (2019) Comparison of MALDI-TOF MS Biotyper and 16S rDNA sequencing for the identification of *Pseudomonas* species isolated from fish. Microb Pathog 132:313-318. https://doi.org/10.1016/j.micpath.2019.04.024

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772-780. https://doi.org/10.1093/molbev/mst010

Khurana H, Sharma M, Verma H, Lopes BS, Lal R, Negi RK (2020) Genomic insights into the phylogeny of *Bacillus* strains and elucidation of their secondary metabolic potential. Genomics 112:3191-3200. https://doi.org/10.1016/j.ygeno.2020.06.005

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120. https://doi.org/10.1007/BF01731581

Kittinger C, Lipp M, Baumert R, Folli B, Koraimann G, Toplitsch D, Liebmann A, Grisold AJ, Farnleitner AH, Kirschner A, Zarfel G (2016) Antibiotic resistance patterns of *Pseudomonas* spp. isolated from the river Danube. Front Microbiol 7:586. https://doi.org/10.3389/fmicb.2016.00586

Kopcakova A, Stramova Z, Kvasnova S, Godany A, Perhacova Z, Pristas P (2014) Need for database extension for reliable identification of bacteria from extreme environments using MALDI TOF mass spectrometry. Chem Pap 68:1435-1442. https://doi.org/10.2478/s11696-014-0612-0

Kostanjšek R, Prodan Y, Stres B, Trontelj P (2019) Composition of the cutaneous bacterial community of a cave amphibian, *Proteus anguinus*. FEMS Microbiol Ecol 95: fiz007. https://doi.org/10.1093/femsec/fiz007

Koubek J, Uhlik O, Jecna K, Junkova P, Vrkoslavova J, Lipov J, Kurzawova V, Macek T, Mackova M (2012) Whole-cell MALDI-TOF: Rapid screening method in environmental microbiology. Int Biodeterior Biodegrad 69:82-86. https://doi.org/10.1016/j.ibiod.2011.12.007

652 Kováč Ľ (2018) Caves as Oligotrophic Ecosystems. In: Moldovan O, Kováč Ľ, Halse
653 S (eds) Cave Ecology. Ecological Studies vol 235. Springer Cham pp 297-307.
654 https://doi.org/10.1007/978-3-319-98852-8_13

655 Kozel P, Delić T, Novak T (2020) *Nemaspela borkoae* sp. nov. (Opiliones:
656 Nemastomatidae), the second species of the genus from the Dinaric Karst. Eur J
657 Taxon 2020:90-107. https://doi.org/10.5852/ejt.2020.717.1103

658 Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics
659 analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870-1874.
660 https://doi.org/10.1093/molbev/msw054

661 Lane DJ (1991) 16S/23S rRNA Sequencing. Nucleic acid techniques in bacterial
662 systematics.

663 Lalucat J, Gomila M, Mulet M, Zaruma A, García-Valdés E (2022) Past, present and
664 future of the boundaries of the *Pseudomonas* genus: Proposal of *Stutzerimonas* gen.
665 Nov. Syst Appl Microbiol 45(1):126289. doi: 10.1016/j.syapm.2021.126289

666 Lavigne JP, Espinal P, Dunyach-Remy C, Messad N, Pantel A, Sotto A (2013) Mass
667 spectrometry: A revolution in clinical microbiology? Clin Chem Lab Med 51:257-270.
668 doi: 10.1515/cclm-2012-0291

669 Lemanceau P, Bakker PAHM, De Kogel WJ, Alabouvette C, Schippers B (1992).
670 Effect of pseudobactin 358 production by *Pseudomonas putida* WCS358 on
671 suppression of fusarium wilt of carnations by nonpathogenic *Fusarium oxysporum*
672 Fo47. Appl Environ Microbiol 58(9):2978-2982.

673 Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new
674 developments. Nucleic Acids Res 47(W1): W256-W259.
675 https://doi.org/10.1093/nar/gkz239

676 Li M, Hong CY, Yan WX, Chao ZS, Gang YC, Ling DJ, Kui XZ, Qin XJ, Liang ZM, He
677 MM (2017) *Bacillus zanthoxyli* sp. nov., a novel nematicidal bacterium isolated from
678 Chinese red pepper (*Zanthoxylum bungeanum* Maxim) leaves in China. Antonie van
679 Leeuwenhoek, 110:1179-1187. doi: 10.1007/s10482-017-0890-y

680 Liu Y, Štefanič P, Miao Y, Xue Y, Xun W, Zhang N, Shen Q, Zhang R, Xu Z, Mandic-
681 Mulec I (2022). Housekeeping gene gyrA, a potential molecular marker for *Bacillus*
682 ecology study. AMB Express 12(1):133. https://doi.org/10.1186/s13568-022-01477-9

683 Mahato NK, Gupta V, Singh P, et al (2017) Microbial taxonomy in the era of OMICS:
684 application of DNA sequences, computational tools and techniques. Antonie van
685 Leeuwenhoek, Int J Gen Mol Microbiol 110:1357-1371.
686 https://doi.org/10.1007/s10482-017-0928-1

687 Marugg JD, Van Spanje MHWP, Hoekstra WP, Schippers B, Weisbeek PJ (1985).
688 Isolation and analysis of genes involved in siderophore biosynthesis in plant-growth-
689 stimulating *Pseudomonas putida* WCS358. J. Bacteriol 164(2):563-570.

690 Matočec SG, Bakran-Petricioli T, Bedek J, et al (2002) An overview of the cave and
691 interstitial biota of Croatia. Natura Croatica: Periodicum Musei Historiae Naturalis
692 Croatici, 11(suppl. 1):1-112.

693 Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M (2022) TYGS and
694 LPSN: a database tandem for fast and reliable genome-based classification and
695 nomenclature of prokaryotes. Nucl Acids Res 50(D1):D801-D807. doi:
696 10.1093/nar/gkab902

697 Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, Iwen P, Dunn J, Hall G,
698 Wilson D, LaSala P, Kostrezewa M, Harmsen D (2008). Evaluation of matrix-assisted
699 laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S
700 rRNA gene sequencing for species identification of nonfermenting bacteria. J Clin
701 Microbiol 46(6):1946-1954. doi: 10.1128/JCM.00157-08

702 Meziane H, Van Der Sluis I, Van Loon LC, Höfte M, Bakker PA (2005). Determinants
703 of *Pseudomonas putida* WCS358 involved in inducing systemic resistance in plants.
704 Mol. Plant Pathol. 6(2):177-185. https://doi.org/10.1111/j.1364-3703.2005.00276.x

705 McDonald D, Price MN, Goodrich J, et al (2012) An improved Greengenes taxonomy
706 with explicit ranks for ecological and evolutionary analyses of bacteria and archaea.
707 ISME J 6:610-618. https://doi.org/10.1038/ismej.2011.139

708 Morimoto Y, Tohya M, Aibibula Z, Baba T, Daida H, Kirikae T (2020) Re-identification
709 of strains deposited as *Pseudomonas aeruginosa*, *Pseudomonas fluorescens* and
710 *Pseudomonas putida* in GenBank based on whole genome sequences. IJSEM
711 70(11):5958-5963. doi: 10.1099/ijsem.0.004468

712 Mudgil D, Paul D, Baskar S, Baskar R, Shouche YS (2022) Cultivable microbial
713 diversity in speleothems using MALDI-TOF spectrometry and DNA sequencing from
714 Krem Soitan, Krem Lawbah, Krem Mawpun, Khasi Hills, Meghalaya, India. Arch
715 Microbiol 204(8):495. https://doi.org/10.1007/s00203-022-02916-8

716 Mulec J (2008) Microorganisms in hypogeon: Examples from Slovenian karst caves.
717 Acta Carsologica 37:153–160. https://doi.org/http://ojs.zrc-
718 sazu.si/carsologica/article/view/167

719 Mulec J, Oarga-Mulec A, Tomazin R, Matos T (2015) Characterization and
720 fluorescence of yellow biofilms in karst caves, southwest Slovenia. Int J Speleol
721 44:107–114. https://doi.org/10.5038/1827-806X.44.2.1

722 Ozimec R, Bedek J, Gottstein S, et al (2009) Crvena Knjiga špiljske faune Hrvatske.
723 Državni zavod za zaštitu prirode, Zagreb

724 Pandey A, Jain R, Sharma A, Dhakar K, Kaira GS, Rahi P, Dhyani A, Pandey N,
725 Adhikari P, Shouche YS (2019) 16S rRNA gene sequencing and MALDI-TOF mass
726 spectrometry based comparative assessment and bioprospection of psychrotolerant
727 bacteria isolated from high altitudes under mountain ecosystem. SN Appl Sci1:1-12.
728 https://doi.org/10.1007/s42452-019-0273-2

729 Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M (2020).List of
730 Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ.
731 IJSEM 70(11):5607-5612. https://doi.org/10.1099/ijsem.0.004332

732 Pašić L, Kovče B, Sket B, Herzog-Velikonja B (2010) Diversity of microbial
733 communities colonizing the walls of a Karstic cave in Slovenia. FEMS Microbiol Ecol
734 71:50-60. https://doi.org/10.1111/j.1574-6941.2009.00789.x

735 Pinar-Méndez A, Fernández S, Baquero D, Vilaró C, Galofré B, González S, et al
736 (2021). Rapid and improved identification of drinking water bacteria using the
737 Drinking Water Library, a dedicated MALDI-TOF MS database. Water Res
738 203:117543. doi: 10.1016/j.watres.2021.117543

739 Pomastowski P, Złoch M, Rodzik A, et al (2019) Analysis of bacteria associated with
740 honeys of different geographical and botanical origin using two different identification
741 approaches: MALDI-TOF MS and 16S rDNA PCR technique. PLoS One 14:1-20.
742 https://doi.org/10.1371/journal.pone.0217078

743 Popović NT, Kazazić SP, Strunjak-Perović I, Čož-Rakovac R (2017) Differentiation of
744 environmental aquatic bacterial isolates by MALDI-TOF MS. Environ Res 152:7-16.
745 doi: 10.1016/j.envres.2016.09.020

746 Rodríguez-Sánchez B, Alcalá L, Marín M, Ruiz A, Alonso E, Bouza E (2016)
747 Evaluation of MALDI-TOF MS (matrix-assisted laser desorption-ionization time-of-
748 flight mass spectrometry) for routine identification of anaerobic bacteria. Anaerobe
749 42:101-107. doi: 10.1016/j.anaerobe.2016.09.009

750 Rossel S, Khodami S, Martínez Arbizu P (2019) Comparison of rapid biodiversity
751 assessment of meiobenthos using MALDI-TOF MS and Metabarcoding. Front Mar
752 Sci 6:659. https://doi.org/10.3389/fmars.2019.00659

753 Saati-Santamaría Z, Peral-Aranega E, Velázquez E, Rivas R, García-Fraile P (2021).
754 Phylogenomic analyses of the genus *Pseudomonas* lead to the rearrangement of
755 several species and the definition of new genera. Biology 10(8):782.
756 https://doi.org/10.3390/biology10080782

757 Samad RA, Al Disi Z, Ashfaq MYM, Wahib SM, Zouari N (2020) The use of principle
758 component analysis and MALDI-TOF MS for the differentiation of mineral forming
759 *Virgibacillus* and *Bacillus* species isolated from sabkhas. RSC advances
760 10(25):14606-14616. https://doi.org/10.1039/D0RA01229G

761 Santos IC, Hildenbrand ZL, Schug KA (2016) Applications of MALDI-TOF MS in
762 environmental microbiology. Analyst 141:2827-2837.
763 https://doi.org/10.1039/C6AN00131A

764 Schmitt BH, Cunningham SA, Dailey AL, Gustafson DR, Patel R (2013) Identification
765 of anaerobic bacteria by Bruker Biotyper matrix-assisted laser desorption ionization–
766 time of flight mass spectrometry with on-plate formic acid preparation. J Clin
767 Microbiol 51(3):782-786. doi: 10.1128/JCM.02420-12

768 Schulthess B, Bloemberg GV, Zbinden A, Mouttet F, Zbinden R, Böttger EC,
769 Hombach M (2016) Evaluation of the Bruker MALDI Biotyper for identification of

770 fastidious Gram-negative rods. J Clin Microbiol 54(3):543-548.
771 https://doi.org/10.1128/jcm.03107-15

772 Seuylemezian A, Aronson HS, Tan J, Lin M, Schubert W, Vaishampayan P (2018)
773 Development of a custom MALDI-TOF MS database for species-level identification of
774 bacterial isolates collected from spacecraft and associated surfaces. Front Microbiol
775 9:780. https://doi.org/10.3389/fmicb.2018.00780

776 Singhal N, Kumar M, Kanaujia PK, Virdi JS (2015) MALDI-TOF mass spectrometry:
777 An emerging technology for microbial identification and diagnosis. Front Microbiol
778 6:1-16. https://doi.org/10.3389/fmicb.2015.00791

779 Sket, B. 2005. Diversity in Dynaric Karst. In: Culver DC and White WB (ed)
780 Encyclopedia of Caves. pp 158-165

781 Strejcek M, Smrhova T, Junkova P, Uhlik O (2018) Whole-cell MALDI-TOF MS
782 versus 16S rRNA gene analysis for identification and dereplication of recurrent
783 bacterial isolates. Front Microbiol 9. https://doi.org/10.3389/fmicb.2018.01294

784 Su C, Lei L, Duan Y, Zhang KQ, Yang J (2012) Culture-independent methods for
785 studying environmental microorganisms: Methods, application, and perspective. Appl
786 Microbiol Biotechnol 93:993-1003. doi: 10.1007/s00253-011-3800-7

787 Surić M, Czuppon G, Lončarić R, Bočić N, Lončar N, Bajo P, Drysdale RN (2020)
788 Stable isotope hydrology of cave groundwater and its relevance for speleothem-
789 based paleoenvironmental reconstruction in Croatia. Water (Switzerland) 12:2386.
790 https://doi.org/10.3390/W12092386

791 Timperio AM, Gorrasi S, Zolla L, Fenice M (2017) Evaluation of MALDI-TOF mass
792 spectrometry and MALDI BioTyper in comparison to 16S rDNA sequencing for the
793 identification of bacteria isolated from Arctic sea water. PLoS One 12:1-15.
794 https://doi.org/10.1371/journal.pone.0181860

795 Topic Popovic N, Kazazic SP, Strunjak-Perovic I, Barisic J, Saureborn Klobucar R,
796 Kepec S, Coz-Rakovac R (2015) Detection and diversity of aeromonads from treated
797 wastewater and fish inhabiting effluent and downstream waters. Ecotoxicol Environ
798 Saf 120:235–242. https://doi.org/10.1016/j.ecoenv.2015.06.011

799 Topić Popović N, Kazazić SP, Strunjak-Perović I, Čož-Rakovac R (2016)
800 Differentiation of environmental aquatic bacterial isolates by MALDI-TOF MS. Environ
801 Res 152:7-16. https://doi.org/10.1016/j.envres.2016.09.020

802 Torres-Sangiao E, Leal Rodriguez C, García-Riestra C (2021) Application and
803 perspectives of MALDI–TOF mass spectrometry in clinical microbiology laboratories.
804 Microorganisms 9(7):1539. https://doi.org/10.3390/microorganisms9071539

805 Uhlik O, Strejcek M, Junkova P, Sanda M, Hroudova M, Vlcek C, Mackova M, Macek
806 T (2011) Matrix-assisted laser desorption ionization (MALDI)-time of flight mass
807 spectrometry- and MALDI biotyper-based identification of cultured biphenyl-
808 metabolizing bacteria from contaminated horseradish rhizosphere soil. Appl Environ
809 Microbiol 77:6858-6866. https://doi.org/10.1128/AEM.05465-11

Velikonja BH, Tkavc R, Pašić L (2014) Diversity of cultivable bacteria involved in the formation of macroscopic microbial colonies (cave silver) on the walls of a cave in Slovenia. Int J Speleol 43:45-56. https://doi.org/10.5038/1827-806X.43.1.5

Viver T, Cifuentes A, Díaz S, Rodríguez-Valdecantos G, González B, Antón J, Rosselló-Móra R (2015) Diversity of extremely halophilic cultivable prokaryotes in Mediterranean, Atlantic and Pacific solar salterns: evidence that unexplored sites constitute sources of cultivable novelty. Syst Appl Microbiol 38:266-275. https://doi.org/10.1016/j.syapm.2015.02.002

Vos M, Quince C, Pijl AS, de Hollander M, Kowalchuk GA (2012) A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS One 7:1-8. https://doi.org/10.1371/journal.pone.0030600

Wang A, Ash GJ (2015) Whole Genome Phylogeny of *Bacillus* by Feature Frequency Profiles (FFP). Sci Rep 5:1-14. https://doi.org/10.1038/srep13644

Wilson DA, Young S, Timm K, et al (2017) Multicenter evaluation of the Bruker MALDI Biotyper CA System for the identification of clinically important bacteria and yeasts. Am J Clin Pathol 147:623-631. https://doi.org/10.1093/ajcp/aqw225

Yilmaz P, Parfrey LW, Yarza P, et al (2014) The SILVA and “all-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res 42:643-648. https://doi.org/10.1093/nar/gkt1209

Zhu S, Ratering S, Schnell S, Wacker R (2011) Matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry, 16S rRNA gene sequencing, and API 32E for identification of *Cronobacter* spp.: a comparative study. J Food Prot 74:2182-2187. https://doi.org/10.4315/0362-028X.JFP-11-205

Zhu HZ, Zhang ZF, Zhou N, Jiang CY, Wang BJ, Cai L, & Liu SJ (2019) Diversity, distribution and co-occurrence patterns of bacterial communities in a karst cave system. Front Microbiol 10:1726. https://doi.org/10.3389/fmicb.2019.01726

Statements and declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

All authors read and approved the final manuscript.

Author contributions

Conceptualization: Ana Bielen, Sandra Hudina; Methodology: Ana Bielen, Snježana Kazazić, Ana Šimatović, Nikolina Udiković-Kolić; Formal analysis and investigation: Ana Bielen, Ivana Babić, Marija Vuk Surjan, Snježana Kazazić, Ana Šimatović, Zrinka Mesić; Writing - original draft preparation: Ana Bielen, Ivana Babić; Writing - review and editing: Ana Bielen, Ivana Babić, Jasna Lajtner, Nikolina Udiković-Kolić, Zrinka Mesić, Sandra Hudina; Supervision: Ana Bielen.

Acknowledgements

The authors are thankful to Sanja Puljas for participating in collecting of the samples, Lidija Buzuk, Lidija Šver, Nives Marčina, Karla Orlić and Dunja Šikić for help with the laboratory work, and prof. Rup Lal for advice on 16S rDNA phylogenetic analysis.