Food Research International, Volume 197, Issue Pt 1, December 2024, Article number 115207
DOI:10.1016/j.foodres.2024.115207

Contribution of data acquired from spectroscopic, genomic and microbiological analyses to enhance mussels' quality assessment.

Anastasia Lytou[1], Léa Saxton[2], Lemonia-Christina Fengou[1], Dimitrios A. Anagnostopoulos[3], Foteini F. Parlapani[3], Ioannis S. Boziaris[3], Fady Mohareb[2] George-John Nychas[1,4]

[1]Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Human Nutrition, School of Food and Nutritional Sciences, Agricultural University of Athens, 11855 Athens, Greece

[2]Bioinformatics Group, Department of Agrifood, School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire MK43 0AL, United Kingdom,

[3]Laboratory of Marketing and Technology of Aquatic Products and Foods, Department of Ichthyology and Aquatic Environment, School of Agricultural Sciences, University of Thessaly, Fytokou Street, 38446 Volos, Greece

[4]International Joint Research Lab (China and Greece) of Digital Transformation as an Enabler for Food Safety and Sustainability, Taian 271018, China

* Corresponding author: gjn@aua.gr

Abstract

In this study, a large amount of heterogeneous data (i.e., microbiological, spectral and Next Generation Sequencing data) were obtained analyzing mussels of different species and origin, to acquire a comprehensive view about the quality and safety of these products. More specifically, spectral data were collected through Fourier transform Infrared (FTIR) spectroscopy, while the overall profile of microorganisms present in these samples, affecting quality and safety of mussels throughout storage, was determined through Next Generation Sequencing (NGS) using 16S rRNA metabarcoding analysis. In parallel, conventional microbiological analysis for the estimation of culturable spoilage microorganisms (total aerobes, *Pseudomonas* spp., *B. thermosphacta*, *Shewanella* spp. and Enterobacteriaceae) was applied. Different machine learning algorithms, namely Partial Least Square (PLS), Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Random Forest (RF) Neural Networks (NN)) were applied accordingly, to assess the potential of FTIR and NGS data to provide useful information about mussels' microbiological quality.

Microbial counts ranged from 3.5 to 9.0 log CFU/g, while NGS revealed several bacterial genera such as *Pseudoalteromonas*, *Psychrobacter*, *Acinetobacter*, *Pseudomonas*, *B. thermosphacta*, *Psychrobacter*, *Kistimonas*, *Psychrilyobacter* to affect the quality of mussels, depending on the mussel species, batch and storage conditions. According to the performance metrics, the SVM algorithm in tandem with FTIR achieved the highest prediction accuracy for microbial counts in *M. chilensis* samples (Rsquared;0.89, RMSE; 0,74), while in the case of predicting the abundance of microbial genera using spectroscopic data, the best performing algorithm varied by bacterial genus. Indicatively, in *M. chilensis*, RF, kNN and NN performed better in predicting *Enterococcus*, *Enhydrobacterium* and *Pseudoalteromonas*, respectively (Rsquared = 0.92, 0.93, 0.99). Associations between genomics data and specific spectral regions were further investigated, revealing certain spectral regions that are associated with mussels' quality and safety.

The application of "multi-omics" in seafood supply chain can provide insightful information about mussels' quality and safety compared to the methodologies followed in current quality and safety management systems.

Keywords: Next Generation Sequencing, FTIR, spoilage, multi-omics, seafood, machine learning

1. Introduction

The current safety and quality controls in the food chain are based on EU legislation 2073/2005. In this regulation, the microbiological criteria for certain micro-organisms (e.g., *E. coli, Salmonella, Listeria monocytogenes, Enterobacteriaceae, Aerobic colony count*) and the implementing rules to be complied with by food business operators when realizing the general and specific hygiene measures referred to in Article 4 of Regulation (EC) No 852/2004, have been defined.

After many years of implementation of this regulation, it is evident that the set measures are lacking or inadequately applied and fail to prevent microbial and/or chemical contamination of food products, which leads to reduced confidence among consumers. This is mainly because (i) the production, supply, and processing sectors of the food chain are fragmented, and this lack of cohesion results in a failure to adopt new and innovative technologies, products, and processes (ii) the complexity of the food systems means that to meet market demands, food business operators (producers, retailers, resellers), and regulators, need to develop and apply structured quality and safety assurance systems based on thorough risk analysis and prevention through monitoring, recording, and controlling critical parameters covering the entire product's life cycle and (iii) the outcome of the conventional microbiological, chemical, or molecular techniques used for the analysis of the end product is a single number from which limited and rather retrospective information can be drawn.

The development and implementation of Microbiological Risk Assessment (MRA), a structured process for determining the risks associated with biological hazards in a food, is considered a fundamental step to fulfil this need (CAC, 1999). Additionally, the methodology of Quantitative MRA (QMRA) which includes hazard identification, hazard characterization, quantification of exposure assessment and finally risk

characterization, has the ability and/or the potential enhancing the reliability of risk estimates. The QMRA approach which is being used worldwide provides a theoretical basis for hazard analysis, determination of the critical control points and formulation of food safety goals and regulations (Jiang et al. 2023; Baydakova et al., 2019; Hiroki & Kawasaki, 2024)

Over the last decade, the potential of omics data (including those derived from non-invasive rapid instruments) for exposure assessment and hazard characterisation in tandem with the tools for this approach, have been extensively discussed in the literature (den Besten et al, 2018; Cocolin et al 2018; Rantsiou et al 2018; Haddad et al., 2018).

In comparison with the current methodology, where the outcome is a single number, the inclusion of omics generates a massive amount of data, that should be processed in tandem with data analytics techniques i.e., their analysis provides an enormous amount of information that the stakeholder should interpret under time constraints.

To effectively handle the vast amount of data, it is important to implement proper data management practices and maintain systematic metadata documentation. Moreover, to extract precise insights from the data, expertise in bioinformatics is indispensable. Such specialized knowledge is essential for conducting thorough analysis and drawing correct interpretations (Kumar et al., 2023; Yap et al., 2022; Nychas et al. 2021). At the same time, the interpretation of the generated results from biological perspective is considered as highly important, and this could upgrade the quality of the final output. To achieve that, experts with biological background play a key role allowing correlations and interpretation of all the information included.

Previous studies focused mainly on the discussion about the integration of genomic, metagenomic and transcriptomic data (Cocolin et al., 2018; Sequino et al., 2022).

Some others included metabolomics data as well, but with focus on targeted analysis of the metabolome data acquired from GC-MS and LC-MS techniques (Borges et al., 2022; Herráiz-Gil et al., 2023; Garcia-Perez et al. 2024). In this study, metabolomic data extracted from rapid and non-destructive IR spectroscopy was used for the prediction of both microbial counts and relative percentages of microbial species in the microbial community of stored mussels.

The aim of this work is to provide - for first time to our knowledge - an insight about how different omics data can be used to improve quality assessment systems and decision-making *via* a. characterizing the mussel's microbiota and its changes over time under different conditions, b. finding the fingerprint of each sample and investigating of how its changes throughout storage can be exploited for the prediction of the microbiological quality and c. correlating results obtained from different analyses and datasets in order to extract as much useful information as possible from these large datasets. This study is a major breakthrough as integrates rapid, non-invasive FTIR analysis with genomic and microbiological data for a more comprehensive and reliable assessment of seafood quality. This novel approach, where the results of one analysis can confirm or reject the findings of another, provides an elevated level of detail and accuracy in at-line quality evaluation. Furthermore, the integration of these diverse analytical techniques allows for a deeper understanding of mussel quality by uncovering insightful details from each individual analysis. These insights based upon omics techniques will increase our knowledge on how changes on certain characteristics (or physical attributes) of food could affect the quality level, and consequently will reduce our uncertainty leading to more accurate quality and safety assessment.

2. Materials and Methods

2.1. Experimental procedure

A graphical abstract of the overall experimental process is illustrated in Figure 1. In brief, it includes the a. collection of different types of mussels (different species of different origin), b. storage at different temperature conditions for certain days to monitor the spoilage, c. acquisition of microbiological -in terms of microbial counts-, NGS and spectral data and d. application of machine learning algorithms to find out any correlations between the different types of data.

For this reason, three different batches of mussels belonging to *Mytilus galloprovincialis* species, kept without the shell in packages with seawater and sold as fresh products, were collected. The term batch refers to a specific quantity of samples produced at a specific date. One of these was of Spanish origin (batch 1) while two of them were from two different Greek producers of two geographical areas (Thermaikos (batch 2) and Strymonikos gulf (batch 3)). Additionally, 2 batches of frozen/thawed *Mytilus chilensis* mussels without the shell of two different brands were also tested (batch 1 and batch 2). Samples were transferred to the laboratory in polystyrene packages under refrigerated conditions. *M. galloprovincialis* products, were then stored at 2 and 4 °C for 9 days, while frozen samples belonging to *M. chilensis* species, were initially kept at 2 °C until thawing (5-6 hours) and then stored at 0, 4 and 6 °C for 10 days, simulating the proposed storage temperature (0 – 4 °C) as well as a slight temperature abuse. Microbiological and FTIR analysis was performed at 6 time points throughout storage. For NGS analysis, samples of both species from the beginning, middle and end of storage were selected. The selection was based on their microbial load, aiming to analyze the microbial profile of fresh, semi-fresh and early

spoiled samples. In Table 1, the number and the type of samples for the two species (per batch) used in each of the analyses is summarized.

2.2. Microbiological analysis

Each sample (a homogenate of two mussels) was weighed (25 g) and 225 mL of quarter strength Ringer's solution was added (Lab M Limited, Lancashire, UK) in a Stomacher bag (400-mL sterile, Seward Medical, London, UK), while the mixture was homogenized in a Stomacher apparatus (Lab Blender 400, Seward Medical, London, UK) for 60 s at room temperature. The analysis was performed under aseptic conditions. Subsequently, serial decimal dilutions were prepared, and aliquots of 0.1 mL from the appropriate decimal dilution was spread on tryptic glucose yeast agar (Plate Count Agar; Biolife, Milan, Italy) followed by incubation at 30 °C for 48 h, for the enumeration of total aerobic counts (TAC).

Pseudomonas spp., *Shewanella* spp. and bacteria of Enterobacteriaceae family counts were also estimated at six certain time points throughout storage. The enumeration of the microbial populations for each group was performed with the standard count method pour or spread plate depending on the examined microbial group and the respective substrate. The substrates, incubation time and temperatures were used, have been described in detail before (Broekaert et al., 2011; Fengou et al., 2022). In particular, a spread method was applied for the enumeration of *Pseudomonas* spp. on Pseudomonas agar base with the selective supplement cephalothin-fucidin-cetrimide (Lab M Limited, Lancashire, UK), and incubated at 25 °C for 48 h. Additionally, 1.0 mL of the appropriate serial decimal dilutions of the mussels homogenate was pour-plated in violet red bile glucose agar (VRBG, Biolife, Milan, Italy) for the enumeration of bacteria belonging to the Enterobacteriaceae family, overlaid with the

same medium and incubated at 37 °C for 24 h, while iron agar (IA) was used for the enumeration of H_2S producing bacteria by counting black colonies, after incubation at 25 °C for 3 d, according to Gram et al. [19]. After incubation, typical colonies for each microbial group were enumerated and colony counts were logarithmically transformed (log CFU/g).

2.3. 16S rRNA metabarcoding analysis

2.3.1. Samples preparation and DNA extraction

Both sample processing and DNA extraction were applied as described previously by Anagnostopoulos et al., (2022b). Briefly, a total of 25 g of each sample (pooled sample from nine samples (n=9) per batch at each sampling point) was transferred aseptically to stomacher bags, followed by the addition of 225 mL sterile saline solution (0.85% w/v) and homogenization for 4 min in a Stomacher. The homogenized suspension was transferred to sterile centrifuge tubes and centrifuged at 136 x *g* for 5 min at 20 °C to remove particles. Then, the supernatant was transferred to sterile centrifuge tubes and centrifuged at 2067 x *g* for 15 min at 20 °C. Finally, the pellet was diluted in 1.0 mL of sterile deionized H_2O. DNA extraction was applied using the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) according to manufacturer's instructions and DNA concentration and quality were measured on a nanodrop Quawell UV-Vis Spectrophotometer Q5000 (Quawell Technology, Inc., USA).

2.3.2. Library preparation, Sequencing and Bioinformatic analysis

High Throughput Sequencing (HTS) analysis was applied using next generation technology (bTEFAP®) according to Dowd et al., (2008), using the

primers 341F (CCTACGGGNGGCWGCAG) and 785R (GACTACHVGGGTTCCATATCC). Each sample was subjected to a Polymerase Chain Reaction (PCR) for a total of 30 cycles, using the HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA, USA). PCR conditions were as follows: initial denaturation at 94 °C for 3 min, followed by 30 cycles of 94 °C for 30 s, 53 °C for 40 s and 72 °C for 1 min. and a final elongation at 72 °C for 5 min (Syropoulou et al., 2021). The amplification was checked on agarose gel (2% w /v) and then the amplicons were mixed in equal concentrations, purified using the Ampure XP beads, and sequenced on an Illumina MiSeq platform according to manufacturer's protocol.

Raw sequences were processed using the MR DNA ribosomal and functional gene analysis pipeline (www.mrdnalab.com, MR DNA, Shallowater, TX), according to Anagnostopoulos et al., (2022b). The obtained zOTUs were taxonomically classified using BLASTn against a curated database derived from National Center for Biotechnology Information (NCBI) and generated into relative abundances at different taxonomic levels. The estimation of alpha diversity was performed as described previously (Dowd et al., 2008; Edgar and Bateman, 2010; Eren et al., 2011; Swanson et al., 2010) using Quantitative Insights Into Microbial Ecology 2 (Qiime 2) (Bolyen et al., 2019). All reads were rarefied to ~10,000 sequences using the DADA2 algorithm and Alpha rarefaction curve was plotted with 10 sampling depths. Finally, raw sequences were deposited in National Centre for Biotechnology Information (NCBI), under the Bioproject PRJNA1091896.

2.4. FTIR analysis

An FTIR-6200 JASCO spectrometer (Jasco Corp., Tokyo, Japan) equipped with a sample chamber, a triglycine sulphate (TGS) detector and a Ge/KBr beamsplitter was

used along with the Spectra Manager™ Code of Federal Regulations (CFR) software version 2 (Jasco Corp., Tokyo, Japan). For spectra acquisition, a ZnSe 45° horizontal attenuated total reflectance (HATR) crystal (PIKE Technologies, Madison, Wisconsin, USA) was utilizied. A 10 g piece of the sample was cut and placed on the HATR crystal. Each measurement was acquired by accumulating 100 scans with a resolution of 4 cm^{-1} within a total integration time of 2 min. Reference spectra were acquired at regular time intervals using the cleaned crystal with no sample. FTIR wavenumbers used for the analysis were in the range of 4000 to 700 cm^{-1}, excluding the noisy signal at the edges of the spectrum from the downstream analysis. This approach is consistent with the methodology used by Argyri et al., 2013.

2.5. Data analysis – Predictive models development, Important features

2.5.1. Prediction of microbial counts using NGS data

A PLS-regression model was used for the prediction of microbial counts using NGS data. This analysis also revealed the important microbial genera for the estimation of microbial population and was performed using the XLSTAT, 2020 software. The dataset was split in training and testing subsets, where 75% of the whole dataset was used to train the models and 25% of it to validate them. Five different training and testing subsets were created and analyzed while, the models' performance was evaluated using the average of Root Mean Squared Error (RMSE) and coefficient of determination (R-squared) values.

2.5.2. Prediction of microbial counts using FTIR data

For the prediction of microbial counts (i.e., total aerobic counts - TAC) using FTIR data, as well as for finding the most informative features the methodology described

in Fengou et al., 2024, was followed. Stratified sampling was applied on the data so 75 – 25% was used to train and test the models (i.e., unknown samples), respectively. PLS was used for the selection and extraction of the features. (i.e., FTIR wavenumbers), which were used as X-variables. PLS was applied on centered and scaled data for the extraction of the PLS components for the training and testing the Support Vector Machines (SVM) models. For further analysis the features with absolute value of regression coefficients >0.05 were selected. SVM with radial basis function (RBF) was applied on the selected and transformed features for the prediction of TAC (Y-variable). Hyperparameters (i.e., cost and gamma) of the SVM models were tuned with grid search method coupled with 10-fold cross validation of the training set. The PLS, SVM models were implemented in R version 4.0.3 and Rstudio version 2022.2.3.492 using the packages 'e1071′ for SVM (Meyer et al., 2021) and'pls' for PLS (Mevik et al., 2020). The models' performance was evaluated using the average of Root Mean Squared Error (RMSE) and coefficient of determination (R-squared) values.

2.5.3. Prediction of the microbial profile (%) using FTIR data

FTIR-NGS data pre-treatment

The analysis was conducted utilizing the open-source software platform *R*. The *R* script utilised several custom-defined functions to automate the process for each mussel's dataset.

The raw FTIR and NGS data underwent several pre-processing steps before analysis to ensure data quality, reliability and consistency. Firstly, rows containing missing values (NA) were removed from the dataset to avoid potential bias in the analysis and helps maintaining the integrity of the dataset. Then, irrelevant columns were

eliminated. Furthermore, baseline correction was performed using the Asymmetric Least Squares (ALS) algorithm from baseline package to mitigate systematic errors in spectral data. Following baseline correction, the occurrences of unique values are computed within the NGS coding column from the FTIR dataset, facilitating the identification of relevant genetic codes. Subsequently, the NGS dataset is filtered based on genus names, ensuring that only pertinent data is retained for further analysis. Finally, through a meticulous process, NGS data rows are duplicated according to the frequency of occurrence of each genus counted from the FTIR dataset, thereby enabling consistency between both datasets for downstream analysis.

Analysis and mathematical modelling of bacterial composition

The analysis focused on understanding the bacterial composition associated with different samples and exploring the potential of predicting bacterial count from NGS data using FTIR features. To achieve this, a second customed-defined function was employed to investigate correlations between spectral features and the presence of specific bacteria. Initially, for illustration purposes, the function prepares the data by extracting relevant columns and reducing feature dimensionality. Subsequently, it computes correlations between the reduced FTIR dataset and each column in the genetic sequencing dataset. The results are formatted into a structured data frame (correlations) to facilitate interpretation. Importantly, visualization is employed to enhance understanding, with the correlation matrix represented as a heatmap using gplots package. This visual depiction enables to discern patterns and relationships between spectral features and bacterial abundances. For the rest of the analysis, the dataset dimensionality was reduced by employing Boruta feature selection algorithm, utilized to identify significant features. Finally, machine learning techniques including

Random Forest (RF), k-Nearest Neighbors (KNN), and Neural Network (NN) models were applied to construct and validate the performance of the developed models to predict genetic sequencing data based on related spectral features. The process involves splitting the data into training and testing sets (80-20), employing cross-validation for model evaluation, and calculating performance metrics such as Root Mean Square Error (RMSE) and R-squared. Iterative procedures are conducted to assess model performance across multiple iterations using various randomized data splitting, with results aggregated to determine mean RMSE and R-squared values. Additionally, plots showing the predicted against the observed bacterial composition are generated for each bacteria species of interest to aid in understanding model predictions and assessing model accuracy.

The described methods were applied to two distinct sample groups: *M.chilensis* and *M.galloprovincialis*. For each group, the respective datasets underwent identical pre-processing and analysis steps.

3. Results

3.1. Microbial changes

The population ranges of total aerobic bacteria and specific spoilage microorganisms in the two mussel species throughout storage are presented in Figure 2. In both mussel species, *Pseudomonas* spp. and *Shewanella* spp. are related to spoilage, covering populations ranging from 1.0 to 8.5 log CFU/g in *M. chilensis* and 1.0 to 7.5 log CFU/g in *M. galloprovincialis*. It needs to be noted that the distribution of microbial counts throughout storage differed significantly between the two species with M. chilensis exhibiting a wider range of microbial counts compared to *M. galloprovincialis*.

3.2. Bacterial diversity of mussels

According to NGS analysis, a total of 1,280,470 raw reads were obtained, while after quality filtering, a total of 781,742 were retained (average of about 27,919 per sample) (Table S1). These high-quality sequences corresponded to 5,973 observed features (ranged from 52 to 421). The rarefaction to ~10,000 sequences for estimation of bacterial diversity was satisfying, since, for example, the Shannon-Wiener Index curves plot (Fig. S1) reached a plateau at ~2,000 sequences, indicating a sufficient sequencing depth to characterize bacterial communities and diversity of the studied samples.

The results of metabarcoding analysis were expressed both at phylum and genus level, but the genus level data provided more specific insights into the objectives of this work.

In M. chilensis samples (Figure 3), the dominance of certain microbial groups throughout storage was more apparent compared to *M. galloprovincialis*, as the microbial diversity was distinctively lower in late storage samples compared to the beginning of storage. In batch 1, *Psychrobacter* and *Pseudoalteromonas* were the dominant genera, while *Pseudomonas*, *B. thermosphacta* and *Shewanella* increased their relative abundance throughout storage. *Psychrobacter* was also the dominant bacterial genus in *M. chilensis* samples of batch 2, followed by *Pseudomonas* and *Acinetobacter*.

M. galloprovincialis samples (Figure 4), displayed a high and complex bacterial diversity at the beginning of storage, while a distinct microbial profile was observed in the three batches throughout storage. At the beginning of storage, several bacterial genera (e.g., *Psychrobacter*, *Pseudoalteromonas*, *Vibrio*, *Psychrilyobacter*, *Exiguobacterium* and *Leuconostoc*) were identified. In batch 1, *Psychrobacter* and

Pseudoalteromonas co-dominated at both 2 and 4 ºC at the end of the shelf life. Batch 2 samples exhibited higher microbial diversity even at late storage, with genera such as *Kistimonas*, *Psychrobacter*, *Pseudomonas*, *Acinetobacter*, *Ralstonia* and *Shewanella* present at a remarkable proportions. In batch 3 *M. galloprovincialis* samples, a different profile was observed at the end of shelf-life, with *Psychrilyobacter* and *Kistimonas* prevailing at 2 ºC and 4 ºC, respectively, followed by *Ralstonia* and *Corynebacterium.*

3.3. FTIR measurements

Typical FTIR spectra of the two species in the ‘’fingerprint’’ area (800-1800 cm^{-1}) are shown in Figure 5. Although, some hidden information may not be immediately apparent from the spectral representation, the differentiation between the two species is still evident. Bands at 1240, 1456, 1340, 1046 cm^{-1} appear to be characteristic not only for the discrimination of the two species, but also for differentiating samples based on their freshness, providing valuable insights into the quality status of these products.

3.4. Correlations between NGS data and microbial counts

Table 2 and Figure 6 present the performance indices for the prediction of microbial counts as well as the standardized coefficients for estimating TAC using as input variables the NGS results. In *M. chilensis* (two batches combined), the PLS model successfully predicted microbial counts for all tested microbial groups. In *M. galloprovincialis* (three batches combined), TAC and *Shewanella* counts were accurately predicted, while the model’s performance was lower for predicting *Pseudomonas* and Enterobacteriaceae populations. As previously indicated, spoilage progression in *M. chilensis* samples was more gradual, with microbial growth

throughout storage following a more typical microbial growth pattern compared to *M. galloprovincialis*. Additionally, deviations among batches both in microbial populations and profiles were smaller in *M. chilensis* compared to *M. galloprovincialis* samples. Based on the standardized coefficients for microbial prediction in Figure 6 in *M. galloprovincialis,* apart from microbial groups that were highly abundant and increased in percentage throughout storage (e.g., *Pseudoalteromonas*, *Psychrobacter*, *Kistimonas*, *Psychlilyobacter*, etc.) there is also significant information in bacteria with lower abundance (relative percentage). For example, *B. thermosphacta*, *Psychromonas*, *Carnobacterium* and *Aeromonas,* which could be critical for estimating microbial quality*,* are some of the bacteria strongly correlated with microbial counts increases despite being present at lower ratios in the microbial community. In *M. chilensis* mussels, aside from *Carnobacterium*, the most informative features (i.e., all the positively correlated bacteria species) were also present in high percentages at the end of the shelf life, as described in section 3.2.

3.5. Correlations between FTIR and conventional microbiological data

The performance indices (i.e., RMSE, R-squared) for predicting TAC using FTIR data showed good results for both mussel species (Table 3) while Figure 7 presents the selected features for the two different species.

The changes caused by microbiological activity were successfully captured by the FTIR spectra, as indicated by RMSE values of <1.0 log for the test sets of both *M. chilensis* and *M. galloprovincialis* models. The bacterial diversity of the mussels did not significantly affect model performance. Although *M. galloprovincialis* exhibited greater bacterial diversity across different batches compared to *M. chilensis*, the FTIR prediction results showed slightly better performance in terms of RMSE and the

train/test performance was very similar indicating no overfitting. The FTIR spectra captured the overall metabolic profile, and when combined with ML methods, accounted for slight differences due to microbial metabolic activity by focusing on the most informative features. The most important regions contributing to the microbial quality characterization of both mussel species were primarily found in the region from 900 to 1300 (i.e., 920, 945, 971, 984, 1080, 1300-1310, 1375-1382 cm^{-1}), with further important wavenumbers identified in the region 1400 to 2000 cm^{-1} (i.e., 1497-1530, 1550-1566, 1623, 1636-1639, 1651, 1660, 1682, 1684, 1696-1699, 1716, 1725-1727, 1919 cm^{-1}). Moreover, highly informative regions/wavenumbers were also found in the range from 2300 to3000 cm^{-1} (i.e., 2301, 2308-2312, 2327, 2345, 2347, 2350, 2355-2364, 2398-2399, 2439, 2664, 2813, 2909-2914, 2923-2940 cm^{-1}). These regions were the most informative among the selected features and were common for both mussel species in predicting microbial population.

3.6. Correlations between FTIR and NGS data

In this study, Fourier Transform Infrared Spectroscopy (FTIR) data was utilized alongside machine learning techniques to predict microbial abundances from Next-Generation Sequencing (NGS) data in two mussels' species. The study aimed to investigate correlations between FTIR wavenumbers and the abundance of various bacterial species identified through metagenomic analysis. Tables 4 and 5 show the performance of each algorithm in predicting the abundance of each microbial genera using FTIR data, while Figures 8 and 9 highlight the most characteristic wavenumbers contributing to this prediction for each microbial genus. According to the results presented in Table 4, the algorithms' performance varied by microbial genus. For instance, in *M. chilensis*, the Random Forest algorithm performed better in predicting

Enterococcus, Staphylococcus, *Rhodococcus, Flavobacterium and Propionibacterium*, whilst the kNN model was more effective at estimating *Enhydrobaterium, Janthinobacterium* and *Shigella*. *Aeromonas, Pseudomonas, Pseudoalteromonas, Vibrio, Leuconostoc* and *Lactococcus* were better estimated using Neural Networks. In *M. galloprovincialis*, RF, kNN and NN performed best in predicting *Clostridium, Bacillus* and *Psychrobacter*, respectively. Although the algorithms showed generally poorer performance in this species compared to *M. chilensis*, all the tested algorithms successfully predicted the abundance *of Pseudomonas* and *Micrococcus* during storage. Figure 8 presents a correlation matrix for *M. chilensis* mussels, highlighting the relationship between FTIR wavenumbers and the abundance of various bacterial pathogens, spoilage microorganisms, and other significant bacterial species. In the heatmap, darker blue color indicates a stronger correlation. Among the most informative wavenumbers, some appear to be associated with the presence of multiple microorganisms while others showcase the presence of certain bacteria. It should be noted that within some large ranges of significant wavenumbers, there are also small areas with less discriminative power.

Figure 8A presents a correlation matrix illustrating the relationship between FTIR wavenumbers (y-axis) and bacterial pathogens (x-axis). Notably, several key findings emerged from the analysis. For *Bacillus*, *Staphylococcus*, and *Vibrio* bacteria, significant correlations were observed in certain wavenumbers within the range of 2531 to 4000 cm^{-1}, suggesting a consistent spectral signature across these pathogens in this spectral region. *Aeromonas* bacteria exhibited high correlation coefficients within the FTIR wavenumber ranges of 1374 to 2145 cm^{-1}, and 3109 to 3206 cm^{-1}, with additional important correlations observed up to 3978 cm^{-1}, indicating distinct spectral features associated with *Aeromonas* in these intervals. *Enterococcus* bacteria

showed a noteworthy correlation between FTIR wavenumbers 3014 and 3882 cm^{-1}, suggesting specific molecular vibrations or functional groups associated with the presence of *Enterococcus*.

Clostridium and *Streptococcus* bacteria displayed strong correlations between the wavenumbers 3207 and 3302 cm^{-1}, implying similarities in biochemical composition or structural features between these bacterial species within this spectral region.

In Figure 8B, focusing on bacteria species identified as spoilers, notable correlations between FTIR wavenumbers and microbial abundances were observed, highlighting distinct spectral signatures associated with each species. *Pseudomonas* and *Acinetobacter* bacteria exhibited strong correlations within the ranges of 796 to 891 cm^{-1}, 3399 to 3688 cm^{-1}, as well as 3785 to 3978 cm^{-1}. Similarly, *Pseudoalteromonas*, *Brochothrix*, displayed significant correlation within narrower ranges, indicatively between 1759 and 1952 cm^{-1} and 2724 to 2820 cm^{-1}. In addition to these ranges, *Lactococcus* and *Shewanella*, showed a broader range of important wavenumbers extending up to 3978 cm^{-1}. Bacteria belonging to *Propionibacterium* genus showcased a correlation peak within the wavenumber range of 2916 to 3978 cm-1. Conversely, *Shigella*, *Actinomyces*, and *Corynebacterium* demonstrated high correlations between the wavenumbers 3207 and 3302 cm^{-1}. Interestingly, there was no common spectral range shared among all spoiler bacteria species, indicating unique biochemical compositions or structural features associated with each.

In Figure 8C, the analysis extends to a list of bacteria of significant importance, revealing also distinct correlations between FTIR wavenumbers and microbial abundances. Notably, *Photobacterium*, *Arcobacter*, *Janthinobacterium*, and *Tenacibaculum* exhibited strong correlations within the wavenumber range of 2531 to 4000 cm^{-1}, suggesting consistent spectral signatures across these bacterial species

within this broad spectral region. Conversely, *Flavobacterium*, *Exiguobacterium*, *Ralstonia*, and *Rhodococcus* displayed significant correlations spanning from 1181 to 2049 cm^{-1}. *Enhydrobacter* and *Kistimonas* demonstrated correlation peaks between FTIR wavenumbers 2531 and 2820, 3496 and 3591 as well as 3978 and 4000 cm^{-1} underscoring distinct spectral features associated with these taxa. Most of the bacterial species under investigation displayed correlations within the wavenumber range of 2821 to 3495 cm^{-1}, suggesting the potential utility of this spectral region for comprehensive microbial detection, as discussed earlier.

All three heatmaps exhibited strong correlations for each bacterial species, regardless of their respective categories. These findings align with the results presented in Table 4**,** which evaluates the performance of kNN, Random Forest, and Neural Network models in predicting bacterial abundances using FTIR data for *M. chilensis*, as demonstrated by their low RMSE and high R-squared values.

In Figure 9, the correlation matrix for *M. galloprovincialis* mussels unveils distinctive patterns in the relationship between FTIR wavenumbers and bacterial abundances, highlighting variations across different bacterial species. In Figure 9A, focusing on pathogenic bacteria, several notable observations emerge. *Enterococcus* and *Aeromonas* display no significant correlations across all wavenumber ranges, consistent with the results in Table 5, which reported high RMSE and low R-squared values for these bacteria across all three predictive models (kNN, RF, and NN). *Staphylococcus* and especially *Vibrio* exhibit pronounced correlations within the wavenumber range of 892 to 2145 cm^{-1}, while *Clostridium* and *Streptococcus* show high correlations at 1952 cm^{-1} and 2049 cm^{-1}, respectively. *Bacillus* demonstrates higher correlation coefficients at wavenumbers above 3495 cm^{-1}.

In Figure 9B, focusing on spoilage bacteria, a less pronounced correlation profile is evident, aligning with the findings in Table 5, in which high RMSE and low R-squared values were reported across all three predictive models. *Pseudomonas,* one of the spoilage bacteria with satisfactory performance metrics*,* is identifiable at specific wavenumbers across the entire spectrum, with higher correlations observed between 1570-1950, 2050-2240, 2820-3300 cm^{-1}. A similar profile was observed for *Pseudoalteromonas* albeit with lower correlation values. *Micrococcus*, *Psychromonas* and *Actinomyces* exhibit relatively high correlations at certain wavenumbers throughout the spectrum.

In the analysis of bacteria of significant importance, illustrated in Figure 9C, for *M. galloprovincialis, Arcobacter*, *Janthinobacterium*, and *Flavobacterium* displayed similar correlation patterns within the wavenumber range of 3592 to 4000 cm^{-1}, suggesting shared spectral signatures indicative of their presence or abundance. *Rhodococcus* showed a profile similar to *Flavobacterium* while *Ralstonia* exhibited a broad range of high correlations, spanning from 1566 to 3301 cm^{-1}. *Kistimonas* demonstrated high correlations across several wavenumber ranges, with a common range shared with *Enhydrobacter* between 1952 and 2049 cm^{-1}.

Figures 10 and 11 present predicted versus observed abundances for indicative bacterial species using the Neural Network (NN) model. Figure 10 highlights the strong agreement between observed and predicted abundances for *Lactococcus, Acinetobacter*, *Shewanella*, and *Psychrobacter* in *M. chilensis* samples. Conversely, the model failed to accurately predict the abundances of *Micrococcus, Psychromonas*, and *Leuconostoc*. Figure 11 demonstrates the model's successful prediction of *Pseudoalteromonas*, *Actinomyces*, and *Micrococcus* abundance in *M.*

galloprovincialis. Additionally, the model's performance was also quite satisfactory for *Lactococcus, Acinetobacter and Psychrobacter*.

4. Discussion

The food sector primarily relies on regulatory inspection to monitor the quality and safety of food products. This involves conducting analyses on the final product, using standard methods specified by regulated authorities such as EFSA (see EU reg. 2073/2005). This procedure is expensive, time-consuming, destructive, and the acquired results are providing retrospective knowledge, preventing the provision of real-time information on the remaining shelf life and safety across the food chain (Nychas et al 2016). For this reason, new strategies have been introduced, such as the Quantitative Microbiological Risk Assessment (QMRA) which can monitor quality and safety almost in real-time. The inclusion of molecular-based '-omics' (Cocolin et al. 2017) in QMRA, although the idea sounds attractive and simple (Nychas et al 2016, 2021), several issues related to different methodologies, e.g. NGS, sensors, that are used to acquire data (structured or unstructured i.e. big data), their integration and interpretations are not yet thoroughly considered, making their application so far limited. Additionally, their prohibitive costs and the complexity of the associated bioinformatics analysis remains the main limitation to their large-scale application, especially in terms of real-time quality and safety monitoring. The need of overcoming these limitations is augmented, since we are living on the digitalization era, where there is an increased perception of using big data among stakeholders, academics, and consumers that stems from diverse technologies and their components (Nychas et al. 2021, Tao et al. 2021).

Bearing this in mind, the last 10-15 years, there is trend to use modernized, molecular and culture-independent techniques such as Next Generation Sequencing (NGS), the

advent of which has brought a revolution in food microbiology studies (Cocolin et al., 2017). The use of this set of methods helps scientists to enrich their molecular understanding about the microbial communities' presence and succession during food storage and/or fermentation (De Filippis et al., 2017). Among others, 16S rRNA metabarcoding analysis is a very appropriate method to study food/seafood bacterial profile, including the non-culturable bacteria that are not detectable with plate-based approach, to monitor how this microbiota is changing during storage conditions, as well as to determine the Specific (Ephemeral) Spoilage Organisms (SSO) fraction that is responsible for the organoleptic rejection of the product (Nychas et al. 2016; Anagnostopoulos et al., 2022).

In the context of this study, the conventional analysis was not only performed as a reference method, but also used as the parameter that need to be predicted, associated, or described, directly or indirectly, using metabolomic/spectral or genomic data. It should be highlighted that the implementation of sensor or NGS technologies is highly dependent on accurate phenotyping records, thus this parameter cannot be completely absent from a modern quality management system, but it should be periodically evaluated to ensure the proper operation of the Quality Management System. It needs to be stressed that the microbial association estimated either with culturable media or molecular tools as well as the type of spoilage i.e. production of off odours such as hydrogen sulphide, trimethylamine (TMA), biogenic amines due to amino acid decarboxylation activity and enzymes such as DNase, lipase and protease, are in line with published studies (Doulgeraki and Nychas, 2013, Parlapani et al., 2015; Antunes-rohling et al., 2019; Anagnostopoulos et al., 2023; Odeyemi et al., 2018b; Yuan et al., 2018; Yuan et al., 2023). Among the consortium members,

pseudomonads and Shewanella were found to be contributors to spoilage, with the latter being more active (Ge et al., 2017).

In this study, for both species (M. galloprovinciallis and M. chilensis) of mussels the conventional microbiological analysis was performed and revealed that pseudomonads and Shewanella were the SSO, which have also been previously reported as spoilage bacteria in seafood products under certain conditions (Anagnostopoulos et al., 2022). Both these bacteria were found to be in high numbers at the end of storage life of mussels. There was always this debate of how the actual numbers of SSO, counted either with conventional microbiological methodologies or provided by NGS can affect the ‘’spoilage phenotype’’ and how their physiology may influence the spoilage evolution, as this may affect the reliability of parameter estimations (Nychas et al. 1988; Nychas and Tassou 1997; Mohareb et al 2015; Pérez-Enciso et al., 2021)

This is the case with bacteria of genus *Psychromonas*, where they were found to be one of the dominant microbial groups in shellfish, although did not show fermentation, CO_2 production, proteolytic, or lipolytic activities (Ratnawati et al., 2024). The role in mussels' spoilage should be further investigated. However, this is a quite interesting finding - given the concept of this study - as it indicates that even if a microorganism does not show significant biological activity, it can still be important for quality evaluation if its growth during storage follows a specific pattern.

Nevertheless, several Pseudoalteromonas species are known for their ability to survive and grow in harsh environmental conditions such as low marine habitat temperatures (Parrilli et al., 2021) and for producing protease, lipase, H_2S, and other sulphur compounds which degrade mussels' quality during storage (Odeyemi et al., 2018; Zou et al., 2020). The psychrotolerant and halotolerant character of *Psychrobacter* enables

their presence in aquatic products being the most dominant group in the mussels cold-chain. Moreover, this genus growth is associated with alcohols production (i.e., 3-methyl-1-butanol) negatively affecting mussels' quality (Zhang et al., 2019). In M. galloprovincialis batch 2), *Kistimonas* was among bacteria that were present at a remarkable ratio at the end of shelf life. A previous study (Schoinas et al., 2023) reported this genus as one of the dominant species in Mediterranean mussels. It is noteworthy that this genus cannot grow without NaCl while are able to grow at relatively high temperatures (>10 °C) and maybe this is why found to be at higher abundance at 4 compared to 2 °C (Glaeser & Kämpfer, 2020).

On this point, it is crucial to be mentioned that *Pseudomonas* and *Shewanella* – the two main culturable spoilage bacteria - were found at remarkable levels only in these products (M. galloprovincialis batch 2). In *M. galloprovincialis* samples of batch 3 *Psychrilyobacter* and *Kistimonas* prevailed in 2^{o} C and 4^{o} C, respectively, at the end of the shelf life. *Psychrilyobacter* is one of the dominant genera in the microbiome of free-living and farmed mussels belonging to *M. chilensis* species (Santibáñez et al., 2022) and degrades proteins and amino acids, implying metabolic activity in the mussel's microbiome. However, further studies are required to discover the role and importance of this genus (Pelican et al., 2020).

In the present study, apart from data obtained through genomic analysis, i.e., NGS, metabarcoding, and classical microbiology, spectroscopy measurements were taken into account (through FTIR measurements). This allowed the acquisition of important knowledge about the seafood system, including the genera of microorganisms that are present, their populations as well as their metabolic fingerprint. The outcome of all the analyses and their correlations has an invaluable impact on seafood quality and safety, assisting the delivery of more prompt and accurate results, the decrease of the food

waste and the establishment of more efficient quality and safety management systems. The novel point of this process was to study the relationships, if any, between results obtained from various analyses being able to predict certain quality and safety characteristics using data from another e.g., predicting the microorganisms' profile using NGS data and FTIR measurement, that is proven to be rapid, accurate and precise as the original one.

Although, the findings from FTIR and other spectroscopic methods that are assessing mussel microbiological quality are limited (if any), nevertheless the FTIR measurements of this study shows similarities, regarding the magnitude of their potential, with other animal-based foods (Fengou et al., 2024, Vargas et al., 2023, Candoğan et al., 2021). Beyond shellfish microbiological quality or freshness, NIR or other spectroscopic methods can rapidly estimate composition, protozoan infection, and authentication (Madigan et al. 2013, Ghidini et al. 2024, Ayvaz et al., 2024, Guévélou et al., 2021, Bartlett et al., 2018).

The present study demonstrates that microbial associations and spectrum reflections exhibit species-specific variations. As a consequence, the theatre of microbial activities i.e., the type of degradation of carbon sources (e.g., saccharolytic or proteolytic) that leads to variability of end-products formation is confirmed by the different acquired spectra (Nychas et al 1988, 2008, Mohareb et al 2015; Feng et al., 2020).

It is well established that most of the microbial metabolic products are the key components of spoilage, and due to the specific frequencies at which they vibrate, they correspond to discrete energy levels and, as such, can bridge the biological with the spectroscopic approach (Argyri et al. 2013; Nychas et al. 2021). For example, the fingerprint region (section 3.3) of the IR spectrum has been associated with the

metabolic activity of microorganisms in meat, fish, and other high perishable food commodities during spoilage. Important regions contributing to the characterization of microbial quality of both mussel species from the region 900 to 1300 cm^{-1} can be attributed to proteins, lipids, phospholipids, and nucleic acids (Candoğan et al., 2021). The characterized as protein region (Amide I and Amide II) showed a high density and as shown from the regression coefficients, high values for both species were observed. Also, in the region 2300-3000 cm^{-1} there were highly informative regions/wavenumbers attributed mainly to lipids and several protein peaks (Candoğan et al., 2021).

Based on the findings of the analysis performed for the prediction of specific bacterial species abundance using FTIR, a broad correlation between FTIR wavelengths 3109 and 3302 cm^{-1} was observed across both pathogenic and spoilage bacterial species, indicating the presence of common spectral features or biochemical markers that transcend individual bacterial taxa. This region is mainly assigned to N-H, CH_2 and CH_3 (i.e., single bond) stretching attributed to alkyl chains primarily presented in lipids and certain amino acids vibrations attributed to proteins (Kassem et al. 2023; Matys et al., 2022). The area between 2900-3000 cm^{-1}, assigned mainly to N-H and C-H stretching, was also among the most informative ones for the prediction of microbial populations (TAC) using FTIR data.

Notably, a common range with elevated correlations is observed across all bacterial species between 892 $cm^{-1}$and 1181 cm^{-1}, suggesting shared spectral features indicative of bacterial presence or abundance i.e., bands related to polysaccharides, lipopolysaccharides and nucleic acids, compounds of bacterial cell wall and membranes (Consumi et al., 2020). Additionally, the same region (1,300–800 cm^{-1}) was also used in other studies to type specific pathogenic bacteria (Quintelas et al.,

2018; Martak et al., 2019) showing the high discriminatory power of this method. However, the prediction of the abundance of microbial abundances or counts it is still a much more complicated task while it should be always considered that the composition of the food matrix may also affect spectral changes at the same regions as well.

Although FTIR has been successfully applied to develop predictive models for general microbial growth such as Total Viable Counts [e.g. (Estelles-Lopez, Ropodi et al. 2017, Spyrelli, Ozcan et al. 2021)], the intricate nature of microbial and food samples can cause spectral signals to overlap, making FTIR for microbial identification difficult at the species level. Thereforre, infrared spectroscopy may not be able to distinguish closely related microbial species or strains, limiting its accuracy in microbial identification (AlMasoud et al., 2021). The integration of metagonimc analysis, which is capable of providing such a high resolution detail about bacterial abundance on the family, or even the species taxonomy level is provides therefore a superior predictive powers to the developed prediction models.

On the other hand, 16S metagenomic analysis only provides insights in the form of relative abundance, and does not provide information about the actual numbers of individual cells (Kallastu et al., 2023), which is a current EU legislation request (Nychas et al. 2016). This may lead to misinterpretations, as an increase in the relative abundance of one taxon results in the concurrent decrease of others. Additional test/methods (e.g., include digital PCR, qPCR, flow cytometry, and culture techniques) are needed to overcome this challenge. As proposed in this work, correlation of metagenomics data with metabolomics could contribute to overcome this issue as the outcome of this analysis is not evaluated individually but relates to several changes occurring on the food matrix.

In this study was evident that diverse food microbiomes can be of great benefit in the process of proof of concept, where in the first place the magnitude of data is limited. Indeed the different tested batches (species, countries, and producers) were found to have different microbial profiles, a trend that is reported in other studies with different products too (Manthou et al. 2021). This should enforce scientists, with different backgrounds to find new ways how to enlighten, through the synthesis of analytics provided by the very same data acquired from completely different analytical instruments, the potential of omics in food safety and quality. This concept can be of great importance i.e., combining different analyses to take a holistic view of outcomes can overcome the weaknesses of the current methodology (Nychas et al. 2008, 2016). Indeed, the simplification of models may result in predictions that do not align with reality Both data-driven and knowledge-based models are accessible, but both are subject to limitations. For instance, knowledge-based models face constraints due to their dependence on field knowledge, while data-driven models are limited by shortcomings in data quality and quantity. Food microbiome exhibits such a wide range of diversity that models may not encompass all genetic components, reactions, or interactions. Despite the robust performance of models in this study, they may overlook vital limitations. Given these constraints, it is advisable to consider the models of predictions better as proof of concept that requires further testing and evaluation by incorporating external data, preventing in this way misinterpretations and erroneous conclusions (Karlsen et al., 2023).

Another point that should be further considered is the fact that the abovementioned analyses, present inconsistencies as they have not yet tested at different environments, using different instruments operated by different persons. As indicated by the performance metrics of the research previously published, AI algorithms have shown

quite satisfactory performance in microbiological studies. However, large-scale industrial applications outside of limited laboratory testing are needed. The accuracy recorded in microbiological quality assessment studies using sensors technology and machine learning may not always imply efficacy in industrial scale (Kumar et al., 2023).

5. Conclusions

The study combined various analytical technologies to examine and monitor seafood quality and safety. It revealed that certain microbial genera, such as Pseudoalteromonas, Psychromonas, Psychrobacter, and Brochothrix, are involved in spoilage in different mussel species. The analysis of FTIR data, combined with microbiological and NGS data, revealed spectral regions correlated with mussels' freshness, including lipid, free fatty acids, protein, peptides, and amino acid changes. Although the results appear promising, it is important to note that due to the limited number of rows in the datasets, the test set used to generate these results and validate the developed models contains relatively few data points. Consequently, the clarity and conclusiveness of the results may be somewhat compromised. Despite this limitation, the outcome offers valuable insights into the predictive capabilities of the tested algorithms for both *M. chilensis* and *M. galloprovincialis* datasets, highlighting the potential utility of machine learning techniques in predicting bacterial abundances using FTIR data. Further validation and exploration of these correlations could enhance our understanding of microbial communities and their ecological roles

Declaration of interest

The authors declare that they have no competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Anastasia Lytou: Investigation, Data curation Formal analysis Writing – original draft and editing, Visualization. Léa Saxton: Investigation, Data curation, Data analysis. Lemonia-Christina Fengou: Methodology, Data analysis, Writing – original draft Visualization. Dimitris Anagnostopoulos: Methodology, Formal analysis, Writing – original draft. Foteini Parlapani: Investigation, Methodology, Data analysis Writing – Editing. Ioannis Boziaris: Supervision, resources, Writing- editing. Fady Mohareb: Conceptualization, Supervision, resources, Methodology, Data analysis, Writing – Original draft. George-John Nychas: Conceptualization, Funding acquisition, Project administration, Supervision resources, Writing-review and editing.

Funding

This study is funded from the European Union's Horizon 2020 research and innovation program with the acronym "DiTECT" under grant agreement No 861915. It is also financial supported from EU Horizon under Grant Agreement No. 10111336542

References

AlMasoud, N., Muhamadali, H., Chisanga, M., AlRabiah, H., Lima, C. A., and Goodacre, R. (2021). Discrimination of bacteria using whole organism fingerprinting: the utility of modern physicochemical techniques for bacterial typing. Analyst 146, 770–788. doi: 10.1039/D0AN01482F

Anagnostopoulos, D.A., Parlapani, F.F., Boziaris I.S. (2022). The evolution of knowledge on seafood spoilage microbiota from the 20th to the 21st century: Have we finished or just begun? Trends in Food Science & Technology 120, 236-247. https://doi.org/10.1016/j.tifs.2022.01.004

Anagnostopoulos, D. A., Syropoulou, F., Parlapani, F. F., Tsiartsafis, A., Exadactylos, A., Nychas, G. J. E., & Boziaris, I. S. (2023). Microbiota profile of filleted gilthead seabream (Sparus aurata) during storage at various conditions by 16S rRNA metabarcoding analysis. Food Research International, 164, 112312.

Antunes-rohling, A., Calero, S., Halaihel, N., Marquina, P., Raso, J., Calanche, J., Beltr, A., Ignacio, Á., Cebri, G. (2019). Characterization of the spoilage microbiota of hake different temperatures. Foods 2, 1–14. https://doi.org/10.3390/foods8100489

Argyri, A. A., Jarvis, R. M., Wedge, D., Xu, Y., Panagou, E. Z., Goodacre, R., & Nychas, G. E. (2013). A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. Food Control, 29(2), 461–470. https://doi.org/10.1016/j.foodcont.2012.05.040

Ayvaz, H., Temizkan, R., Kaya, B., Salman, M., Menevseoglu, A., Ayvaz, Z., ... & Mortas, M. (2024). Machine Learning-Assisted Near-and Mid-Infrared spectroscopy for rapid discrimination of wild and farmed Mediterranean mussels (*Mytilus galloprovincialis*). Microchemical Journal, *196*, 109669. https://doi.org/10.1016/j.microc.2023.109669

Bartlett, J.K., Maher, W.A., Purss, M.B.J. (2018). Near infra-red spectroscopy quantitative modelling of bivalve protein, lipid and glycogen composition using single-species versus multi-species calibration and validation sets. Spectrochim. Acta A Mol. Biomol. Spectrosc. 193, 537–557. https://doi.org/10.1016/j.saa.2017.12.046.

Baydakova, Elena Unguryanu, Tatiana N. Mikhaylova, Rufina I. (2019). On quantitative assessment of microbe risk caused by exposure to enteric viruses in drinking water. Health Risk Analysis Open Access. Issue 2, Pages 108 – 1142019. https://doi.org/10.21668/health.risk/2019.2.12.eng

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. Nature Biotechnology, 37(8):852-857 https://doi.org/10.7287/peerj.preprints.27295v2.

Borges F, Briandet R, Callon C, Champomier-Verges MC, Christieans S, Chuzeville S, Denis C, Desmasures N, Desmonts MH, Feurer C, Leroi F, Leroy S, Mounier J, Passerini D, Pilet MF, Schlusselhuber M Stahl V, Strub C, Talon R, Zagorec M. (2022). Contribution of omics to bio preservation: Toward food microbiome

engineering. Frontiers in Microbiology, 13, 35983334. https://doi.org/10.3389/fmicb.2022.951182

Broekaert, K., Heyndrickx, M., Herman, L., Devlieghere, F., & Vlaemynck, G. (2011). Seafood quality analysis: Molecular identification of dominant microbiota after ice storage on several general growth media. Food Microbiology, 28(6), 1162–1169. https://doi.org/10.1016/j.fm.2011.03.009

Codex Alimentarius Commission, 1999 Principles and guidelines for the conduct of microbiological risk assessment - CAC/GL-30, 1999 - cir.nii.ac.jp;FAO/WHO, 2002 FAO/WHO (2002). Principles and guidelines for incorporating microbiological risk assessment in the development of food safety standards, guidelines and related texts. In: Report of a Joint FAO/WHO Consultation, Kiel, Germany, 18–22 March 2002, Retrieved from. http://www.fao.org/3/a-y4302e.pdf

Candoğan, K., Altuntas, E.G. and İğci, N. (2021). Authentication and quality assessment of meat products by fourier-transform infrared (FTIR) spectroscopy. Food Engineering Reviews, *13*, pp.66-91. Doi: 10.1007/s12393-020-09251-y

Cocolin, L.S., Mataragas, M., Bourdichon, F., Doulgeraki, A.I., Pilet, M.F., Jagadeesan, B., Rantsiou, K., & Phister, T.G. (2018). Next generation microbiological risk assessment meta-omics: The next need for integration. International Journal of Food Microbiology, 287, 10-17. doi: 10.1016/j.ijfoodmicro.2017.11.008

Consumi, M., Jankowska, K., Leone, G., Rossi, C., Pardini, A., Robles, E., Wright, K., Brooker, A., Magnani, A. (2020). Non-Destructive Monitoring of P. fluorescens and S. epidermidis Biofilm under Different Media by Fourier Transform Infrared Spectroscopy and Other Corroborative Techniques. Coatings, *10*, 930. https://doi.org/10.3390/coatings10100930

De Filippis, F., Parente, E., & Ercolini, D. (2017). Metagenomics insights into food fermentations. Microbial Biotechnology, 10(1), 91–102. https://doi.org/10.1111/1751-7915.12421

Doulgeraki, A. and Nychas, G-J. (2013). Monitoring the succession of the biota of a selective growth medium for pseudomonads with different approaches. Food Microbiology 34,62-69. https://doi.org/10.1016/j.fm.2012.11.017

Jiang, X., Zhang, Y,, Luo, X., Zhu, L., Mao, Y., Liang, R., Yang, X., Han, G., Dong, P., (2023). A review of quantitative risk assessment of pathogenic bacteria in chilled meat. Shipin Kexue/Food Science, 44, 288-295 doi: 10.7506/spkx1002-6630-20220122-225

Dowd, S. E., Callaway, T. R., Wolcott, R. D., Sun, Y., McKeehan, T., Hagevoort, R. G., et al. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol., 8, 1–8. https://doi.org/10.1186/1471-2180-8-125/TABLES/2

den Besten, H.M.W., Amézquit,A., Bover-Cid,S., Dagnas,S., Ellouze, M., Guillou, S., Nychas, G., O'Mahony, C., Pérez-Rodriguezi,F., Membré, J-M., (2018) Next generation of microbiological risk assessment: Potential of omics data for exposure

assessment Int. J. Food Microbiology 287, 18-27
https://doi.org/10.1016/j.ijfoodmicro.2017.10.006

Edgar, R. C., & Bateman, A. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26, 2460–2461. https://doi.org/10.1093/BIOINFORMATICS/BTQ461

Eren, A. M., Zozaya, M., Taylor, C. M., Dowd, S. E., Martin, D. H., & Ferris, M. J. (2011). Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. PLoS One, 6, e26732. doi: 10.1371/journal.pone.0026732

Estelles-Lopez, L., A. Ropodi, D. Pavlidis, J. Fotopoulou, C. Gkousari, A. Peyrodie, E. Panagou, G.-J. Nychas and F. Mohareb (2017). An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling. Food Research International (Ottawa, Ont.). **99:** 206-215.

Feng, B., Shi, H., Xu, F., Hu, F., He, J., Yang, H., et al. (2020). FTIR-assisted MALDI-TOF MS for the identification and typing of bacteria. Anal. Chim. Acta 1111, 75–82. doi: 10.1016/j.aca.2020.03.037

Fengou, L. C., Liu, Y., Roumani, D., Tsakanikas, P., & Nychas, G. J. E. (2022). Spectroscopic data for the rapid assessment of microbiological quality of chicken burgers. Foods, 11(16). https://doi.org/10.3390/foods11162386

Fengou, L.C., Lytou, A.E., Tsekos, G., Tsakanikas, P. and Nychas, G.J.E. (2024). Features in visible and Fourier transform infrared spectra confronting aspects of meat quality and fraud. Food Chemistry, 440, p.138184. https://doi.org/10.1016/j.foodchem.2023.138184

Galazzo, G., van Best, N., Benedikter, B.J., Janssen, K., Bervoets, L., Driessen, C., Oomen, M., Lucchesi, M., van Eijck, P.H., Becker, H.E.F., Hornef, M.W., Savelkoul, P.H., Stassen, F.R.M, Wolffs, P.F. and Penders, J. (2020). How to count our microbes? The effect of different quantitative microbiome profiling approaches. Front. Cell. Infect. Microbiol. 10:403 doi: 10.3389/fcimb.2020.00403

García-Perez,P, Becchi,P.P., Zhanga, L., Rocchetti, G., Lucini, L., (2024) Metabolomics and chemometrics: The next-generation analytical toolkit for the evaluation of food quality and authenticity. Trends in Food Science and Technology, 147, 104481 https://doi.org/10.1016/j.tifs.2024.104481

Ge, Y., J. Zhu, X. Ye, Y. Yang, Spoilage potential characterization of *Shewanella* and *Pseudomonas* isolated from spoiled large yellow croaker (*Pseudosciaena crocea*). (2017). Letters in Applied Microbiology, Volume 64, Issue 1, Pages 86–93, https://doi.org/10.1111/lam.12687

Ghidini, S., Varrà, MO, Bersellini, D., Conter, M., Fabrile, MP, Ianieri, A., Emanuela Zanardi, E. (2024). Real-time and non-destructive control of the freshness and viability of live mussels through portable near-infrared spectroscopy, Food Control, Volume 160, https://doi.org/10.1016/j.foodcont.2024.110353 .

Glaeser, Stefanie & Kämpfer, Peter. (2020). Kistimonas. Published by John Wiley & Sons, Inc., in association with Bergey's Manual of Systematics of Archaea and Bacteria (pp.1-7) https://doi.org/10.1002/9781118960608.gbm01541

Gram, L., Trolle, G., Huss, H.H. (1987). Detection of specific spoilage bacteria from fish stored at low (0 ◦C) and high (20 ◦C) temperatures. Int. J. Food Microbiol. 1987, 4, 65–72 https://doi.org/10.1016/0168-1605(87)90060-2

Guévélou, E., Carnegie, R. B., Whitefleet-Smith, L., Small, J. M., & Allen Jr, S. K. (2021). Near infrared reflectance spectroscopy to quantify Perkinsus marinus infecting Crassostrea virginica. Aquaculture, *533*, 736063. https://doi.org/10.1016/j.aquaculture.2020.736063

Haddad, N., Johnson, N., Kathariou, S., Métris, A., Phister, T., Pielaat, A., Tassou, C., Wells-Bennik, M. H. J., & Zwietering, M. H. (2018). Next generation microbiological risk assessment—Potential of omics data for hazard characterisation. International Journal of Food Microbiology, 287, 28-39. https://doi.org/10.1016/j.ijfoodmicro.2018.04.015

Herraiz-Gil, S., del Carmen de Arriba, M., Escamez, M. J., & Leon, C. (2023). Multi-omic data integration in food science and analysis. Current Opinion in Food Science, 52, 101049. https://doi.org/10.1016/j.cofs.2023.101049

Hiroki, A., Susumu, K. (2024). Modeling strain variability in *Campylobacter jejuni* thermal inactivation by quantifying the number of strains required. International Journal of Food Microbiology, 110618. https://doi.org/10.1016/j.ijfoodmicro.2024.110618

Kallastu, A., Malv, E., Aro, V., Meikas, A., Vendelin, M., Kattel, A., Nahku, R., Kazantseva, J. (2023). Absolute quantification of viable bacteria abundances in food by next-generation sequencing: Quantitative NGS of viable microbes. Current Research in Food Science. Jan 12;6:100443. doi: 10.1016/j.crfs.2023.100443. PMID: 36691592; PMCID: PMC9860258.

Karlsen, S.T., Rau, M.H., Sánchez, B.J., Jensen, K., Zeidan, A.A. (2023). From genotype to phenotype: computational approaches for inferring microbial traits relevant to the food industry. FEMS Microbiological Reviews. Jul 5;47(4):fuad030. doi: 10.1093/femsre/fuad030. PMID: 37286882; PMCID: PMC10337747.

Kassem, A., Abbas, L., Coutinho, O., Opara, S., Najaf, H., Kasperek, D., Pokhrel, K., Li. X. and Tiquia, A. S. (2023). Applications of Fourier Transform-Infrared spectroscopy in microbial cell biology and environmental microbiology: advances, challenges, and future perspectives. Frontiers in Microbiology. 14:1304081. doi: 10.3389/fmicb.2023.1304081

Kumar, R., Yadav, G., Kuddus, M. et al. (2023). Unlocking the microbial studies through computational approaches: how far have we reached? . Environmental Science Pollution Research 30, 48929–48947. https://doi.org/10.1007/s11356-023-26220-0

Lane, H.M., Murray, S.C. (2021). High throughput can produce better decisions than high accuracy when phenotyping plant populations. Crop Science. 61(5):3301–13. https://doi.org/10.1002/csc2.20514

Manthou, E., Coeuret, G., Chaillou, S., Nychas, G-J.E., (2021) Evolution of fungal community associated with ready-to-eat pineapple during storage under different temperature conditions. Food Micro 97, 103736

Madigan, T., Kiermeier, A., Carragher, J., de Barros Lopes, M., & Cozzolino, D. (2013). The use of rapid instrumental methods to assess freshness of half shell Pacific oysters, Crassostrea gigas: A feasibility study. Innovative Food Science & Emerging Technologies, 19, 204-209. https://doi.org/10.1016/j.ifset.2013.04.005

Martak, D., Valot, B., Sauget, M., Cholley, P., Thouverez, M., Bertrand, X., et al. (2019). Fourier-transform infrared spectroscopy can quickly type gram-negative Bacilli responsible for hospital outbreaks. Frontiers in Microbiology. 10:1440. doi: 10.3389/fmicb.2019.01440

Matys, J., Turska-Szewczuk, A., Gieroba, B., Kurzylewska, M., Pękala-Safińska, A., Sroka-Bartnicka, A. (2022). Evaluation of Proteomic and Lipidomic Changes in Aeromonas-Infected Trout Kidney Tissue with the Use of FT-IR Spectroscopy and MALDI Mass Spectrometry Imaging. International Journal of Molecular Science. 23, 12551. https://doi.org/10.3390/ijms232012551

Mevik, B-H.R. Wehrens, K.H., Liland, A. (2020). PLS: Partial Least Squares and Principal Component Regression (2.7-3).

Meyer, D.E., Dimitriadou, K., Hornik, A., Weingessel, F., Leisch, C.-C., Chang, (2021). e1071: Misc Functions of the Department of Statistics Probability Theory Group (Formerly E1071) (1.7-5).

Mohareb, F., Iriondoa, M., Doulgeraki, A.I, Van Hoekc, A., Aarts, H., Cauchia, M, and Nychas, G-J (2015) Identification of meat spoilage gene biomarkers in *Ps. putida* using gene profiling, Food Control 57, 152-160

Nychas, G-J. E & Tassou, Ch.Ch (1997) Spoilage process and proteolysis in chicken as noted by HPLC method *Journal of Science of Food and Agriculture* 74, 199-208

Nychas G-J**.** E**.**, Dillon D.M. Board R.G (1988) Glucose: a Key substrate in microbial spoilage of meat and meat products. *Biotechnology and Applied Biochemistry* 10, 203-231.

Nychas, G-J. E., P Skandamis, C.C. Tassou, & K. Koutsoumanis (2008) Meat spoilage during distribution. Meat Science 78, 77-89

Nychas, G-J. E, Panagou, E.Z. & Mohareb, F. (2016) "Novel Approaches for Food Safety Management and Communication" Current Opinion in Food Science 12, 13-20).

Nychas, G.J., Sims, E., Tsakanikas, P., Mohareb, F. (2021). Data science in the food industry. Annu Rev Biomed Data Sci. Jul 20;4:341-367. doi: 10.1146/annurev-biodatasci-020221-123602. Epub 2021 May 13. PMID: 34465171.

Odeyemi, O.A., Burke, C.M., Bolch, C.C.J., Stanley, R. (2018a). Seafood spoilage microbiota and associated volatile organic compounds at different storage temperatures and packaging conditions. International Journal of Food Microbiology 280, 87–99. https:// doi.org/10.1016/j.ijfoodmicro.2017.12.029

Odeyemi, O.A., Burke, C.M., Bolch, C.J.S., Stanley, R. (2018b). Evaluation of spoilage potential and volatile metabolites production by Shewanella baltica isolated from modified atmosphere packaged live mussels. Food Research International. 103, 415–425. https://doi.org/10.1016/j.foodres.2017.10.068.

Parlapani, F.F., Kormas, K.A., Boziaris, I.S. (2015). Microbiological changes, shelf life and identification of initial and spoilage microbiota of sea bream fillets stored under various conditions using 16S rRNA gene analysis. J. Sci. Food Agric. 95, 2386–2394. https://doi.org/10.1002/jsfa.6957

Parrilli, E., Tedesco, P., Fondi, M., Tutino, M.L., Lo Giudice, A., de Pascale, D., Fani, R. (2021). The art of adapting to extreme environments: the model system *Pseudoalteromonas*. Phys. Life Rev. 36, 137–161 https://doi.org/10.1016/j.plrev.2019.04.003.

Pelikan, C., Wasmund, K., Glombitza, C., Hausmann, B., Herbold, C.W., Flieder, M., Loy, A. (2020). Anaerobic bacterial degradation of protein and lipid macromolecules in subarctic marine sediment. ISME J. 153 (15), 833–847. https://doi.org/10.1038/ s41396-020-00817-6

Pérez-Enciso, M., Zingaretti, L.M., Ramayo-Caldas, Y. et al. (2021). Opportunities and limits of combining microbiome and genome data for complex trait prediction. Genet Sel Evol 53, 65. https://doi.org/10.1186/s12711-021-00658-7

Quintelas, C., Ferreira, E. C., Lopes, J. A., and Sousa, C. (2018). An overview of the evolution of infrared spectroscopy applied to bacterial typing. Biotechnology Journal. 13:1700449. doi: 10.1002/biot.201700449

Rantsiou, K., Kathariou, S., Winkler, A., Skandamis, P., Saint-Kyr, M.J., Rouzeau-Szynalski, K., Amézquita, A. (2018). Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment. International Journal of Food Microbiology, 287, pp. 3–9. https://doi.org/10.1016/j.ijfoodmicro.2017.11.007

Ratnawati, S.E., Kuuliala, L., Verschuere, N., Cnockaert, M., Vandamme, P., Devlieghere, F. (2014). The exploration of dominant spoilage bacteria in blue mussels (*Mytilus edulis*) stored under different modified atmospheres by MALDI-TOF MS in combination with 16S rRNA sequencing. Food Microbiology. Apr;118:104407. doi: 10.1016/j.fm.2023.104407. Epub 2023 Oct 20. PMID: 38049269.

Santibáñez, P., Romalde, J., Maldonado, J., Fuentes, D., Figueroa, J. (2022). First characterization of the gut microbiome associated with *Mytilus chilensis* collected at a

mussel farm and from a natural environment in Chile. Aquaculture 548, 737644. https://doi.org/10.1016/J.AQUACULTURE.2021.737644

Schoinas, K., Konstantou, V., Bompou, E., Floros, G., Chatziplis, D., Imsiridou, A., Loukovitis, D. (2023). Microbiome Profile of the Mediterranean Mussel (*Mytilus galloprovincialis*) from Northern Aegean Sea (Greece) Culture Areas, Based on a 16S rRNA Next Generation Sequencing Approach. Diversity. 15(3):463. https://doi.org/10.3390/d15030463

Sequino, G., Valentino, V., Villani, F., De Filippis, F. (2022). Omics-based monitoring of microbial dynamics across the food chain for the improvement of food safety and quality. Food Research International (Ottawa, Ont.). Jul;157:111242. DOI: 10.1016/j.foodres.2022.111242. PMID: 35761554

Spyrelli, E. D., O. Ozcan, F. Mohareb, E. Z. Panagou and G. E. Nychas (2021). "Spoilage assessment of chicken breast fillets by means of fourier transform infrared spectroscopy and multispectral image analysis." Curr Res Food Sci 4: 121-131.

Swanson, K.S., Dowd, S.E., Suchodolski, J.S., Middelbos, I.S., Vester, B.M., Barry, K.A., Nelson, K.E., Torralba, M., Henrissat, B., Coutinho, P.M., Cann, I.K., White, B.A., Fahey, G.C. (2010). Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. ISME J. 2011 54 5, 639–649. https://doi.org/10.1038/ismej.2010.162.

Syropoulou, F., Parlapani, F.F., Anagnostopoulos, D.A., Stamatiou, A., Mallouchos, A., Boziaris, I.S. (2021). Spoilage investigation of chill stored meagre (Argyrosomus regius) using modern microbiological and mnalytical techniques. Foods 2021, Vol. 10, Page 3109 10, 3109. https://doi.org/10.3390/FOODS10123109

Tao, Q., Ding, H., Wang, H., Cui, X. (2021). Application Research: Big Data in Food Industry. Foods, 10, 2203. https://doi.org/10.3390/foods10092203

Tuggle, C.K., Clarke, J.L., Murdoch, B.M. et al. (2024). Current challenges and future of agricultural genomes to phenomes in the USA. Genome Biol 25, 8. https://doi.org/10.1186/s13059-023-03155-w

Vargas, L.R.C., Notario, K.A., Espinosa, H.R., Macuil, R.D., González, H.R., Frómeta, A.E.N. and Lozano, L.S.A. (2023). FT-IR analysis of tilapia fillets: Developing PLS models for the prediction of storage days, aerobic plate count, and lactobacilli. *Vibrational Spectroscopy*, *129*, p.103619. https://doi.org/10.1016/j.vibspec.2023.103619

Yap, M., Ercolini, D., Álvarez-Ordóñez, A., O'Toole, P.W., O'Sullivan, O., Cotter, P.D. (2022). Next-Generation Food Research: Use of Meta-Omic Approaches for Characterizing Microbial Communities Along the Food Chain. Annual Reviews in Food Science and Technology. Mar 25;13:361-384. doi: 10.1146/annurev-food-052720-010751. Epub 2021 Oct 22. PMID: 34678075.

Yuan, L., Sadiq, F.A., Liu T.,J, Li, Y., Gu, J.S., Yang, H.Y., He, G.Q. (2018). Spoilage potential of psychrotrophic bacteria isolated from raw milk and the thermo-

stability of their enzymes. J Zhejiang Univ Sci B. 2018 Aug.;19(8):630-642. doi: 10.1631/jzus.B1700352. PMID: 30070086; PMCID: PMC6102184.

Yuan, L., Dai, H., He, G., Yang, Z., Jiao, X. (2023). Invited review: Current perspectives for analyzing the dairy biofilms by integrated multiomics. *J. Dairy Sci.* 106, 8181–8192. doi: 10.3168/jds.2023-23306

Zhang, Y., Wei, J., Yuan, Y., Yue, T. (2019). Diversity and characterization of spoilage associated psychrotrophs in food in cold chain. International Journal of Food Microbiology. 290, 86–95. https://doi.org/10.1016/j.ijfoodmicro.2018.09.026

Zou, Y., Robbens, J., Heyndrickx, M., Debode, J., Raes, K. (2020). Quantification of extracellular proteases and chitinases from marine bacteria. Curr. Microbiol. 77, 3927–3936. https://doi.org/10.1007/s00284-020-02216-8.

TABLES

Table 1. Samples examined in each of the individual analyses

	M. chilensis			***M. galloprovincialis***			
Analyses	Batch* 1 (Brand 1)	Batch 2 (Brand 2)	Total	Batch 1 (Spanish)	Batch 2 (Greek – Producer 1)	Batch 3 (Greek – Producer 2)	Total
Microbiological (cultures)	30	40	70**	45	58	39	142
FTIR	50	60	110	90	117	77	284
NGS	3	3	6 ***	5	5	5	15

*In *M. chilensis*, the two batches are differentiated by the brand and the selling point (both brands are originated from Chile). In *M. galloprovincialis*, the three batches are differentiated by geographical origin and producers)
** In conventional microbiological analysis each sample was a homogenate of 2 individual organisms (mussels)
*** Each of the samples was a homogenate of 9 individual mussels which were stored under the same conditions

Table 2. Performance indices of PLS regression model for the prediction of spoilage bacteria counts using the NGS data

		TAC	*Pseudomonas*	*Shewanella*	Enterobacteriaceae
PLS-R					
M. galloprovincialis	R^2	0,98	0,77	0,94	0,72
	RMSE	0,23	0,43	0,41	0,71
M. chilensis	R^2	0,94	0,95	0,91	0,97
	RMSE	0,52	0,42	0,78	0,22

Table 3. Performance indices of PLS feature extraction - SVM regression for the prediction of spoilage bacteria counts using FTIR data

Species		RMSE	R^2
M. galloprovincialis	Training	0,54	0,77
	Test	0,59	0,71
M. chilensis	Training	0,59	0,91
	Test	0,91	0,80

Table 4. Performance indices of different machine learning algorithms for the prediction of bacteria relative abundance in mussels, *M. chilensis*, using FTIR data

Species	Range of values (%)	RMSE RF	Rsquared RF	RMSE kNN	Rsquared kNN	RMSE NN	Rsquared NN
Aeromonas	0.004-2.65	0.41	0.99	0.50	0.89	0.13	0.99
Bacillus	0.00-0.78	0.004	0.66	0.003	0.69	0.13	0.12
Clostridium	0.00-0.016	0.002	0.97	0.002	0.98	0.12	0.26
Enterococcus	0.002-0.37	0.05	0.92	0.07	0.75	0.15	0.47
Staphylococcus	0.009-0.34	0.01	0.99	0.02	0.91	0.10	0.37
Streptococcus	0.016-13.8	1.78	0.95	1.52	0.99	0.4	0.99
Vibrio	0.038-10.9	0.86	0.86	1.24	0.84	0.35	0.98
Arcobacter	0.002-1.75	0.13	0.92	0.15	0.76	0.1	0.9
Enhydrobacterium	0.004-0.43	0.11	0.81	0.053	0.93	0.12	0.7
Exiguobacterium	0.056-7.51	0.74	0.99	2.47	0.87	0.33	0.99
Flavobacterium	0.026-0.15	0.01	0.97	0.04	0.18	0.08	0.29
Janthinobacterium	0.00-0.014	0.002	0.93	0.0008	0.99	0.17	0.12
Kistimonas	0.026-0.83	0.11	0.98	0.06	0.98	0.16	0.82
Marinobacterium	0.004-0.013	0.0008	0.85	0.002	0.36	0.064	0.11
Photobacterium	0.00-0.26	0.02	0.71	0.01	0.77	0.06	0.28
Ralstonia	0.015-0.49	0.06	0.91	0.08	0.84	0.064	0.91
Rhodococcus	0.00-0.102	0.004	0.81	0.02	0.01	0.119	0.28
Tenacibaculum	0.00-0.19	0.055	0.71	0.03	0.89	0.09	0.47
Acinetobacter	0.20-13.90	0.91	0.99	1.53	0.98	0.48	0.99
Actinomyces	0.00-6.70	0.63	0.97	0.78	0.99	0.44	0.98
Brochothrix	0.003-3.30	0.34	0.97	0.35	0.95	0.24	0.97
Corynebacterium	0.01-7.90	0.19	0.98	0.31	0.99	0.2	0.98
Lactococcus	0.00-6.70	0.95	0.94	1.2	0.88	0.29	0.99
Leuconostoc	0.007-15.9	0.87	0.8	0.47	0.91	0.28	0.98
Micrococcus	0.00-2.90	0.002	0.10	0.002	0.02	0.09	0.07
Propionibacterium	0.004-0.26	0.009	0.90	0.03	0.2	0.13	0.36
Pseudoalteromonas	0.20-18.20	4.11	0.94	12.05	0.35	0.51	0.99

Pseudomonas	0.17-27.1	1.94	0.99	4.37	0.99	0.24	0.99
Psychrobacter	0.80-66.70	9.48	0.96	9.07	0.87	1.91	0.99
Psychromonas	0.00-4.50	0.22	0.9	0.57	0.16	0.17	0.94
Shewanella	0.006-5.70	0.16	0.98	0.53	0.42	0.14	0.95
Shigella	0.00-0.82	0.11	0.95	0.1	0.99	0.06	0.98

Table 5. Performance indices of different machine learning algorithms for the prediction of bacteria relative abundance in mussels, *M. galloprovincialis*, using FTIR data

Species	Range of values (%)	RMSE RF	Rsquared RF	RMSE kNN	Rsquared kNN	RMSE NN	Rsquared NN
Aeromonas	0.00-1.12	0.38	0.06	0.29	0.16	0.38	0.31
Bacillus	0.00-1.61	0.05	0.88	0.02	0.97	0.07	0.73
Clostridium	0.00-1.23	0.03	0.98	0.05	0.95	0.1	0.86
Enterococcus	0.00-0.97	0.09	0.66	0.09	0.64	0.17	0.3
Staphylococcus	0.54-9.47	0.40	0.66	0.52	0.56	0.39	0.7
Streptococcus	1.22-8.42	0.46	0.94	0.69	0.74	0.63	0.77
Vibrio	0.006-2.05	0.24	0.49	0.19	0.68	0.28	0.51
Arcobacter	0.00-0.98	0.01	0.88	0.01	0.79	0.07	0.23
Enhydrobacterium	0.00-3.51	0.66	0.83	0.75	0.67	0.54	0.79
Exiguobacterium	0.00-0.88	0.19	0.89	0.32	0.77	0.24	0.61
Flavobacterium	0.002-0.61	0.17	0.32	0.16	0.25	0.16	0.49
Janthinobacterium	0.00-0.65	0.43	0.11	0.42	0.0099	0.34	0.43
Kistimonas	0.01-30.2	0.004	0.75	0.005	0.66	0.054	0.15
Marinobacterium	0.00-1.12	0.01	0.26	0.006	0.92	0.055	0.02
Photobacterium	0.00-2.77	0.02	0.99	0.02	0.99	0.12	0.76
Ralstonia	0.099-20.22	2.06	0.96	2.05	0.97	1.4	0.95
Rhodococcus	0.00-0.102	0.06	0.001	0.08	0.000012	0.116	0.15
Tenacibaculum	0.00-0.98	0.016	0.94	0.01	0.96	0.1	0.5
Acinetobacter	0.25-7.85	1.32	0.48	1.73	0.008	1.35	0.34
Actinomyces	0.00-6.72	0.38	0.7	0.45	0.59	0.49	0.56
Brochothrix	0.003-3.30	0.29	0.39	0.28	0.47	0.3	0.42
Corynebacterium	0.01-7.91	0.81	0.47	0.90	0.48	0.91	0.58
Lactococcus	0.00-6.71	0.44	0.74	0.47	0.67	0.65	0.59

Leuconostoc	0.007-15.90	4.7	0.32	4.9	0.36	3.57	0.66
Micrococcus	0.00-2.94	0.08	0.99	0.055	0.99	0.1	0.98
Propionibacterium	0.48-3.99	0.06	0.53	0.49	0.66	0.49	0.65
Pseudoalteromonas	0.22-27.58	8.7	0.53	9.11	0.39	6.22	0.7
Pseudomonas	0.30-13.01	0.55	0.99	0.7	0.99	0.49	0.99
Psychrobacter	0.84-31.09	7.23	0.51	7.62	0.31	3.66	0.90
Psychromonas	0.00-1.09	0.31	0.64	0.36	0.48	0.26	0.75
Shewanella	0.007-5.58	0.25	0.43	0.30	0.23	0.35	0.33
Shigella	0.00-0.043	0.001	0.82	0.002	0.99	0.09	0.27

FIGURES

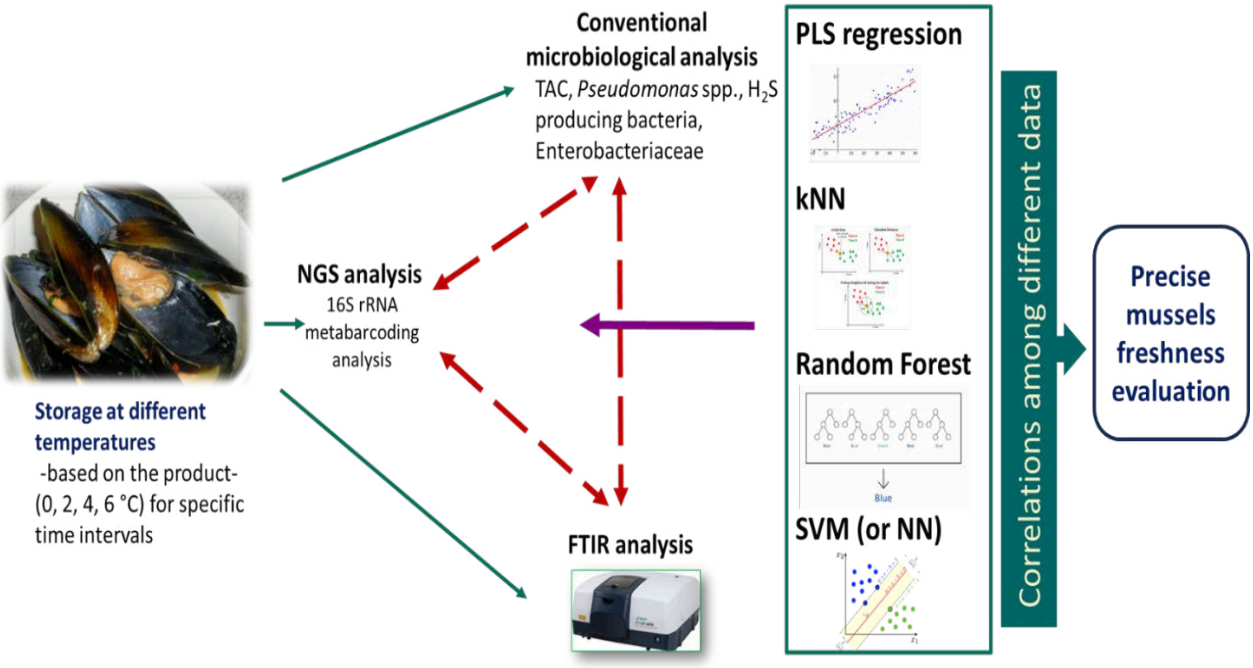


Figure 1. Graphical abstract of the experimental procedure

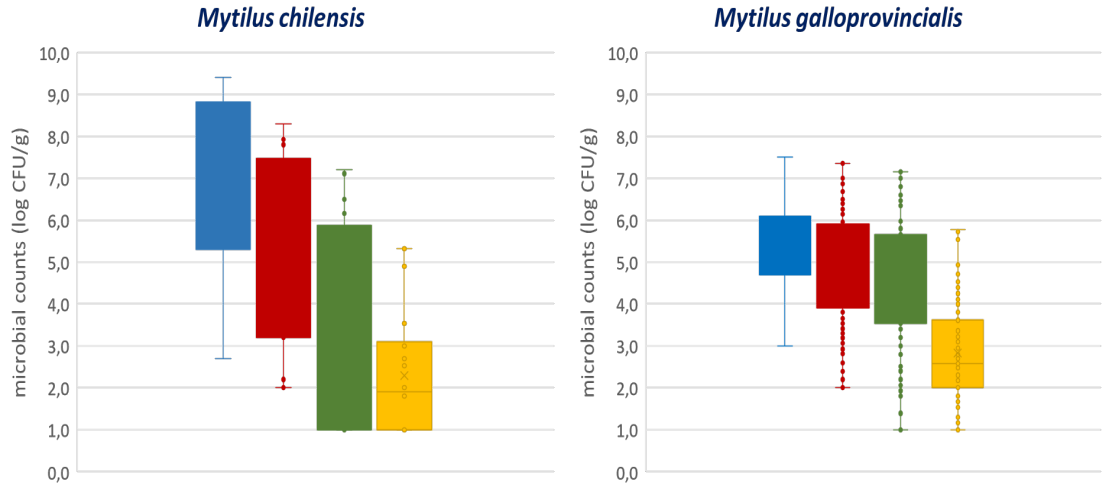


Figure 2. Microbial populations in the different mussel species throughout storage

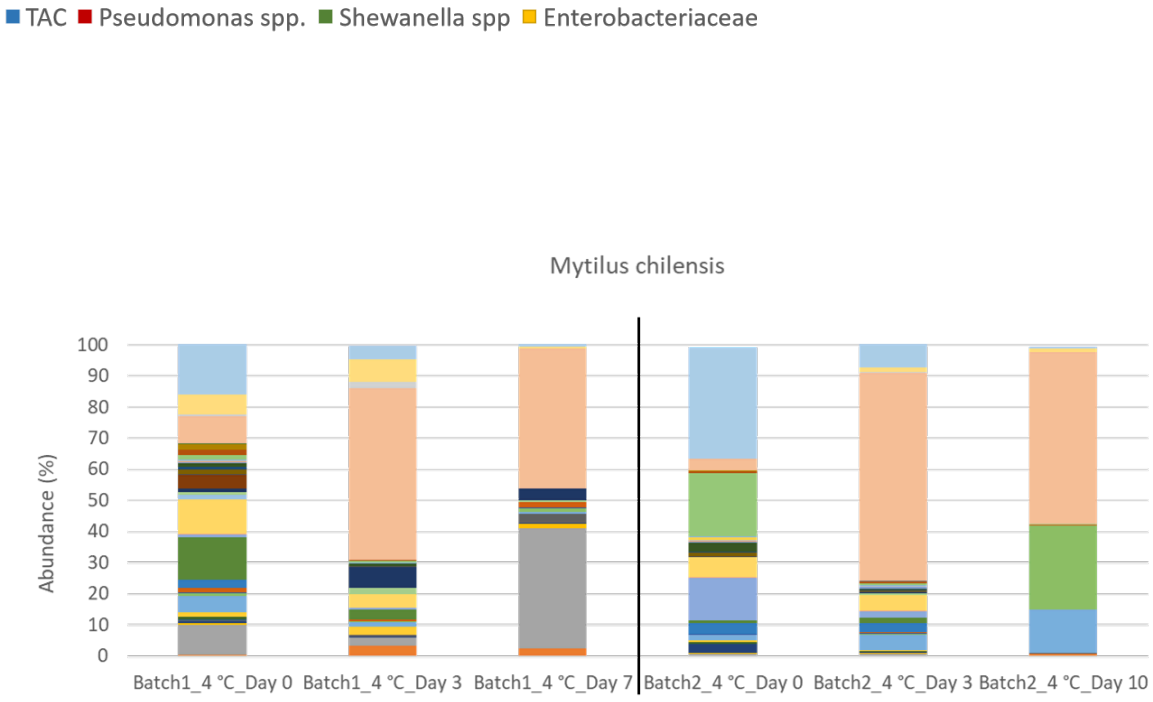


Figure 3. Changes in the relative abundances (%) of bacterial genera found in 2 different batches of *M. chilensis* samples throughout storage at 4 °C

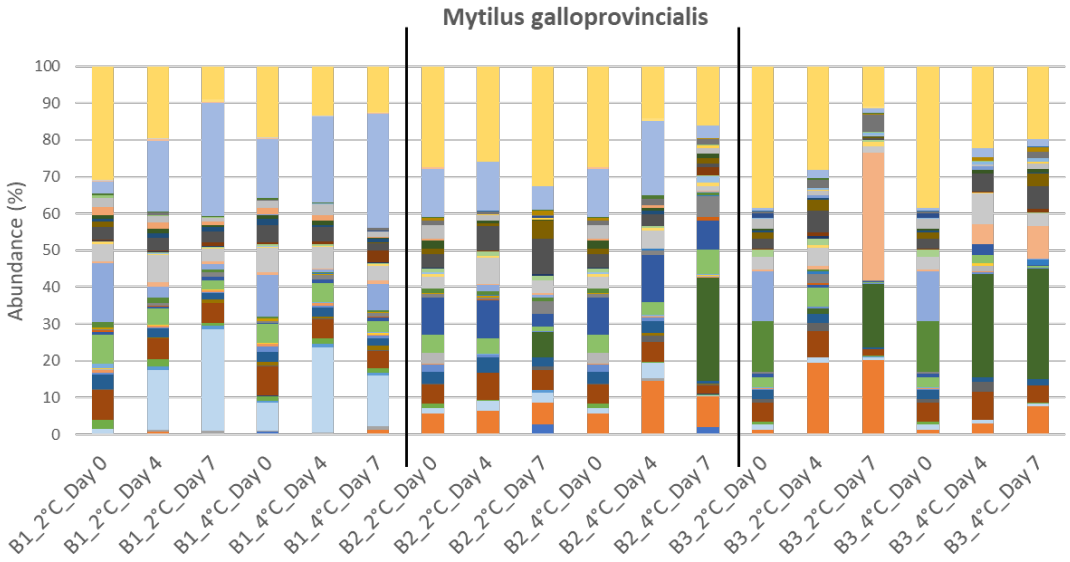
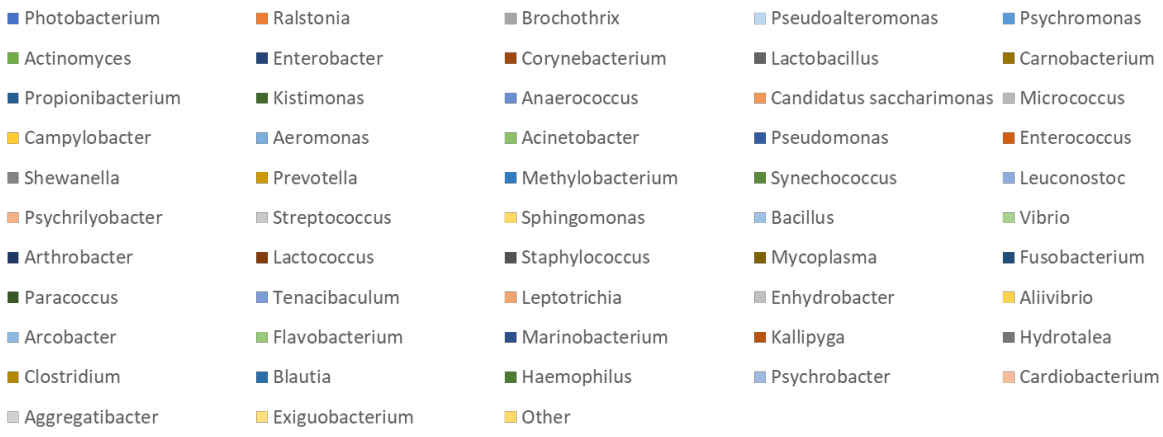


Figure 4. Changes in the relative abundances (%) of bacterial genera found in 3 different batches of *M. galloprovincialis* samples throughout storage at 2 and 4 °C



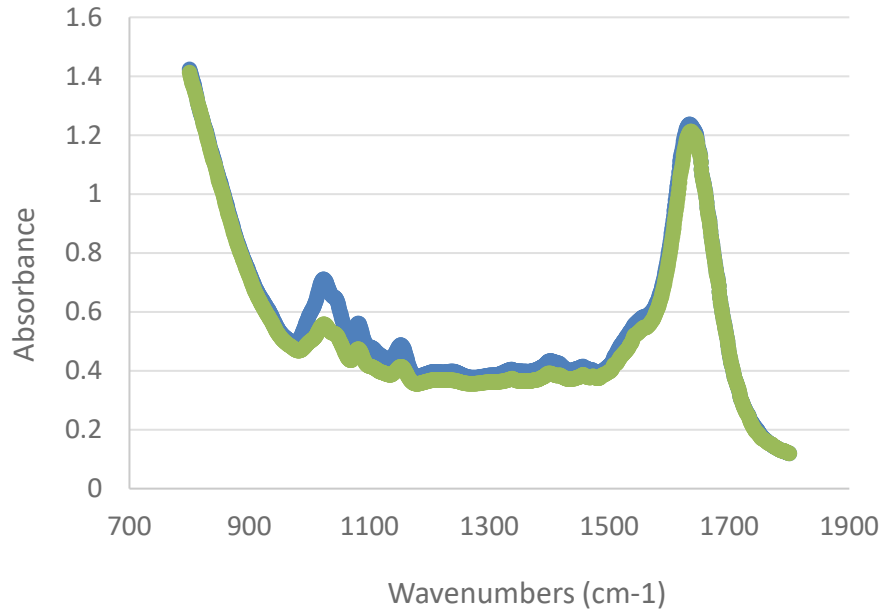


Figure 5. Indicative FTIR area spectra of the two mussel species

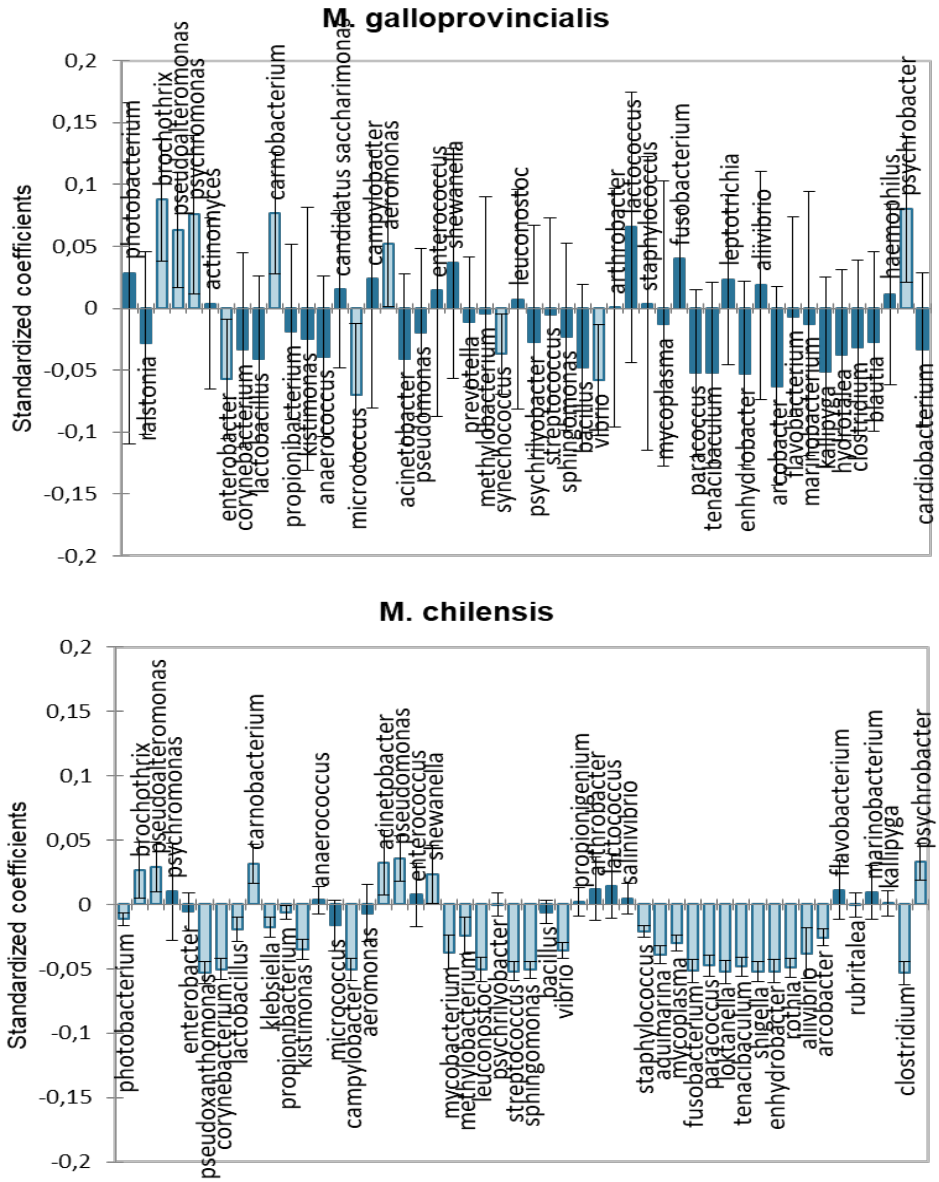


Figure 6. Standardized coefficients for the estimation of Total Aerobic Counts using NGS data in *M. galloprovincialis* and *M. chilensis* samples

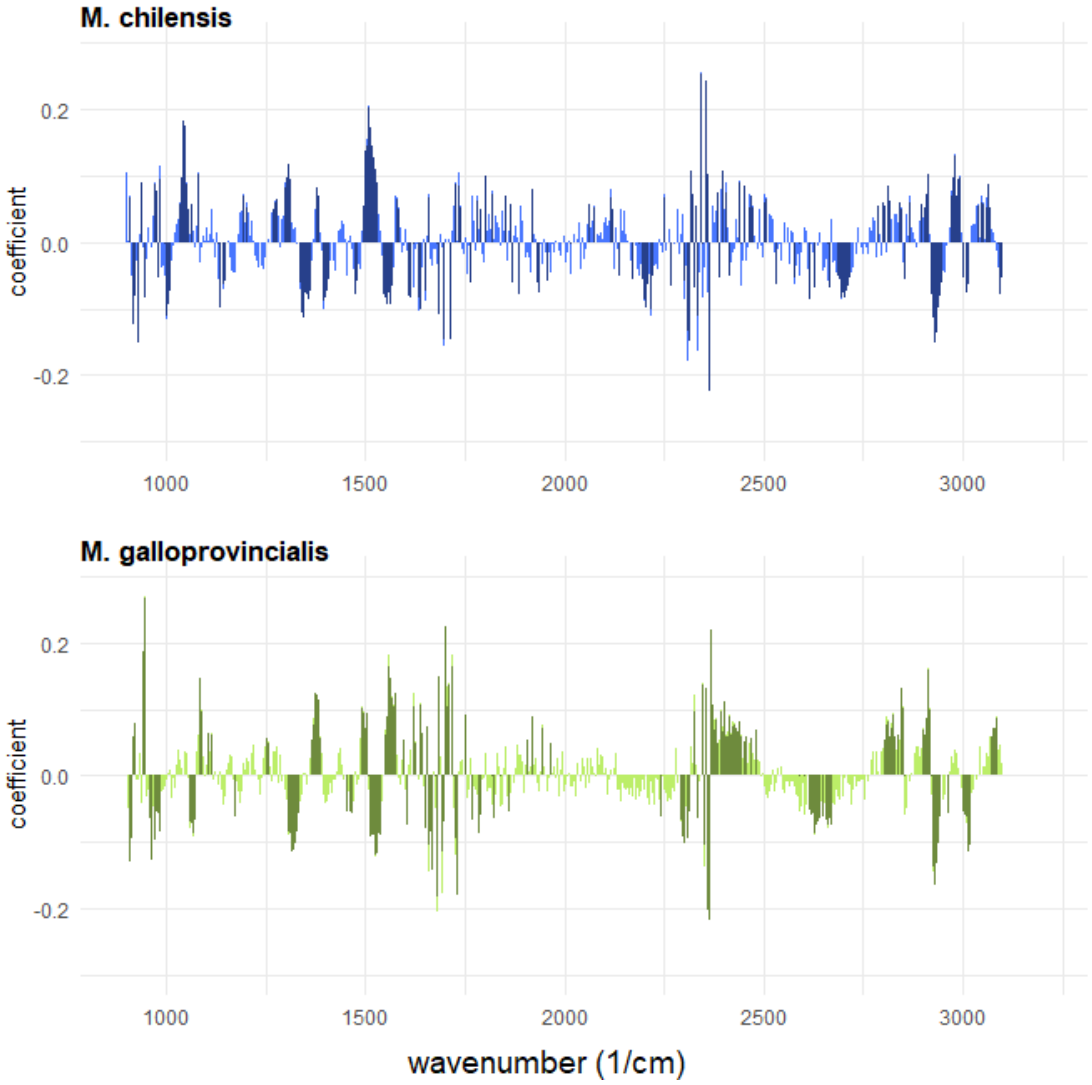


Figure 7. Most informative features (dark shaded) for the estimation of Total Aerobic Counts using FTIR data in *M. galloprovincialis M. chilensis* samples

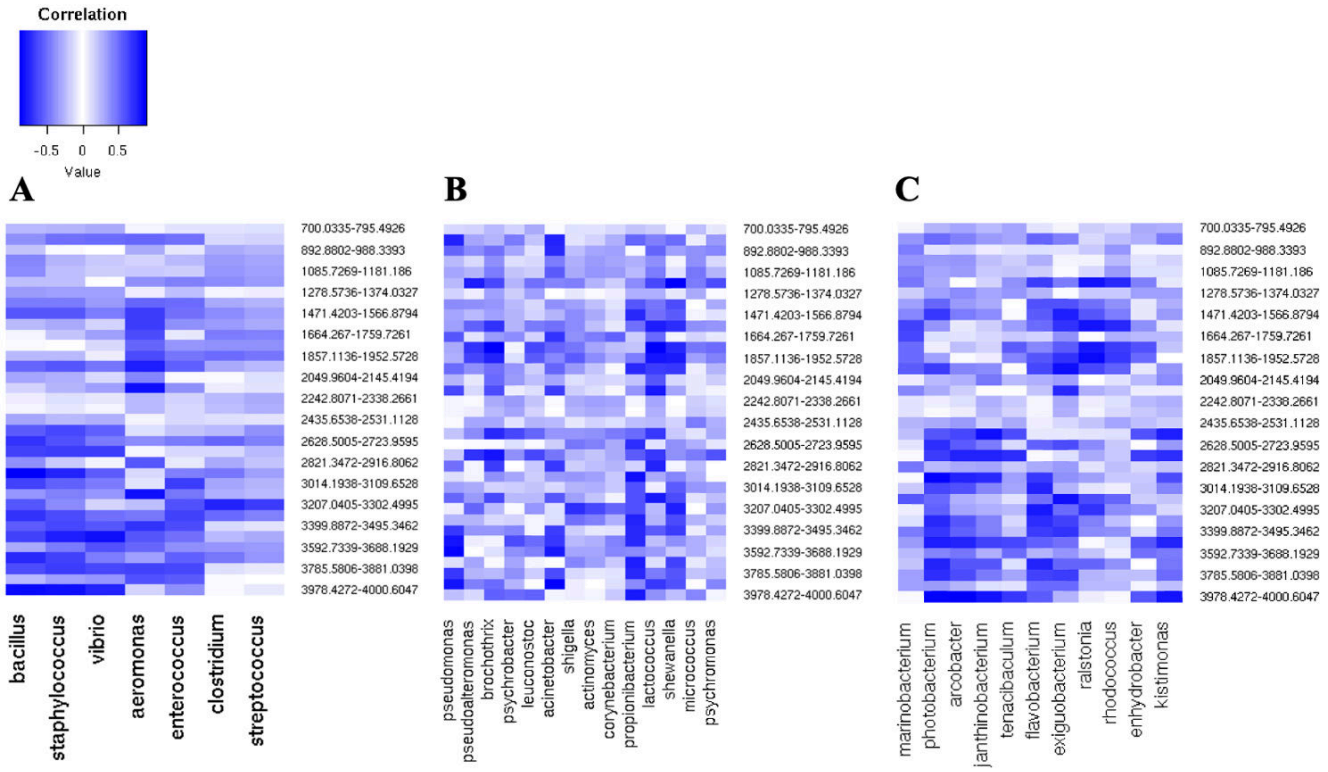


Figure 8. Correlation matrix between FTIR measures and bacterial abundance for *M.chilensis* for 3 different groups of bacteria species A) Correlations between bacteria species identified as pathogens, B) Correlations for bacteria species identified as spoilers and C) Correlations for other bacteria species.

*Photobacterium is potentially spoiler genus

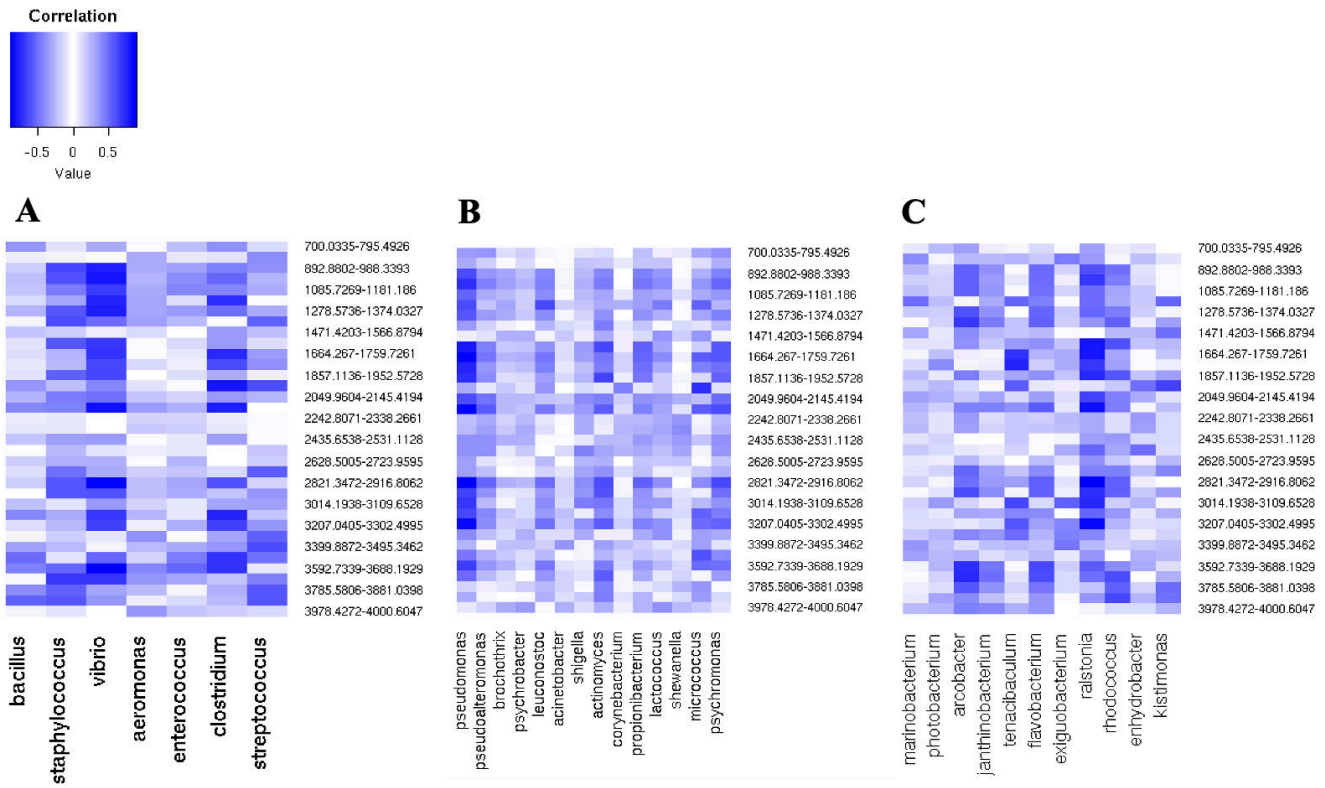


Figure 9. Correlation matrix between FTIR measures and bacterial abundance for *M.galloprovincialis* for 3 different groups of bacteria species A) Correlations between bacteria species identified as pathogens, B) Correlations for bacteria species identified as spoilers and C) Correlations for other bacteria species

*Photobacterium is potentially spoiler genus

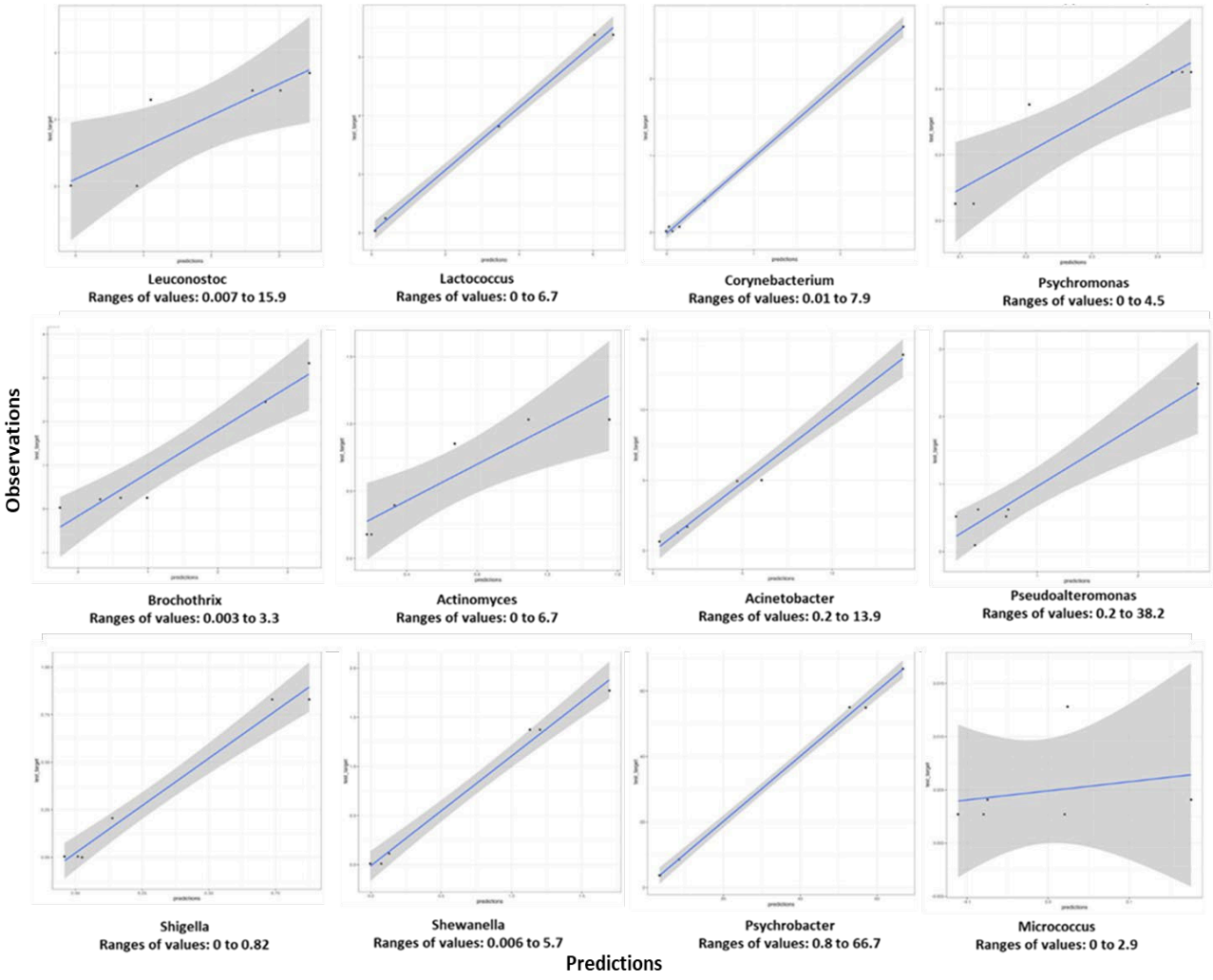


Figure 10. Predicted vs observed values of genera related to spoilage in *M. chilensis* samples using FTIR data for Neural Network model

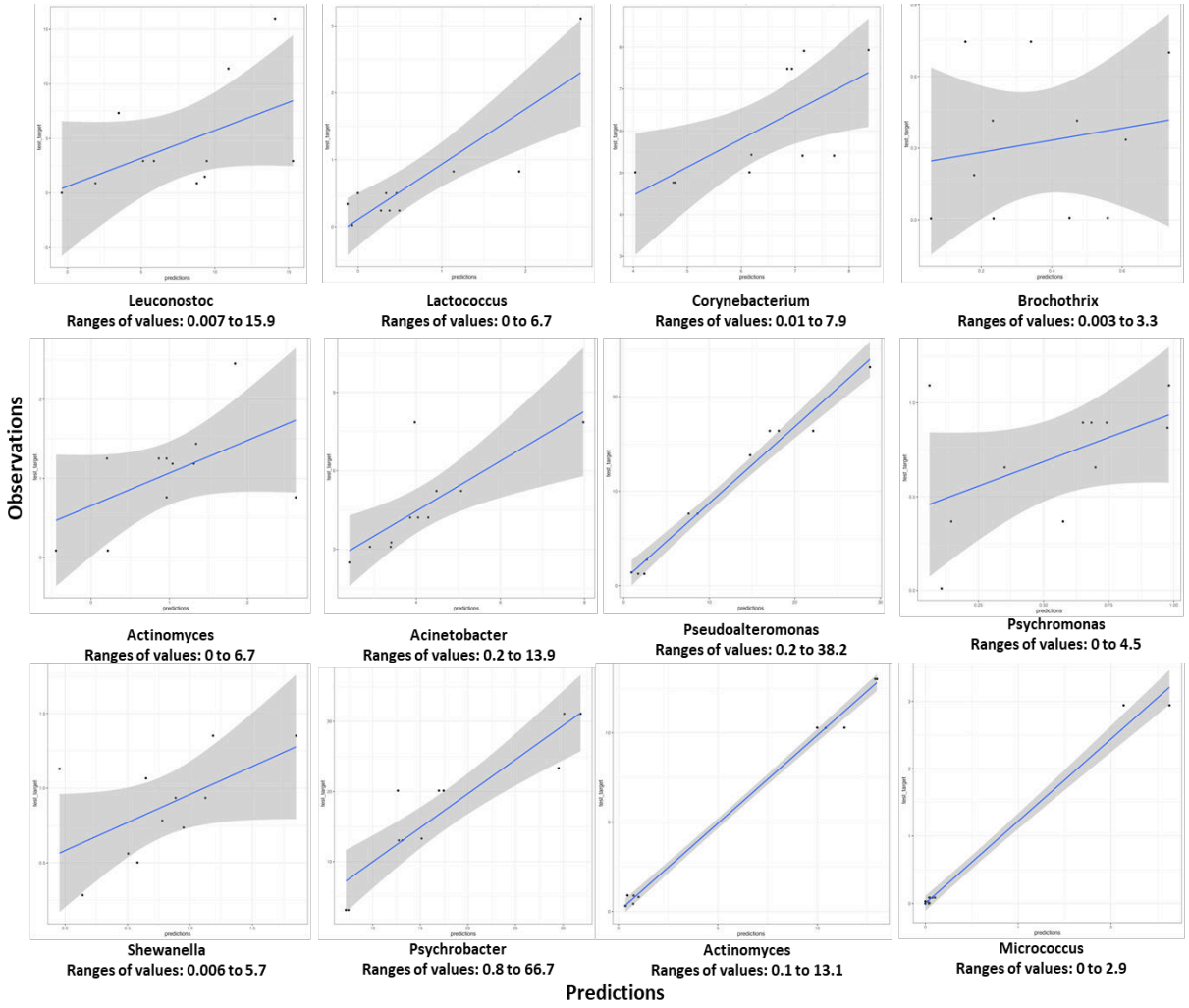


Figure 11. Predicted vs observed values of genera related to spoilage in *M. galloprovincialis* samples using FTIR data for Neural Network model

Supplementary files

Table S1. Number of reads and alpha diversity indices of different mussels (MS) during storage days (D) at different temperatures (2 ºC, 4 ºC).

Sample*	Coding used in Figures	Raw reads	Filtered/denoised reads	Shannon	Observed Features
MS1.4.D0	Batch1_4°C_day 0	58,027	33,178	6.59	383
MS1.4.D3	Batch1_4°C_day 3	54,926	33,222	4.32	143
MS1.4.D7	Batch1_4°C_day 7	49,374	29,380	3.04	60
MS2.2.D1	Batch3_2°C_day 1	55,834	37,648	6.94	346
MS2.2.D3	Batch3_2°C_day 3	42,681	28,315	6.09	282
MS2.2.D6	Batch3_2°C_day 6	39,995	26,567	4.91	212
MS2.4.D1	Batch3_4°C_day 1	49,543	32,782	6.4	303
MS2.4.D3	Batch3_4°C_day 3	48,131	32,924	5.61	266
MS2.4.D6	Batch3_4°C_day 6	56,449	39,215	5.61	272
MS4.4.D0	Batch2_4°C_day 0	49,281	24,372	5.98	247
MS4.4.D3	Batch2_4°C_day 3	39,708	18,773	4.6	150
MS4.4.D10	Batch2_4°C_day 10	43,029	15,080	3.74	52
MS5.4.D0	Batch4_4°C_day 0	18,579	12,834	6.19	139
MS5.4.D4	Batch4_4°C_day 4	25,905	17,566	5.78	106

MS5.4.D7	Batch4_4°C_day 7	43,625	26,682	6.15	238
MS5.2.D4	Batch4_2°C_day 4	53,859	23,359	6.18	148
MS5.2.D7	Batch4_2°C_day 7	53,903	33,445	6.35	147
MS7.2.D0	Batch5_2°C_day 0	55,636	36,212	6.36	195
MS7.4.D4	Batch5_4°C_day 4	34,544	24,566	3.4	87
MS7.2.D4	Batch5_2°C_day 4	47,004	32,023	4.06	116
MS7.2.D7	Batch5_2°C_day 7	36,433	25,105	6.22	141
MS7.4.D7	Batch5_4°C_day 7	43,713	22,682	5.52	132

*as deposited in National Centre for Biotechnology Information (NCBI), under the Bioproject PRJNA1091896

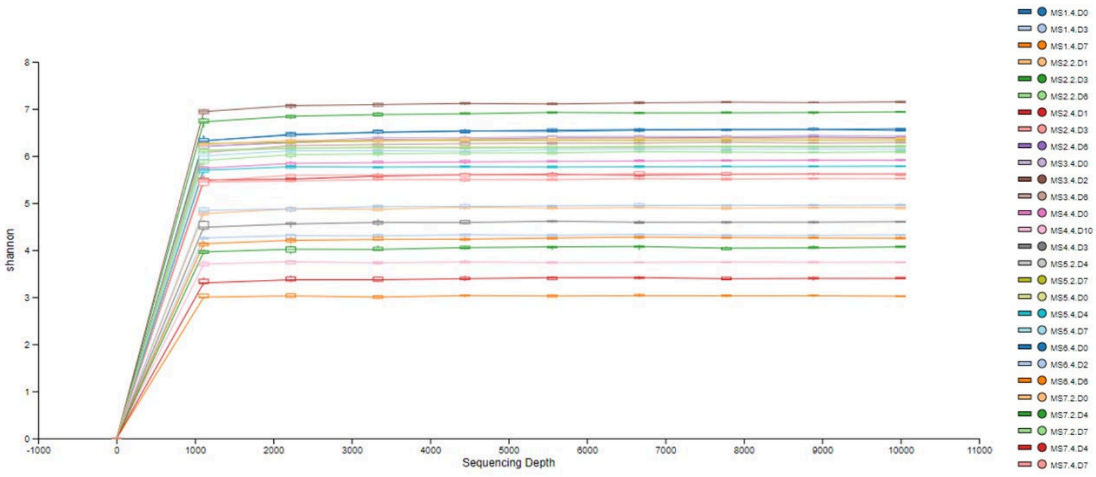


Fig. S1. Shannon-Wiener rarefaction curves of different mussels (MS) at different temperatures (2 ºC, 4 ºC) at intervals of storage time (D), obtained through 10 sampling depths.

Contribution of data acquired from spectroscopic, genomic and microbiological analyses to enhance mussels' quality assessment

Lytou, Anastasia

2024-12-01

Lytou A, Saxton L, Fengou L-C, et al., (2024) Contribution of data acquired from spectroscopic, genomic and microbiological analyses to enhance mussels' quality assessment. Food Research International, Volume 197, Issue Pt 1, December 2024, Article number 115207

https://doi.org/10.1016/j.foodres.2024.115207