Animal Microbiome

RESEARCH Open Access

Reveal your microbes, and i'll reveal your origins: geographical traceability via *Scomber colias* intestinal tract metagenomics

Bernardo Duarte[1,2*], Eduardo Feijão[1,3], Ana Cruz-Silva[1,3], Pedro Pascoal[4], Mónica Nunes[4], Marcelo Pereira[3], Andreia Figueiredo[2,3], Ricardo P. Dias[3,4], Susanne E. Tanner[1,5] and Vanessa F. Fonseca[1,5]

Abstract

The commercial demand for small pelagic fish, such as Atlantic chub mackerel (*Scomber colias*), renders them susceptible to provenance fraud. *Scomber colias* specimens intestinal tract bacteriome from five distinct fishing areas along the Portuguese Atlantic coastline were analyzed by 4th-generation sequencing. Bacteria diversity indices and differential abundance revealed dissimilarities in operational taxonomic unit (OTU) abundance among specimens from distinct fishing sites. Random forest-based model yielded an 85% accuracy rate in attributing sample provenance based on intestinal tract bacteriome OTU relative abundance. Further refinement of microbial features using Indicator Species Analysis, Linear Discriminant Analysis Effect Size (LEfSe) and OTU Gini scores enabled the identification of 3–5 bacterial OTU location biomarkers per fishing site. The intestinal tract bacteriome revealed sequences linked to pathogenic bacteria, particularly in specimens from Center-North and Center-South fishing areas. While this doesn't imply active pathogens, it highlights potential public health concerns and complements efforts to improve seafood microbiological quality and traceability.

Keywords Atlantic Chub mackerel, Small pelagic fish species, Intestinal tract metagenomics, Traceability, Provenance

Introduction

The pursuit of robust traceability tools for food commodities is pivotal for both food producers and regulatory agencies to uphold elevated food quality standards and ensure consumer safety [1, 2]. In this context, an escalating demand for dependable food traceability mechanisms has emerged, driven by the imperative to verify the geographical origin of food items and mitigate instances of commercial malpractices and adulteration [3]. This highlights the need for advanced, sophisticated labeling methodologies that enhance traceability fidelity through the integration of autonomous analytical diagnostic tools capable of validating the geographic provenance of food products [4]. In the realm of seafood products, this imperative assumes heightened significance given the critical role of traceability in the seafood industry and

*Correspondence:
Bernardo Duarte
baduarte@fc.ul.pt
[1]MARE—Marine and Environmental Sciences Centre & ARNET—Aquatic Research Infrastructure Network Associated Laboratory, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal
[2]Departamento de Biologia Vegetal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal
[3]BioISI—Instituto de Biossistemas e Ciências Integrativas, Departamento de Biologia Vegetal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal
[4]cE3c—Center for Ecology, Evolution and Environmental Changes & CHANGE—Global Change and Sustainability Institute, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal
[5]Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal

the prevention of illegal, unreported, and unregulated (IUU) fishing activities [5]. Although traceability achievements for sedentary or resident seafood species have been notable [6–11], similar success has been elusive for pelagic, highly mobile species due to their acquisition of environmental signatures across multiple habitats during their lifecycle up to the point of capture. Various analytical methodologies have been explored for this purpose, including total elemental fingerprints (TEF) derived from muscle or carbonate structures [8, 12, 13], stable isotope ratios [14, 15], fatty acids [10, 16, 17], amino acid profiles [18], and more recently, animal microbial communities [19–22]. The decreasing costs of sequencing and enhanced equipment accessibility now facilitate cost-effective high-throughput analysis of microbial communities across several matrices [23].

Recent research into fish gut microbiota has provided novel insights into the intricate interactions between microbes and their host organisms, emphasizing the core role of these microbial communities in modulating host digestive and immune systems, while also highlighting their susceptibility to modulation by a myriad of host-associated factors [24]. Habitat has been identified as the primary determinant shaping the gut microbial community in fish, surpassing trophic and taxonomic influences [25]. This underscores the potential utility of microbial communities as tracers of geographical provenance, as demonstrated in recent studies focusing on the gill microbiomes of the pelagic fish *Scomber scombrus* [26] and sessile bivalves [27, 28]. As mentioned above, it is critical to assess the efficacy and accuracy of metagenomic approaches, particularly at smaller geographical scales, and among pelagic and mobile species, which can pose challenges to conventional traceability methodologies. In this study, we aimed to leverage the potential influence of habitat-shaping fish microbial communities. According to a recent study, diverse microbial landscapes have been associated with varying degrees of anthropogenic pressure along the Portuguese coastline [29]. These heterogeneous microbial environments have the potential to significantly affect the composition of fish intestinal tract microbiomes, along with direct habitat effects on feeding and ecological behaviors.

Atlantic chub mackerel (*S. colias*) is a small pelagic fish of the Atlantic Ocean (west and east), Mediterranean Sea, and southern Black Sea [30, 31]. In the eastern Atlantic, the bulk of catches are in the northwest waters of Africa, and it was previously considered a bycatch species with low commercial value in Atlantic European waters. However, in recent years (around 20), there has been an increase in the commercial landings of this species in Atlantic waters, which is partly due to the catching efforts led by the Spanish and Portuguese purse seine fleets in the Cantabrian Sea and Atlantic Iberian waters, partly driven by efforts to offset sardine (*Sardina pilchardus*) [32]. Additionally, target campaigns to increase public interest and consumption have been implemented (e.g., by the public company responsible for the first sale of fish at auctions and to support the fishing sector in mainland Portugal), reinforcing the nutritional benefits of this species and that the consumption of this species, which is not regulated in terms of total allowed catches, contributes to an increase in the profitability of the fleet and to responsible consumption [31]. In mainland Portugal, it is the second species in terms of commercial landings (20 640 t corresponding to 19%) and the fifth species in terms of commercial value (9 057 000 € corresponding to 3%) [33]. Consequently, the escalating demand for and market prominence of *S. colias* underscores its susceptibility to food fraud concerning geographical provenance.

Against this backdrop, the development of efficient and reliable traceability tools is of paramount importance for delineating the geographical provenance of *S. colias* specimens along the Portuguese coastline. This study aimed to assess the potential of microbiome bacterial communities in the intestinal tract (in this study, intestinal tract only) of *S. colias* specimens from five sampling sites along the Portuguese mainland west and southern coasts as tracers of geographic provenance. Specifically, a machine learning approach trained and tested with samples from the aforementioned sites aimed to discern fish provenance while identifying a reduced set of bacterial taxa for future traceability studies.

Materials and methods

Sampling area and animal processing

Atlantic chub mackerel (*S. colias*) post-mortem specimens were obtained from commercial fishing landings operating in five distinct fishing zones along the North-Atlantic Portuguese coast from September to November 2020 (Fig. 1). Thirty individuals were sourced from each sampling area, with a total of 150 individuals analyzed. Refrigerated transport of specimens to the laboratory was followed immediately by individual dissection to isolate the intestinal tract and extract 1 cm of the intestine just above the anus for subsequent metagenomic analyses. All individuals were processed within a maximum of 24 h after capture. Prior to dissection, the specimens were subjected to measurements of total length (cm) and weight (g). Fulton condition index (K) is calculated using the following equation:

$$K = (Weight / Length^3) \times 1000$$

All samples were stored in sterilized microtubes at -80 °C. All materials used for animal dissection were sterilized between animals.

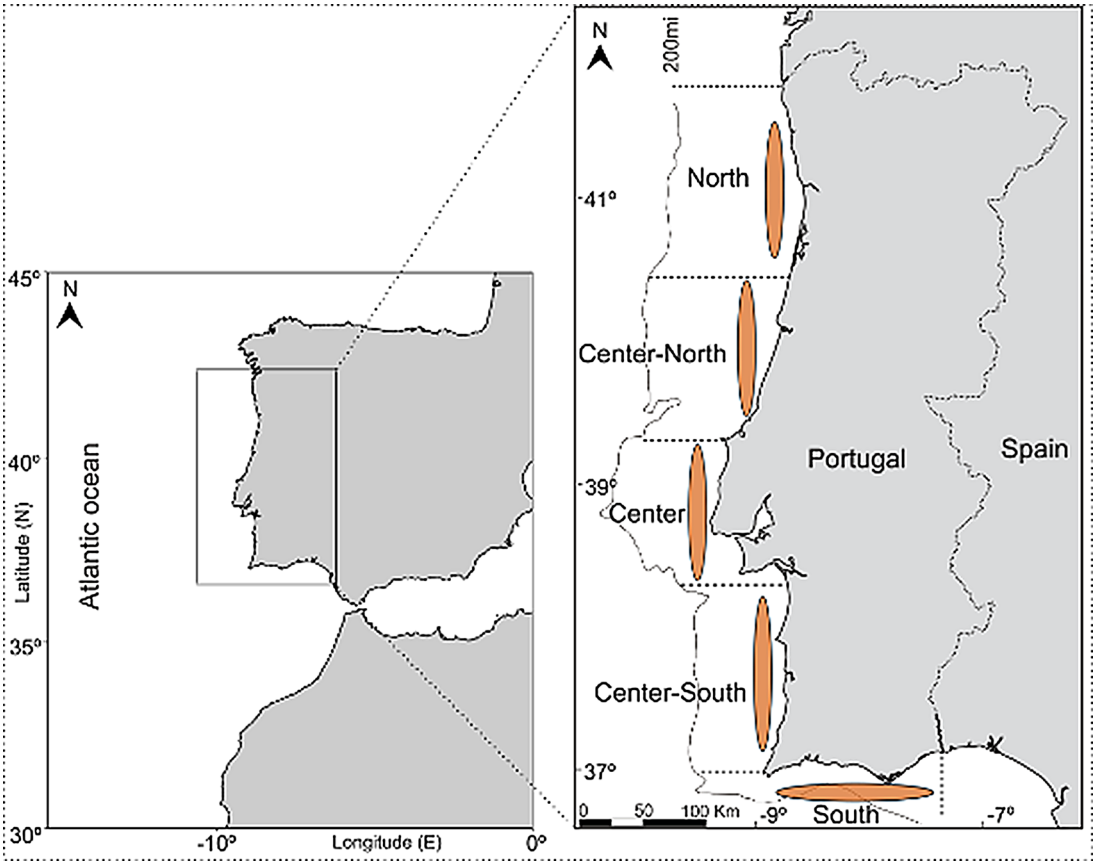

Fig. 1 Location of the five sampling areas along the west and south coasts of Portugal

Intestinal tract bacterial Microbiome 3rd generation sequencing

DNA extraction

For microbiome characterization, 20 mg of each intestine was weighed and subjected to DNA extraction using the MGISP-960 High-throughput Automated Sample Preparation System and the MGI Easy Nucleic Acid Extraction Kit. The manufacturer's instructions were followed with minor modifications regarding the incubation time with lysis buffer and proteinase K. DNA quality and concentration were assessed using a NanoDrop1000 spectrophotometer (Thermo Fisher Scientific, USA) and QubitTM 4 Fluorometer (Fisher Scientific, USA) analysis.

Amplification of 16 S rDNA gene and sequencing library Preparation

For rDNA gene amplification, Long Amp Hot Start Taq 2× Master Mix (New England Biolabs, MA, USA) was used at 1X concentration along with 50 ng/µL of genomic DNA from swab samples. To amplify the full-length 16 S rDNA bacterial gene, 0.25 µM of the primer pair 27 F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1492R (5′-CGGTTACCTTGTTACGACTT-3′) were used. PCRs were conducted on a Biometra UNO II using the following conditions: 1 cycle at 94 °C for 1 min, 35 cycles at 94 °C for 20 s, 55 °C for 30 s, and 65 °C for 2 min, with a final extension at 65 °C for 5 min. Amplification products were visualized by gel electrophoresis and purified using the Solid Phase Reversible Immobilization (SPRI) technique with magnetic beads [34]. Quantification was performed using the dsDNA HS assay for Qubit. DNA was end-repaired (New England BioLabs, MA, USA), cleaned with Agencourt AMPure XP Beads (Beckman Coulter, High Wycombe, UK), and dA-tailed (New England BioLabs, MA, USA). The library was prepared from 200 fmol of input DNA from each sample using native barcoding (SQK-LSK109 with EXP-NBD196) (Oxford Nanopore Technologies, Oxford, UK), according to the manufacturer's protocol.

Sequencing by long-read nanopore

The library was quantified and prepared for PromethION sequencing using a Flow Cell (R9.4.1) FLO-MIN106D, MinKNOW 18.12.4 software, and a standard 48-h run script with active channel selection enabled. The mean length of the sequenced reads was 1500 bp, and the mean quality score was 13.2. Data were stored in FASTQ files with a total of 38,794,398 reads. Reads with a length ≥ 300 bp, accompanied by quality scores ≥ 7, were selected for subsequent taxonomic analysis. For

taxonomic reconstruction, a custom in-house pipeline relying on k-mer taxonomic classification was employed. The analytical pipeline starts with Prinseq-lite version 0.20.4 [35] to remove the reads with less than 300 bps and with a dust score, for reduction of read complexity, below 7. Taxonomic classification was performed using Kraken2 (version 2.1.2), running on default options and with the NCBI Refseq database (Bacteria) to merge assignments into the highest common taxonomic rank (LCA) [36]. Each unique taxonomic classification assigned to the reads by Kraken2 was considered as an Operational Taxonomic Unit (OTU). This allows for a direct interpretation of taxonomic diversity without the need for traditional clustering methods.

Data analysis

Only OTUs with prevalence in at least 20% of the samples analysed (1/n, where n is the number of sampling sites) were retained for analysis (Total 3495 OTUs). OTUs are widely used in microbial community analysis, providing a classification tool for microorganisms based on genetic sequence similarity and serving as practical and flexible units for studying microbial diversity [37, 38]. OTUs serve as proxies for bacterial taxa in metagenomic studies, enabling the exploration of microbial diversity without culturing [37]. This dataset was the basis for all subsequent analyses. All statistical analyses were performed using R-Studio Version 1.4.1717. Rarefaction was performed to standardize the sequencing depth across the samples to the lowest number of reads. Alpha diversity metrics, including the Shannon diversity index and observed species count, were calculated to assess species richness and evenness within individual samples. For beta diversity analysis, Bray-Curtis dissimilarity was employed to measure the compositional differences between samples. Principal Coordinates Analysis (PCoA) was subsequently conducted to visualize these differences in a reduced dimensional space. All analyses were performed using the *phyloseq* package [39] in R-Studio Version 1.4.1717. Statistical evaluation of diversity metrics and taxa abundance data across sampling sites was performed using t-tests with the *ggpubr* package [40].

Heat trees were constructed using the *metacoder* package [41] to visualize the differential abundance between taxonomic branches along the sampling sites. Only OTUs with a complete taxonomic lineage were used to mitigate taxonomic bias at higher taxonomic levels. Standard data preprocessing procedures, as per the package guidelines, were enacted. Specifically, zero- and low-abundance OTUs (counts < 5) were filtered out and data normalization for uneven sampling was conducted [41]. The Davidson-Harel visualization layout was adopted, with the Reingold-Tilford algorithm employed for node placement. The subsequent assessment of significant differences in taxonomic abundance between sample groups was performed using the built-in Wilcoxon test within the *metacoder* package.

A Biological Observation Matrix (BIOM) with only the filtered set of OTUs fulfilling the abovementioned prevalence threshold was generated utilizing the *animalcules* package [42]. This matrix was then employed to conduct a Linear Discriminant Analysis Effect Size (LEfSe) analysis using MicrobiomeAnalyst 2.0 [43], aiming to determine the features (OTUs) most likely to explain differences between classes (fishing areas) by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance.

To evaluate the efficiency of the intestinal tract bacteriome dataset as a predictor of the animal capture area, Random Forests (RF) were implemented using the *randomForest* package [44]. The value of mtry, representing the number of variables randomly sampled at each split, was determined as the square root of the total number of features introduced in the model (3495 OTUs). The maximum number of trees was determined by evaluating the classification error using a fixed mtry value. The optimal number of trees (3500) was selected to stabilize the analysis error, while achieving the lowest training and testing error values. The variable importance within the model was assessed using Gini index values. No threshold was applied for filtering the attained Gini values. Model performance analysis was conducted by randomly partitioning the dataset into training (70% of the samples) and testing (30% of the samples) subsets. Model efficiency parameters (e.g., Sensitivity, Specificity, Positive Prediction Value, Negative Prediction Value, Prevalence, Detection Rate, Detection Prevalence, Balanced Accuracy, and Accuracy) were computed using the *randomForest* package [44].

Indicator OTUs for each intestinal tract microbiome category (fish provenance area) were identified using a point biserial correlation analysis via the *indicspecies* package [45, 46] using only the filtered set of OTUs fulfilling the abovementioned prevalence threshold. A significance level of $p < 0.05$ was used to denote significant point biserial correlations.

To derive a reduced dataset of potential biomarkers (OTUs) indicative of specimen provenance, a combined approach utilizing LEfSe LDA-scores, random forest Gini importance values, and point biserial correlation statistical values from the Indicator OTU analysis was adopted. First, only the OTUs highlighted by the OTU analysis indicator for each fishing area were retained. Subsequently, the Gini importance values and LEfSe LDA scores were assigned to the selected OTUs. A weighted combined biomarker score was computed using the three importance values according to the following formula:

$$Biomarker\ score = 0.50 \times indicspecies\ point\ biserial\ correlation\ value + 0.25 \times (Gini\ importance\ value) + 0.25 \times LEfSe\ LDA - score$$

A weighted approach was employed to assign importance to biomarker candidates, with particular emphasis on the point biserial correlation values derived from the *indicspecies* analysis. Given its intrinsic suitability for identifying category-exclusive indicator OTUs, this statistical method assumes a heightened significance in the selection process. Additionally, the incorporation of LEfSe LDA scores facilitated secondary dataset filtration utilizing a statistical framework similar to that employed in the indicator analysis. Furthermore, the inclusion of the Gini importance values augmented the biomarker score to reflect the accuracy of the resultant traceability model. Following the computation of biomarker scores, OTUs were ranked based on this composite value, with the top five OTUs exhibiting the highest scores earmarked as potential biomarkers for each site. The abundance of these selected OTUs within each provenance site, as well as their cumulative abundance, were assessed and compared across different provenance sites using the *agricolae* package [47].

Additionally, the presence of genetic sequences originating from potentially pathogenic microorganisms within the fish intestinal tract microbiome was determined using a functional analysis of the bacterial community via the *microeco* R package in conjunction with the prokaryotes database FAPROTAX [48, 49].

Results

Fish morphometrics

Significant variations in the total length and weight of the fish were observed across all sampling sites (Fig. 2A and B). Specifically, the specimens collected from the Center-North region exhibited the highest recorded values for both parameters, whereas those obtained from the Center region displayed comparatively lower values in terms of length and weight. Furthermore, analysis of the Fulton condition index revealed higher values among fish sourced from northern fishing areas (Fig. 2C).

Intestinal tract Microbiome ecology

Site-associated microbiomes of the intestinal tract (in this study, the intestinal tract only) were sequenced using long-read Oxford Nanopore sequencing technology. A total of 38,794,398 million reads were obtained, with reads meeting the criteria of length ≥ 300 base pairs and quality scores ≥ 7 selected for subsequent taxonomic analysis. Taxonomic reconstruction relied on 36,486,322 million reads and employed a custom in-house pipeline utilizing k-mer taxonomic classification, which identified 36,117,996 million reads. After rarefaction (Fig. 3A), 3495 distinct bacterial taxa were enumerated, with 3344 taxa classified at the species level. The analyzed prokaryotic community demonstrated taxonomic diversity spanning 19 Phyla, 37 Classes, 93 Orders, 223 Families, and 2021 Genera.

Regarding microbiome diversity, microbial communities exhibited significantly higher specific diversity (Shannon index) in the intestines of fish captured in the Center fishing area compared to those captured in the Center-North and South fishing areas (Fig. 3B). In contrast, the Chao1 index, a metric sensitive to rare OTUs and indicative of species richness, was significantly lower in intestinal microbial communities from fish collected in the Center-South fishing area than those from the North and Center-North. Conversely, this index was considerably higher in the southern fishing area than in fish collected at the Center and Center-South. Moreover, intestinal microbiomes of specimens from the Center area exhibited higher equitability (Pielou index) compared to the communities analyzed from samples obtained at the Center-North, Center-South, and South.

Taxonomic composition (Fig. 3C) revealed dominance in terms of abundance of *Pseudomonadota* and *Bacillota*

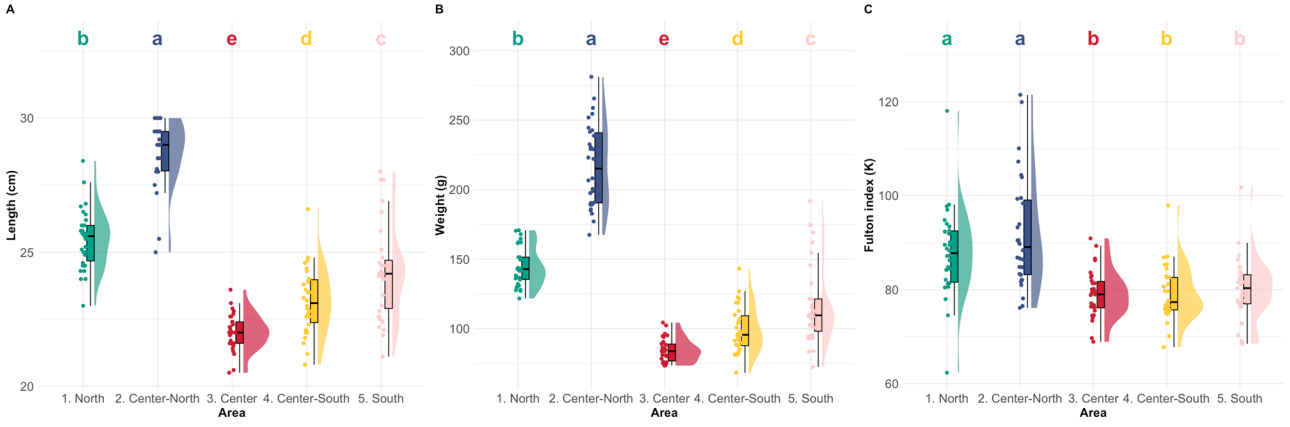

Fig. 2 *Scomber colias* total length (**A**), weight (**B**), and Fulton condition index (**C**) of the specimens collected from the five sampling areas along the Portuguese west and south coast ($N = 30$ replicates per site; letters denote statistically significant differences at $p < 0.05$)

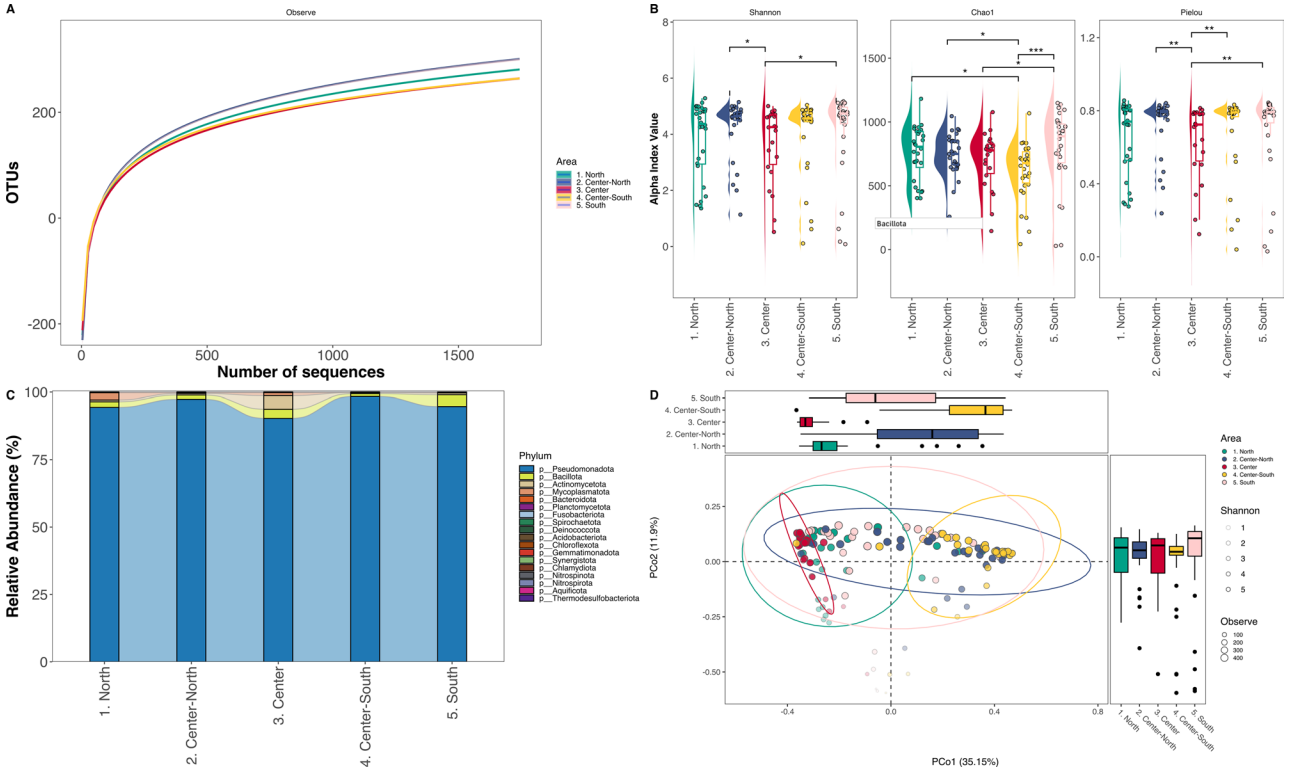

Fig. 3 Rarefaction curves (**A**), alpha diversity (**B**), phylum relative abundance (**C**), and principal coordinate analysis (beta-diversity PCoA, D) Operational Taxonomic Units of the intestinal tract microbiome of the *S. colias* specimens captured at the five surveyed fishing areas ($N = 30$ replicates per site; asterisks denote significant differences at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) using the Kruskal-Wallis test)

across samples, except for fish captured in the North and Center fishing areas, where *Mycoplasmatota* and *Actinomycetota*, respectively, exhibited higher prevalence. Regarding β-diversity (Fig. 3D), sample group separation was most pronounced along the first axis of Principal Coordinates Analysis (PCoA). The two clusters that separate along the horizontal axis are Center-south on one side and Center and North on the other side, while Center-North and South are dispersed among both clusters. Intestinal tract microbiome samples from individuals captured in the Center region formed the most cohesive cluster, with only minor dispersion attributed to outlier samples with low observation numbers and specific diversity values. Intestinal tract microbial communities from specimens captured in the North and Center-South regions also constituted distinct clusters.

Analysis of OTU abundance through comparative phylogenetic heat trees (Fig. 4) revealed notable differences in branch abundance across phylogenetic trees derived from samples originating from various fishing areas. Notably, certain branches exhibited significantly higher abundance, quantified by the number of operational taxonomic units (OTUs), in samples sourced from distinct fishing areas. Specifically, *Bacteria*, *Pseudomonadota*, and *Alphaproteobacteria* OTUs demonstrated significantly greater abundance in the intestinal tract microbiomes of individuals captured in the North area compared to those from other fishing areas, except for the Center. Conversely, *Gammaproteobacteria* exhibited an inverse trend. A comparison of the intestinal tract microbiomes between individuals from the North and Center regions revealed a lower degree of dissimilarity. In contrast, dissimilarities in OTU numbers between the intestinal tract microbiomes of individuals captured in the South region and those from other fishing areas were most pronounced when compared with samples from the Center and Center-South regions. Specifically, a significantly higher number of *Bacteria*, *Pseudomonadota*, and *Alphaproteobacteria* OTUs, along with a lower number of *Gammaproteobacteria* OTUs, were detected in samples originating from the South fishing region compared to those collected from fish intestines in the Center-South. Conversely, an inverse trend was observed when comparing the microbiomes of individuals from the South and Center fishing areas. Notably, when comparing OTU abundance in the intestinal tract microbiomes of animals captured at the Center-North location with those from the Center area, a significantly higher abundance of *Gammaproteobacteria* phylum OTUs was evident, alongside a lower number of *Bacteria* and *Pseudomonadota* OTUs in the intestinal tract microbiomes of animals captured at the Center-North. In contrast, an

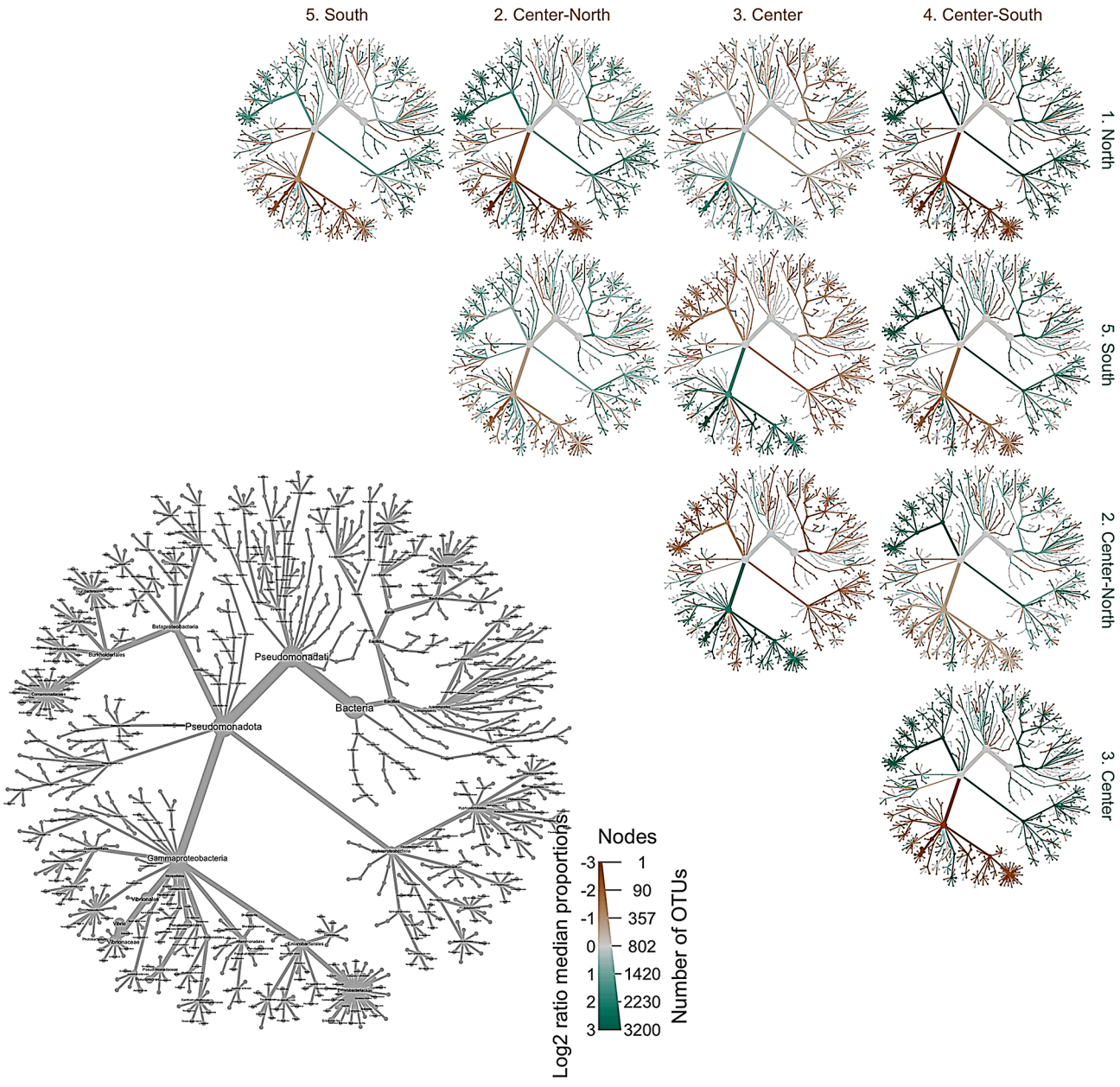

Fig. 4 Phylogenetic heat tree comparison between the intestinal tract microbiome of *S. colias* specimens captured in the five surveyed fishing areas ($N = 30$ replicates per site; colored branches denote significantly different abundance of a specific taxon at $p < 0.05$ through a Kruskal-Wallis test)

inverse trend was observed when comparing *Gammaproteobacteria*, *Bacteria*, and *Pseudomonadota* phylum OTU numbers between the intestinal tract microbiomes of individuals from the Center and Center-South areas.

Metagenomic-based traceability model

Utilizing the relative abundance data of assessed OTUs from each intestinal tract microbiome sample across distinct provenance areas, a Random Forest approach was employed to evaluate the discriminatory potential of these OTUs in the fish capture area (Fig. 5). The resulting random forest biplot (Fig. 5A) distinctly delineates three primary and well-defined sample clusters corresponding to the most divergent areas (North, Center, and South), alongside two overlapping clusters representative of intermediate fishing areas (Center-North and Center-South). Upon random partitioning of the samples into training and testing subsets for the random forest model, an overall accuracy of 76.8% and 85.0% was achieved during the training and testing phases, respectively (Fig. 5B, Table S1). Except for the Center-North and South areas, the microbiome features exhibited high (> 85%) accuracy in determining sample provenance during the model training phase. Subsequently, these

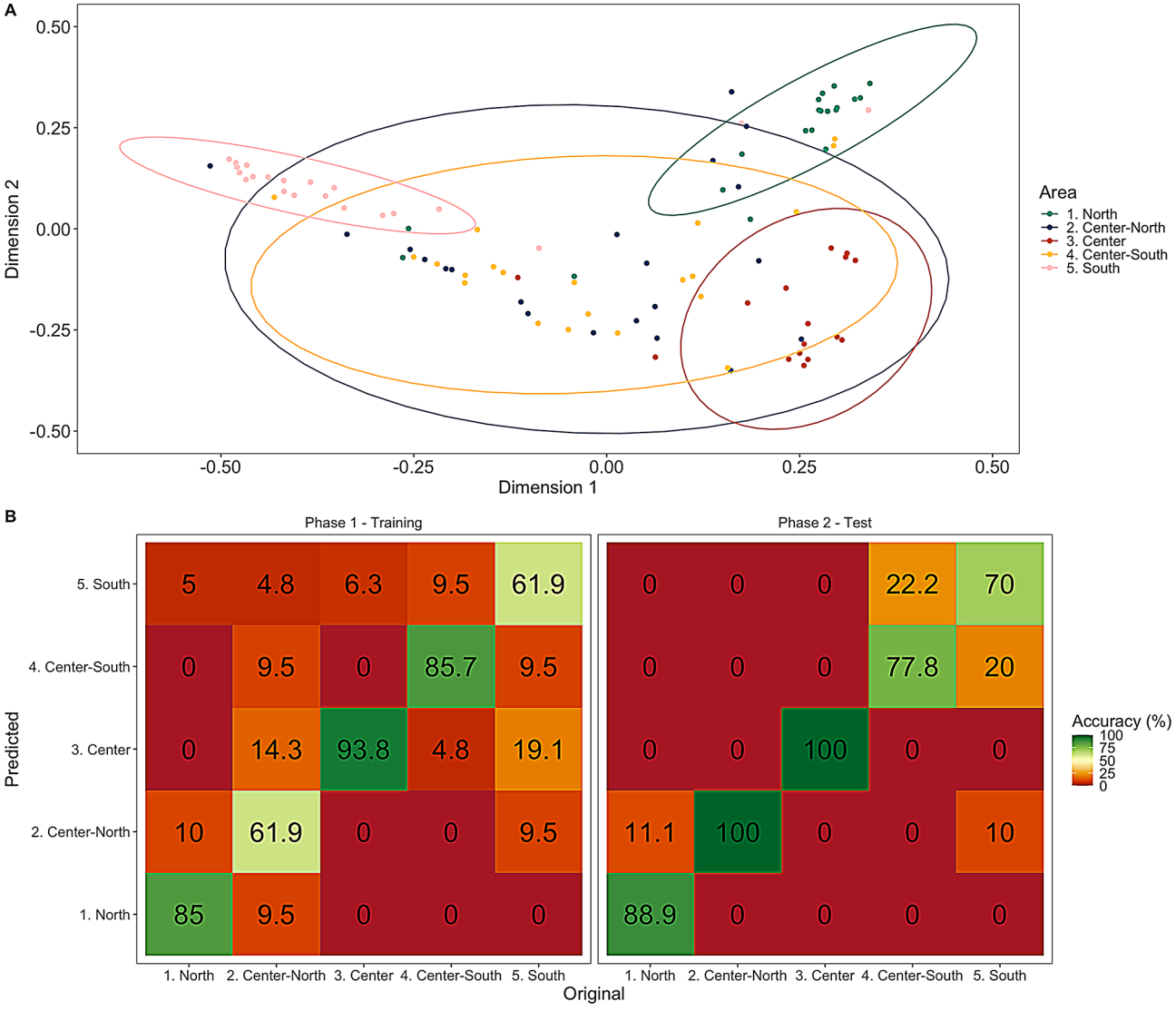

Fig. 5 Random Forest biplot (**A**) and model training and testing accuracy heatmaps (**B**) based on the *S. colias* intestinal tract microbiome OTU relative abundances assessed for the animals captured in the five surveyed fishing areas

partial accuracies were further enhanced during the testing phase, albeit with a slight decline in the classification accuracy observed for samples originating from the Center-South area. Additionally, this approach facilitated the determination of the relevance of each OTU to the model accuracy (Gini value), which subsequently informed biomarker selection.

Site-specific metagenomic biomarkers

To elucidate potential metagenomic indicators of fish provenance, two complementary approaches (LEfSe and Indicator Species Analysis) were employed and integrated with the Gini importance values derived from Random Forest modeling.

Initially, Linear Discriminant Analysis Effect Size (LEfSe) identified 102, 69, 142, 172, and 15 operational taxonomic units (OTUs) as the most likely to elucidate differences in fish provenance across the North, Center-North, Center, Center-South, and South fishing areas, respectively. Fifteen OTUs with the highest LDA scores for each site were selected (Fig. 6A).

The intestinal tract microbiomes from fish captured in the South region, which were primarily (>95% of the top 15 LEfSe OTUs) dominated by *Pseudomonadota* (Fig. 6B), all other microbiomes exhibited similar dominance by *Pseudomonadota*. In contrast, the intestinal tract microbiomes of fish captured in the South fishing area displayed high abundances of OTUs from the *Bacillota* and *Pseudomonadota* phyla (40% and 60% of the top 15 LEfSe OTUs, respectively). Furthermore, phylogenetic diversity revealed that intestinal tract communities from fish captured in the North and Center areas

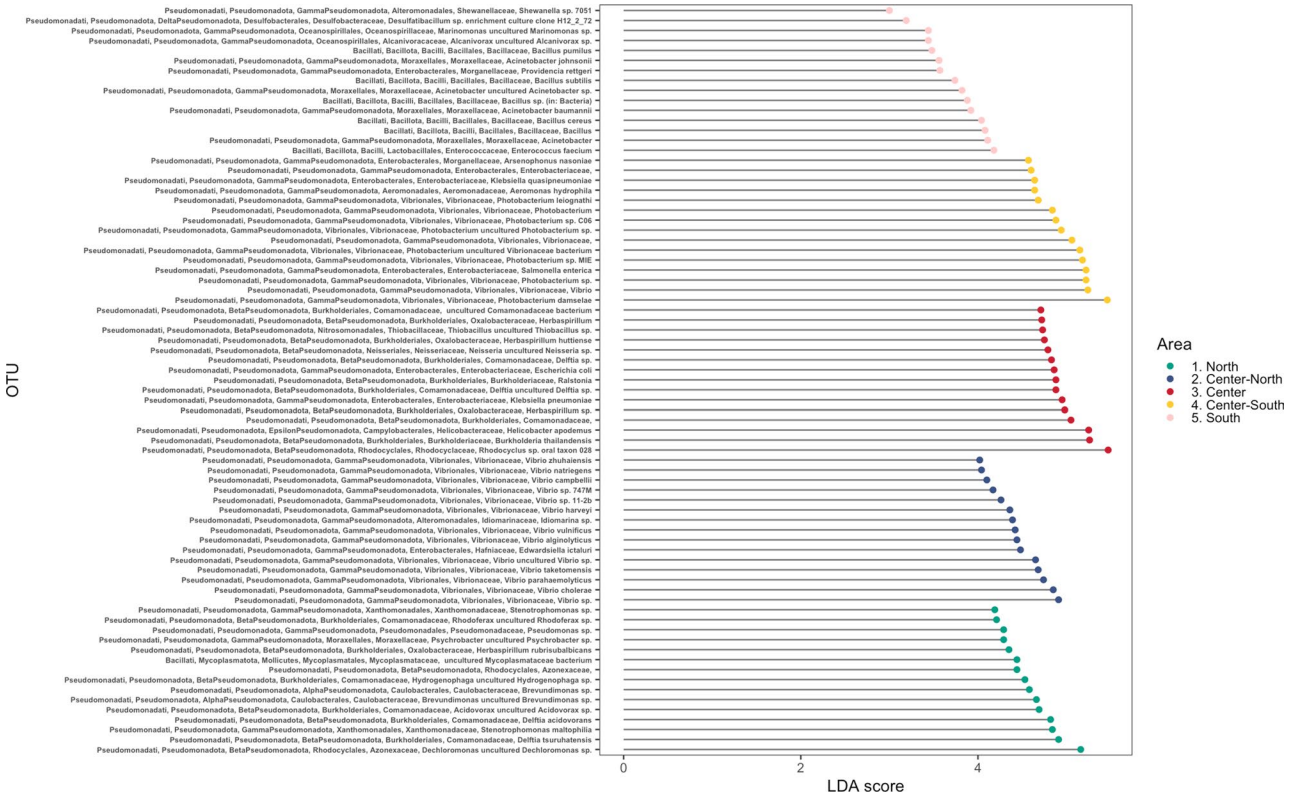

Fig. 6 *Scomber colias* intestinal tract microbiome LEfSe (Linear Discriminant analysis Effect Size) Top 15 OTU Linear Discriminant (LDA) score features by site

were predominantly characterized by discriminant OTUs from the *Alphaproteobacteria* and *Betaproteobacteria* classes. Conversely, intestinal tract microbiomes from fish captured in the Center-North and Center-South areas were dominated by *Gammaproteobacteria* class-discriminant OTUs. The intestinal tract microbiome of the fish captured in the southern region was dominated by OTUs from the *Bacilli* and *Gammaproteobacteria* classes. Specifically, the order *Burkholderiales* played a dominant role among discriminant OTUs in the intestinal tract microbiomes of fish captured in the North and Center areas, whereas the order *Vibrionales* dominated in those captured in the Center-North and Center-South areas. In contrast, intestinal tract microbial communities from the southern area exhibited a more uniform distribution of OTUs from various orders. Notably, intestinal tract microbiomes from the Center-North and Center-South areas were characterized by the dominance of the Family *Vibrionaceae* among discriminant OTUs, while those from the South fishing area were dominated by OTUs from the *Bacillaceae* family. Furthermore, intestinal tract microbiomes from the Center-North and South areas exhibited a high prevalence of discriminant OTUs belonging to the *Vibrio* and *Bacillus* genera, whereas no specific genus prevalence was discernible in microbiomes from other surveyed fishing areas.

Indicator Species Analysis (*indicspecies*) revealed 226, 167, 272, 224, and 26 site-specific indicator species ($p < 0.05$) within the intestinal tract microbiomes of specimens captured in the North, Center-North, Center, Center-South, and South fishing areas, respectively (Fig. 7). Among these, 799, 46, 41, and 15 selected indicator operational taxonomic units (OTUs) belonged to the *Pseudomonadota*, *Bacillota*, *Actinomycetota*, and *Bacteroidota* phyla, respectively, representing the most prevalent phyla among the indicator species. Nonetheless, distinct patterns emerged regarding the specificity of indicator phyla across the surveyed sites. Notably, *Actinomycetota* OTUs were selected as indicator species for all sites except those sampled in the South area, with the highest point biserial statistic value observed for OTUs from the intestinal tract microbiomes of fish captured in the Center-South area. Regarding *Bacteroidota* and *Mycoplasmatota*, indicator OTUs from these phyla were exclusively observed in the microbiomes of the fish captured in the northern and central areas. Additionally, *Planctomycetota* and *Thermodesulfobacteriota* phyla were each represented by one OTU, specifically associated with the intestinal tract microbiomes of animals captured in the North and Center-South sampling areas, respectively. While *Bacillota* OTUs were selected as indicator species across all sampling areas, their prevalence

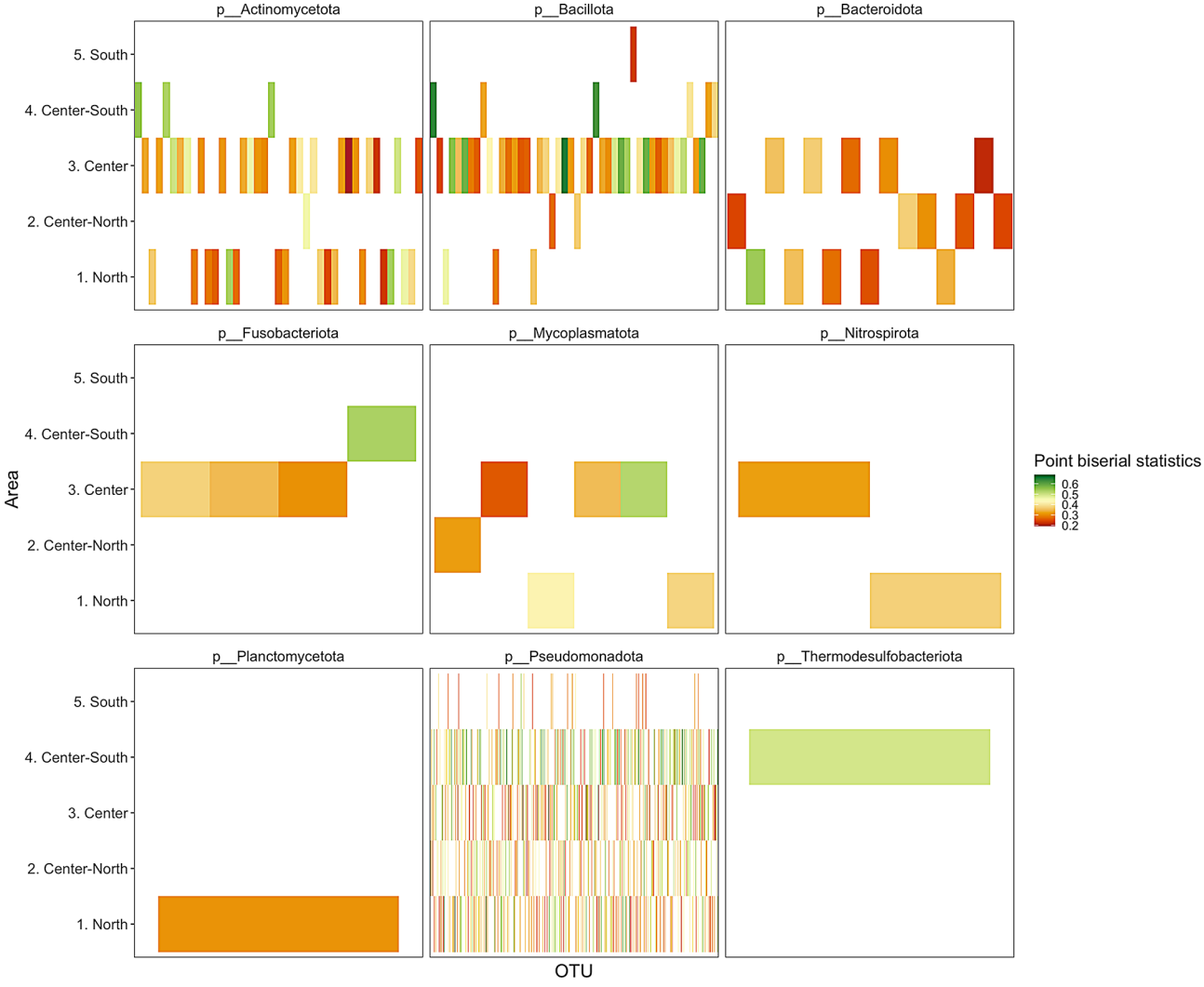

Fig. 7 *Scomber colias* intestinal tract microbiome indicator species point biserial statistical value heatmap per phylum of the five fishing areas considered

was notably higher in the intestinal tract microbiomes of fish captured in the central area than in other species. Two *Nitrospirota* OTUs were identified as potential indicators of fish captured in the North and Center areas. Lastly, the *Fusobacteriota* phylum exhibited four OTUs as potential indicator species for the Center and Center-South fish intestinal tract microbiomes, with notably high point biserial statistical values, particularly observed in the indicator OTUs assessed for the Center-South intestinal tract microbiome.

By integrating three variable selection methodologies, a refined set of candidate biomarker OTUs was identified to facilitate discrimination of the capture areas of *S. colias* specimens. This involved calculating a biomarker score and employing a weighted average of statistical values [(1) Gini importance value from Random Forest models, (2) LDA score from LEfSe, and (3) point biserial statistical value from Indicator Species Analysis] to rank OTUs according to their biomarker potential. Subsequently, the top five capture site biomarker OTUs were extracted (Fig. 8, Table S2). The relative abundance analysis of OTUs selected as biomarkers for microbiomes from fish captured in the North (Fig. 8A-E) demonstrated high specificity, with consistently elevated abundance in the North area compared to other areas. Aggregating the abundance of all North candidate biomarkers (Fig. 8F) revealed significantly higher levels than in the intestinal tract microbiomes of fish captured elsewhere. Notably, three of the five candidate OTU biomarkers for this site belonged to the *Psychrobacter* genus (Fig. 8A-C). Four out of five candidate biomarkers for animals captured in the Center-North area exhibited similar abundance values to those in microbiomes from animals captured in the Center-South area, except for OTU 76,258 *Vibrio rumoiensis* (Fig. 8J), which displayed comparable abundance in the intestinal tract microbiomes of fish from the Center-North and Center-South areas. Similarly, the combined abundance of the five candidate biomarkers

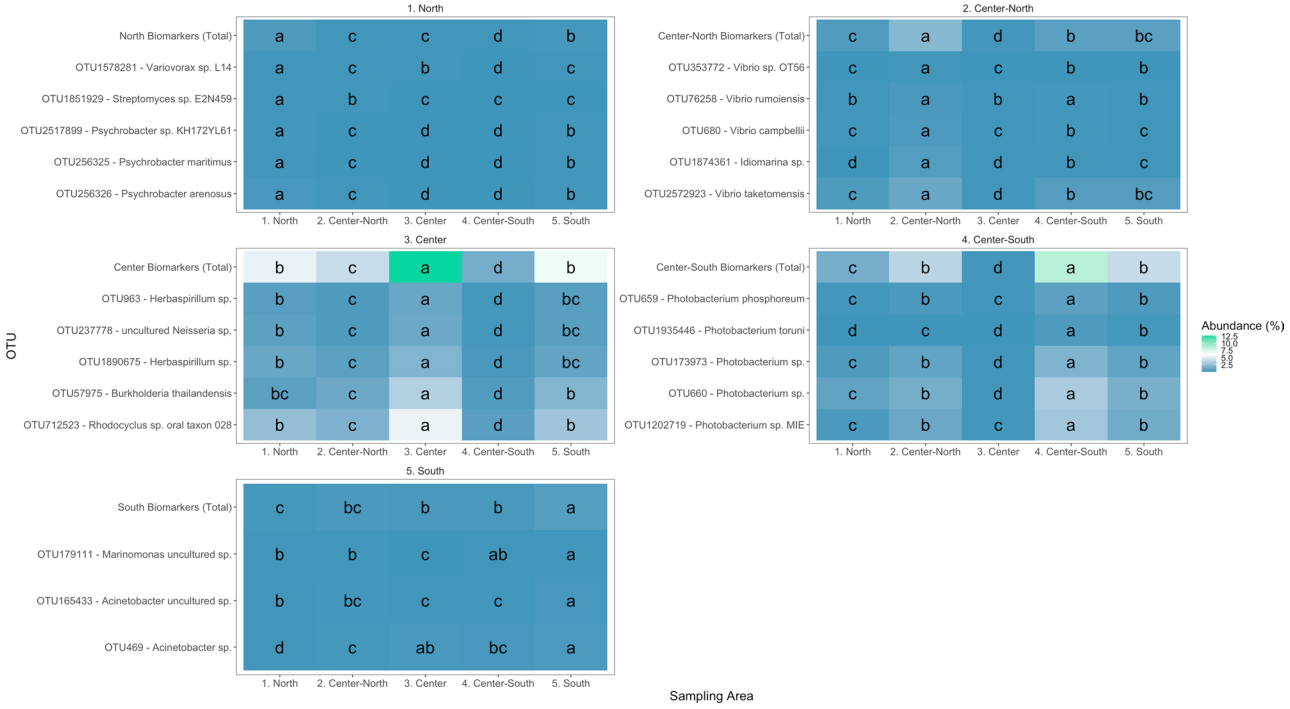

Fig. 8 *Scomber colias* intestinal tract microbiome biomarker OTUs relative abundance and overall total OTU abundance were assessed for each sampling area

(Fig. 8L) was significantly higher in the microbiomes of fish captured in the Center-North area than in those from other surveyed areas. Notably, four of the assessed biomarkers in this region belonged to the *Vibrio* genus (Fig. 8G, I-K). All OTUs selected as potential biomarkers for fish captured in the Center and Center-South regions consistently exhibited higher abundances in intestinal tract microbiomes in these areas compared to other specimens (Fig. 8M-Q and S-W). Furthermore, all candidate OTU biomarkers assessed for fish captured in the Center-South region belonged to the genus *Photobacterium* (Fig. 8S-W). The combined abundance of Center and Center-South candidate biomarkers (Fig. 8R and X, respectively) resulted in significantly higher pooled abundances among intestinal tract microbial communities of animals fished in each area. After statistical filtration (OTU selection by Indicator Species Analysis through its point biserial correlation statistical values and ranking of the selected indicators using Gini importance values and LEfSe LDA-scores), candidate OTU biomarkers for animal intestinal tract microbiomes captured in the southern region were reduced to three taxonomic units. Among these, only one (Fig. 8Z) exhibited significantly higher abundance in fish captured in this area, while the remaining two showed similar abundance values to those in the intestinal tract microbiomes of animals from the Center and Center-South areas, respectively, for OTU 469– *Acinetobacter sp.* and OTU 179,111– *Marinomonas* uncultured sp. (Fig. 8Y and A1). Nonetheless, the pooled abundance of these three candidate biomarkers was significantly higher in animals captured in the southern region than in those from other provenances.

Potential presence of pathogenic microorganisms

Assessment of bacterial diversity within the fish intestinal tract microbiome (in this study, the intestinal tract only) sampled from different coastal regions also revealed the presence of genetic material associated with potentially pathogenic microorganisms (Fig. 9). All intestinal tract metagenomes revealed the presence of potentially pathogenic bacteria. The intestinal tracts of fish captured in the Center-North, Center-South, and South regions exhibited significantly higher abundances of genetic material from microorganisms linked to human septicemia and nosocomial infections than in the other two sampled regions (Fig. 9A and C, respectively). Moreover, the intestinal tract microbiomes of fish captured in the Center-South area had significantly higher levels of genetic material associated with microorganisms implicated in human pneumonia than those in the other sampled regions (Fig. 9B). Conversely, fish originating from the Center region displayed higher abundances of genetic material indicative of meningitis-related microorganisms compared to fish captured in other surveyed areas (Fig. 9D). Intestinal tract microbiomes of fish from the intermediate sampling areas, Center-North and Center-South, exhibited significantly higher abundances of genetic sequences linked to enteric pathogens causing

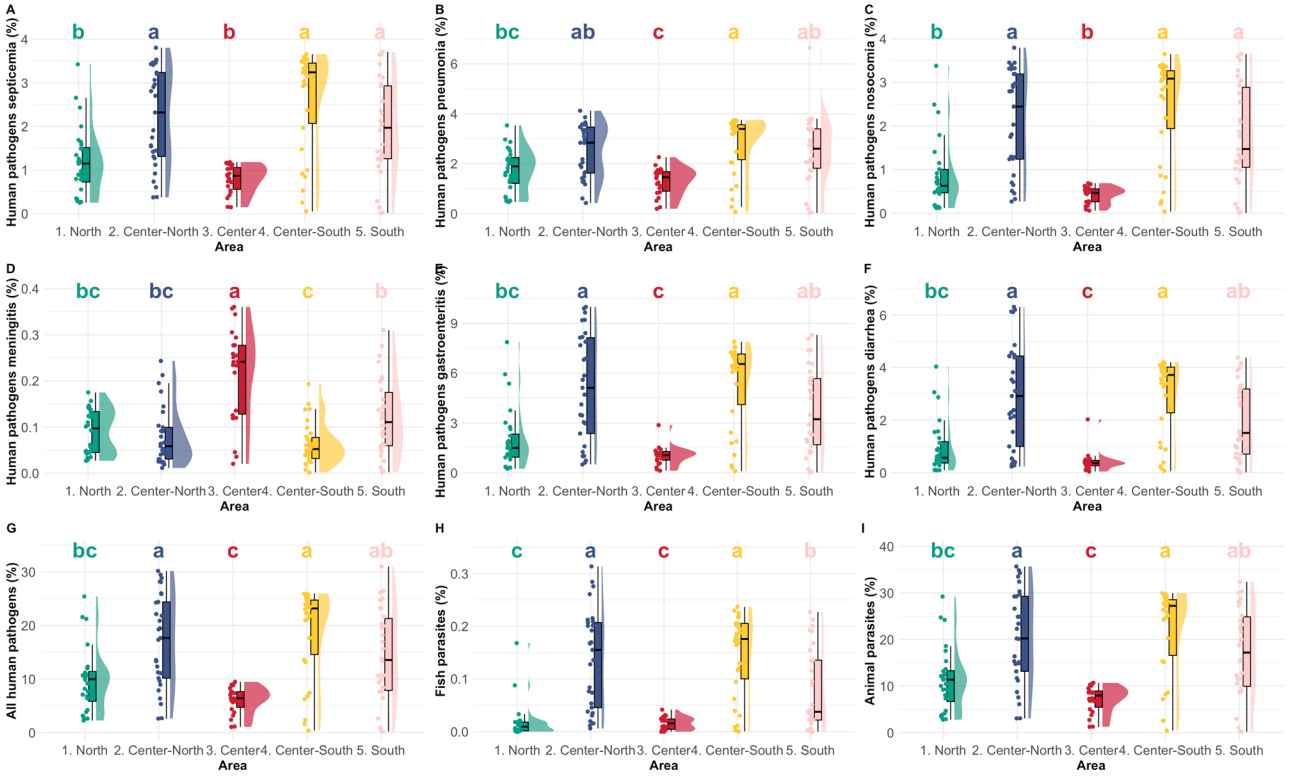

Fig. 9 The abundance of potentially pathogenic OTUs in the intestinal tract microbiome of *S. colias* in the five fishing areas considered

gastroenteritis (for example, *Campylobacter jejuni*, *Escherichia coli*, *Providencia rettgeri*, among others) and diarrhea (for example, *Morganella morganii*, *Proteus mirabilis*, *Vibrio cholerae*, and others) than fish from the other sampled regions (Fig. 9E and F), along with an overall higher abundance of genetic elements from known potentially pathogenic microorganisms compared to the other sampled regions (Fig. 9G). Additionally, fish captured in these areas displayed a significantly higher abundance of genetic material associated with fish and animal parasites within their intestinal tract microbiomes than fish from the other sampled regions (Fig. 9H and I).

Discussion

The present investigation delved into the bacterial communities inhabiting the intestinal tract of *S. colias* specimens collected in five distinct fishing areas along a 600 km long coastal stretch. Previous works already pointed out the use of animal microbiomes as powerful tools for enhancing traceability, mostly focusing on seafood production/capture methods or sessile organisms [19–22]. Thus, to our knowledge, the present paper reports for the first time an application of fish intestinal tract microbiomes to efficiently trace the fish's geographical origin, especially in highly mobile species such as *S. colias*. Analysis of ecological indices unveiled potential variations in the intestinal tract microbiome composition across sampling sites. Remarkably, the predominant bacterial phyla observed in these intestinal tract microbiomes align with previous research focusing on *S. colias* gill microbiomes [50], as well as with studies on the intestinal tract microbiomes of other seafood species, which reported a prevalence of *Pseudomonadota*, *Bacillota*, and *Bacteroidota* [28]. This trend is consistent with the general pattern observed in coastal marine waters, particularly along the Portuguese coast [29]. This also supports previous findings suggesting that the host habitat serves as the primary driver shaping fish intestinal tract microbial communities [25]. Furthermore, this study evaluated the potential of using *S. colias* intestinal tract bacterial community analysis to discern geographical origins. The results demonstrated the feasibility of discriminating the area of origin of specimens along the Portuguese mainland coast, even within small distances between the most distant sites and contiguous sampling areas. Following model training and testing, a high degree of accuracy was achieved in classifying individuals from most of the surveyed areas. The lowest classification accuracy was observed for the intestinal tract microbiome samples from the Center-South and South fishing areas. Notably, misclassification events in the Center-South area samples often resulted from classification as South area intestinal tract microbiomes, and vice versa. Considering the proximity and similarity in physicochemical characteristics

between these two fishing areas, this phenomenon may indicate either the movement of individuals between these areas or the influence of similar abiotic features shaping the intestinal tract microbiome, thereby increasing their resemblance.

As suggested by previous research [27, 28], the utilization of a traceability approach based on intestinal tract bacterial metagenomic profiling entails comparing the results of individuals to be traced with a database containing the reference bacterial fingerprint for each relevant location. However, databases specific to the intestinal tract microbiomes of local fish are currently lacking. To address this gap, machine learning techniques such as random forest modeling can be employed. This involved using a subset of reference samples (in this case, 70% of the sample population) to construct a reference model, against which the remaining samples were blindly compared. This process evaluates the efficiency of the model and the selected features when classifying unknown samples. Despite the high replication effort required, this approach not only yields a highly accurate model for classifying samples according to their origin but also provides an initial reference to intestinal tract microbiome with spatial resolution for this important species. The introduction of metagenomic sequencing approaches as tools for detecting potential pathogens offers valuable insights into the microbiological safety of seafood products, as well as into the environmental quality of coastal environments, with fish intestinal tract microbiome being proposed as a bioindicator of chemical contamination (e.g. wastewater and polycyclic aromatic hydrocarbons) in the marine environment [51, 52]. Previous studies [53] have advocated the use of cutting-edge techniques to clarify the loading, transport, and fate of pathogens in coastal environments and establish comprehensive datasets to provide reliable information for improved public health risk assessment associated with recreational water use and seafood consumption. However, the implementation of novel control measures must ensure that these are cost- and environmentally friendly and sustainable in both the pre- and post-harvest stages [54]. Effectively, our approach could lead to a continuous update of reference intestinal tract microbial landscapes from different fishing areas and other seafood species [28]. Furthermore, our study underscores this potential by employing a straightforward method based on existing databases (e.g., FAPROTAX [49]) to classify the detected Operational Taxonomic Units (OTUs) into pathogenic groups. Notably, this approach revealed a lower abundance of human pathogen-related sequences in the intestinal tract of fish captured from the North and Center fishing areas, compared to specimens captured in the Center-North and Center-South areas. The detection of genetic sequences associated with pathogenic microorganisms does not imply the presence of active pathogens. However, it and can help guide food safety surveillance and control efforts, obviating the need to inspect samples lacking genetic material from these pathogenic groups and directing focus toward sample groups or areas with higher abundances of pathogen-related sequences, facilitating further identification of potential pathogens and assessing their consumption risk. This is particularly important for seafood species that are consumed raw in the form of sushi or sashimi [55], which is not the general case of the species in this study. *S. colias*, like other small pelagic species, is a valuable fisheries resource and a valuable nutritious seafood, typically consumed cooked and without the intestinal tract. Therefore, the presence of human pathogen-related sequences in the intestinal tract is likely to pose minimal to no consumption risk.

Another innovative aspect of this study was the use of conventional microbial ecology methodologies, including Linear Discriminant Analysis (LDA) score derived from Linear Discriminant Analysis Effect Size (LEfSe) and point biserial statistical values from Indicator Species Analysis, in conjunction with machine learning models, to identify a condensed set of OTU biomarkers capable of accurately delineating the geographic origin of the animals based on the abundance of specific taxonomic traits. Machine learning techniques capitalize on large datasets, such as those provided by metagenomic approaches, to leverage numerous features with discriminatory potential [56]. However, seafood traceability based on a large number of OTUs may require specialized expertise in microbial ecology and bioinformatics. Thus, reducing the number of traits evaluated for precise classification of seafood samples according to their geographical origin adds value by enabling targeted analyses, such as quantitative PCR (qPCR), at lower analytical costs. From an ecological perspective, it was intriguing to observe that fish captured in specific fishing areas exhibited candidate biomarkers that predominantly belonged to site-specific taxonomic branches. For instance, biomarkers selected for the intestinal tract microbiomes of fish captured in the North (three out of five candidate OTU biomarkers) belonged to the *Psychrobacter* genus, whereas those from the Center-North (four out of the assessed biomarkers) were from the *Vibrio* genus, and all candidate OTU biomarkers from the Center-South were from the genus *Photobacterium*. This selection, incorporating a site-specific statistical approach (indicator species point biserial statistical value), underscores habitat specificity in shaping the animal microbiome [25]. Moreover, these taxa are known to respond to distinct environmental factors. For example, the genus *Psychrobacter* has previously been detected in northern waters gut fish species [57, 58], whereas *Photobacterium* has been identified as

a potential biomarker for polycyclic aromatic hydrocarbon (PAH) exposure [59]. The Center-South fishing area, characterized by a coastal region with the largest commercial cargo harbor in Portugal (and within the top 15–20 in Europe in terms of container cargo) and recent hydrocarbon refinery activity, represents a unique environmental setting. The *Vibrio* genus, which is typically abundant in fish intestinal tract microbiomes [60], exhibited high abundance and potential biomarker roles in individuals captured in the Center-North area. However, the specificity of OTUs from this genus as biomarkers for this specific area, coupled with the absence of potential biomarkers from the *Vibrio* genus across all areas, warrants further investigation.

Despite the obvious added value of traceability efforts and the potential for early warning signaling and monitoring of environmental and seafood microbiological quality, the utilization of metagenomic biomarkers also presents certain limitations. In the present study, individuals were sampled immediately upon landing and did not undergo the usual first sale at the fishing harbor market (usually by auction sale, but in the case of purse seiners, landings often by direct sale) and transport chain processes from auctions to consumers. This may lead to cross-contamination events and alterations in the intestinal tract microbiome due to processing and preservation techniques applied before commercialization [14]. Additionally, the occurrence of extreme and short-term events, such as sudden salinity variations, heatwaves, and algal blooms, can profoundly impact fish microbiomes owing to their rapid turnover. The influence of these factors on microbial composition remains inadequately understood and the current study design cannot predict whether such events could override geographical clustering.

Conclusions

The intestinal tract microbiomes of *S. colias* serve as effective tracers of the capture location. By utilizing machine-learning techniques, our model achieved an 85% accuracy rate for blind-sample classification. A set of location-specific operational taxonomic unit (OTU) biomarkers were developed by integrating variable importance data from these models with conventional biomarker selection methods, such as the Linear Discriminant Analysis (LDA) score derived from Linear Discriminant Analysis Effect Size (LEfSe) and the point biserial statistical value from Indicator Species Analysis. These biomarkers enable the identification of specimen origins using targeted and cost-effective technologies. Furthermore, this metagenomic approach applied to the intestinal tract microbiome of *S. colias* along the Portuguese coast also allowed the detection of genetic material indicative of human pathogenic bacteria in an untargeted manner. Although it does not confirm the presence of active pathogens in a seafood produce, that is typically cooked and consumed without the intestinal tract, it still underscores the necessity for a broader monitoring system to ensure human health safety concerning coastal water use. This methodology can also contribute to food safety surveillance and control efforts to focus on confirming pathogen contamination solely in samples in which these genetic sequences have already been identified. In summary, the metagenomic traceability tools provided herein are crucial for accurately determining sample origin and safety with high precision and resolution, even in cases involving highly mobile small pelagic fish species within closely located and contiguous sampling areas.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42523-025-00398-9.

Supplementary Material 1

Supplementary Material 2

Author contributions

B.D. Formal analysis, Writing - Original Draft, Project administration; E.F. Writing - Review & Editing; A.C.S. Formal analysis, Writing - Review & Editing; P.P., M.N. and M.P. Investigation, Writing - Review & Editing; A.F. Supervision, Writing - Review & Editing; R.P.D. Supervision, Resources, Writing - Review & Editing; S.E.T. Investigation, Writing - Review & Editing; V.F.F. Investigation, Writing - Review & Editing, Funding acquisition.

Funding

The authors would like to thank Fundação para a Ciência e a Tecnologia (FCT) for funding MARE (Marine and Environmental Sciences Centre, https://doi.org/10.54499/UIDB/04292/2020 and https://doi.org/10.54499/UIDP/04292/2020), ARNET (Aquatic Research Network Associated Laboratory, https://doi.org/10.54499/LA/P/0069/2020), BioISI (Biosystems and Integrative Sciences Institute, https://doi.org/10.54499/UIDB/04046/2020 and https://doi.org/10.54499/UIDP/04046/2020), Ce3C (Centre for Ecology, Evolution and Environmental Changes, https://doi.org/10.54499/UIDB/00329/2020 and https://doi.org/10.54499/UIDP/00329/2020) and CHANGE (Global Change and Sustainability Institute, https://doi.org/10.54499/LA/P/0121/2020). Work was also funded via the Project SEA.ID (d055db08-ebb0-4d06-b0fc-c6d54d91d552). E. Feijão was supported by FCT Ph.D. Grant (2022.11260.BD). A. Cruz-Silva was funded by a fellowship from BioSys PhD program PD65-2012 (UI/BD/153050/2022). S.E. Tanner was supported by a researcher contract (DOI: https://doi.org/10.54499/2021.02710.CEECIND/CP1650/CT0011).

Data availability

Data can be provided upon reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 19 November 2024 / Accepted: 17 March 2025
Published online: 06 May 2025

References

1. Aung MM, Chang YS. Traceability in a food supply chain: safety and quality perspectives. Food Control. 2014;39:172–84.
2. Leal MC, Pimentel T, Ricardo F, Rosa R, Calado R. Seafood traceability: Current needs, available tools, and biotechnological challenges for origin certification. 2015.
3. Milan M, Maroso F, Dalla Rovere G, Carraro L, Ferraresso S, Patarnello T, et al. Tracing seafood at high Spatial resolution using NGS-generated data and machine learning: comparing Microbiome versus SNPs. Food Chem. 2019;286:413–20.
4. Parreño-Marchante A, Alvarez-Melcon A, Trebar M, Filippin P. Advanced traceability system in aquaculture supply chain. J Food Eng. 2014;122:99–109.
5. Chuaysi B, Kiattisin S. Fishing vessels behavior identification for combating IUU fishing: enable traceability at sea. Wirel Pers Commun. 2020;115:2971–93.
6. Duarte B, Mamede R, Duarte IA, Caçador I, Tanner SE, Silva M, et al. Elemental chemometrics as tools to depict stalked barnacle (Pollicipes pollicipes) harvest locations and food safety. Molecules. 2022;27:1298.
7. Forleo T, Zappi A, Melucci D, Ciriaci M, Griffoni F, Bacchiocchi S, et al. Inorganic elements in mytilus galloprovincialis shells: geographic traceability by multivariate analysis of ICP-MS data. Molecules. 2021;26:2634.
8. Mamede R, Duarte IA, Caçador I, Tanner SE, Silva M, Jacinto D, et al. Elemental fingerprinting of sea urchin (Paracentrotus lividus) gonads to assess food safety and trace its geographic origin. J Food Compos Anal. 2022;114:104764.
9. Mamede R, Ricardo F, Santos A, Díaz S, Santos SAO, Bispo R, et al. Revealing the illegal harvesting of Manila clams (Ruditapes philippinarum) using fatty acid profiles of the adductor muscle. Food Control. 2020;118:107368.
10. Ricardo F, Gonçalves D, Pimentel T, Mamede R, Rosário M, Domingues M, Lillebø AI, et al. Prevalence of phylogenetic over environmental drivers on the fatty acid profiles of the adductor muscle of marine bivalves and its relevance for traceability. Ecol Ind. 2021;129:108017.
11. Varrà MO, Husáková L, Patočka J, Ghidini S, Zanardi E. Multi-element signature of cuttlefish and its potential for the discrimination of different geographical provenances and traceability. Food Chem. 2021;356:129687.
12. Duarte B, Carreiras J, Mamede R, Duarte IA, Caçador I, Reis-Santos P, et al. Written in ink: elemental signatures in Octopus ink successfully trace geographical origin. J Food Compos Anal. 2022;109:104479.
13. Varrà MO, Ghidini S, Husáková L, Ianieri A, Zanardi E. Advances in troubleshooting fish and seafood authentication by inorganic elemental composition. Foods. 2021;10:270.
14. del Rio-Lavín A, Weber J, Molkentin J, Jiménez E, Artetxe-Arrate I, Pardo MÁ. Stable isotope and trace element analysis for tracing the geographical origin of the mediterranean mussel (Mytilus galloprovincialis) in food authentication. Food Control. 2022;139:109069.
15. Portarena S, Farinelli D, Lauteri M, Famiani F, Esti M, Brugnoli E. Stable isotope and fatty acid compositions of monovarietal Olive oils: implications of ripening stage and climate effects as determinants in traceability studies. Food Control. 2015;57:129–35.
16. Fonseca VF, Duarte IA, Matos AR, Reis-Santos P, Duarte B. Fatty acid profiles as natural tracers of provenance and lipid quality indicators in illegally sourced fish and bivalves. Food Control. 2021; 134:108735.
17. Mottese AF, Albergamo A, Bartolomeo G, Bua GD, Rando R, De Pasquale P, et al. Evaluation of fatty acids and inorganic elements by multivariate statistics for the traceability of the Sicilian capparis spinosa L. J Food Compos Anal. 2018;72:66–74.
18. Zhao X, Liu Y, Li Y, Zhang X, Qi H. Authentication of the sea cucumber (*Apostichopus japonicus*) using amino acids carbon stable isotope fingerprinting. Food Control. 2018;91:128–37.
19. Cohen FPA, Pimentel T, Valenti WC, Calado R. First insights on the bacterial fingerprints of live seahorse skin mucus and its relevance for traceability. Aquaculture. 2018;492:259–64.
20. Pimentel T, Marcelino J, Ricardo F, Soares AMVM, Calado R. Bacterial communities 16S rDNA fingerprinting as a potential tracing tool for cultured Seabass dicentrarchus labrax. Sci Rep. 2017;7:11862.
21. Meriggi N, Russo A, Renzi S, Cerasuolo B, Nerini M, Ugolini A, et al. Enhancing seafood traceability: tracking the origin of Seabass and seabream from the Tuscan Coast area by the analysis of the gill bacterial communities. Anim Microbiome. 2024;6:13.
22. Nerini M, Russo A, Decorosi F, Meriggi N, Viti C, Cavalieri D, et al. A microbial phenomics approach to determine metabolic signatures to enhance seabream Sparus aurata traceability, differentiating between Wild-Caught and farmed. Foods. 2024;13:2726.
23. Foster JA, Bunge J, Gilbert JA, Moore JH. Measuring the microbiome: perspectives on advances in DNA-based techniques for exploring microbial life. Brief Bioinform. 2012;13:420–9.
24. Talwar C, Nagar S, Lal R, Negi RK. Fish gut microbiome: current approaches and future perspectives. Indian J Microbiol. 2018;58:397–414.
25. Kim PS, Shin N-R, Lee J-B, Kim M-S, Whon TW, Hyun D-W, et al. Host habitat is the major determinant of the gut Microbiome of fish. Microbiome. 2021;9:166.
26. Piredda R, Mottola A, Lorusso L, Ranieri L, Catanese G, Cipriano G, et al. Microbiome-based study in wild-caught *Scomber scombrus* fish products at the end of the supply chain. LWT. 2023;186:115264.
27. Liu X, Teixeira JS, Ner S, Ma KV, Petronella N, Banerjee S et al. Exploring the potential of the Microbiome as a marker of the geographic origin of fresh seafood. Front Microbiol. 2020;11.
28. del Rio-Lavín A, Monchy S, Jiménez E, Pardo MÁ. Gut microbiota fingerprinting as a potential tool for tracing the geographical origin of farmed mussels (Mytilus galloprovincialis). PLoS ONE. 2023;18:e0290776.
29. Duarte B, Figueiredo A, Ramalhosa P, Canning-Clode J, Caçador I, Fonseca VF. Unravelling the Portuguese coastal and transitional waters' microbial resistome as a biomarker of differential anthropogenic impact. Toxics. 2022; 10:613.
30. ICES. Workshop on Atlantic chub mackerel (Scomber colias). 2020. https://doi.org/10.17895/ICES.PUB.5970
31. Muniz AA, Moura A, Triay-Portella R, Moreira C, Santos PT, Correia AT. Population structure of the Chub mackerel (Scomber colias) in the North-east Atlantic inferred from otolith shape and body morphometrics. Mar Freshw Res. 2020;72:341–52.
32. Villamor B, Carrera P, Castro J, Ramos F, Velasco F, Sobrino I et al. The Chub Mackerel (Scomber Colias) in the Atlantic Spanish Waters (ICES Divisions 8.c and 9.a): Biological, fishery and survey data. 2017.
33. DGRM, Recursos da Pesca. / 2022. Lisbon: Direção Geral de Recursos Marinhos (DGRM); 2023.
34. Stortchevoi A, Kamelamela N, Levine SS. SPRI Beads-based size selection in the range of 2-10 kb. J Biomol Tech. 2020;31:7–10.
35. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.
36. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. Genome Biol. 2019;20:257.
37. Bhat AH, Department of Biotechnology K S. Rangasamy College of Technology, Tiruchengode, India, Prabhu P. Department of Biotechnology K. S. Rangasamy College of Technology, Tiruchengode, India. OTU Clustering A window to analyse uncultured microbial world. IJSRCSE. 2017;5:62–8.
38. Zalewski S, Stewart CJ, Embleton ND, Berrington JE. Brief guide to the analysis, interpretation and presentation of microbiota data. Arch Dis Child Educ Pract Ed. 2018;103:327-330.
39. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of Microbiome census data. PLoS ONE. 2013;8:e61217.
40. Kassambara A, ggpubr. ggplot2 Based Publication Ready Plots. 2023.
41. Foster ZSL, Sharpton TJ, Grünwald NJ. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. PLoS Comput Biol. 2017;13:e1005404.
42. Zhao Y, Federico A, Faits T, Manimaran S, Segrè D, Monti S, et al. Animalcules: interactive Microbiome analytics and visualization in R. Microbiome. 2021;9:76.
43. Lu Y, Zhou G, Ewald J, Pang Z, Shiri T, Xia J. MicrobiomeAnalyst 2.0: comprehensive statistical, functional and integrative analysis of Microbiome data. Nucleic Acids Res. 2023;51:W310–8.
44. Liaw A, Wiener M. Classification and regression by randomforest. R News. 2002;2:5.
45. Cáceres MD, Legendre P. Associations between species and groups of sites: indices and statistical inference. Ecology. 2009;90:3566–74.
46. De Cáceres M, Legendre P, Moretti M. Improving indicator species analysis by combining groups of sites. Oikos. 2010;119:1674–84.
47. De Mendiburu F, Simon R. Agricolae - Ten years of an open source statistical tool for experiments in breeding, agriculture and biology. 2015.
48. Liu C, Cui Y, Li X, Yao M. Microeco: an R package for data mining in microbial community ecology. FEMS Microbiol Ecol. 2021;97:fiaa255.
49. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean Microbiome. Science. 2016;353:1272–7.
50. Itay P, Shemesh E, Ofek-Lalzar M, Davidovich N, Kroin Y, Zrihan S et al. An insight into gill Microbiome of Eastern mediterranean wild fish by applying next generation sequencing. Front Mar Sci. 2022;9.

51. Al-Bahry SN, Mahmoud IY, Al-Belushi KIA, Elshafie AE, Al-Harthy A, Bakheit CK. Coastal sewage discharge and its impact on fish with reference to antibiotic resistant enteric bacteria and enteric pathogens as bio-indicators of pollution. Chemosphere. 2009;77:1534–9.
52. Walter JM, Bagi A, Pampanin DM. Insights into the potential of the Atlantic Cod gut Microbiome as biomarker of oil contamination in the marine environment. Microorganisms. 2019;7:209.
53. Stewart JR, Gast RJ, Fujioka RS, Solo-Gabriele HM, Meschke JS, Amaral-Zettler LA, et al. The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. Environ Health. 2008;7(Suppl 2):S3.
54. Sheng L, Wang L. The microbial safety of fish and fish products: recent advances in Understanding its significance, contamination sources, and control strategies. Comp Rev Food Sci Food Safe. 2021;20:738–86.
55. Lehel J, Yaucat-Guendi R, Darnay L, Palotás P, Laczay P. Possible food safety hazards of ready-to-eat Raw fish containing product (sushi, sashimi). Crit Rev Food Sci Nutr. 2021;61:867–88.
56. Mathieu A, Leclercq M, Sanabria M, Perin O, Droit A. Machine learning and deep learning applications in metagenomic taxonomy and functional annotation. Front Microbiol. 2022;13.
57. Spanggaard B, Huber I, Nielsen J, Nielsen T, Appel KF, Gram L. The microflora of rainbow trout intestine: a comparison of traditional and molecular identification. Aquaculture. 2000;182:1–15.
58. Wilson B, Danilowicz BS, Meijer WG. The diversity of bacterial communities associated with Atlantic Cod Gadus Morhua. Microb Ecol. 2008;55:425–34.
59. Spilsbury F, Foysal MJ, Tay A, Gagnon MM. Gut Microbiome as a potential biomarker in fish: dietary exposure to petroleum hydrocarbons and metals, metabolic functions and cytokine expression in juvenile lates calcarifer. Front Microbiol. 2022;13.
60. Burtseva O, Kublanovskaya A, Fedorenko T, Lobakova E, Chekanov K. Gut Microbiome of the white sea fish revealed by 16S rRNA metabarcoding. Aquaculture. 2021;533:736175.

Publisher's note