Theses of the PhD Dissertation

Surányi Botond Bendegúz

Budapest

2023

Hungarian University of Agriculture and Life Sciences

Characterization of the microbiota of irrigation water used in vegetable farming applying MALDI-TOF MS

Doctoral thesis

Botond Bendegúz Surányi

Budapest

2023

Doctoral School

Name: Doctoral School of Food Science

Field: Food Science

Head: Livia Simon Sarkadi, DSc

Department of Nutrition Science

Institute of Food Science and Technology

Hungarian University of Agriculture and Life Sciences

Supervisors:

Csilla Mohácsi-Farkas, PhD

Department of Food Microbiology, Hygiene and Safety, Institute of Food Science and Technology, Hungarian University of Agriculture and Life Sciences

Tekla Engelhardt, PhD

Digital Food Chain Education, Research, Development and Innovation Institute, University of Veterinary Medicine Budapest

The applicant met the requirement of the Ph.D. regulations of the Hungarian University of Agriculture and Life Sciences and the thesis is accepted for the defense process.

.....

Head of the Doctoral School

....

Supervisor Supervisor

Table of contents

1 INTRODUCTION

In order to intensify and guarantee the agricultural productivity and thereby to be able to feed the world's rapidly growing population, the accessibility of high-quality irrigation water has become very important. Recycled and microbiologically non-characterized waters are increasingly applied as irrigation water in cultivation in order to cope with water limitation due to climate change and to support rapid population growth. Crops can be contaminated with potentially harmful microorganisms at any of the several steps in the food production chain, during primary production, at processing stage and during preparation as in each step water plays a crucial role. At farm level, one of the major sources of food-borne pathogens is insufficient quality irrigation water which can be contaminated by sewage overflows, polluted storm- and agricultural runoffs or even by fecal pollution of wildlife (Uyttendaele et al. 2015). Moreover, serious bacterial pathogens such as *Listeria monocytogenes*, verotoxigenic *Escherichia coli*, *Salmonella* spp., *Escherichia coli* O157:H7 are able to survive and even grow in contaminated irrigation water (Cevallos-Cevallos et al. 2014;), while the reported numbers of food- and waterborne outbreaks are also increasing. According to the European Food Safety Authority (EFSA) growing numbers of outbreaks, cases, hospitalizations and deaths related to food of non-animal origin were observed (European Food Safety Authority 2019). Leafy vegetables irrigated with contaminated water are considered to be a common cause of human gastroenteritis, due to the presence of microbial pathogens. In the United States, 22.8%-46% of foodborne illnesses were related to fresh produce such as fruits, fungi, leafy vegetables, sprout vegetables in the period of 1998-2008 (Uyttendaele et al. 2015). Moreover, Turner et al. (2019) analyzed two decades (1996-2016) in the US where they observed 46 outbreaks causing 2240 confirmed cases, where the affected food matrices were romaine lettuce and spinach. In addition, according to the EFSA, 31 outbreaks related to vegetables and juices causing 626 cases. Moreover, 48 waterborne outbreaks were connected to the consumption of tap and well water resulting 1969 cases in 2019 (EFSA 2019, EFSA 2021a). Besides, the interest in consuming fresh produce has been exponentially grown due to the promotion of healthy eating involving the consumption of 5-7 portion of such foods per day (Betts 2014). Therefore, identifying and characterizing the microorganisms or even the complete bacterial community of irrigation water used for food production and its environment can prevent the increasing numbers of the cases. In order to ensure that, fast and reliable detection and identification of food- and waterborne bacteria should be an attainable option. Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS), a rapid

microbiology technique involving laser energy absorbing matrix to create ions from larger molecules, has been applied in the field of microbiology for its fast, accurate and inexpensive nature. Its use in microbiology is mainly centered around microbial diagnosis in clinical bacteriology but it is increasingly utilized not only in environmental bacteriology to identify food- and waterborne bacteria but to detect antibiotic resistance in bacteria and for bacterial strain typing as well. As identifying bacteria by 16S rRNA gene sequencing, the golden standard of bacterial identification, requires trained personnel and lengthy processes not to mention its higher cost of identification therefore it is not intended to use for fast identification. For this reason, MALDI-TOF MS can be a promising tool for environmental monitoring of the irrigation water used for food production and its environment.

2 OBJECTIVES

The aim of my PhD thesis is to form a comprehensive picture about the bacterial quality of irrigation waters and its surrounding environment in Hungary. To achieve this, MALDI-TOF MS was used to identify bacteria from different environmental matrices such as ground water, running water, lakes, manures and vegetables. Besides, some technical attributes of MALDI-TOF MS and the best culture media to identify waterborne bacteria by this technique were also analyzed. Furthermore, the efficacy of MALDI-TOF MS regarding identifying waterborne bacteria was tested and compared with Sanger sequencing. In addition, monitoring of irrigation water via culture-dependent and culture independent techniques were also performed.

To achieve these objectives, I set the following tasks:

- Isolating and identifying bacteria from different samples from the food production chain with MALDI-TOF MS (well, running- and still water used for irrigation, vegetables and manure).
- Testing sample preparation methods of MALDI-TOF MS of extended direct transfer and direct transfer procedure to identify waterborne bacteria.
- Examining different culture media to identify waterborne bacteria by MALDI-TOF MS.
- Applying multivariate statistical methods to differentiate bacterial strains using MALDI-TOF MS.
- Analyzing the effect of culture media via multivariate statistical methods on the mass spectra of bacterial strains.
- Testing the efficacy of MALDI-TOF MS regarding waterborne isolates against Sanger sequencing.
- Analyzing the whole (cultivable and uncultivable) bacterial microbiome of irrigation water by both MALDI-TOF MS and 16S rRNA amplicon sequencing.

3 MATERIALS AND METHODS

3.1 Microorganisms

3.1.1 Bacterial strains used for testing the culture media's impact on mass spectra

The impact of culture media on the protein mass fingerprint (PMF) of isolates was tested on Gram-negative and -positive bacteria. The bacterial strains, used to test culture media's impact on the PMF of Gram-negative bacteria, were *E. coli* DSM 11250 and *E. coli* ATCC 13706. Both *E. coli* strains were obtained from the Department of Food Science and Technology (DLWT), University of Natural Resources and Life Sciences, Vienna, Austria.

The former was isolated from human feces, while the latter is a strain used for water testing. Furthermore, *S. aureus* ATCC 25923 and *S. aureus* ATCC 43300 were used to test culture media's impact on the PMF of Gram-positive bacteria.

3.1.2 Sampling and bacterial isolation

Sampling sites were chosen due to their utilization as irrigation water. Sampling was also done in different regions, where irrigation water contamination (e.g., by manure) could have occurred or from irrigated crops (e.g., corn, lettuce, onion, sorrel, spinach, and tomato) where irrigation water might have transmitted the microbes on to them. One sample was taken from each sampling site. Altogether, 21 samples were collected from different areas of Hungary. The analyzed samples involved in which still water (2), running water (5), irrigation waters from wells (8), manure (4) and vegetables (2). Experiments analyzing samples from Karcag, Kengyel, Rákóczifalva and Szolnok were done in the Fall of 2021 at Institute of Food Science at University of Natural Resources and Life Sciences, Vienna (BOKU). The MALDI-TOF MS instrument (Bruker MALDI Biotyper) was provided by the EQ-BOKU VIBT GmbH and the BOKU Core Facility Food & Bio Processing. Experiments analyzing all the other samples were done in the period of 2019-2022 at Department of Food Microbiology, Hygiene and Safety at MATE Institute of Food Science and Technology. Water samples were collected into sterile bottles to minimize microbial contamination while vegetables samples were collected into sterile plastic bags, and both were transported in cooled state.

Two groups were formed from the vegetables in accordance with their origins as Vegetables1 were collected from Soroksár while Vegetables2 were collected from Debrecen. Group of Vegetables1 includes onion, corn, lettuce, spinach, while group of Vegetables2 contains tomato, spinach and sorrel. Sampling of each group of samples (water, manure, vegetables) was performed from the Summer of 2019 to the Spring of 2022. Bacterial isolates used in this study

included 311 isolates from different water, vegetables, and manure samples. The isolates were chosen based on their different morphological characteristics.

3.2 Culture media and cultivation of isolates

3.2.1 Culture media used to isolate bacteria from environmental samples

Bacterial isolation was performed after preparing a ten-fold serial dilution in buffered peptone water (BPW) (Thermo Fisher Scientific Inc., Oxoid Ltd., Basingstoke, UK) up to dilution 10^{-3}. The dilutions of samples were plated in duplicates on Trypticase Soy Agar (TSA) (Merck KGaA, Darmstadt, Germany), Reasoner's 2A agar (R2A agar) (Merck KGaA, Darmstadt, Germany), Violet Red Bile Dextrose agar (VRBD) (Merck KGaA, Darmstadt, Germany), and Yeast Extract Agar (YEA) (Merck KGaA, Darmstadt, Germany) plates by spread plate method. Agar plates were incubated at 30 °C for 24-48 h. Plates of *E. coli* and *S. aureus* strains were incubated for 24 h on 37 °C.

3.3 Methods

3.3.1 MALDI-TOF MS to identify bacteria from environmental samples

To identify the isolates, extended direct transfer procedure was used, therefore each colony of isolates was placed onto the Bruker's ground steel target plate, overlaid with 1 µL of 70% formic acid after airdrying, overlaid with 1 µL of α-cyano-4 hydroxycinnamic acid matrix solution (HCCA). Each bacterial colony was measured two times. The identification process was done using MALDI Biotyper (Bruker Daltonics GmbH & Co, Billerica, Massachusetts, USA). FlexControl and FlexAnalysis (Bruker Daltonics GmbH & Co, Bremen, Germany) were used for data acquisition and data processing, respectively. FlexAnalysis was used to preprocess mass spectra which involves baseline subtraction, smoothing and peak picking. MALDI-TOF MS identification results were accepted at genus or species level according to Bruker's instructions.

3.3.2 DNA Extraction and Sanger sequencing of waterborne isolates

DNA extraction of the previously cultured isolates was performed by Chelex Method. After extracting the DNA of isolates, 16S rRNA gene specific PCR was performed. PCR products were evaluated by 1% agarose gel electrophoresis. Samples were purified with the peqGOLD Cycle-Pure Kit (VWR International, Radnor, Pennsylvania, USA) following the manufacturer's instructions. 3 µL of 27F gene primer were added to 12 µL of purified DNA, then DNA Sanger sequencing was performed by Microsynth AG (Balgach, Switzerland). Sequences of the isolates were blasted against the NCBI RefSeq RNA sequence database to identify them.

As for 16S rRNA gene sequencing, in accordance with previous findings (Strejcek et al. 2018) 98.65% sequence similarity threshold was accepted to bacterial species demarcation.

3.3.3 DNA-extraction and next-generation sequencing of irrigation water samples

For the isolation of microbial genomic DNA from irrigation water samples DNeasy PowerSoil Pro Kit (QIAGEN, Hilden, Germany) was used using the instructions of the manufacturer. Amplicon library generation, quality control and sequencing were performed at the Vienna Biocenter Core Facilities NGS Unit (www.vbcf.ac.at). The V3–V5 hypervariable region of the 16S rRNA gene was amplified and sequenced using a MiSeq Illumina platform with a 300 bp paired-end read protocol (Illumina, Inc., San Diego, California, USA). Raw sequence data is available in the European Nucleotide Archive under accession number PRJEB56665.

3.4 Data analysis

3.3.1 Analysis of data obtained from MALDI-TOF MS identification

The identification process was done using MALDI Biotyper (Bruker Daltonics GmbH & Co, Billerica, Massachusetts, USA). FlexControl and FlexAnalysis (Bruker Daltonics GmbH & Co, Bremen, Germany) were used for data acquisition and data processing, respectively.

3.4.2 Analysis of data obtained from biomarker detection experiments

The settings of the MALDI-TOF MS instrument were the same as detailed in the subchapter of '*3.4.1 Analysis of data obtained from MALDI-TOF MS identification'*. FlexControl and FlexAnalysis (Bruker Daltonics GmbH & Co, Bremen, Germany) were used for data acquisition and data processing, respectively. Each raw spectrum was converted to a .csv file by the software and a list of intensities with the corresponding m/z data was created. To differentiate the two *E. coli* isolates and *S. aureus* isolates, Mass-Up was used to analyze further the bacterial mass spectra and to perform principal component analysis (López-Fernández et al. 2015).Differentiation of mass spectra of the two *E. coli* strains (DSM 11250 and ATCC 13706) and *S. aureus* strains (ATCC 25923 and ATCC 43300) cultivated on different culture media was made by Discriminant analysis (DA) using IBM SPSS Statistics 27 (IBM Corp. 2020).

3.4.3 Analysis of data obtained from culture media experiments

The settings of the MALDI-TOF MS instrument were the same as detailed in the subchapter of '*3.4.1 Analysis of data obtained from MALDI-TOF MS identification'*. FlexControl and FlexAnalysis (Bruker Daltonics GmbH & Co, Bremen, Germany) were used for data acquisition and data processing, respectively.

One-way ANOVA was applied to compare the identification score values of *S. aureus* isolates cultivated on the four culture media (IBM SPSS Statistics 27, Armonk, New York, USA). Based on the values of Skewness and Kurtosis, model residuals had normal distribution. Based on Levene's, homogeneity of variance was violated ($p<0.001$). ANOVA was significant ($F=22.164$; $p<0.001$), therefore Games-Howell test (Post hoc) was used because of the error variances was violated (IBM Corp. 2020).

One-way ANOVA was also applied on data obtained from *E. coli* culture media experiments and showed that there was no significant difference between the log score values of the identification ($p>0.05$) (IBM SPSS Statistics 27, Armonk, New York, USA). Kolmogorov–Smirnov test ($p>0.05$) proved that model residuals had normal distribution and homogeneity of variance was checked by Levene's test ($p>0.05$) (IBM Corp. 2020).

3.4.4 Data analysis of comparing MALDI-TOF MS and 16S rRNA gene sequencing

Paired t-test was used to compare the efficacy of identification of the MALDI-TOF MS and 16S rRNA gene sequencing (IBM SPSS Statistics 27, Armonk, New York, U.S.) (IBM Corp 2020).

3.4.5 Bioinformatics and sequence analysis of next-generation sequencing data

Primers were removed from the raw sequences using cutadapt v2.1 (Martin 2011). Raw sequences were further processed with the dada2 v1.14.1 pipeline in R v3.6.3 (Callahan et al. 2016; R Core Team 2021). Taxonomic assignment was performed via the SILVA rRNA database SSU 138 using the 'assignTaxonomy' command (Quast et al. 2013). MicrobiomeAnalyst was used to analyze data derived from 16S rRNA amplicon sequencing (Dhariwal et al. 2017).

4. RESULTS AND DISCUSSION

4.1 Identifying bacteria from environmental samples using MALDI-TOF MS

MALDI-TOF MS was used to identify bacteria from different types of environmental samples. Altogether, 21 samples were used for bacterial isolation of 311 isolates were analyzed. The samples contained a higher number of Gram-negative isolates (225) of which 178 (79.1%) and 92 (40.9%) were identified at genus and species level, respectively. Moreover, only 47 (20.9%) Gram-negative isolates were categorized as unidentified. The number of Gram-positive isolates (86) were lower compered to Gram-negative ones, 22 (25.6%) and 62 (72.1%) of them were identified at species and genus level, respectively. However, 24 of 86 (27.9%) Gram-positive isolates were not identified at any level. However, in general, taking into consideration the results of both Gram-positive and Gram-negative environmental isolates, better identification results were obtained both at genus (77.2%) and species level (36.6%).

4.2 Comparing sample preparation techniques of MALDI-TOF MS

Two sample preparation methods, extended direct transfer procedure (on-target extraction) and direct transfer procedure was compared to determine which is the best suited one for environmental isolates. The former involves an additional step of formic acid extraction before adding the matrix while the latter uses only the matrix.

The results of the formic acid extraction are clearly visible as only 2 of the 62 isolates were not identified at any level whereas the unidentified number of isolates lacking formic acid extraction were 16. By applying extended direct transfer procedure, 60 (96.8%) isolates and 31 (50%) of the isolates were identified at genus and species level. However, using direct transfer procedure only 46 (73%) and 21 (33.9%) of the isolates were identified at genus and species level.

The benefit of applying the extended direct transfer procedure is well-marked from the results. It should also be noted the average log scores of the extended direct transfer procedure were above 2 therefore reaching the species level threshold, while the average log scores of the direct transfer procedure was only 1.85. Besides, it was proved by Paired t-test ($t=16.09$, $p<0.001$) that the difference of the average log scores of the two sample preparation methods was significant.

4.3 The effect of culture media on MALDI-TOF MS identification regarding waterborne isolates

The effect of different culture media was tested on the identification performance of MALDI-TOF MS; therefore, the results of the validated culture medium (TSA) were compared to the results of R2A and Yeast Extract Agar (YEA). The 23 isolates chosen for these measurements have been isolated previously from different aqueous environment (lakes, rivers). The isolates contain mostly Gram-negative waterborne bacteria from the genera of *Aeromonas*, *Acinetobacter*, *Pseudomonas*, *Shewanella* and Gram-positive genera such as *Bacillus* and *Micrococcus*.

The results on TSA agar, the validated culture medium for MALDI-TOF MS identifications, shows that 11 (47.8%) and 19 (82.6%) of the isolates were identified at species and genus level. In addition, only four isolates were categorized as unsuccessful identifications. On R2A agar, 12 (52.2%) out of 23 isolates were identified at species level, which is higher than the result achieved on TSA agar. However, the number of unidentified isolates is also higher than on TSA agar, as five (21.8%) of the 23 isolates could not be identified even at genus level, compared to four unidentified isolates obtained on TSA. On R2A agar, 18 (78.3%) isolates were identified at genus level. The results of Yeast extract agar (YEA) were similar to those obtained on TSA and R2A, as 11 (47.8%) of 23 isolates were identified at species level. Yeast Extract Agar proved to be the best culture medium, since the lowest number of unidentified isolates, only three (13%), was also obtained on this culture medium.

In general, it was proved by ANOVA (F=0.10, p= 0.90, F_{crit}= 3.13) that R2A and Yeast Extract Agar are also suitable to identify waterborne bacteria as no significant difference was observed regarding the averages of log scores. However, the average log scores were below 2, the species level identification threshold, in each case (TSA, 1.97; R2A, 1.95; YEA, 1.97). This also highlights the limitation of MALDI-TOF MS Biotyper to identify environmental bacteria at species level therefore this measurement also necessitates the enrichment of the database by a wide range of environmental isolates. It can also be recommended to parallel identify waterborne isolates on TSA and Yeast Extract Agar to achieve better identification at species level.

4.4 Discriminating bacterial strains while simultaneously testing the effect of culture media on PMF of strains using MALDI-TOF MS

4.4.1 Discriminating bacterial strains via multivariate statistical methods and proteomics

The ability of MALDI-TOF MS to discriminate bacterial strains was tested and simultaneously the effect of culture media on the PMF of strains was also analyzed. Two *S. aureus* strains were tested as *S. aureus* ATCC 25923, a quality control strain while *S. aureus* ATCC 43300, a methicillin-resistant one. A specific peak was observed at m/z 5868 in the mass spectrum of *S. aureus* ATCC 43300 on all four culture media which peak can be considered as a specific biomarker for this strain as it was present only in the mass spectra of the aforementioned strain.

Principal component analysis (PCA) was used to differentiate the strains of *S. aureus* ATCC 25923 and *S. aureus* ATCC 43300. PCA separated eight groups as each group contained mass spectra obtained on them separately per strains. By applying PCA it was possible to differentiate *S. aureus* ATCC 43300 from *S. aureus* ATCC 25923 as each group were well separated from each other. However, it can be observed that groups of culture media containing the mass spectra of *S. aureus* ATCC 43300 and *S. aureus* ATCC 25923 were separated but R2A agar seemed to be the best one. Because the groups of those were separated the most regarding each *S. aureus* strain.

In the next part of experiments, PCA was used to differentiate the strains of *E. coli* ATCC 13706 and *E. coli* DSM 11250. Six groups were clearly separated as each group contain the mass spectra obtained on them separately per strains. Culture media applied in this study coupled with By applying PCA on the data of mass spectra it was possible to differentiate the two aforementioned strains as each group were well separated from each other.

Moreover, two strain specific peaks at m/z 6640 and m/z 8912 were also detected in the mass spectra of *E. coli* ATCC 13706. These peaks made it possible to distinguish the two *E. coli* strains used in this study as these were only present in the mass spectra of *E. coli* ATCC 13706 on all three culture media. The presence of these peaks in the mass spectra of *E. coli* isolates identified by MALDI-TOF MS can lead to the rapid identification of *E. coli* ATCC 13706, so these can be considered as important biomarkers for this species.

4.4.2 Analyzing the effect of culture media on the PMF of bacterial strains

Discriminant analysis (DA) was used to analyze the effect of culture media on the protein mass fingerprint (PMF) of *S. aureus* strains. Regarding *S. aureus* ATCC 25923 DA clearly separated Baird-Parker Agar from the rest of the groups which was also visible in the identification results

as the lowest scores were achieved on that culture medium. Moreover, the groups of all four culture media were distinguished from each other. The groups of TSA, R2A and YEA were closer to each other because those spectra were similar to each other and their identification scores were also closer as well. Regarding *S. aureus* ATCC 43300, DA showed that the spectra obtained on the two high-nutrient culture media, TSA and YEA, were similar to each other. However, not only the group of Baird-Parker Agar was further away from the two high nutrient culture media groups but the group of R2A, a low-nutrient culture medium, as well.

Interestingly, Baird-Parker Agar, the selective and differential culture medium for the isolation and enumeration of *S. aureus* in foods, environmental and clinical specimens, was the least effective of the tested culture media in terms of the confidence of identification. The best culture medium for both *S. aureus* strains was found to be the TSA as the highest log scores were achieved on that medium. The log scores of *S. aureus* ATCC 25923 were higher on Baird-Parker Agar, YEA, but on R2A agar *S. aureus* ATCC 43300 achieved better results. However, it is also important to add that *S. aureus* ATCC 25923 can be found in the Bruker's database which fact could explain the higher log scores of identifications. Given the fact that the other *S. aureus* strain (ATCC 43300) is not part of the database as an added reference strain this also highlights that the database enhancement is inevitable thus better results could be achieved.

One-way ANOVA was used to compare the identification score values of *S. aureus* isolates cultivated on the four culture media. ANOVA was significant ($F=22.164$; $p<0.001$), therefore Games-Howell test (Post hoc) was used as the error variances were violated. Games-Howell post hoc test showed that the averages of spectra obtained on Baird-Parker Agar were significantly different from the TSA agar ($p<0.001$). Thus, demonstrating that TSA agar can also be a good fit to identify *S. aureus* isolates as the highest identification score values were obtained on that culture medium. The averages of spectra obtained on R2A and Yeast Extract Agar were not differed significantly from the average of TSA agar. Therefore, these measurements showed that R2A and YEA agars are also suitable to identify *S. aureus* by MALDI-TOF MS as the application of both culture media generated species level identifications.

In the next part, by DA three groups were formed considering the type of culture media used to cultivate two *E. coli* strains (ATCC 13706, DSM 11250). The created groups, Group1 (R2A agar), Group2 (Yeast Extract Agar) and Group3 (TSA), could be distinguished for each bacterial strain. DA produced a clear separation of bacterial mass spectra generated on the different culture media for both strains. The three media differ in terms of composition. R2A

agar is a low-nutrient medium used for microbial monitoring of treated potable water, whereas YEA is a nutrient rich culture medium used for the plate count of organisms in water. TSA is a nonselective culture medium providing enough nutrients to cultivate a wide variety of microbes. The differences of generated spectra by the applied culture media were displayed by DA as groups of two high-nutrient media (YEA, TSA) were closer to each other, while the group of low-nutrient R2A agar was distant.

Although, there was a minor difference regarding the confidence of identifications, MALDI-TOF MS was able to identify both *E. coli* strains at species level on all three culture media used in this study. The results show that secure species identification can be achieved on the three examined culture media (YEA, TSA, R2A). Thus, the best culture media proved to be the Yeast Extract Agar because the application of it generated the highest identification scores. One-way ANOVA showed that there was no significant difference between the score values of identification, therefore the confidence of identification of *E. coli* isolates was not differed ($p>0.05$). Kolmogorov–Smirnov test ($p>0.05$) proved that model residuals had normal distribution and homogeneity of variance was checked by Levene's test ($p>0.05$).

4.5 Comparing MALDI-TOF MS and 16S rRNA gene sequencing to identify waterborne bacteria

In this subchapter of the thesis, the identification efficiency of MALDI-TOF MS against Sanger sequencing was tested on 42 bacteria from water. Bacterial isolates were cultivated from five irrigation water samples (Kengyel, Karcag, Rákóczifalva, Szolnok1, Szolnok2). MALDI-TOF MS and 16S rRNA gene sequencing identified 11 (26.2%) of the 42 isolates with different results. Different results were obtained for two of the 20 isolates belonging to the genus *Acinetobacter*, the most frequently isolated genus in these experiments. In both cases, MALDI-TOF MS identified *Acinetobacter junii*, while Sanger sequencing identified these isolates as *Acinetobacter schindleri* species. None of the techniques could identify an isolate belonging to the genus *Rhodococcus* at species level. Differences were also observed between *Enterobacter* and *Pseudomonas* isolates. MALDI-TOF MS identified the isolates marked #6 and #9 in the thesis as *E. hormaechei* and *E. cloacae* species, but Sanger sequencing could not distinguish the former, since the sequences of both *E. cloacae* and *E. hormaechei* showed 99.9% similarity. Although isolate #9 was identified as *E. cloacae* by MALDI-TOF MS, the result of Sanger sequencing showed 99.48% sequence similarity to an *E. hormaechei* isolate. Isolates marked #10 and #11 were identified by MALDI-TOF MS as *Pseudomonas veronii* isolates, while in the case of the former, Sanger sequencing showed 100% similarity also for *Pseudomonas*

veronii/ *Pseudomonas extremaustralis* species. In the case of the latter isolate, the best match was an uncharacterized *Pseudomonas* species.

The applied methods generated similar identification results, as both Sanger sequencing and MALDI-TOF MS identified more than 60% of the 42 bacteria isolated from water at species level. Using MALDI-TOF MS, 95.2% genus level and 66.7% species level identification results were achieved, whereas by applying Sanger sequencing, 90.5% genus level identification and 64.3% species level results were achieved. However, the paired t-test showed that there was no significant difference between the identification results of the two methods $t(41)=2.02$; $p=0.57$).

4.6 16S rRNA amplicon sequencing to characterize the community of irrigation waters from the county of Jász-Nagykun-Szolnok

Five irrigation water samples (Kengyel as Sample1, Karcag as Sample2, Rákóczifalva as Sample3, Szolnok1 as Sample4, Szolnok2 as Sample5) which were used for the comparison of MALDI-TOF MS and 16S rRNA gene sequencing were applied in these experiments also to monitor the uncultivable part of the irrigation water regarding microbial water quality.

The microbial community of water samples was diversified as the five most abundant genera were different in each sample. In Sample1, the most abundant genus was *Tepidimonas* followed by *Flavobacterium*, *Methylococcus*, *Methylophilaceae* UBA6140 and *Nocardioides*. In Sample2, the most abundant genus was *Sideroxydans* which were followed by genus *Brevundimonas*, *Terrimonas*, *Mycobacterium*, and *Candidatus Omnitrophus*. In Sample3, the most abundant genus was *Nitrosomonas*, an ammonia-oxidizing genus, which was followed by *Candidatus Nitrotoga*, an uncultured nitrite-oxidizing and naturally occurring bacterial genus in aqueous ecosystems and *Permianibacter*. Genus *Hydrogenophaga*, a hydrogen oxidizing genus, and *Pseudohongiella* were also common. The abundance of genera monitored in Sample4 was comparable to Sample3 as nitrifying-bacterial genera such as *Nitrosomonas* and *Candidatus Nitrotoga* were the second and fourth most abundant genera. However, the most abundant genus was *Gordonia* and the third most abundant genus was *Sphingobium* while the fifth was genus *Rhodococcus*. In Sample5 the dominance of Comamonadaceae family could be noticed as four of the five most abundant genera belonged to that family. The most abundant genus was *Rhodoferax* followed by *Acidovorax*, *Hydrogenophaga*, *Aquabacterium* and *Dechloromonas*.

Although genus *Nitrosomonas* was the most dominant in terms of relative abundance (11.04%), zero isolates were cultivated from it. In contrast, *Acinetobacter*, the most dominant genus

regarding cultivated isolates, was only the 31st in terms of relative abundance (0.64%) in the entire bacterial community. Similarly, despite five isolates had been isolated from genera *Pseudomonas* and *Enterobacter*, their relative abundance was only 0.24% and 0.04%, respectively. Although genus *Brevundimonas* had the highest relative abundance (2.18%) among cultivated genera, only three isolates of it were cultivated. Furthermore, two isolates of the genus *Rhodococcus* were cultivated, which also had the second highest relative abundance (0.81%) value among cultivated genera. Although only one isolate was cultivated and identified as a member of the genus *Chryseobacterium* its relative abundance (0.35%) was the fourth highest among cultivated genera.

5 CONCLUSIONS AND RECOMMENDATIONS

In the first part of the thesis, to provide fast and trustworthy identification of food- and waterborne bacteria, Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was applied to identify microorganisms from agricultural environment involving irrigation water, running water, manure and vegetables. A comprehensive picture about the identification performance of MALDI-TOF MS (Biotyper) was provided regarding environmental microbiology by analyzing 311 isolates. The results suggest that Gram-positive bacteria is more difficult to identify by MALDI-TOF MS (Biotyper) as lower (25.5%) species identification scores were achieved compared to Gram-negative isolates (40.9%). However, data obtained from the measurements indicates that MALDI-TOF MS (Biotyper) can be a reliable technique to identify bacteria from agricultural environment at genus level, but to realize similar results at species level, on which 36.6% identification scores were achieved, a database expansion with environmental isolates is inevitable.

Two types of MALDI-TOF MS sample preparation techniques, extended direct transfer procedure involving formic acid extraction and direct transfer procedure, were compared to determine which the best one is for environmental isolates. The average log scores of the former were above 2 therefore reached the species level threshold while the average log scores of the latter was only 1.85.

The effect of culture media on MALDI-TOF MS identification regarding waterborne bacteria was also tested using Tryptic Soy Agar, R2A and Yeast Extract Agar. It was proved by ANOVA that all three culture media are suitable to identify waterborne bacteria as no significant difference was observed regarding the averages of log scores. However, the species level identification threshold was not achieved using any of the culture media (TSA, 1.97; R2A, 1.95; YEA, 1.97). In these measurements, even higher identification scores could have been achieved by applying lower cutoff scores from 2 to ≥1.9, however that necessitates more measurements.

In the next part, it was shown that applying discriminant analysis on the mass spectra of isolates can help understanding the effect of culture media on the identification. Therefore, groups according to spectra obtained on specific culture media could be created which explained which culture medium the best is to use for the identification of *E. coli* and *S. aureus* isolates.

Moreover, it was possible to differ strains of *E. coli* from each other by applying principal component analysis on mass spectral data. Besides, using the same method it was also possible to distinguish MRSA from non-resistant *S. aureus*. Therefore, MALDI-TOF MS was able to

strain type of waterborne and antibiotic-resistant bacteria. In the future, a study expanded with more strains with an emphasis on antibiotic-resistant bacteria could be performed.

In the following section, a comparison of MALDI-TOF MS to 16S rRNA gene sequencing regarding waterborne isolates were performed. Both methods resulted similar identification outcomes as more than 60% of the waterborne isolates were identified at species level by Sanger sequencing and MALDI-TOF MS as well. The application of MALDI-TOF MS made it possible to identify more isolates at both species and genus level, but a paired t-test showed that the identification results of the two methods did not differ significantly. Interestingly, 11 of 42 (26.2%) isolates were identified differently with the only discrepancies were observed at species level in the case of 10 isolates. Therefore, it was successfully demonstrated that MALDI-TOF MS (Biotyper) can act as an alternative to 16S rRNA gene sequencing of isolates to identify waterborne bacteria due its rapid and accurate nature.

In the last section, MALDI-TOF MS was coupled with 16S rRNA gene amplicon sequencing to monitor the quality of irrigation water. As 99% of bacteria are not culturable, a lower concentration of microbial pathogens might not be cultivated which hinder the occurrence of those pathogens in irrigation waters. Interestingly, genera that were cultivated the most (*Acinetobacter*, *Pseudomonas*, *Enterobacter*, *Brevundimonas*) had a low relative abundance in the amplicon dataset. This highlights the fact that most of the environmental bacteria are uncultivable and the need for culture-independent methods to monitor natural waters. Based on this fact it would probably be worth to test more Hungarian irrigation waters by these methods.

6 THESES – NEW SCIENTIFIC RESULTS

1. For the first time, MALDI-TOF MS was utilized to identify bacteria from Hungarian agricultural environment involving irrigation water, running water, lakes, manure and vegetables. Thus, a broad picture about the identification performance of MALDI-TOF MS (Biotyper) was given regarding environmental microbiology by analyzing 311 bacteria with the genera of *Pseudomonas*, *Acinetobacter*, *Bacillus* and *Aeromonas* were the most frequently occurring ones. Moreover, the results of identifications also highlight the fact that environmental isolates of Gram-positive bacteria are more difficult to identify as lower (25.6%) species identification scores were obtained compared to Gram-negative isolates (40.9%).

2. I could demonstrate that extended direct transfer procedure was superior to identify environmental bacteria compared to direct transfer procedure as the averages of log scores of the former were 2.01, reaching the species level threshold while the latter only averaged 1.85. It was proven by paired t-test ($t=16.09$, $p<0.001$) that the difference was significant.

3. I could prove that by using discriminant analysis MALDI-TOF MS mass spectra can be separated based on the culture medium on which the given microbe was cultured. In addition, I found that Yeast Extract Agar and R2A agar can also be used to identify *E. coli* and *S. aureus* isolates using MALDI-TOF MS.

4. For the first time it was demonstrated that R2A and Yeast extract agar were also appropriate to identify waterborne isolates as no significant disparity was observed regarding the averages of log scores. However, the species level identification threshold was not accomplished using any of the culture media (TSA, 1.97; R2A, 1.95; YEA, 1.97).

5. I could demonstrate that it is possible to differentiate bacterial strains by combining MALDI-TOF MS with principal component analysis (PCA). New biomarkers were found for *E. coli* ATCC 13706 at m/z 6640 and m/z 8912 therefore it was possible to distinguish from *E. coli* DSM 11250. Moreover, a specific peak was detected at m/z 5868 in the mass spectrum of the methicillin-resistant *S. aureus* ATCC 43300 therefore it was possible to differentiate the antibiotic-resistant strain from the non-resistant *S. aureus*. Thus, present results bring a new point of view on identifying and discriminating bacteria by the MALDI-TOF MS technique utilizing PCA.

6. For the first time, the identification performance of MALDI-TOF MS using Hungarian waterborne bacterial isolates was tested against Sanger sequencing. Sanger sequencing and

MALDI-TOF MS generated almost identical results as 64.3% and 66.7% of the isolates were identified at species level with the methods. However, using MALDI-TOF more isolates were identified at both species and genus level, but a paired t-test proved that the identification results of the two methods did not differ significantly ($t(41)=2.02$; $p=0.57$). By comparing MALDI-TOF MS to 16S rRNA gene sequencing regarding waterborne isolates it was successfully proven that MALDI-TOF MS (Biotyper) can be a great option to 16S rRNA gene sequencing of isolates to identify waterborne bacteria due its fast and accurate nature.

7. For the first time MALDI-TOF MS was coupled with next-generation sequencing to monitor the irrigation waters in Eastern-Hungary. Results suggested that genera that were cultivated the most (*Acinetobacter*, 0.64%; *Pseudomonas*, 0.24%; *Enterobacter*, 0.04%; *Brevundimonas*, 2.18%) had a low relative abundance in the amplicon dataset. Therefore, highlighting the need for the culture-dependent techniques to be supplemented with culture-independent methods to monitor natural waters as only 8 of 188 (4.25%) genera, which had a relative abundance above 0.01%, were cultivated.

7 REFERENCES

Betts, R. (2014): 'Microbial update: fruit and salad', *International Food Hygiene*, 25(3); 9–12. doi: https://assets.thermofisher.com/TFS-Assets/LSG/brochures/International-Food-Hygiene-Vol25-No3-Fruit-and-Fresh-Produce.pdf

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., & Holmes, S.P. (2016): 'DADA2: high-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. doi: https://doi.org/10.1038/nmeth.3869

Cevallos-Cevallos, J.M., Gu, G., Richardson, S.M., Hu, J., & van Bruggen, A.H. (2014): 'Survival of *Salmonella enterica* Typhimurium in water amended with manure', *Journal of Food Protection*, 77(12), pp. 2035-42. doi: 10.4315/0362-028X.JFP-13-472

Dhariwal, A., Chong, J., Habib, S., King, I., Agellon, L.B., & Xia., J. (2017): 'MicrobiomeAnalyst – a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data', *Nucleic Acids Research*, 45(1), pp. 80-188. doi: https://doi.org/10.1093/nar/gkx295

European Food Safety Authority, European Centre for Disease Prevention and Control (EFSA and ECDC). (2019): 'The European Union One Health 2018 Zoonoses Report', *EFSA Journal*, 17(12), e05926. doi: https://doi.org/10.2903/j.efsa.2019.5926

European Food Safety Authority, European Centre for Disease Prevention and Control (EFSA and ECDC). (2021): 'The European Union One Health 2019 Zoonoses Report', *EFSA Journal*, 19(2), e06406. doi: https://doi.org/10.2903/j.efsa.2021.6406

Falardeau, J., Johnson, R.P., Pagotto, F., & Wang, S. (2017): 'Occurrence, characterization, and potential predictors of verotoxigenic *Escherichia coli*, *Listeria monocytogenes*, and *Salmonella* in surface water used for produce irrigation in the Lower Mainland of British Columbia, Canada', *PloS One*, 12(9), e0185437. doi: https://doi.org/10.1371/journal.pone.0185437

Gu, G., Luo, Z., Cevallos-Cevallos, J.M., Adams, P., Vellidis, G., Wright, A., & van Bruggen, A.H. (2013): 'Factors affecting the occurrence of *Escherichia coli* O157 contamination in irrigation ponds on produce farms in the Suwannee River Watershed', *Canadian Journal of Microbiology*, 59(3), pp. 175-82. Epub 2012 Dec 3. PMID: 23540335. doi: https://doi.org/10.1139/cjm-2012-0599

IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp. https://ibm.com

López-Fernández, H., Santos, H.M., Capelo, J.L., Fdez-Riverola, F., Glez-Peña, D., & Reboiro-Jato, M. (2015): 'Mass-Up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery', *BMC Bioinformatics*, 5(16), 318. doi: https://doi.org/10.1186/s12859-015-0752-4

Martin, M. (2011): 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet. journal*, 17(1), 10 doi: https://doi.org/10.14806/ej.17.1.200

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F.O. (2013): 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research*, 41(D1), pp. 590-596. doi: https://doi.org/10.1093/nar/gks1219

R Core Team. (2021): R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. 2021. https://www.r-project.org.

Schloss, P.D., & Handelsman, J. (2005): 'Metagenomics for studying unculturable microorganisms: cutting the Gordian knot', *Genome Biology*, 6(8), 229. doi: https://doi.org/10.1186/gb-2005-6-8-229

Strejcek, M., Smrhova, T., Junkova P., & Uhlik, O. (2018): 'Whole-Cell MALDI-TOF MS Versus 16S rRNA Gene Analysis for Identification and Dereplication of Recurrent Bacterial Isolates', *Frontiers in Microbiology*, 9, 1294. doi: https://doi.org/10.3389/fmicb.2018.01294

Turner, K., Moua, C.N., Hajmeer, M., Barnes, A., & Needham, M. (2019): 'Overview of Leafy Greens–Related Food Safety Incidents with a California Link: 1996 to 2016', *Journal of Food Protection*, 82(3), pp. 405–414. doi: https://doi.org/10.4315/0362-028X.JFP-18-316

Uyttendaele, M., Jaykus, L-A., Amoah, P., Chiodini, A., Cunliffe, D., Jacxsens, L., Holvoet, K., Korsten, L., Lau, M., McClure, P., Medema, G., Sampers, I., & Jasti, P.R. (2015): 'Microbial Hazards in Irrigation Water: Standards, Norms, and Testing to Manage Use of Water in Fresh Produce Primary Production', *Comprehensive Reviews in Food Science*, 14(4), pp. 336-356. doi: https://doi.org/10.1111/1541-4337.12133

8 PUBLICATIONS

Botond Bendegúz Surányi, Benjamin Zwirzitz, Csilla Mohácsi-Farkas 1, Tekla Engelhardt 3 and Konrad J. Domig **(2023)** Comparing the efficacy of MALDI-TOF MS and sequencing-Based identification techniques (Sanger and NGS) to monitor the microbial community of irrigation water. *Microorganisms,* 11(2), 287. **Q2, IF: 4.926.** doi: https://doi.org/10.3390/microorganisms11020287

Anna Jánosity, József Baranyi, **Botond Bendegúz Surányi**, Sonja Smole Možina, Andrea Taczman-Brückner, Gabriella Kiskó, Anja Klančnik. **(2023)** Estimating the optimal efflux inhibitor concentration of carvacrol as a function of the bacterial physiological state. *Frontiers in Microbiology*, 14, 1073798. **Q1, IF: 6.064.** doi: https://doi.org/10.3389/fmicb.2023.1073798

Botond Bendegúz Surányi, Andrea Taczman-Brückner, Csilla Mohácsi-Farkas, Tekla Engelhardt. **(2023)** Rapid identification of bacteria from agricultural environment using MALDI-TOF MS. *Acta Alimentaria: An International Journal of Food Science*, 52, 1. **Q3, IF: 1.000. doi:** https://doi.org/10.1556/066.2022.00202

CONFERENCE PRESENTATIONS

Botond Surányi, Tekla Engelhardt, Csilla Mohácsi-Farkas. **(2022)** Analyzing the effect of culture media composition on Gram-positive bacteria's peptide mass fingerprint. 4th FoodConf - International Conference on Food Science and Technology. Book of Abstracts, p. 79. Budapest, Magyarország. ISBN: 9786150154220.

Andrea Taczman Brückner, Ivett Juhász, Vivien Dancs, Hajnalka Erdős, **Botond Surányi**, Tamás Kocsis, Gabriella Kiskó. **(2022)** Removal of Pseudomonas aeruginosa biofilm in plastic bottles filled with different beverages. 4th FoodConf - International Conference on Food Science and Technology. Book of Abstracts, p. 29. 1 p. ISBN: 9786150154220.

Botond Surányi, Csilla Mohácsi-Farkas, Tekla Engelhardt. **(2022)** Comparing a rapid- and a molecular biology identification method using environmental samples. MTA Kémiai Tudományok Osztálya, Élelmiszer-tudományi Bizottság, Élelmiszer-mikrobiológiai és Élelmiszer-biztonsági kollokvium.

Botond Surányi, Tekla Engelhardt, Csilla Mohácsi-Farkas. **(2021)** Mass spectrometry-based rapid identification of food- and waterborne microorganisms. MiCEnt 2021: Integrative

Biology Symposium: Microbiology, Enteric Nervous System, Central Nervous System (oral peresentation)

Botond Surányi, Engelhardt Tekla, Mohácsiné Farkas Csilla. **(2021)** Víz- és élelmiszer-eredetű mikroorganizmusok gyors módszeres azonosítása. MTA Élelmiszertudományi Tudományos Bizottság 382. Tudományos Kollokviuma (Online conference)

Taczmanné Brückner Andrea, Juhász Ivett, Dancs Vivien, Erdős Hajnalka, **Surányi Botond**, Kocsis Tamás, Kiskó Gabriella. **(2021)** Biofilm kialakulása és háztartási módszerekkel történő eltávolításának hatékonysága többször használatos műanyag ivópalackban. Lippay János – Ormos Imre – Vas Károly (LOV) Tudományos Ülésszak : Összefoglalók, Budapest, Magyarország, pp 129-129. ISBN: 9786150137384

Botond Surányi, Christoph Schönher, Philipp Proksch, Marija Zunabovic-Pichler, Ivett Juhász, Andrea Taczmann-Brückner, Engelhardt Tekla. **(2021)** Analysis of water- and foodborne bacterial mass spectra. FoodMicro2022 (abstract)

Botond Surányi, Tamás Kocsis, Tekla Engelhardt, Csilla Mohácsiné Farkas. **(2019)** Identification of microorganisms of irrigation water samples with MALDI-TOF MS. BiosysFoodEng 2019 - Proceedings : 3rd International Conference on Biosystems and Food Engineering. Budapest, Magyarország (poster presentation)

Botond Surányi, Tamás Kocsis, Tekla Engelhardt, Csilla Mohácsiné Farkas. **(2019)** Detecting microbes with MALDI-TOF MS from irrigation water samples. SZIENtific meeting for young researchers - Ifjú Tehetségek Találkozója (ITT), Budapest (poster presentation)