






Article

Diet and Genotype Shape the Intestinal Microbiota of European Sea Bass (*Dicentrarchus labrax*): Insights from Long-Term In Vivo Trials and Machine Learning

Silvio Rizzi ¹, Giulio Saroglia ², Violeta Kalemi ¹, Simona Rimoldi ¹ and Genciana Terova ^{1,*}

¹ Department of Biotechnology and Life Sciences, University of Insubria, Via Jean Henry Dunant 3, 21100 Varese, Italy; vkalemi@uninsubria.it (V.K.); simona.rimoldi@uninsubria.it (S.R.)

² Medical Devices Research Area, Institute of Digital Technologies for Personalized Healthcare (MeDiTech), University of Applied Sciences and Arts of Southern Switzerland, Via la Santa 1, 6962 Lugano, Switzerland; giulio.saroglia@supsi.ch

* Correspondence: genciana.terova@uninsubria.it

Abstract

To reduce dependence on oceanic resources, poultry-based ingredients and fortified feeds have become valid alternatives to fish meal (FM) and fish oil (FO). While their impact on growth performance is well established, effects on host-associated microbiota remain less characterized. This study examines the gut microbiota of European sea bass (*Dicentrarchus labrax*) following FM and FO replacement with poultry- and additive-based diets, applying machine learning (ML) to evaluate diet and genotype effects. A secondary analysis of microbial profiles from two prior trials employed classification models to determine associations between microbial abundance and categorical groupings, and regression models to assess the predictive power of ingredient variations on microbial abundance. Regressors showed limited predictive capacity, whereas classifiers performed better, particularly when genotype was considered. For poultry-based diets, average accuracy was approximately 0.4 for synergistic effects, 0.6 for diet effects, and 0.8 for genotype effects; for fortified-feed diets, average accuracy was approximately 0.2, 0.4, and 0.5, respectively. Feature selection detected microbial genera encompassing beneficial (*Brevundimondas*, *Clostridium*, *Idiomarina*, *Lactobacillus*, *Marinobacter*, *Pseudoalteromonas*, *Salinisphaera*), neutral (*Enterovibrio*, *Flavobacterium*, *Photobacterium*), opportunistic (*Acinetobacter*, *Escherichia-Shigella*, *Streptococcus*), and undercharacterized (*Acholeplasma*, *Cutibacterium*, *Enhydrobacter*, *Micrococcus*, *Peptoniphilus*, *Salegentibacter*) taxa. ML techniques thus reveal diet- and genotype-specific signatures, underlining the importance of integrated computational-microbiological pipelines.

Keywords: aquaculture nutrition; gut microbiota; machine learning; genotype–diet interactions; sustainable aquafeeds



check for updates

Academic Editor: Maria Pia Ferraz

Received: 12 November 2025

Revised: 5 December 2025

Accepted: 8 December 2025

Published: 10 December 2025

Citation: Rizzi, S.; Saroglia, G.; Kalemi, V.; Rimoldi, S.; Terova, G. Diet and Genotype Shape the Intestinal Microbiota of European Sea Bass (*Dicentrarchus labrax*): Insights from Long-Term In Vivo Trials and Machine Learning. *Appl. Sci.* **2025**, *15*, 13029. <https://doi.org/10.3390/app152413029>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid expansion of the aquaculture industry, which has significantly outpaced livestock and crop production over the last two decades [1], has increased the demand for fish-derived feed inputs, particularly fish meal (FM) and fish oil (FO), leading to disproportionate fish harvesting and raising concerns for marine ecosystems and food security [2]. Owing to its favorable functional (e.g., digestibility and palatability) and nutritional (e.g., essential amino acids and essential fatty acids) properties, FM remains a key dietary component, especially for carnivorous fish species; however, its limited availability has prompted

efforts to reduce reliance on it. As harvested wild fish consumption continues to grow, the aquaculture sector thus faces increasing pressure to define environmentally sustainable alternatives to mitigate the overexploitation of oceanic resources while, at the same time, meeting the nutritional requirements for farmed fish, without compromising growth performance or nutrient utilization [3]. Although FM replacement has usually relied on plant as well as animal by-products to reduce trophic transfer inefficiencies, new growth-promoting additives are now being explored for feed enhancement.

Among animal-rendered by-products, FM and FO replacement has been commonly achieved with poultry meal (PM) and poultry oil (PO). On the one hand, PM shares functional and nutritional properties with FM, offering advantages such as stable availability and lower cost. However, like other animal-based ingredients, PM composition is subject to variability, often resulting in nutritional deficiencies (especially essential amino acids) that might hinder the fulfillment of species-specific nutritional requirements [4]. Despite these shortcomings, which have usually been addressed with the supplementation of deficient nutrients or with the combined use of alternative sources, PM has been utilized as an FM substitute at different inclusion levels in aquafeed formulations, influencing production efficiency in a species- and environment-dependent manner. This has prompted investigations into its effects on the gut microbiota across aquaculture species, including large yellow croaker (*Larimichthys crocea*) [5], gilthead seabream (*Sparus aurata*) [6,7], Japanese abalone (*Haliotis discus hannai*) [8], rainbow trout (*Oncorhynchus mykiss*) [9–11], and European sea bass (*Dicentrarchus labrax*) [12]. On the other hand, PO has emerged as a promising candidate for FO replacement, in spite of differences in chemical composition. FO contains long-chain polyunsaturated fatty acids (LC-PUFAs), particularly $n - 3$ LC-PUFAs (e.g., eicosapentaenoic acid, EPA; docosapentaenoic acid, DPA; docosahexaenoic acid, DHA); in turn, PO is abundant in monounsaturated fatty acids (MUFAs) and $n - 6$ LC-PUFAs [13]. Nevertheless, FO substitution with PO has been less explored, with existing studies focusing on gilthead seabream [6] and yellowtail kingfish (*Seriola lalandi*) [13].

Beyond FM and FO replacement to enhance growth performance and feed efficiency, aquafeeds have also been supplemented with functional additives, notably organic acids (OAs). While widely adopted in livestock production, their application in aquaculture has been limited and is currently focused on high-value species. As nonantibiotic compounds with defined antimicrobial activity spectra, OAs are commonly administered as blends (OABs) to overcome the inconsistencies observed with single-acid applications. Similar to their growth-promoting effects, which are influenced by biological variables (e.g., species, physiological traits, and age) and extrinsic factors (e.g., rearing environment, dosage, and blend concentration), OABs have also been shown to modulate the gut microbiota in species- and formulation-specific manners, primarily by suppressing potentially pathogenic microbial populations [14]. Moreover, for enhanced functional outcomes, OABs have often been combined with other bioactive feed components, such as essential oils, plant extracts, and enzymes. Microbiome studies have reported these synergistic effects in multiple species, including European sea bass [15], rainbow trout [16], and Nile tilapia (*Oreochromis niloticus*) [17,18].

With the rise in antibiotic resistance, plant-derived compounds have progressively gained prominence as environmentally sustainable feed additives for the control of disease outbreaks in aquaculture farming thanks to successful applications in livestock production [19]. Indeed, phytogenic feed additives (PFAs) are generally recognized for their therapeutically favorable characteristics, including minimal side effects, reduced drug resistance, and economic feasibility. Furthermore, PFAs exhibit anabolic (biomass enhancement, nutrient utilization, energy retention), immunoprotective (disease resistance, stress

response), and cytoprotective (antioxidant, anti-inflammatory) properties [20]. Their bioactive components (e.g., polysaccharides, flavonoids, polyphenols, alkaloids, saponins, and terpenoids) have especially demonstrated efficacy against a wide range of fish diseases, including bacterial, viral, fungal, and parasitic infections [21]. Although PFA effects vary with concentration and preparation methods, phytochemicals could promote beneficial intestinal microbiota through fermentation, while suppressing pathogenic microbes [20]. In this regard, investigations into PFA-mediated microbiota modulation have been conducted in several fish species, including largemouth bass (*Micropterus salmoides*) [22], barbel chub (*Acrossocheilus fasciatus*) [23], common carp (*Cyprinus carpio*) [24], Nile tilapia [25], and rainbow trout [26]. While the precise mechanisms behind PFA function in fish remain to be thoroughly elucidated, dietary intervention continues to offer one of the most practical approaches to promote aquatic animal health [27].

In light of growing restrictions on chemotherapeutic use in aquaculture due to bioaccumulation risks, probiotics, together with plant-derived compounds, have emerged as suitable feed additives with multiple beneficial properties [28]. Indeed, probiotic bacteria enhance growth, nutrient digestion, pathogen resistance, and stress tolerance, while probiotic yeasts support gut microbiota modulation, immune stimulation, and growth improvement. In contrast, probiotic filamentous fungi primarily promote the production of digestive enzymes, including amylases, xylanases, cellulases, β -glucanases, lipases, and proteases. Although typically of non-fish origin, multi-strain probiotics are increasingly derived from the native gastrointestinal microbiota of the fish host [29]. In this regard, a dosage of 1×10^6 colony-forming units (CFUs) per gram of feed is generally suggested to ensure viable colonization [30], despite optimal efficacy depending on biological (species, developmental stage) and non-biological (rearing environment, administration period, administration method) factors [28]. Common probiotic strains usually include lactic acid bacteria (e.g., *Bacillus*, *Enterococcus*, *Lactobacillus*, and *Lactococcus*) and yeasts, such as *Saccharomyces cerevisiae* [31–33]. Due to their influence on intestinal microbiota composition and homeostasis, probiotics have thus been investigated in various fish species, including Atlantic salmon (*Salmo salar*) [34], gilthead seabream [35,36], rainbow trout [37], and Nile tilapia [38]. Nevertheless, commercial strain selection must account for potential risks, including dysbiosis and zoonotic transmission [28].

As a commercially valuable carnivorous marine species, European sea bass has been selectively bred to improve productivity (e.g., growth, feed efficiency, processing yields) and robustness (e.g., disease resistance, stress endurance, hypoxia tolerance) traits [39]. In particular, breeding programs have expanded considerably in recent years through the application of parentage assignment via microsatellite markers, enabling accurate pedigree reconstruction without the need for physical tagging. This genetic approach has facilitated family-based selection for quantitative traits while simultaneously supporting genetic parameter estimation and inbreeding management. When combined with reproduction control through artificial fertilization, fish farmers have been able to establish controlled mating designs enhancing genetic gain and promoting the long-term sustainability of sea bass aquaculture [40]. Given the selection of high-performing genotypes, it has become increasingly necessary to assess how dietary regimens influence intestinal microbiota composition in improved strains, as investigated in Torrecillas et al. [41] and Rimoldi et al. [42].

Building on the findings of Torrecillas et al. [41] (hereafter referred to as Study 1) and Rimoldi et al. [42] (hereafter referred to as Study 2), this study seeks to deepen the analysis of the intestinal microbiota of European sea bass by applying artificial intelligence (AI), with particular emphasis on machine learning (ML), to uncover patterns within the datasets obtained from both investigations. While machine learning has become essential to microbiome studies in human and livestock research, its application in aquaculture has remained

limited [43]. In this regard, machine learning has been successfully applied in aquaculture monitoring and analytics (e.g., species classification, biomass estimation, size measurement, behavior analysis, disease prevention, feeding frequency optimization, and water quality monitoring), spanning image-based and non-image-based implementations [44,45]. However, its use in fish-related microbiome studies has been scarce, with existing work largely restricted to the investigation of microbiota composition in relation to environmental conditions [46,47] and network modeling of microbial interactions with biotic and abiotic rearing parameters [48]. To the best of our knowledge, machine learning has not yet been applied to investigate the relationship between microbial profiles and dietary formulations, especially in the context of feeds derived from more environmentally sustainable sources. Such applications are urgently needed in the aquaculture sector to minimize the depletion of marine stocks, reduce the disruption of food webs, and reduce the pressure on other intensive production systems [2,49]. This knowledge gap is particularly relevant given the growing interest in ML-driven design of environmentally efficient aquafeeds [50] and the promising outcomes reported in livestock nutrition [51]. While several bioinformatic tools are currently available for microbiome analysis, such as the Analysis of Compositions of Microbiomes with Bias Correction (ANCOM-BC) for differential abundance testing [52], the Statistical Analysis of Metagenomic Profiles (STAMP) for taxonomic and functional profile testing [53,54], the Linear Discriminant Analysis Effect Size (LEfSe) for biomarker discovery [55], they are not intended for predictive modeling, being restricted to specific domains and generally unsuitable for commercial-scale applications. In contrast, machine learning delivers the scalability and adaptability needed to harness high-dimensional multi-omics datasets generated via next-generation sequencing (NGS) technologies [56]. Unlike conventional bioinformatic methods that often analyze features separately, machine learning also enables joint feature analysis, thus offering a more comprehensive perspective on microbial communities. In light of the considerable advantages afforded by machine learning, we developed a structured pipeline to reveal associations that may have been overlooked by conventional bioinformatic approaches routinely used in microbiota research. Accordingly, the following research objectives (ROs) were established: RO1, to investigate the presence of a distinct subset of microbial taxa exerting significant influence on predictive outcomes; RO2, to assess potential associations between gut microbiota composition and categorical combinations; RO3, to evaluate the extent to which dietary ingredient composition variations can predict shifts in microbial abundance.

2. Materials and Methods

2.1. Data Source and Experimental Procedures

Torrecillas et al. [41] examined the interplay between genotype and diet in the intestinal microbiota of European sea bass using isolipidic, isonitrogenous formulations: a control diet (20% FM/5–9% FO) and an experimental diet (10% FM/10% PM/0% FO/1–3% PO/2–3% DHA oil). Diets were tailored to three developmental stages and produced in three pellet sizes to meet nutritional demands. Following a 42-week pre-trial period, fish were reared for 87 weeks in a flow-through system (FTS) under natural photoperiod conditions. At trial completion, autochthonous intestinal microbiota was sampled and analyzed via 16S rRNA gene sequencing (V4 hypervariable region). The authors found that, when considering high-growth (HG) and wild-type (WT) genotypes, genotype exerted a stronger influence on microbiota composition than diet. In particular, HG fish exhibited lower inter-individual variability, suggesting greater adaptive capacity to dietary changes as a result of selective breeding; conversely, genotype effects were largely confined to specific bacterial taxa, regardless of the diet.

Building on the formulation developed by Torrecillas et al. [41], Rimoldi et al. [42] used this diet as a control and supplemented it with individual functional additives: probiotics (*Bacillus subtilis*, *Bacillus licheniformis*, and *Bacillus pumilus* at 2×10^{10} CFU/g in equal ratios), organic acids (sodium butyrate), and phytogetic compounds (garlic extract with medium-chain fatty acids). After 42 weeks to reach the juvenile stage, fish were reared for 12 weeks in a recirculating aquaculture system (RAS), with feeding consisting of a high-dose phase (10 g/kg probiotics, 7.5 g/kg organic acids, 7.5 g/kg phytoGENICS for 2 weeks) followed by a low-dose phase (2 g/kg probiotics, 3 g/kg organic acids, 5 g/kg phytoGENICS for 10 weeks). After each phase, fish underwent a pathogen challenge (10^5 CFU *Vibrio anguillarum* per fish via anal inoculation) and a stress test (high stocking density). At the end of the feeding trial, autochthonous intestinal microbiota was sampled and analyzed via 16S rRNA gene sequencing (V4 hypervariable region). Despite supplementation, the results showed no clear separation between control and additive-treated groups, although the relative abundance of specific taxa differed between genotypes.

2.2. Machine Learning Pipeline Implementation

Starting from the findings reported in Torrecillas et al. [41] and Rimoldi et al. [42], we used the microbial counts derived from 16S rRNA gene sequencing on the autochthonous intestinal microbiota of European sea bass from both studies; in this regard, the sequencing data had previously been deposited as FASTQ files into the European Nucleotide Archive (ENA) under accession numbers PRJEB47388 [41] and PRJEB61519 [42], with the preprocessing pipeline for metabarcoding sequencing data documented in the respective paper. Taxonomic classifications assigned to the amplicon sequence variants (ASVs) from both datasets were utilized without further modification or subset pooling, with SILVA as the reference database for taxa identification. In this study, we leveraged ML to uncover hidden patterns behind individual and synergistic relationships between diet and genotype across both datasets. To this end, ML analyses were conducted using ad hoc encodings to assess the influence of the categorical combinations considered across both studies (Table 1 and Figure 1). While microbial counts were available at different taxonomic levels (phylum, class, order, family, genus, and species), we focused on lower resolutions, namely family and genus, to ensure higher specificity and biological relevance [57]; in particular, the species level was excluded because taxonomic resolution at this rank was insufficient in our 16S rRNA gene amplicon data generated with Illumina technology, whose short-read amplicon sequencing often lacks the sequence length and unique variation required to reliably discriminate closely related species [58,59].

To implement an efficient pipeline, we took both framework selection and model assessment concerns into consideration. On the one hand, we selected Python 3.13.3 as the reference programming language owing to its extensive support for data analysis libraries, which, in the proposed implementation, included: Pandas 2.2.3, NumPy 2.2.5, SciPy 1.15.2, scikit-learn 1.6.1, XGBoost 3.0.0, CatBoost 1.2.8, Matplotlib 3.10.1, and SHAP 0.46.0. On the other hand, given the small size of the original datasets, we applied leave-one-out cross-validation (LOOCV) to maximize information utilization and minimize prediction bias, despite the increasing number of microbial features over taxonomic levels. Furthermore, we applied evaluation metrics tailored to balanced datasets to assess model performance: for classification, we used the accuracy (ACC) and the Matthews correlation coefficient (MCC); for regression, we used the coefficient of determination (R^2), the mean absolute error (MAE), which assigns equal weight to errors, and the root mean squared error (RMSE), which assigns greater weight to larger discrepancies. While other performance metrics exist, particularly for classification, they are more frequently applied when dealing with imbalanced datasets. Common metrics include precision (i.e., the fraction of true positives

among all predicted positives), recall (i.e., the fraction of true positives among all actual positives), and F1-score (i.e., the harmonic mean of precision and recall). However, for balanced datasets, accuracy offers more immediate interpretability, while MCC is considered more informative than F1-score as it takes into account the balanced ratio of confusion matrix categories (true positives, false positives, true negatives, false negatives) [60]. Given the exploratory nature of this study, the highest performing model within each categorical grouping was determined through a comparative assessment of performance metrics rather than formal inferential testing. Accuracy was prioritized for classification tasks, while R^2 was emphasized for regression tasks, as both are widely accepted indicators of predictive reliability. The use of this strategy enabled the recognition of promising models while avoiding additional statistical analyses, with the understanding that more rigorous validation will be needed in future confirmatory studies. With these implementation choices, we sought to strike an effective balance between model reliability and output informativeness.

As a fundamental prerequisite for predictive modeling, both datasets underwent preprocessing to retain samples pertinent to the research objectives. Specifically, the original datasets, which consisted of 42 instances for Study 1 (130 families and 221 genera) and 60 instances for Study 2 (189 families and 316 genera), were filtered to exclude measurements related to feed and pre-feeding conditions. Following filtering, the datasets were reduced to 24 samples for Study 1 and 48 samples for Study 2. While this reduction improved the consistency and reliability of the input data by ensuring that genotype and/or diet were the sole influencing factors during feature selection, it nevertheless constrained the effective sample size available for modeling. Because smaller datasets can lead to models being overly tailored to a specific dataset, we mitigated potential effects on statistical power and model generalizability by consistently leveraging LOOCV and systematically comparing ML-derived results with conventional microbiome analyses reported in our prior studies, thereby confirming the consistency of the retained datasets with established findings. Additional preprocessing procedures also included handling missing values, removing null columns, and resolving data inconsistencies to ensure data integrity and reliability. These steps formed the foundational components of the analytical framework employed to pursue the research objectives outlined below.

Table 1. Overview of the categorical combinations considered in Study 1 and Study 2. (WT, wild-type genotype; HG, high-growth genotype; CTRL, control diet; F, future diet with FM/FO replacement; ORG, experimental diet supplemented with organic acids; PHYTO, experimental diet supplemented with phytogetic extracts; PROB, experimental diet supplemented with probiotics; CTRL-WT/CTRL-HG, control diet fed to wild-type/high-growth genotype; F-WT/F-HG, future diet fed to wild-type/high-growth genotype; ORG-WT/ORG-HG, organic-acid-supplemented diet fed to wild-type/high-growth genotype; PHYTO-WT/PHYTO-HG, phytogetic-extract-supplemented diet fed to wild-type/high-growth genotype; PROB-WT/PROB-HG, probiotic-supplemented diet fed to wild-type/high-growth genotype; G1, genotype class group for Study 1; G2, genotype class group for Study 2; D1, diet class group for Study 1; D2, diet class group for Study 2; DG1, diet \times genotype class group for Study 1; DG2, diet \times genotype class group for Study 2).

| | Factor | Class | Group |
|---------|------------------------|--|-------|
| Study 1 | Genotype | WT, HG | G1 |
| | Diet | CTRL, F | D1 |
| | Diet \times Genotype | CTRL-WT, CTRL-HG, F-WT, F-HG | DG1 |
| Study 2 | Genotype | WT, HG | G2 |
| | Diet | CTRL, ORG, PHYTO, PROB | D2 |
| | Diet \times Genotype | CTRL-WT, CTRL-HG, ORG-WT, ORG-HG, PHYTO-WT, PHYTO-HG, PROB-WT, PROB-HG | DG2 |

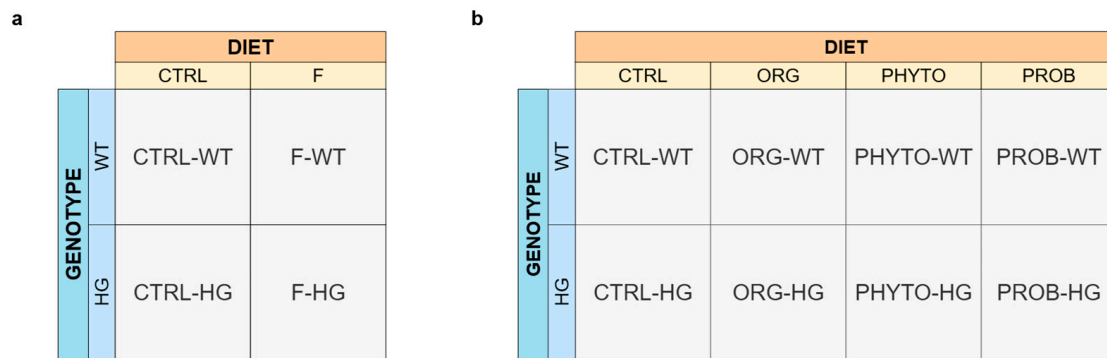


Figure 1. Schematic depiction of the categorical combinations considered in Study 1 (a) and Study 2 (b). (WT, wild-type genotype; HG, high-growth genotype; CTRL, control diet; F, future diet with FM/FO replacement; ORG, experimental diet supplemented with organic acids; PHYTO, experimental diet supplemented with phytogetic extracts; PROB, experimental diet supplemented with probiotics; CTRL-WT/CTRL-HG, control diet fed to wild-type/high-growth genotype; F-WT/F-HG, future diet fed to wild-type/high-growth genotype; ORG-WT/ORG-HG, organic-acid-supplemented diet fed to wild-type/high-growth genotype; PHYTO-WT/PHYTO-HG, phytogetic-extract-supplemented diet fed to wild-type/high-growth genotype; PROB-WT/PROB-HG, probiotic-supplemented diet fed to wild-type/high-growth genotype).

RO1. In an effort to reduce dataset complexity, feature selection was performed to determine the most significant microbial taxa, enhancing model generalizability and mitigating overfitting. In the case of RO2, we applied recursive feature elimination (RFE) using ensemble-based, margin-based, and generalized linear models as supervised estimators. To ensure a balanced selection between permissive and conservative algorithms, we retained microbial taxa with the greatest overlap across estimators, under the constraint that, given n as the number of features and m as the number of samples, $1 \leq n < m$. In the case of RO3, we identified statistically significant microbial taxa through point-biserial correlation analysis and analysis of variance (ANOVA), each considering class grouping—namely, diet \times genotype combinations and, where applicable, diet alone. In order to ensure robust selection, we imposed threshold criteria of $|r| > 0.5$ and $p < 0.05$ for point-biserial correlation analyses, and a significance threshold of $p < 0.05$ for ANOVA-based comparisons; in the latter case, taxa selection was conducted using a series of independent one-way ANOVAs. Although such feature selection approaches can reduce high dimensionality, they may lead to overfitting on small datasets and therefore select dataset-specific predictors: on the one hand, repeated model fitting on limited samples can inflate variance in feature-importance estimates for RFE, while, on the other hand, small sample sizes can destabilize correlation estimates and p -values in point-biserial correlation and ANOVA. Despite this overfitting potential, RFE, ANOVA, and point-biserial correlation offer methodological advantages over alternative selection approaches such as mutual information and Least Absolute Shrinkage and Selection Operator (LASSO). RFE provides a model-specific, iterative framework that systematically removes less informative predictors, returning a ranked subset of variables optimized for predictive performance and reducing the risk of model overfitting. By contrast, point-biserial correlation is particularly well suited for binary classification tasks, as it furnishes a straightforward and interpretable measure of linear association between continuous predictors and binary outcomes, with minimal computational overhead. ANOVA complements these methods by enabling the statistical evaluation of mean differences across categorical groups and providing interpretable inferential outputs. Taken together, these techniques highlight the importance of statistical robustness and model-specific optimization, offering advantages in settings where balanced datasets and explicit

feature relevance are prioritized, as opposed to the coefficient penalization of LASSO or the less interpretable nonlinear dependencies captured by mutual information. In both research objectives, nevertheless, any non-microbial contaminants or ambiguous taxa were manually removed.

RO2–RO3. Both research objectives were addressed with a three-stage pipeline tailored to meet task-specific needs. In the first stage, models were evaluated on full datasets with default implementations to establish baseline performance. In the second stage, both datasets were reduced to include only the microbial taxa identified during feature selection, and models were evaluated using the same default settings to assess potential overfitting. Given the limited size of our datasets, we used cross-validation (CV) in both stages instead of traditional train–test splitting. This approach allowed to reduce variance in performance estimates arising from different split proportions and enhance generalizability. In the third stage, models were evaluated on the reduced datasets using nested cross-validation (NCV) to ensure unbiased performance evaluation while, at the same time, optimizing hyperparameters. In particular, the NCV function was designed to return both the average model performance (measured as accuracy in classification tasks and as root mean squared error in regression tasks) and the optimal parameter configuration. The latter was applied to evaluate models on the reduced datasets, producing out-of-fold (OOF) predictions able to approximate performance on unseen data. While both research objectives share a common analytical framework, their respective implementation peculiarities were addressed accordingly (Figure 2). In the case of RO2, SHAP and rank analyses were performed to visualize the individual contributions of the most relevant taxa and to identify the most influential features across different class groupings. In the case of RO3, regression analysis required extending both datasets to include the respective quantities of diet ingredients reported in Torrecillas et al. [41] (Supplementary Table S1) and Rimoldi et al. [42] (Supplementary Table S2).

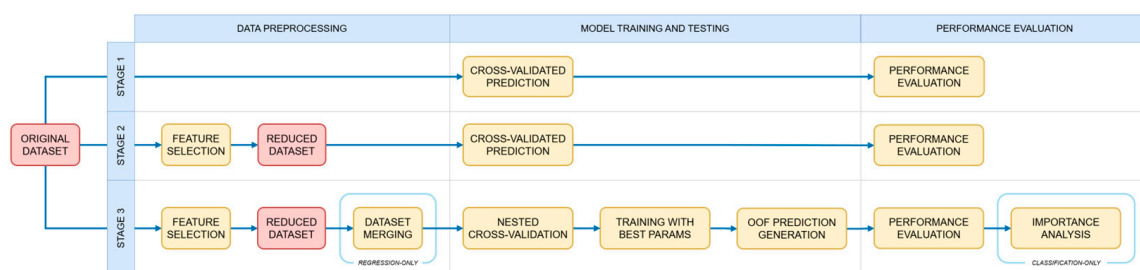


Figure 2. Schematic representation illustrating the implementation workflow for the proposed pipeline. Rows correspond to the implementation stages, and columns correspond to the internal phases within each stage.

Since most algorithms are now capable of handling both classification and regression, we leveraged a diverse set of algorithmic families, each with distinct underlying mechanisms: tree-based (decision trees, DTs), ensemble-based (random forests, RFs; extra trees, ETs; gradient boosting, GB; extreme gradient boosting, XGB; categorical boosting, CB), margin-based (support vector machines, SVMs), probability-based (Naïve Bayes, NB), distance-based (k -nearest neighbors, k -NNs), neural network-based (multilayer perceptrons, MLPs), and generalized linear models (GLMs). In the case of RO2, as a classification task, we selected the following algorithms: DT classifier (DTC), RF classifier (RFC), ET classifier (ETC), GB classifier (GBC), XGB classifier (XGBC), CB classifier (CBC), multinomial NB (MNB), SVM classifier (SVC), k -NN classifier (KNC), MLP classifier (MLPC), and logistic regression (LREG). In the case of RO3, as a regression task, we selected the following algorithms: DT regressor (DTR), RF regressor (RFR), ET regressor (ETR), GB regressor

(GBR), XGB regressor (XGBR), SVM regressor (SVR), k -NN regressor (KNN), and MLP regressor (MLPR). In both cases, SVM-based models were evaluated using various kernel functions (linear, LK; polynomial, PK; radial basis function, RK; sigmoid, SK) during the first two pipeline stages. However, in the final stage, LK was preferred due to its superior generalizability, mitigating the overfitting risks associated with PK and RK as well as the unstable behavior observed in SK. Ultimately, to guarantee proper learning and unbiased performance, datasets were standardized when using scale-sensitive algorithms.

3. Results

To comprehensively analyze intestinal microbial composition under various categorical groupings across both studies, we carried out task-specific feature selection and used a structured three-stage pipeline across biologically informative taxonomic levels (family, genus) on microbial profiles derived from the gut microbiota of European sea bass in two previous studies. For conciseness purposes, we report the findings from the final pipeline stage, where selected models underwent complete optimization. The detailed assessment of previous phases, including baseline performance and overfitting probing, and feature importance evaluation analysis, including SHAP analysis and rank analysis, are provided separately as Supplementary Materials (Study 1: Supplementary Files S1, S2 and S5 for the family level; Supplementary Files S3, S4 and S6 for the genus level. Study 2: Supplementary Files S7, S8 and S11 for the family level; Supplementary Files S9, S10 and S12 for the genus level). To better contextualize ML-derived results, we summarize the main findings from traditional microbiome analyses, as extensively detailed in Torrecillas et al. [41] and Rimoldi et al. [42], to provide a necessary reference for interpreting the outcomes of ML-based approaches.

3.1. Study 1: Evaluating Alternative Formulations from Poultry-Based Sources

3.1.1. Prior Results on Intestinal Microbial Communities

As reported in Torrecillas et al. [41], when considering the most representative taxa, the microbial community profile of the autochthonous intestinal microbiota consisted of 20 families and 20 genera. When comparing wild-type and high-growth sea bass groups, the authors delineated a core microbiota comprising the following genera: *Acinetobacter*, *Brevundimonas*, *Clostridium sensu stricto 1*, *Corynebacterium*, *Cutibacterium*, *Enhydrobacter*, *Escherichia-Shigella*, *Kocuria*, *Lactobacillus*, *Micrococcus*, *Paracoccus*, *Pseudomonas*, *Staphylococcus*, *Stenotrophomonas*, *Streptococcus*, *Vibrio*, as well as unidentified members of *Clostridiaceae* and *Vibrionaceae*. Differential abundance analysis identified genotype effect across families (*Caulobacteraceae*, *Flavobacteriaceae*, *Lactobacillaceae*, *Micrococcaceae*, *Moraxellaceae*, *Neisseriaceae*, *Propionibacteriaceae*, *Rhodobacteraceae*, *Stenotrophomonas*, *Staphylococcaceae*, *Streptococcaceae*, *Xanthomonadaceae*, *Weeksellaceae*, and unknown *Peptostreptococcales-Tissierellales*) and genera (*Corynebacterium*, *Cutibacterium*, *Enhydrobacter*, *Lactobacillus*, *Micrococcus*, *Paracoccus*, *Psychrobacter*, *Staphylococcus*, and *Streptococcus*). In contrast, an interaction effect was identified exclusively for *Clostridium sensu stricto 1*, while no statistically significant diet influence was detected.

3.1.2. Classification Modeling for Dietary Regime Inference from Microbial Abundance

The integration of RFE-based feature selection with nested cross-validation allowed to observe the influence of different class groupings across both taxonomic levels (Tables 2 and 3). It is worth noticing that classification performance improves when moving from combined to individual groupings, with genotype exhibiting the highest values and diet demonstrating intermediate results.

Table 2. Performance of optimized classification models on the autochthonous intestinal microbiota of European sea bass at the family level for Study 1. Models are marked in bold based on the highest out-of-fold accuracy as the most suitable indicator of performance generalization; in the case of equal values, average accuracy was applied as secondary criterion to distinguish the most stable performer. (AVG NCV, average model performance during nested cross-validation; OOF ACC, accuracy of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MCC, Matthews correlation coefficient of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTC, decision tree classifier; RFC, random forest classifier; ETC, extra tree classifier; GBC, gradient boosting classifier; XGBC, extreme gradient boosting classifier; CBC, categorical boosting classifier; MNB, multinomial Naïve Bayes; SVC, support vector machine classifier; KNC, *k*-nearest neighbor classifier; LREG, logistic regression; MLPC, multilayer perceptron classifier).

| | | DTC | RFC | ETC | GBC | XGBC | CBC | MNB | SVC | KNC | LREG | MLPC |
|-----------|---------|------|-------|-------------|------|-------|------|-------------|-------|-------------|-------|------|
| Group DG1 | AVG NCV | 0.33 | 0.21 | 0.38 | 0.42 | 0.25 | 0.38 | 0.25 | 0.08 | 0.38 | 0.08 | 0.29 |
| | OOF ACC | 0.33 | 0.21 | 0.33 | 0.42 | 0.25 | 0.42 | 0.25 | 0.12 | 0.46 | 0.08 | 0.29 |
| | OOF MCC | 0.11 | −0.06 | 0.11 | 0.22 | 0.00 | 0.23 | 0.00 | −0.18 | 0.33 | −0.23 | 0.12 |
| Group D1 | AVG NCV | 0.50 | 0.46 | 0.58 | 0.50 | 0.42 | 0.54 | 0.62 | 0.12 | 0.33 | 0.25 | 0.12 |
| | OOF ACC | 0.54 | 0.46 | 0.50 | 0.54 | 0.42 | 0.58 | 0.62 | 0.25 | 0.58 | 0.38 | 0.50 |
| | OOF MCC | 0.08 | −0.08 | 0.00 | 0.08 | −0.17 | 0.17 | 0.25 | −0.51 | 0.17 | −0.25 | 0.00 |
| Group G1 | AVG NCV | 0.83 | 0.83 | 0.88 | 0.83 | 0.79 | 0.75 | 0.25 | 0.79 | 0.71 | 0.79 | 0.79 |
| | OOF ACC | 0.83 | 0.83 | 0.88 | 0.83 | 0.88 | 0.75 | 0.29 | 0.79 | 0.79 | 0.79 | 0.79 |
| | OOF MCC | 0.67 | 0.67 | 0.75 | 0.67 | 0.75 | 0.51 | −0.43 | 0.64 | 0.59 | 0.64 | 0.64 |

Table 3. Performance of optimized classification models on the autochthonous intestinal microbiota of European sea bass at the genus level for Study 1. Models are marked in bold based on the highest out-of-fold accuracy as the most suitable indicator of performance generalization; in the case of equal values, average accuracy was applied as secondary criterion to distinguish the most stable performer. (AVG NCV, average model performance during nested cross-validation; OOF ACC, accuracy of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MCC, Matthews correlation coefficient of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTC, decision tree classifier; RFC, random forest classifier; ETC, extra tree classifier; GBC, gradient boosting classifier; XGBC, extreme gradient boosting classifier; CBC, categorical boosting classifier; MNB, multinomial Naïve Bayes; SVC, support vector machine classifier; KNC, *k*-nearest neighbor classifier; LREG, logistic regression; MLPC, multilayer perceptron classifier).

| | | DTC | RFC | ETC | GBC | XGBC | CBC | MNB | SVC | KNC | LREG | MLPC |
|-----------|---------|-------------|-------------|-------------|-------------|-------------|-------------|------|------|-------------|------|-------|
| Group DG1 | AVG NCV | 0.46 | 0.38 | 0.21 | 0.33 | 0.38 | 0.50 | 0.33 | 0.21 | 0.29 | 0.38 | 0.46 |
| | OOF ACC | 0.46 | 0.38 | 0.29 | 0.33 | 0.38 | 0.50 | 0.33 | 0.25 | 0.42 | 0.46 | 0.46 |
| | OOF MCC | 0.28 | 0.17 | 0.06 | 0.11 | 0.17 | 0.34 | 0.11 | 0.00 | 0.22 | 0.29 | 0.36 |
| Group D1 | AVG NCV | 0.58 | 0.58 | 0.62 | 0.58 | 0.67 | 0.58 | 0.50 | 0.58 | 0.58 | 0.58 | 0.62 |
| | OOF ACC | 0.58 | 0.58 | 0.62 | 0.58 | 0.67 | 0.58 | 0.50 | 0.62 | 0.62 | 0.67 | 0.62 |
| | OOF MCC | 0.22 | 0.22 | 0.31 | 0.22 | 0.38 | 0.22 | 0.00 | 0.38 | 0.38 | 0.45 | 0.38 |
| Group G1 | AVG NCV | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.50 | 0.67 | 0.92 | 0.67 | 0.92 |
| | OOF ACC | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.50 | 0.67 | 0.92 | 0.71 | 0.42 |
| | OOF MCC | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 | 0.00 | 0.45 | 0.85 | 0.51 | −0.30 |

Family level (Table 2). For the DG1 group, GBC achieved the highest average performance, consistent with OOF predictions, where CBC and KNC attained similar results. In the D1 group, MNB showed the best average performance, also reflected in OOF predictions, where CBC and KNC emerged as suitable alternatives. For the G1 group, ETC

performed best in terms of average performance and OOF predictions, while XGBC and, to a lesser extent, DTC, RFC, and GBC produced comparable outcomes in the latter.

Genus level (Table 3). For the DG1 group, CBC reached the highest average and OOF performance, with DTC and MLPC yielding similar results. In the D1 group, XGBC performed best, while ETC and MLPC achieved comparable outcomes. For the G1 group, all models showed extremely high performance except MNB, SVC, and LREG.

At both levels, the RFE function enhanced model performance across different categorical groupings (Supplementary Files S1 and S3) by condensing the most influential features into minimal subsets (Table 4), enabling more detailed insights through SHAP and rank analysis (Supplementary Files S2 and S4). At the family level, the following patterns emerged: for the DG1 group, *Neisseriaceae* had the strongest influence for CTRL-WT, CTRL-HG, F-HG, and, to a lesser extent, F-WT; for the D1 group, *Moraxellaceae*, *Mycoplasmataceae*, and *Pseudoalteromonadaceae* showed the greatest impact; for the G1 group, *Neisseriaceae* and *Streptococcaceae* exerted near-equivalent influence. At the genus level, the following trends were observed: in the DG1 group, *Enhydrobacter* and *Clostridium sensu stricto 1* contributed similarly for CTRL-WT and F-WT, while *Clostridium sensu stricto 1* and *Enhydrobacter* dominated CTRL-HG and F-HG, respectively; in the D1 and G1 groups, *Pseudoalteromonas* and *Flavobacterium* were the sole features identified by RFE.

Table 4. Microbial taxa selected through RFE across taxonomic levels in the autochthonous intestinal microbiota of European sea bass for Study 1.

| Taxonomic Level | Class Grouping | Minimal Feature Subset Size | Most Influential Taxa |
|-----------------|----------------|-----------------------------|--|
| Family | Group DG1 | 6 | <i>Neisseriaceae</i> , <i>Peptostreptococcaceae</i> , <i>Rhodobacteraceae</i> , <i>Staphylococcaceae</i> , <i>Streptococcaceae</i> , <i>Xanthomonadaceae</i> , <i>Corynebacteriaceae</i> , <i>Enterobacteriaceae</i> , <i>Exiguobacteraceae</i> , <i>Micrococcaceae</i> , <i>Moraxellaceae</i> , <i>Mycoplasmataceae</i> , |
| | Group D1 | 13 | <i>Neisseriaceae</i> , <i>Propionibacteriaceae</i> , <i>Pseudoalteromonadaceae</i> , <i>Sphingomonadaceae</i> , <i>Staphylococcaceae</i> , <i>Weeksellaceae</i> , <i>Xanthomonadaceae</i> |
| | Group G1 | 2 | <i>Neisseriaceae</i> , <i>Streptococcaceae</i> |
| Genus | Group DG1 | 2 | <i>Clostridium sensu stricto 1</i> , <i>Enhydrobacter</i> |
| | Group D1 | 1 | <i>Pseudoalteromonas</i> |
| | Group G1 | 1 | <i>Flavobacterium</i> |

3.1.3. Regression Analysis for Predicting Microbial Shifts from Feed Composition Changes

The incorporation of statistical techniques inside nested cross-validation produced comparable results across both taxonomic levels (Tables 5 and 6). However, regression performance metrics were available only for the DG1 group since the feature selection procedure detected no correlation for the D1 group (Table 7). However, the pipeline results were unsatisfactory, as indicated by both goodness-of-fit and error metrics, thereby warranting additional investigations into the relationship between diet ingredient variation and microbial abundance shift.

Table 5. Performance of optimized regression models on the autochthonous intestinal microbiota of European sea bass at the family level for Study 1. (AVG NCV, average model performance during nested cross-validation; OOF R², coefficient of determination of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MAE, mean absolute error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF RMSE, root mean squared error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTR, decision tree regressor; RFR, random forest regressor; ETR, extra tree regressor; GBR, gradient boosting regressor; XGBR, extreme gradient boosting regressor; SVR, support vector machine regressor; KNR, *k*-nearest neighbor regressor; MLPR, multilayer perceptron regressor).

| | | DTR | RFR | ETR | GBR | XGBR | SVR | KNR | MLPR |
|-----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Group DG1 | AVG NCV | 1.07×10^6 | 1.04×10^6 | 1.07×10^6 | 1.07×10^6 | 1.07×10^6 | 6.83×10^5 | 1.02×10^6 | 6.83×10^5 |
| | OOF R ² | -0.05 | -0.04 | -0.05 | -0.05 | -0.05 | -0.14 | -0.05 | -0.14 |
| | OOF MAE | 1.07×10^6 | 1.03×10^6 | 1.07×10^6 | 1.07×10^6 | 1.07×10^6 | 6.83×10^5 | 1.07×10^6 | 6.83×10^5 |
| | OOF RMSE | 1.99×10^6 | 1.97×10^6 | 1.99×10^6 | 1.99×10^6 | 1.99×10^6 | 2.05×10^6 | 1.99×10^6 | 2.05×10^6 |

Table 6. Performance of optimized regression models on the autochthonous intestinal microbiota of European sea bass at the genus level for Study 1. (AVG NCV, average model performance during nested cross-validation; OOF R², coefficient of determination of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MAE, mean absolute error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF RMSE, root mean squared error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTR, decision tree regressor; RFR, random forest regressor; ETR, extra tree regressor; GBR, gradient boosting regressor; XGBR, extreme gradient boosting regressor; SVR, support vector machine regressor; KNR, *k*-nearest neighbor regressor; MLPR, multilayer perceptron regressor).

| | | DTR | RFR | ETR | GBR | XGBR | SVR | KNR | MLPR |
|-----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Group DG1 | AVG NCV | 1.51×10^8 | 1.51×10^8 | 1.51×10^8 | 1.51×10^8 | 1.51×10^8 | 1.17×10^8 | 1.42×10^8 | 1.45×10^8 |
| | OOF R ² | -0.08 | -0.08 | -0.08 | -0.08 | -0.08 | -0.16 | -0.06 | -0.16 |
| | OOF MAE | 1.51×10^8 | 1.51×10^8 | 1.51×10^8 | 1.51×10^8 | 1.51×10^8 | 1.17×10^8 | 1.40×10^8 | 1.28×10^8 |
| | OOF RMSE | 1.82×10^8 | 1.83×10^8 | 1.82×10^8 | 1.82×10^8 | 1.82×10^8 | 1.91×10^8 | 1.79×10^8 | 1.77×10^8 |

Table 7. Microbial taxa selected via statistical techniques across taxonomic levels in the autochthonous intestinal microbiota of European sea bass for Study 1 (*p*, *p*-value; *max*, highest microbial count within the dataset).

| Taxonomic Level | Class Grouping | Minimal Feature Subset Size | Most Influential Taxa |
|-----------------|----------------|-----------------------------|---|
| Family | Group DG1 | 3 | <i>Mycoplasmataceae</i> ($p = 0.037$, $max = 9.50 \cdot 10^6$), <i>Planococcaceae</i> ($p = 0.030$, $max = 3.50 \cdot 10^6$), <i>Pseudoalteromonadaceae</i> ($p = 0.048$, $max = 1.05 \cdot 10^7$) |
| Genus | Group DG1 | 6 | <i>Clostridium sensu stricto 1</i> ($p = 0.023$, $max = 3.02 \cdot 10^9$), <i>Curvibacter</i> ($p = 0.046$, $max = 1.10 \cdot 10^6$), <i>Mycoplasma</i> ($p = 0.037$, $max = 9.50 \cdot 10^6$), <i>Pseudoalteromonas</i> ($p = 0.048$, $max = 1.05 \cdot 10^7$), <i>Ulvibacter</i> ($p = 0.030$, $max = 4.30 \cdot 10^6$), unknown <i>Planococcaceae</i> ($p = 0.030$, $max = 3.50 \cdot 10^6$) |

3.2. Study 2: Investigating Dietary Alternatives Through Fortified Feeds

3.2.1. Previous Findings on Intestinal Microbial Communities

As detailed in Rimoldi et al. [42], when focusing on the most representative taxa, the autochthonous intestinal microbiota comprised 28 families and 25 genera, with substantial taxonomic overlap across genotypes, regardless of dietary plan, while, when considering genotype alone, only a limited number of taxa were exclusive to either group. Although alpha diversity analysis could not detect statistically significant differences in species richness

or overall diversity between experimental groups, diet and genotype effects (individual and combined) were present for selected taxa, with diet emerging as the prominent driver. Diet-related effects were observed at family (*Moraxellaceae*, *Pseudomonadaceae*, *Sphingomonadaceae*, *Streptococcaceae*, and *Weeksellaceae*) as well as genus (*Acinetobacter*, *Novosphingobium*, *Pseudomonas*, *Sphingobium*, and *Streptococcus*) levels, while genotype-related effects were confined to the genus *Photobacterium*; instead, interaction effects were identified at family (*Weeksellaceae*) and genus (*Acinetobacter* and *Enterovibrio*) levels.

3.2.2. Microbiota-Based Classification Models for Dietary Regime Discrimination

The processing methodology used to analyze microbial data from Study 1 facilitated the investigation of class grouping effects across taxonomic levels (Tables 8 and 9). Within this context, while the increase in classification performance was less pronounced than in Study 1, noticeable improvements appeared when shifting from combined to individual groupings; among these, genotype achieved overall superior performance, while diet remained intermediate.

Family level (Table 8). In the DG2 group, DTC achieved higher average performance, confirmed by OOF predictions, where MLPC produced similar results. For the D2 group, GBC delivered the best overall performance. In the G2 group, MNB showed the highest average performance, with DTC and SVC as feasible competitors; OOF predictions further indicated CBC as a viable alternative.

Genus level (Table 9). For the DG2 group, XGBC achieved the highest average performance, followed by MNB; OOF predictions confirmed these outcomes, with DTC performing comparably to XGBC. In the D2 group, GBC showed the highest accuracy across both average and OOF predictions. For the G2 group, MNB attained the best overall performance, with DTC and SVC producing similar outcomes; these findings were supported by OOF predictions, which confirmed the superior performance of MNB and, to a lesser extent, DTC, while also highlighting CBC as an interesting alternative.

Table 8. Performance of optimized classification models on the autochthonous intestinal microbiota of European sea bass at the family level for Study 2. Models are marked in bold based on the highest out-of-fold accuracy as the most suitable indicator of performance generalization; in the case of equal values, average accuracy was applied as secondary criterion to distinguish the most stable performer. (AVG NCV, average model performance during nested cross-validation; OOF ACC, accuracy of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MCC, Matthews correlation coefficient of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTC, decision tree classifier; RFC, random forest classifier; ETC, extra tree classifier; GBC, gradient boosting classifier; XGBC, extreme gradient boosting classifier; CBC, categorical boosting classifier; MNB, multinomial Naïve Bayes; SVC, support vector machine classifier; KNC, *k*-nearest neighbor classifier; LREG, logistic regression; MLPC, multilayer perceptron classifier).

| | | DTC | RFC | ETC | GBC | XGBC | CBC | MNB | SVC | KNC | LREG | MLPC |
|-----------|---------|-------------|------|------|-------------|-------|-------------|------|------|------|------|-------------|
| Group DG2 | AVG NCV | 0.29 | 0.23 | 0.17 | 0.21 | 0.15 | 0.23 | 0.15 | 0.15 | 0.17 | 0.15 | 0.23 |
| | OOF ACC | 0.29 | 0.25 | 0.23 | 0.25 | 0.17 | 0.19 | 0.15 | 0.15 | 0.25 | 0.17 | 0.29 |
| | OOF MCC | 0.19 | 0.14 | 0.12 | 0.14 | 0.05 | 0.07 | 0.02 | 0.03 | 0.14 | 0.05 | 0.19 |
| Group D2 | AVG NCV | 0.56 | 0.44 | 0.46 | 0.62 | 0.56 | 0.50 | 0.52 | 0.44 | 0.33 | 0.38 | 0.33 |
| | OOF ACC | 0.60 | 0.50 | 0.48 | 0.65 | 0.54 | 0.52 | 0.52 | 0.46 | 0.44 | 0.42 | 0.42 |
| | OOF MCC | 0.47 | 0.33 | 0.32 | 0.53 | 0.39 | 0.36 | 0.36 | 0.30 | 0.26 | 0.23 | 0.22 |
| Group G2 | AVG NCV | 0.56 | 0.52 | 0.54 | 0.06 | 0.48 | 0.56 | 0.54 | 0.54 | 0.50 | 0.54 | 0.56 |
| | OOF ACC | 0.56 | 0.52 | 0.54 | 0.06 | 0.48 | 0.56 | 0.54 | 0.54 | 0.52 | 0.56 | 0.56 |
| | OOF MCC | 0.20 | 0.09 | 0.13 | −0.88 | −0.15 | 0.20 | 0.21 | 0.21 | 0.15 | 0.20 | 0.20 |

Table 9. Performance of optimized classification models on the autochthonous intestinal microbiota of European sea bass at the genus level for Study 2. Models are marked in bold based on the highest out-of-fold accuracy as the most suitable indicator of performance generalization; in the case of equal values, average accuracy was applied as secondary criterion to distinguish the most stable performer. (AVG NCV, average model performance during nested cross-validation; OOF ACC, accuracy of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MCC, Matthews correlation coefficient of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTC, decision tree classifier; RFC, random forest classifier; ETC, extra tree classifier; GBC, gradient boosting classifier; XGBC, extreme gradient boosting classifier; CBC, categorical boosting classifier; MNB, multinomial Naïve Bayes; SVC, support vector machine classifier; KNC, *k*-nearest neighbor classifier; LREG, logistic regression; MLPC, multilayer perceptron classifier).

| | | DTC | RFC | ETC | GBC | XGBC | CBC | MNB | SVC | KNC | LREG | MLPC |
|-----------|---------|------|------|-------|-------------|-------------|------|-------------|------|-------|------|------|
| Group DG2 | AVG NCV | 0.21 | 0.25 | 0.15 | 0.15 | 0.29 | 0.21 | 0.27 | 0.12 | 0.23 | 0.19 | 0.17 |
| | OOF ACC | 0.29 | 0.23 | 0.19 | 0.17 | 0.29 | 0.25 | 0.27 | 0.15 | 0.21 | 0.21 | 0.21 |
| | OOF MCC | 0.19 | 0.12 | 0.07 | 0.05 | 0.19 | 0.14 | 0.17 | 0.02 | 0.10 | 0.10 | 0.10 |
| Group D2 | AVG NCV | 0.40 | 0.40 | 0.27 | 0.50 | 0.40 | 0.33 | 0.35 | 0.38 | 0.40 | 0.25 | 0.31 |
| | OOF ACC | 0.44 | 0.38 | 0.33 | 0.52 | 0.42 | 0.40 | 0.35 | 0.38 | 0.40 | 0.29 | 0.40 |
| | OOF MCC | 0.26 | 0.17 | 0.11 | 0.36 | 0.22 | 0.19 | 0.14 | 0.18 | 0.20 | 0.06 | 0.20 |
| Group G2 | AVG NCV | 0.58 | 0.50 | 0.42 | 0.46 | 0.50 | 0.56 | 0.62 | 0.58 | 0.27 | 0.54 | 0.40 |
| | OOF ACC | 0.60 | 0.56 | 0.48 | 0.48 | 0.54 | 0.60 | 0.62 | 0.56 | 0.42 | 0.56 | 0.56 |
| | OOF MCC | 0.21 | 0.13 | −0.04 | −0.04 | 0.08 | 0.21 | 0.25 | 0.14 | −0.17 | 0.13 | 0.13 |

Consistent with the findings reported for Study 1, feature selection played a crucial role in enhancing model performance across categorical groupings (Supplementary Files S7 and S9). Indeed, by distilling microbial features into minimal informative subsets (Table 10), this approach facilitated deeper interpretability through SHAP and rank analysis (Supplementary Files S8 and S10). At the family level, the following patterns were observed: (i) in the DG2 group, *Acholeplasmataceae* emerged as top-performing feature for CTRL-WT, CTRL-HG, ORG-WT, PRO-WT, and PRO-HG, with contributions coming from *Weeksellaceae* for CTRL-HG, ORG-WT, ORG-HG, PHYTO-WT, PRO-WT, PRO-HG, and *Caulobacteraceae* for PHYTO-HG; (ii) in the D2 group, *Weeksellaceae* was the feature with the highest influence across ORG, PHYTO, and PRO, while *Streptococcaceae* and *Pseudomonadaceae* were notably associated with CTRL; (iii) in the G2 group, *Intrasporangiaceae* was identified as the sole determinant feature. At the genus level, distinct trends were identified: (i) for the DG2 group, *Streptococcus*, *Cutibacterium*, and *Salinisphaera* for CTRL-WT; *Acinetobacter*, *Cutibacterium*, and *Streptococcus* for CTRL-HG; *Enterovibrio*, *Streptococcus*, *Marinobacter*, and *Acinetobacter* for ORG-WT; *Cutibacterium*, *Acinetobacter*, and *Streptococcus* for ORG-HG; *Lactobacillus* and *Streptococcus* for PHYTO-WT; *Enterovibrio*, *Streptococcus*, and *Brevundimonas* for PHYTO-HG; *Cutibacterium*, *Acinetobacter*, and *Marinobacter* for PRO-WT; *Acinetobacter* and *Lactobacillus* for PRO-HG; (ii) for the D2 group, *Streptococcus* was the primary feature for CTRL and ORG, while *Lactobacillus* dominated PHYTO and PRO, as well as playing an important role in ORG; (iii) for the G2 group, *Photobacterium* was identified as the principal discriminative feature.

Table 10. Microbial taxa selected through RFE across taxonomic levels in the autochthonous intestinal microbiota of European sea bass for Study 2.

| Taxonomic Level | Class Grouping | Minimal Feature Subset Size | Most Influential Taxa |
|-----------------|----------------|-----------------------------|---|
| Family | Group DG2 | 4 | <i>Acholeplasmataceae, Alcanivoracaceae, Caulobacteraceae, Weeksellaceae</i> |
| | Group D2 | 3 | <i>Pseudomonadaceae, Streptococcaceae, Weeksellaceae</i> |
| | Group G2 | 2 | <i>Intrasporangiaceae, Pseudohongiellaceae</i> |
| Genus | Group DG2 | 12 | <i>Acinetobacter, Brevundimonas, Cutibacterium, Enterovibrio, Escherichia-Shigella, Idiomarina, Lactobacillus, Marinobacter, Peptoniphilus, Salegentibacter, Salinisphaera, Streptococcus</i> |
| | Group D2 | 3 | <i>Acholeplasma, Lactobacillus, Streptococcus</i> |
| | Group G2 | 3 | <i>Micrococcus, Photobacterium, unknown Flavobacteriaceae</i> |

3.2.3. Modeling Microbial Responses to Dietary Modulation

As with the analyses conducted on data derived from Study 1, the integration of statistical techniques with nested cross-validation yielded similar results across multiple taxonomic levels (Tables 11 and 12) for both categorical groupings (Table 13). However, regression analyses, consistent with those performed for Study 1, produced overall unsatisfactory results, characterized by low goodness-of-fit metrics and elevated error values. These findings further underscore the need for an in-depth investigation of the dynamics linking dietary ingredient variations to microbial abundance shifts.

Table 11. Performance of optimized regression models on the autochthonous intestinal microbiota of European sea bass at the family level for Study 2. (AVG NCV, average model performance during nested cross-validation; OOF R², coefficient of determination of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MAE, mean absolute error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF RMSE, root mean squared error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTR, decision tree regressor; RFR, random forest regressor; ETR, extra tree regressor; GBR, gradient boosting regressor; XGBR, extreme gradient boosting regressor; SVR, support vector machine regressor; KNR, *k*-nearest neighbor regressor; MLPR, multilayer perceptron regressor).

| | | DTR | RFR | ETR | GBR | XGBR | SVR | KNR | MLPR |
|-----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Group DG2 | AVG NCV | 1.92×10^6 | 1.88×10^6 | 1.92×10^6 | 1.92×10^6 | 1.92×10^6 | 1.62×10^6 | 1.93×10^6 | 1.72×10^6 |
| | OOF R ² | −0.01 | 0.00 | −0.01 | −0.01 | −0.01 | −0.13 | 0.00 | −0.21 |
| | OOF MAE | 1.92×10^6 | 1.88×10^6 | 1.92×10^6 | 1.92×10^6 | 1.92×10^6 | 1.62×10^6 | 1.91×10^6 | 1.72×10^6 |
| | OOF RMSE | 2.98×10^6 | 2.95×10^6 | 2.98×10^6 | 2.98×10^6 | 2.98×10^6 | 3.27×10^6 | 2.96×10^6 | 3.50×10^6 |
| Group D2 | AVG NCV | 2.64×10^7 | 2.63×10^7 | 2.64×10^7 | 2.64×10^7 | 2.64×10^7 | 2.33×10^7 | 2.55×10^7 | 2.82×10^7 |
| | OOF R ² | 0.04 | 0.05 | 0.04 | 0.04 | 0.04 | −0.13 | −0.03 | 0.04 |
| | OOF MAE | 2.64×10^7 | 2.63×10^7 | 2.64×10^7 | 2.64×10^7 | 2.64×10^7 | 2.33×10^7 | 2.55×10^7 | 2.64×10^7 |
| | OOF RMSE | 4.28×10^7 | 4.27×10^7 | 4.28×10^7 | 4.28×10^7 | 4.28×10^7 | 4.69×10^7 | 4.30×10^7 | 4.28×10^7 |

Table 12. Performance of optimized regression models on the autochthonous intestinal microbiota of European sea bass at the genus level for Study 2. (AVG NCV, average model performance during nested cross-validation; OOF R², coefficient of determination of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF MAE, mean absolute error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; OOF RMSE, root mean squared error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions; DTR, decision tree regressor; RFR, random forest regressor; ETR, extra tree regressor; GBR, gradient boosting regressor; XGBR, extreme gradient boosting regressor; SVR, support vector machine regressor; KNR, *k*-nearest neighbor regressor; MLPR, multilayer perceptron regressor).

| | | DTR | RFR | ETR | GBR | XGBR | SVR | KNR | MLPR |
|-----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Group DG2 | AVG NCV | 1.87×10^6 | 1.83×10^6 | 1.87×10^6 | 1.87×10^6 | 1.87×10^6 | 1.58×10^6 | 1.91×10^6 | 1.67×10^6 |
| | OOF R ² | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.10 | 0.00 | -0.12 |
| | OOF MAE | 1.87×10^6 | 1.83×10^6 | 1.87×10^6 | 1.87×10^6 | 1.87×10^6 | 1.58×10^6 | 1.87×10^6 | 1.67×10^6 |
| | OOF RMSE | 3.23×10^6 | 3.20×10^6 | 3.23×10^6 | 3.23×10^6 | 3.23×10^6 | 3.57×10^6 | 3.23×10^6 | 3.73×10^6 |
| Group D2 | AVG NCV | 1.13×10^7 | 1.13×10^7 | 1.13×10^7 | 1.13×10^7 | 1.13×10^7 | 9.85×10^6 | 1.08×10^7 | 1.03×10^7 |
| | OOF R ² | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | -0.15 | -0.05 | -0.20 |
| | OOF MAE | 1.13×10^7 | 1.13×10^7 | 1.13×10^7 | 1.13×10^7 | 1.13×10^7 | 9.85×10^6 | 1.08×10^7 | 1.05×10^7 |
| | OOF RMSE | 1.96×10^7 | 1.96×10^7 | 1.96×10^7 | 1.96×10^7 | 1.96×10^7 | 2.16×10^7 | 1.93×10^7 | 2.25×10^7 |

Table 13. Microbial taxa selected via statistical techniques across taxonomic levels in the autochthonous intestinal microbiota of European sea bass for Study 2 (*p*, *p*-value; *max*, highest microbial count within the dataset).

| Taxonomic Level | Class Grouping | Minimal Feature Subset Size | Most Influential Taxa |
|-----------------|----------------|-----------------------------|--|
| Family | Group DG2 | 9 | <i>Dermabacteraceae</i> ($p = 0.012$, $max = 6.20 \cdot 10^6$), <i>Oleiphilaceae</i> ($p = 0.002$, $max = 1.40 \cdot 10^6$), <i>Pasteurellaceae</i> ($p = 0.029$, $max = 3.60 \cdot 10^6$), <i>Sphingomonadaceae</i> ($p = 0.011$, $max = 6.38 \cdot 10^7$), <i>Streptococcaceae</i> ($p = 0.024$, $max = 2.57 \cdot 10^7$), <i>Thalassospiraceae</i> ($p = 0.035$, $max = 1.70 \cdot 10^6$), <i>Veillonellaceae</i> ($p = 0.032$, $max = 6.00 \cdot 10^6$), <i>Weeksellaceae</i> ($p = 0.012$, $max = 1.67 \cdot 10^7$), unknown <i>Burkholderiales</i> ($p = 0.032$, $max = 1.50 \cdot 10^6$) |
| | Group D2 | 6 | <i>Dermabacteraceae</i> ($p = 0.048$, $max = 6.20 \cdot 10^6$), <i>Moraxellaceae</i> ($p = 0.014$, $max = 5.76 \cdot 10^8$), <i>Oleiphilaceae</i> ($p = 0.012$, $max = 1.40 \cdot 10^6$), <i>Pseudomonadaceae</i> ($p = 0.042$, $max = 5.60 \cdot 10^8$), <i>Saprospiraceae</i> ($p = 0.035$, $max = 1.61 \cdot 10^7$), <i>Sphingomonadaceae</i> ($p = 0.009$, $max = 6.38 \cdot 10^7$) |
| Genus | Group DG2 | 11 | <i>Brachybacterium</i> ($p = 0.012$, $max = 6.20 \cdot 10^6$), <i>Haemophilus</i> ($p = 0.022$, $max = 3.60 \cdot 10^6$), <i>Lactococcus</i> ($p = 0.010$, $max = 3.60 \cdot 10^6$), <i>Marixanthomonas</i> ($p = 0.046$, $max = 3.20 \cdot 10^6$), <i>Oleiphilus</i> ($p = 0.002$, $max = 1.40 \cdot 10^6$), <i>Persicirhabdus</i> ($p = 0.010$, $max = 1.80 \cdot 10^6$), <i>Porphyrobacter</i> ($p = 0.037$, $max = 1.20 \cdot 10^6$), <i>Sphingobium</i> ($p = 0.017$, $max = 5.07 \cdot 10^7$), <i>Thalassospira</i> ($p = 0.035$, $max = 1.70 \cdot 10^6$), <i>Veillonella</i> ($p = 0.017$, $max = 4.20 \cdot 10^6$), unknown <i>Enterobacteriaceae</i> ($p = 0.047$, $max = 1.20 \cdot 10^8$) |
| | Group D2 | 7 | <i>Brachybacterium</i> ($p = 0.048$, $max = 6.20 \cdot 10^6$), <i>Novosphingobium</i> ($p = 0.014$, $max = 2.32 \cdot 10^7$), <i>Oleiphilus</i> ($p = 0.012$, $max = 1.40 \cdot 10^6$), <i>Pseudomonas</i> ($p = 0.042$, $max = 5.60 \cdot 10^8$), <i>Sphingobium</i> ($p = 0.021$, $max = 5.07 \cdot 10^7$), <i>Veillonella</i> ($p = 0.032$, $max = 4.20 \cdot 10^6$), unknown <i>Enterobacteriaceae</i> ($p = 0.013$, $max = 1.20 \cdot 10^8$) |

4. Discussion

As aquaculture progressively shifts from reliance on oceanic resources toward more environmentally and economically sustainable practices, multiple feed ingredient alternatives have been evaluated as substitutes [61]. Among these, poultry-based ingredients have historically represented a common replacement for FM and FO [62]. Their viability as

alternative protein and lipid sources at different inclusion levels has been well established over time [63,64], particularly in crustaceans and marine fish [4]. In European sea bass, for instance, poultry-rendered ingredients have been combined with insect-derived ingredients within the framework of circular economy, offering nutritional potential without compromising growth performance, feed conversion efficacy, and overall health [12]. In parallel, feed additives have emerged as important components of modern aquafeeds. Originally used to compensate for deficiencies in essential nutrients, additives are now increasingly recognized as functional ingredients capable of modulating metabolism and physiology, enhancing growth and health while also contributing to environmental and economic sustainability [65]. Modern additives encompass different compounds, including probiotics, which have demonstrated beneficial effects on fish performance and health [66]. While ingredient replacements and functional additives are gaining prominence as strategies to reduce environmental footprint and improve economic sustainability, their efficacy remains strongly dependent on species-specific factors and rearing conditions. To date, most studies have focused on growth performance and feed utilization, whereas the impacts of replacement diets on host-associated microbiota remain comparatively underexplored.

In our companion studies, we examined the dietary effects of poultry by-product ingredients (Study 1) and additive-based fortification (Study 2) in feeds on the gut microbiota of European sea bass. Microbial community composition was assessed in both cases with metabarcoding-derived abundance data. While both studies supported the validity of the proposed alternative dietary strategies, the relative contributions of diet and host genotype differed between them. Specifically, genotype exerted the strongest influence in Study 1, whereas diet was the predominant factor in Study 2, although these effects were restricted to a limited number of microbial taxa. Building on these findings, we conducted an in-depth analysis of microbial abundance patterns across both studies using artificial intelligence, specifically machine learning, to better characterize the interrelationship between host-associated microbiota and the individual as well as synergistic effects of diet and genotype, in line with our research objectives.

Although routinely applied in human microbiome research and, more recently, livestock studies due to its capacity to process and interpret large-scale datasets, artificial intelligence has witnessed limited application in fish microbiome studies [43]. In the present work, artificial intelligence enabled to perform classification (RO2) and regression (RO3) tasks, as well as to identify minimal feature subsets associated with each dietary treatment (RO1), yielding informative results across both studies. In classification analyses, both studies demonstrated improved model accuracy when shifting from synergistic to individual factors for data categorization, with the combined grouping exhibiting the lowest predictive performance and genotype providing the highest discriminative ability. Indeed, the stronger effect of genotype, which was also reported in the reference studies, is supported by the larger number of specimens attributable to this grouping compared to the combined one. This ensured a sufficiently robust sample size for representative inference on data derived from metabarcoding analyses, despite the determination of a suitable sample size remaining contested. Indeed, recent combinatorial studies on European sea bass and gilthead seabream suggest that nine individuals per group may suffice to capture approximately 90% of bacterial species richness, thus offering representative insights into microbial shifts [67]. In this study, the sample size for the combined grouping did not permit confirmatory outcomes, in agreement with the conclusions of Panteli et al. [67] on microbiome investigations with comparable biological replicates [36,68,69]; nevertheless, comparable small sample sizes have also been used in pilot microbiome studies on livestock [70–72] and humans [73,74]. When considering the individual groupings, on the other hand, sample size proved sufficient to provide preliminary support for the

feasibility of our predictive pipeline. At the same time, the transition from multiclass to binary classification likely contributed to the improved performance, given the reduced algorithmic complexity of binary classification. By contrast, regression analyses in both studies produced unsatisfactory results, highlighting the need for further methodological refinement to enhance predictive power. Such improvements will be essential to enable the development of dietary interventions at the level of individual feed ingredients. While the regression outcomes underscore the necessity for optimization at the implementation stage, the classification results support the feasibility of our approach and thus provide a foundation for the development of ML pipelines capable of interrogating more complex datasets and uncovering more intricate host-microbiota-diet relationships.

Predictive performance differed markedly between regression and classification. Regression analyses yielded unsatisfactory results, likely due to the limited sample size and the high degree of similarity between diets in each study, differing only in a limited number of ingredients and in minor quantities. This lack of variability may have hindered the identification of patterns linking formulation differences to shifts in microbial abundance. Greater diversification in ingredient quantities could potentially enhance the ability to detect such relationships. In addition, future studies may benefit from the application of more advanced regression techniques like penalized regression models, such as LASSO (based on L1 regularization), Ridge (based on L2 regularization), and Elastic Net (combining L1 and L2 regularization) [75]. Despite these methodological possibilities, the regression models applied in this study did not achieve sufficient predictive power to support meaningful biological interpretation. By contrast, classification analyses produced promising results, thus warranting further investigation. Across both studies, tree-based (DTC) and ensemble-based (RFC, ETC, CBC, GBC, XGBC) models demonstrated higher performance, with ensemble models proving particularly effective. This advantage could be attributed to the capacity of ensemble methods to mitigate overfitting as well as maintain robustness to noise, albeit at the expense of higher computational resources and reduced interpretability relative to single-tree models [56,76,77]. The enhanced performance of ensemble-based models has also been reported in studies applying machine learning to fish microbiota, although such investigations have usually focused on associations with contaminated environments rather than dietary influences [46,78]; nevertheless, research leveraging machine learning on fish microbiota remains scarce. Comparable outcomes were observed with probability-based (MNB) and distance-based (KNC) models, which are valued for rapid training and high interpretability, although they exhibit sensitivity to correlated or irrelevant features. Margin-based (SVC), neural-network-based (MLPC), and generalized linear (LREG) models also merit consideration, despite attaining comparable performance only within specific categorical groupings. These limitations are consistent with their inherent characteristics: margin-based models are optimized for high-dimensional spaces, whereas neural networks generally require large datasets to achieve optimal training. Nevertheless, SVC and LREG remain widely common in microbiome research [56,77], including the few studies applying machine learning to fish microbiota [46,78]. In the case of SVC, reduced performance may be attributed to the use of the linear kernel, which has been selected due to its generalizability and stability, rather than the radial basis function, which is frequently adopted as a default kernel despite its higher risk of overfitting. For MLPC, performance decrease likely reflects the small sample size, which constrains the capacity of the model to detect meaningful patterns. Despite outcome variability across both studies, machine learning has demonstrated efficacy in extracting information at multiple taxonomic levels [79], underscoring the need for further refinement of the implemented analytical pipeline.

Classification results were strongly influenced by the feature selection process, which proved instrumental in both mitigating overfitting and enhancing model generalizability.

Building on these favorable results, the microbial taxa identified with feature selection in the classification task were further investigated to assess their biological relevance. Given the high taxonomic resolution afforded by the SILVA database, together with the findings reported in both companion studies, result interpretation focused on the genus level, for which, based on the most prevalent taxa described in Study 1 and Study 2, high-frequency (HF) and low-frequency (LF) genera were defined. For Study 1, HF genera included *Clostridium sensu stricto* 1 and *Enhydrobacter*, which were part of the study-specific core microbiota established by the authors; LF genera included *Pseudoalteromonas* and *Flavobacterium*. For Study 2, HF genera included *Acinetobacter*, *Cutibacterium*, *Enterovibrio*, *Escherichia-Shigella*, *Idiomarina*, and *Streptococcus*; LF genera included *Acholeplasma*, *Brevundimonas*, *Lactobacillus*, *Marinobacter*, *Micrococcus*, *Peptoniphilus*, *Photobacterium*, *Salegentibacter*, *Salinisphaera*, and an unclassified member of the family *Flavobacteriaceae*.

From an ecological perspective, the microbial genera comprising HF and LF groups in both studies are known to influence the gut health of fish hosts through beneficial, neutral, and pathogenic roles. *Clostridium sensu stricto* 1, *Lactobacillus*, and *Brevundimonas* are frequently associated with fermentation of dietary substrates, production of short-chain fatty acids, and probiotic effects capable of improving immunity and growth [34,80,81]. Marine-associated taxa like *Idiomarina*, *Marinobacter*, *Pseudoalteromonas*, and *Salinisphaera* contribute to ecological stability by degrading complex polysaccharides or hydrocarbons and by producing antimicrobial compounds that suppress pathogens [82–85]. *Enterovibrio*, *Flavobacterium* and *Photobacterium* include commensal and pathogenic species, reflecting the dual role of some genera in nutrient cycling and disease [86–88]. Opportunistic or conditionally pathogenic taxa such as *Acinetobacter*, *Escherichia-Shigella*, and *Streptococcus* can compromise gut homeostasis under stress conditions [89], while genera like *Acholeplasma*, *Cutibacterium*, *Enhydrobacter*, *Micrococcus*, *Peptoniphilus*, and *Salegentibacter* are less characterized but may contribute to microbial diversity and niche occupation. Collectively, these genera shape gut health by mediating digestion, modulating immune responses, and influencing resilience against environmental stressors, with their impact depending on host species, diet, and microbial balance.

In the case of Study 1, HF genera were associated with the DG1 grouping, with *Clostridium* as the only genus for which a statistically significant interaction effect was detected through conventional microbiota analyses; moreover, using our approach, *Enhydrobacter* was identified as an additional genus with influence on the interaction effect. In contrast, LF genera included *Pseudoalteromonas* and *Flavobacterium*, which, in the companion study, showed no statistically significant effect with D1 or G1, respectively. For the G1 grouping, while *Flavobacterium* was not among the genera identified as significant in conventional microbiota analyses, statistical significance was observed at the family level for *Flavobacteriaceae*. In light of the high accuracy attained by the selected models, however, *Flavobacterium* alone was sufficient to predict the dietary regime, without the inclusion of additional genera, thereby suggesting a functionally significant role for this taxon. For the D1 grouping, instead, conventional analyses did not identify any statistically significant associations with microbial taxa; in contrast, our feature selection approach identified *Pseudoalteromonas* as the sole genus required to achieve accurate diet-based classification.

In the case of Study 2, for the DG2 grouping, conventional techniques identified *Acinetobacter* and *Enterovibrio* as genera demonstrating statistically significant interaction effects; furthermore, both genera were classified as HF taxa by the feature selection procedure, which, however, also highlighted other influential genera, particularly among the LF group. For the D2 grouping, feature selection identified a smaller set of predictive genera compared with those detected as statistically significant by conventional techniques. *Streptococcus* was the only genus shared between approaches, while *Acholeplasma* and *Lactobacillus* were

uniquely identified through feature selection and not detected by conventional methods. For the G2 grouping, conventional analyses found statistical significance only for *Photobacterium*, which was likewise included in the feature set returned by feature selection, while *Micrococcus* was restricted to the ML-derived feature set.

In both studies, the use of machine learning for feature selection deepened our understanding of the microbial genera impacted by diet, genotype, and diet \times genotype interactions. In this regard, while ML-based feature selection was largely consistent with conventional microbiota analyses, with statistically significant taxa also captured within the ML-derived feature set, it additionally identified microbial taxa not originally reported as such. For Study 1, the following additional genera were identified: *Enhydrobacter* (DG1), *Pseudoalteromonas* (D1), and *Flavobacterium* (G1). For Study 2, the following additional genera were detected: *Brevundimonas*, *Cutibacterium*, *Escherichia-Shigella*, *Idiomarina*, *Lactobacillus*, *Marinobacter*, *Peptoniphilus*, *Salegentibacter*, *Salinisphaera*, and *Streptococcus* (DG2); *Acholeplasma* and *Lactobacillus* (D2); *Micrococcus* (G2). When these findings were compared with HF and LF genera reported in both studies, a greater number of LF genera was observed. This suggests that low-abundance taxa may exert a stronger influence on gut microbiota profile than previously estimated with conventional microbial community analyses, despite not being recognized as core microbiota components [90,91]. These results benefited from the capacity of machine learning to mine data more extensively than conventional techniques, owing to the broader range of available models and the possibility to evaluate features concurrently rather than independently. Moreover, the independence of ML models from biological assumptions enabled the recognition of subtle patterns that may be overlooked by conventional methods.

Despite the generally unsatisfactory predictive performance observed in regression models, the results of the feature selection procedure warrant brief consideration. In Study 1, statistically significant features were identified only within the DG1 grouping, whereas no features were selected for the D1 grouping. In Study 2, by contrast, the procedure detected significant features for the DG2 and D2 groupings. These findings find support in the microbial diversity analyses reported in the reference studies, although the observed effects were generally restricted to specific taxa. Specifically, Study 1 revealed interaction effects without evidence of a diet effect, while Study 2 identified both interaction and diet effects. Notably, however, the microbial taxa highlighted by the feature selection procedure did not fully overlap with those reported in the original analyses, thus underscoring the need for further investigation. This discrepancy suggests that, while the selection algorithm was capable of detecting statistically significant features, the biological relevance of these findings remains uncertain.

4.1. Methodological Considerations and Research Horizons

The present study offers valuable insights into the individual and synergistic effects of multiple factors (diet and genotype) on the autochthonous intestinal microbiota of European sea bass; however, certain limitations must be addressed to ensure scientific completeness, transparency, and rigor.

4.1.1. Computational Considerations

Compared to research studies on fish growth performance, the sample sizes used in the present work were relatively limited (24 fish for Study 1, 48 fish for Study 2), though numerically consistent with those reported in microbiota-focused studies [46,92,93]. While larger sample sizes are usually recommended to enhance statistical robustness and diet-induced microbiota shift detection [94–96], to minimize dataset-specific bias, we implemented computational strategies including feature selection to reduce dimensionality,

LOOCV to maximize data utilization, and NCV to separate hyperparameter tuning from model evaluation [97]. We also decided against data augmentation, given that the artificial manipulation of biological datasets may result in the introduction of unrealistic patterns and thus compromise biological validity. Nonetheless, current research is investigating targeted augmentation strategies to address these challenges, particularly in the context of microbial datasets [98–100]. Moreover, we acknowledge certain limitations associated with the feature selection procedure: on the one hand, feature selection was performed on the entire datasets prior to modeling, thereby introducing a potential risk of data leakage; on the other hand, taxa identification is inherently dependent on the algorithms used to implement the selection procedure, which could return feature sets differing from those reported in this study when using different learning estimators, selection functions, and inclusion criteria. While our choice to define a common feature set prioritized interpretability and consistency, future work could address overfitting by performing feature selection within each fold of nested cross-validation, followed by consensus procedures to derive a unified feature set for better interpretability. With larger datasets, feature selection could also be limited to a dedicated subset since such data volumes generally ensure representativeness [77,101]. Furthermore, taxonomy-aware feature engineering approaches, which leverage phylogenetic hierarchy to create compact and informative feature spaces, may provide even more insightful outcomes in supervised contexts [102]. Despite the present constraints, the alignment between ML-derived findings and those obtained through conventional methods reinforces the reliability of our approach and highlights its potential for future applications.

In microbiome studies with small sample sizes, informative feature selection has generally relied on established bioinformatic tools (e.g., ANCOM-BC, STAMP, LEfSe), which, while offering interpretability through conventional statistical metrics (e.g., p -values, confidence intervals, effect sizes), remain domain-specific and unsuitable for predictive modeling in supervised contexts. By contrast, machine learning offers the ability to jointly assess features, quantify feature importance, capture interaction effects and co-occurrence patterns among taxa, and manage high-dimensional data. In line with our research objectives, we implemented ML models capable of evaluating multiple features simultaneously and integrating with explanatory frameworks, thus balancing predictive capability with biological interpretability and complementing conventional microbiome analyses. However, considering the considerable shift toward ML-based approaches in microbiome research, it is important to acknowledge that the black-box nature of ML models poses interpretability challenges for non-experts. To overcome this, hybrid strategies integrating domain-specific methods with ML models within unified analytical pipelines have been proposed. For instance, in a recent colorectal cancer study, LEfSe and ANCOM-BC were used to characterize individual taxa with altered abundance in affected patients, followed by a Bayesian machine learning model to substantiate compositional differences observed in individual microbial markers [103]. Within such combined frameworks, machine learning contributes scalability and flexibility for handling high-dimensional, multi-omics datasets, while domain-specific tools provide biologically grounded insights based on transparent assumptions. Leveraging transdisciplinary knowledge transfer [104], these hybrid implementations could be applied to aquaculture research, fostering intelligent fish farming and providing deeper insights into microbiota alterations driven by different dietary and rearing conditions. Building on the research of Zakaria et al. [105], who demonstrated the integration of artificial intelligence with gut microbiota modulation for disease detection in shrimp aquaculture, a comparable hybrid pipeline can be envisioned for precision aquaculture in fish farming. Conventional bioinformatic techniques could serve as a first step in filtering and prioritizing candidate microbial features, thus reducing dimensionality and ensuring biological

relevance. Selected features can then be incorporated into ML models capable of modeling complex nonlinear interactions with environmental and dietary parameters. This pipeline could facilitate early detection of stress or disease, optimize feed formulations to support beneficial microbial communities, and provide real-time decision support for farmers. Ultimately, the convergence of machine learning and bioinformatic tools represents a promising paradigm for precision aquaculture, enabling microbiota-informed predictions that support the sustainable management of dietary and rearing practices while preserving interpretability.

4.1.2. Biological Considerations

While ML modeling in this study was performed on microbiota abundance data, it is important to consider the rearing system used, namely FTS in Study 1 and RAS in Study 2, as they exert direct effects on microbial communities [106]. In FTS, fish are continuously exposed to a constant influx of externally renewed water, resulting in higher environmental microbial input and greater temporal variability in intestinal communities. In contrast, RAS limit external microbial inputs through water treatment processes, while relying on biofilter-associated microbiota to maintain water quality. Indeed, these controlled rearing configurations enhance host-mediated selective pressures that foster the establishment of facility-specific microbial assemblages, thereby acting as physical environments as well as ecological filters that shape the microbiota communities colonizing the host gut. As such, it is reasonable to hypothesize that the specificity of microbiota associated with the rearing system may have influenced the features selected in each study. This raises the important methodological consideration that the discriminative features identified through machine learning may not solely reflect diet and/or genotype effects but also capture signatures of the rearing environment itself. This influence is especially relevant when comparing studies across facilities or rearing systems, as the microbial baseline established by RAS or FTS may bias feature selection toward taxa that are characteristic of the system rather than universally associated with diet or genotype. Therefore, it should not be definitively ruled out that the selected features are partly representative of the rearing system, in addition to the combined and/or individual effects under examination. The acknowledgment of the role of rearing-system-specific microbiota enhances the interpretive framework of aquaculture microbiome studies, highlighting the importance of distinguishing between host-driven and environment-driven microbial signals and thus ensuring that predictive pipelines capture biologically meaningful patterns rather than artifacts of facility-specific microbial ecology. Beyond rearing conditions, it is important to acknowledge that 16S rRNA gene sequencing can introduce methodological biases depending on experimental design choices, potentially affecting abundance estimation and thus compromising community diversity assessment [107,108]. Together with other factors influencing microbiota analysis (e.g., species, life stage, captivity status), these considerations highlight the importance of rigorous experimental design and methodological consistency in microbiome research to ensure reliability, reproducibility, and knowledge transferability across studies.

As this investigation was conducted within a ML framework, the validation of both the selected models and the feature subsets identified in each study on external datasets constitutes a fundamental step in ensuring pipeline robustness. However, it is important to notice that, although poultry-derived and fortified feeds have frequently been used as FM and FO substitutes, both studies utilized specifically developed feed formulations. As such, the present study focused on the implementation of diet-specific predictive models able to learn patterns unique to each dietary formulation. Consequently, robust external validation would necessitate datasets derived from European sea bass reared under comparable conditions and fed identical diets to avoid feature misattribution, reduced predictive

performance, or misleading conclusions; however, to the best of our knowledge, no such datasets are currently available that would enable the external validation of our models without introducing potential bias.

As publicly accessible microbiome datasets become increasingly common, large-scale research programs and consortia are simultaneously emerging to expand this knowledge. Although still modest in scale compared to landmark initiatives such as the Human Microbiome Project, these efforts are oriented toward mapping, engineering, and applying microbiomes in aquaculture under diverse environmental and rearing conditions. A central focus lies in the development of integrated multitrophic aquaculture (IMTA) systems, which provide experimental platforms for investigating innovative and environmentally sustainable feeds, thus positioning aquaculture practices within the broader paradigms of the circular economy and the One Health approach. Importantly, the emergence of these initiatives underscores both the scientific potential of microbiome-informed aquaculture and the urgent need to reform data-sharing policies [109]. Establishing clear guidelines for equitable reuse and interoperability of datasets will be crucial to ensure that the benefits of these programs extend across disciplines, regions, and stakeholders, ultimately fostering a more collaborative and sustainable research ecosystem.

5. Conclusions

With the ongoing transition in aquaculture toward environmentally sustainable and economically efficient feed strategies, poultry-based ingredients and fortified feeds have emerged as established alternatives to reduce reliance on FM and FO. In the present study, we applied machine learning to investigate the gut microbiota of European sea bass using abundance data from two prior trials. While conventional microbiome analyses revealed limited effects for diet and genotype, machine learning provided more nuanced insights. Regression models performed unsatisfactorily, owing to small sample sizes, low dietary variability, and inherent model constraints, underscoring the need for methodological refinement to improve predictive capacity. In contrast, classification analyses showed consistency across studies and with conventional microbiome evaluations, supporting the robustness of our pipeline, particularly with ensemble-based models and, to a lesser extent, probability-based and margin-based models. Taken together, these results underscore the importance of considering both genetic background and dietary formulation when evaluating microbiota composition in aquaculture species. Nevertheless, additional validation is needed to confirm the discriminative power and biological consistency of ML models. Despite existing limitations, this work marks an early application of machine learning to fish microbiota and provides a valuable platform for advancing for future studies. Hybrid approaches integrating AI-driven methods with conventional microbiome analyses hold particular promise, offering system-level perspectives on microbiota–host interactions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app152413029/s1>, Supplementary File S1: Performance metrics of classification models for the gut microbiota of European sea bass at the family level (Study 1). Supplementary File S2: Classification feature importance variation in the gut microbiota of European sea bass at the family level (Study 1). Supplementary File S3: Performance metrics of classification models for the gut microbiota of European sea bass at the genus level (Study 1). Supplementary File S4: Classification feature importance variation in the gut microbiota of European sea bass at the genus level (Study 1). Supplementary File S5: Performance metrics of regression models for the gut microbiota of European sea bass at the family level (Study 1). Supplementary File S6: Performance metrics of regression models for the gut microbiota of European sea bass at the genus level (Study 1). Supplementary File S7: Performance metrics of classification models for the gut microbiota of European sea bass at the family level (Study 2). Supplementary File S8: Classification feature importance

variation in the gut microbiota of European sea bass at the family level (Study 2). Supplementary File S9: Performance metrics of classification models for the gut microbiota of European sea bass at the genus level (Study 2). Supplementary File S10: Classification feature importance variation in the gut microbiota of European sea bass at the genus level (Study 2). Supplementary File S11: Performance metrics of regression models for the gut microbiota of European sea bass at the family level (Study 2). Supplementary File S12: Performance metrics of regression models for the gut microbiota of European sea bass at the genus level (Study 2). Supplementary Table S1: ingredient composition (percentage) of control and experimental diets in Study1. Supplementary Table S2: ingredient composition (percentage) of control and experimental diets in Study 2.

Author Contributions: Conceptualization, G.T. and S.R. (Silvio Rizzi); Methodology, S.R. (Silvio Rizzi), S.R. (Simona Rimoldi) and G.S.; Data Curation, S.R. (Silvio Rizzi), S.R. (Simona Rimoldi), G.S. and V.K.; Writing—Original Draft Preparation, S.R. (Silvio Rizzi) and G.T.; Writing—Review & Editing, S.R. (Silvio Rizzi), S.R. (Simona Rimoldi), G.S., V.K. and G.T.; Funding Acquisition, G.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by I-FISH. Protocol Number: 414352 (7 December 2023). Area di Orientamento Occupazionale (AOO)—Fondo per la Crescita Sostenibile (FCS)—Accordi per l’innovazione (D.M. 31 December 2021 and D.D. 14 November 2022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the sequencing data used in this study were previously deposited as FASTQ files into the European Nucleotide Archive (ENA) database under accession numbers PRJEB47388 and PRJEB61519. No new sequencing data were generated in this study.

Acknowledgments: G.S. and V.K. are doctoral students enrolled in the Ph.D. program in Life Sciences and Biotechnology at the University of Insubria, Varese, Italy.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---------|---------------------------------------|
| FM | fish meal |
| FO | fish oil |
| PM | poultry meal |
| PO | poultry oil |
| DHA | docosahexaenoic acid |
| DPA | docosapentaenoic acid |
| EPA | eicosapentaenoic acid |
| LC-PUFA | long-chain polyunsaturated fatty acid |
| MUFA | monounsaturated fatty acid |
| OA | organic acid |
| OAB | organic acid blend |
| PFA | phytogenic feed additive |
| ASV | amplicon sequence variant |
| CFU | colony-forming unit |
| IMTA | integrated multitrophic aquaculture |
| FTS | flow-through system |
| RAS | recirculating aquaculture system |
| NGS | next-generation sequencing |
| AI | artificial intelligence |
| ML | machine learning |
| RO | research objective |
| ENA | European Nucleotide Archive |

| | |
|----------------|--|
| HF | high-frequency |
| LF | low-frequency |
| WT | wild-type genotype |
| HG | high-growth genotype |
| CTRL | control diet |
| CTRL-WT | control diet fed to wild-type genotype |
| CTRL-HG | control diet fed to high-growth genotype |
| F | future diet |
| F-WT | future diet fed to wild-type genotype |
| F-HG | future diet fed to high-growth genotype |
| ORG | experimental diet supplemented with organic acids |
| ORG-WT | organic-acid-supplemented diet fed to wild-type genotype |
| ORG-HG | organic-acid-supplemented diet fed to high-growth genotype |
| PHYTO | experimental diet supplemented with phytogetic extracts |
| PHYTO-WT | phytogetic-extract-supplemented diet fed to wild-type genotype |
| PHYTO-HG | phytogetic-extract-supplemented diet fed to high-growth genotype |
| PROB | experimental diet supplemented with probiotics |
| PROB-WT | probiotic-supplemented diet fed to wild-type genotype |
| PROB-HG | probiotic-supplemented diet fed to high-growth genotype |
| D1 | diet class group for Study 1 |
| D2 | diet class group for Study 2 |
| G1 | genotype class group for Study 1 |
| G2 | genotype class group for Study 2 |
| DG1 | diet × genotype class group for Study 1 |
| DG2 | diet × genotype class group for Study 2 |
| CV | cross-validation |
| NCV | nested cross-validation |
| LOOCV | leave-one-out cross-validation |
| OOF | out-of-fold |
| ACC | accuracy |
| MCC | Matthews correlation coefficient |
| R ² | coefficient of determination |
| MAE | mean absolute error |
| RMSE | root mean squared error |
| RFE | recursive feature elimination |
| ANOVA | analysis of variance |
| LASSO | least absolute shrinkage and selection operator |
| LEfSe | linear discriminant analysis effect size |
| ANCOM-BC | analysis of compositions of microbiomes with bias correction |
| STAMP | statistical analysis of metagenomic profiles |
| DT | decision tree |
| DTC | decision tree classifier |
| DTR | decision tree regressor |
| RF | random forest |
| RFC | random forest classifier |
| RFR | random forest regressor |
| ET | extra tree |
| ETC | extra tree classifier |
| ETR | extra tree regressor |
| GB | gradient boosting |
| GBC | gradient boosting classifier |
| GBR | gradient boosting regressor |
| XGB | extreme gradient boosting |
| XGBC | extreme gradient boosting classifier |

| | |
|--------------------|--|
| XGBR | extreme gradient boosting regressor |
| CB | categorical boosting |
| CBC | categorical boosting classifier |
| SVM | support vector machine |
| SVC | support vector machine classifier |
| SVR | support vector machine regressor |
| NB | Naïve Bayes |
| MNB | multinomial Naïve Bayes |
| <i>k</i> -NN | <i>k</i> -nearest neighbors |
| KNC | <i>k</i> -nearest neighbor classifier |
| KNR | <i>k</i> -nearest neighbor regressor |
| MLP | multilayer perceptron |
| MLPC | multilayer perceptron classifier |
| MLPR | multilayer perceptron regressor |
| GLM | generalized linear model |
| LREG | logistic regression |
| LK | linear kernel |
| PK | polynomial kernel |
| RK | radial basis function kernel |
| SK | sigmoid kernel |
| AVG NCV | average model performance during nested cross-validation |
| OOF ACC | accuracy of the model using the best parameters returned by nested cross-validation on out-of-fold predictions |
| OOF MCC | Matthews correlation coefficient of the model using the best parameters returned by nested cross-validation on out-of-fold predictions |
| OOF R ² | coefficient of determination of the model using the best parameters returned by nested cross-validation on out-of-fold predictions |
| OOF MAE | mean absolute error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions |
| OOF RMSE | root mean squared error of the model using the best parameters returned by nested cross-validation on out-of-fold predictions |

References

1. Troell, M.; Naylor, R.L.; Metian, M.; Beveridge, M.; Tyedmers, P.H.; Folke, C.; Arrow, K.J.; Barrett, S.; Crépin, A.S.; Ehrlich, P.R.; et al. Does aquaculture add resilience to the global food system? *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13257–13263. [\[CrossRef\]](#)
2. Olsen, R.L.; Hasan, M.R. A limited supply of fishmeal: Impact on future increases in global aquaculture production. *Trends Food Sci. Technol.* **2012**, *27*, 120–128. [\[CrossRef\]](#)
3. Naylor, R.; Burke, M. Aquaculture and ocean resources: Raising tigers of the sea. *Annu. Rev. Environ. Resour.* **2005**, *30*, 185–218. [\[CrossRef\]](#)
4. Galkanda-Arachchige, H.S.C.; Wilson, A.E.; Davis, D.A. Success of fishmeal replacement through poultry by-product meal in aquaculture feed formulations: A meta-analysis. *Rev. Aquac.* **2020**, *12*, 1624–1636. [\[CrossRef\]](#)
5. Wang, X.; Luo, H.; Zheng, Y.; Wang, D.; Wang, Y.; Zhang, W.; Chen, Z.; Chen, X.; Shao, J. Effects of poultry by-product meal replacing fish meal on growth performance, feed utilization, intestinal morphology and microbiota communities in juvenile large yellow croaker (*Larimichthys crocea*). *Aquac. Rep.* **2023**, *30*, 101547. [\[CrossRef\]](#)
6. Fontinha, F.; Magalhães, R.; Moutinho, S.; Santos, R.; Campos, P.; Serra, C.R.; Aires, T.; Oliva-Teles, A.; Peres, H. Effect of dietary poultry meal and oil on growth, digestive capacity, and gut microbiota of gilthead seabream (*Sparus aurata*) juveniles. *Aquaculture* **2021**, *530*, 735879. [\[CrossRef\]](#)
7. Psafakis, P.; Meziti, A.; Berillis, P.; Mente, E.; Kormas, K.A.; Karapanagiotidis, I.T. Effects of dietary fishmeal replacement by poultry by-product meal and hydrolyzed feather meal on liver and intestinal histomorphology and on intestinal microbiota of gilthead seabream (*Sparus aurata*). *Appl. Sci.* **2021**, *11*, 8806. [\[CrossRef\]](#)
8. Wu, Z.; Yu, X.; Fu, Y.; Guo, J.; Pan, M.; Guo, Y.; Liu, J.; Mai, K.; Zhang, W. Impacts of replacing dietary fish meal with poultry by-product meal on growth, digestive enzymes and gut microbiota, biomarkers of metabolic and immune response, and resistance to *Vibrio* challenge in abalone (*Haliotis discus hannai*). *Aquaculture* **2023**, *576*, 739871. [\[CrossRef\]](#)

9. Hasan, I.; Rimoldi, S.; Chiofalo, B.; Oteri, M.; Antonini, M.; Armone, R.; Kalemi, V.; Gasco, L.; Terova, G. Effects of poultry by-product meal and complete replacement of fish oil with alternative oils on growth performance and gut health of rainbow trout (*Oncorhynchus mykiss*): A FEEDNETICS™ validation study. *BMC Vet. Res.* **2024**, *20*, 472. [[CrossRef](#)] [[PubMed](#)]
10. Gaudioso, G.; Marzorati, G.; Faccenda, F.; Weil, T.; Lunelli, F.; Cardinaletti, G.; Marino, G.; Olivotto, I.; Parisi, G.; Tibaldi, E.; et al. Processed animal proteins from insect and poultry by-products in a fish meal-free diet for rainbow trout: Impact on intestinal microbiota and inflammatory markers. *Int. J. Mol. Sci.* **2021**, *22*, 5454. [[CrossRef](#)]
11. Rimoldi, S.; Terova, G.; Ascione, C.; Giannico, R.; Brambilla, F. Next generation sequencing for gut microbiome characterization in rainbow trout (*Oncorhynchus mykiss*) fed animal by-product meals as an alternative to fishmeal protein sources. *PLoS ONE* **2018**, *13*, e0193652. [[CrossRef](#)]
12. Rimoldi, S.; Di Rosa, A.R.; Armone, R.; Chiofalo, B.; Hasan, I.; Saroglia, M.; Kalemi, V.; Terova, G. The replacement of fish meal with poultry by-product meal and insect exuviae: Effects on growth performance, gut health and microbiota of the European seabass, *Dicentrarchus labrax*. *Microorganisms* **2024**, *12*, 744. [[CrossRef](#)] [[PubMed](#)]
13. Bowyer, J.N.; Qin, J.G.; Smullen, R.P.; Stone, D.A.J. Replacement of fish oil by poultry oil and canola oil in yellowtail kingfish (*Seriola lalandi*) at optimal and suboptimal temperatures. *Aquaculture* **2012**, *356*, 211–222. [[CrossRef](#)]
14. Ng, W.K.; Koh, C.B. The utilization and mode of action of organic acids in the feeds of cultured aquatic animals. *Rev. Aquac.* **2017**, *9*, 342–368. [[CrossRef](#)]
15. Busti, S.; Rossi, B.; Volpe, E.; Ciulli, S.; Piva, A.; D'Amico, F.; Soverini, M.; Candela, M.; Gatta, P.P.; Bonaldo, A.; et al. Effects of dietary organic acids and nature identical compounds on growth, immune parameters and gut microbiota of European sea bass. *Sci. Rep.* **2020**, *10*, 21321. [[CrossRef](#)] [[PubMed](#)]
16. Huyben, D.; Chiasson, M.; Lumsden, J.S.; Pham, P.H.; Chowdhury, M.A.K. Dietary microencapsulated blend of organic acids and plant essential oils affects intestinal morphology and microbiome of rainbow trout (*Oncorhynchus mykiss*). *Microorganisms* **2021**, *9*, 2063. [[CrossRef](#)]
17. da Silva, V.G.; Favero, L.M.; Mainardi, R.M.; Ferrari, N.A.; Chideroli, R.T.; Di Santis, G.W.; de Souza, F.P.; da Costa, A.R.; Gonçalves, D.D.; Nuez-Ortin, W.G.; et al. Effect of an organic acid blend in Nile tilapia growth performance, immunity, gut microbiota, and resistance to challenge against francisellosis. *Res. Vet. Sci.* **2023**, *159*, 214–224. [[CrossRef](#)]
18. Addam, K.G.S.; Pereira, S.A.; Jesus, G.F.A.; Cardoso, L.; Syracuse, N.; Lopes, G.R.; Lehmann, N.B.; da Silva, B.C.; de Sá, L.S.; Chaves, F.C.M.; et al. Dietary organic acids blend alone or in combination with an essential oil on the survival, growth, gut/liver structure and de hemato-immunological in Nile tilapia *Oreochromis niloticus*. *Aquac. Res.* **2019**, *50*, 2960–2971. [[CrossRef](#)]
19. Wang, J.; Deng, L.; Chen, M.; Che, Y.; Li, L.; Zhu, L.; Chen, G.; Feng, T. Phyto-genic feed additives as natural antibiotic alternatives in animal health and production: A review of the literature of the last decade. *Anim. Nutr.* **2024**, *17*, 244–264. [[CrossRef](#)]
20. Gherescu, P.; Mihailov, S.; Grozea, A. The influence of some phyto-additives on bio-productive performances and the health of the farmed fish—Review. *Sci. Pap. Anim. Sci. Biotechnol.* **2023**, *56*, 128.
21. Zhang, W.; Zhao, J.; Ma, Y.; Li, J.; Chen, X. The effective components of herbal medicines used for prevention and control of fish diseases. *Fish Shellfish Immunol.* **2022**, *126*, 73–83. [[CrossRef](#)]
22. Zhu, C.B.; Shen, Y.T.; Ren, C.H.; Yang, S.; Fei, H. A novel formula of herbal extracts regulates growth performance, antioxidant capacity, intestinal microbiota and resistance against *Aeromonas veronii* in largemouth bass (*Micropterus salmoides*). *Aquaculture* **2024**, *583*, 740614. [[CrossRef](#)]
23. Shen, Y.T.; Ding, Z.L.; Wang, X.Y.; Chen, W.Q.; Xia, R.X.; Yang, S.; Fei, H. Combination of herbal extracts regulates growth performance, liver and intestinal morphology, antioxidant capacity, and intestinal microbiota in *Acrossocheilus fasciatus*. *Aquaculture* **2025**, *594*, 741428. [[CrossRef](#)]
24. Feher, M.; Fauszt, P.; Tolnai, E.; Fidler, G.; Pesti-Asboth, G.; Stigel, A.; Szucs, I.; Biro, S.; Remenyik, J.; Paholcsek, M.; et al. Effects of phytonutrient-supplemented diets on the intestinal microbiota of *Cyprinus carpio*. *PLoS ONE* **2021**, *16*, e0248537. [[CrossRef](#)]
25. Soares, M.P.; Cardoso, I.L.; Araújo, F.E.; De Angelis, C.F.; Mendes, R.; Mendes, L.W.; Fernandes, M.N.; Jonsson, C.M.; de Queiroz, S.C.D.N.; Duarte, M.C.T.; et al. Influences of the alcoholic extract of *Artemisia annua* on gastrointestinal microbiota and performance of Nile tilapia. *Aquaculture* **2022**, *560*, 738521. [[CrossRef](#)]
26. Karataş, B. Dietary *Cyanus depressus* (M. Bieb.) Soják plant extract enhances growth performance, modulates intestinal microbiota, and alters gene expression associated with digestion, antioxidant, stress, and immune responses in rainbow trout (*Oncorhynchus mykiss*). *Aquac. Int.* **2024**, *32*, 7929–7951. [[CrossRef](#)]
27. Ahmadifar, E.; Pourmohammadi Fallah, H.; Yousefi, M.; Dawood, M.A.O.; Hoseinifar, S.H.; Adineh, H.; Yilmaz, S.; Paolucci, M.; Doan, H.V. The gene regulatory roles of herbal extracts on the growth, immune system, and reproduction of fish. *Animals* **2021**, *11*, 2167. [[CrossRef](#)]
28. Melo-Bolívar, J.F.; Ruiz-Pardo, R.Y.; Hume, M.E.; Sidjabat, H.E.; Villamil-Díaz, L.M. Probiotics for cultured freshwater fish. *Microbiol. Aust.* **2020**, *41*, 105–108. [[CrossRef](#)]

29. Melo-Bolívar, J.F.; Ruiz-Pardo, R.Y.; Hume, M.E.; Nisbet, D.J.; Rodriguez-Villamizar, F.; Alzate, J.F.; Junca, H.; Villamil-Díaz, L.M. Establishment and characterization of a competitive exclusion bacterial culture derived from Nile tilapia (*Oreochromis niloticus*) gut microbiomes showing antibacterial activity against pathogenic *Streptococcus agalactiae*. *PLoS ONE* **2019**, *14*, e0215375. [[CrossRef](#)]
30. Merrifield, D.L.; Bradley, G.; Harper, G.M.; Baker, R.T.M.; Munn, C.B.; Davies, S.J. Assessment of the effects of vegetative and lyophilized *Pediococcus acidilactici* on growth, feed utilization, intestinal colonization and health parameters of rainbow trout (*Oncorhynchus mykiss* Walbaum). *Aquac. Nutr.* **2011**, *17*, 73–79. [[CrossRef](#)]
31. Hasan, I.; Rimoldi, S.; Saroglia, G.; Terova, G. Sustainable fish feeds with insects and probiotics positively affect freshwater and marine fish gut microbiota. *Animals* **2023**, *13*, 1633. [[CrossRef](#)] [[PubMed](#)]
32. Alonso, S.; Carmen Castro, M.; Berdasco, M.; de la Banda, I.G.; Moreno-Ventas, X.; de Rojas, A.H. Isolation and partial characterization of lactic acid bacteria from the gut microbiota of marine fishes for potential application as probiotics in aquaculture. *Probiotics Antimicrob. Proteins* **2019**, *11*, 569–579. [[CrossRef](#)]
33. Abareethan, M.; Amsath, A. Characterization and evaluation of probiotic fish feed. *Int. J. Pure Appl. Zool.* **2015**, *3*, 148–153.
34. Gupta, S.; Fečkaninová, A.; Lokesh, J.; Koščová, J.; Sørensen, M.; Fernandes, J.; Kiron, V. *Lactobacillus* dominate in the intestine of Atlantic salmon fed dietary probiotics. *Front. Microbiol.* **2019**, *9*, 3247. [[CrossRef](#)]
35. Moroni, F.; Naya-Català, F.; Piazzon, M.C.; Rimoldi, S.; Caldach-Giner, J.; Giardini, A.; Martínez, I.; Brambilla, F.; Pérez-Sánchez, J.; Terova, G. The effects of nisin-producing *Lactococcus lactis* strain used as probiotic on gilthead sea bream (*Sparus aurata*) growth, gut microbiota, and transcriptional response. *Front. Mar. Sci.* **2021**, *8*, 659519. [[CrossRef](#)]
36. Cerezuela, R.; Fumanal, M.; Tapia-Paniagua, S.T.; Meseguer, J.; Moriñigo, M.Á.; Esteban, M.Á. Changes in intestinal morphology and microbiota caused by dietary administration of inulin and *Bacillus subtilis* in gilthead sea bream (*Sparus aurata* L.) specimens. *Fish Shellfish Immunol.* **2013**, *34*, 1063–1070. [[CrossRef](#)]
37. Ramos, M.A.; Weber, B.; Gonçalves, J.F.; Santos, G.A.; Rema, P.; Ozório, R.O.A. Dietary probiotic supplementation modulated gut microbiota and improved growth of juvenile rainbow trout (*Oncorhynchus mykiss*). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **2013**, *166*, 302–307. [[CrossRef](#)]
38. Standen, B.T.; Rodiles, A.; Peggs, D.L.; Davies, S.J.; Santos, G.A.; Merrifield, D.L. Modulation of the intestinal microbiota and morphology of tilapia, *Oreochromis niloticus*, following the application of a multi-species probiotic. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 8403–8417. [[CrossRef](#)] [[PubMed](#)]
39. Vandeputte, M.; Gagnaire, P.A.; Allal, F. The European sea bass: A key marine fish model in the wild and in aquaculture. *Anim. Genet.* **2019**, *50*, 195–206. [[CrossRef](#)]
40. Vandeputte, M.; Dupont-Nivet, M.; Haffray, P.; Chavanne, H.; Cenadelli, S.; Parati, K.; Vidal, M.O.; Vergnet, A.; Chatain, B. Response to domestication and selection for growth in the European sea bass (*Dicentrarchus labrax*) in separate and mixed tanks. *Aquaculture* **2009**, *286*, 20–27. [[CrossRef](#)]
41. Torrecillas, S.; Rimoldi, S.; Montero, D.; Serradell, A.; Acosta, F.; Fontanillas, R.; Allal, F.; Haffray, P.; Bajek, A.; Terova, G. Genotype × nutrition interactions in European sea bass (*Dicentrarchus labrax*): Effects on gut health and intestinal microbiota. *Aquaculture* **2023**, *574*, 739639. [[CrossRef](#)]
42. Rimoldi, S.; Montero, D.; Torrecillas, S.; Serradell, A.; Acosta, F.; Haffray, P.; Hostins, B.; Fontanillas, R.; Allal, F.; Bajek, A.; et al. Genetically superior European sea bass (*Dicentrarchus labrax*) and nutritional innovations: Effects of functional feeds on fish immune response, disease resistance, and gut microbiota. *Aquac. Rep.* **2023**, *33*, 101747. [[CrossRef](#)]
43. Rizzi, S.; Saroglia, G.; Kalemí, V.; Rimoldi, S.; Terova, G. Artificial intelligence in microbiome research and beyond: Connecting human health, animal husbandry, and aquaculture. *Appl. Sci.* **2025**, *15*, 9781. [[CrossRef](#)]
44. Zhao, S.; Zhang, S.; Liu, J.; Wang, H.; Zhu, J.; Li, D.; Zhao, R. Application of machine learning in intelligent fish aquaculture: A review. *Aquaculture* **2021**, *540*, 736724. [[CrossRef](#)]
45. Gladju, J.; Kamalam, B.S.; Kanagaraj, A. Applications of data mining and machine learning framework in aquaculture and fisheries: A review. *Smart Agric. Technol.* **2022**, *2*, 100061. [[CrossRef](#)]
46. Turner, J.W., Jr.; Cheng, X.; Saferin, N.; Yeo, J.Y.; Yang, T.; Joe, B. Gut microbiota of wild fish as reporters of compromised aquatic environments sleuthed through machine learning. *Physiol. Genom.* **2022**, *54*, 177–185. [[CrossRef](#)]
47. Zhang, B.; Xiao, J.; Liu, H.; Zhai, D.; Wang, Y.; Liu, S.; Xiong, F.; Xia, M. Vertical habitat preferences shape the fish gut microbiota in a shallow lake. *Front. Microbiol.* **2024**, *15*, 1341303. [[CrossRef](#)] [[PubMed](#)]
48. Soriano, B.; Hafez, A.I.; Naya-Català, F.; Moroni, F.; Moldovan, R.A.; Toxqui-Rodríguez, S.; Piazzon, M.C.; Arnau, V.; Llorens, C.; Pérez-Sánchez, J. SAMBA: Structure-learning of aquaculture microbiomes using a Bayesian approach. *Genes* **2023**, *14*, 1650. [[CrossRef](#)] [[PubMed](#)]
49. Hua, K.; Cobcroft, J.M.; Cole, A.; Condon, K.; Jerry, D.R.; Mangott, A.; Praeger, C.; Vucko, M.J.; Zeng, C.; Zenger, K.; et al. The future of aquatic protein: Implications for protein sources in aquaculture diets. *One Earth* **2019**, *1*, 316–329. [[CrossRef](#)]
50. Cooney, R.; Wan, A.H.L.; O'Donncha, F.; Clifford, E. Designing environmentally efficient aquafeeds through the use of multicriteria decision support tools. *Curr. Opin. Environ. Sci. Health* **2021**, *23*, 100276. [[CrossRef](#)]

51. Garcia-Launay, F.; Dusart, L.; Espagnol, S.; Laisse-Redoux, S.; Gaudré, D.; Méda, B.; Wilfart, A. Multiobjective formulation is an effective method to reduce environmental impacts of livestock feeds. *Br. J. Nutr.* **2018**, *120*, 1298–1309. [[CrossRef](#)]
52. Lin, H.; Peddada, S.D. Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* **2020**, *11*, 3514. [[CrossRef](#)]
53. Parks, D.H.; Beiko, R.G. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **2010**, *26*, 715–721. [[CrossRef](#)]
54. Parks, D.H.; Tyson, G.W.; Hugenholtz, P.; Beiko, R.G. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **2014**, *30*, 3123–3124. [[CrossRef](#)] [[PubMed](#)]
55. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011**, *12*, R60. [[CrossRef](#)]
56. Li, P.; Luo, H.; Ji, B.; Nielsen, J. Machine learning for data integration in human gut microbiome. *Microb. Cell Fact.* **2022**, *21*, 241. [[CrossRef](#)]
57. Wakita, Y.; Shimomura, Y.; Kitada, Y.; Yamamoto, H.; Ohashi, Y.; Matsumoto, M. Taxonomic classification for microbiome analysis, which correlates well with the metabolite milieu of the gut. *BMC Microbiol.* **2018**, *18*, 188. [[CrossRef](#)]
58. Biada, I.; Santacreu, M.A.; González-Recio, O.; Ibáñez-Escriche, N. Comparative analysis of Illumina, PacBio, and nanopore for 16S rRNA gene sequencing of rabbit's gut microbiota. *Front. Microbiomes* **2025**, *4*, 1587712. [[CrossRef](#)]
59. Macip, G.; Soler-Comas, A.; Palomeque, A.; Motos, A.; Llonch, B.; Canseco-Ribas, J.; Bueno-Freire, L.; Calabretta, D.; Kiarostami, K.; Cabrera, R.; et al. Comparative analysis of illumina and oxford nanopore sequencing platforms for 16S rRNA profiling of respiratory microbial communities. *Sci. Rep.* **2025**, *15*, 33688. [[CrossRef](#)] [[PubMed](#)]
60. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **2017**, *10*, 35. [[CrossRef](#)] [[PubMed](#)]
61. Boyd, C.E.; D'Abramo, L.R.; Glencross, B.D.; Huyben, D.C.; Juarez, L.M.; Lockwood, G.S.; McNevin, A.A.; Tacon, A.G.J.; Teletchea, F.; Tomasso, J.R., Jr.; et al. Achieving sustainable aquaculture: Historical and current perspectives and future needs and challenges. *J. World Aquac. Soc.* **2020**, *51*, 578–633. [[CrossRef](#)]
62. Gallagher, M.L.; Degani, G. Poultry meal and poultry oil as sources of protein and lipid in the diet of European eels (*Anguilla anguilla*). *Aquaculture* **1988**, *73*, 177–187. [[CrossRef](#)]
63. Friesen, E.; Balfry, S.K.; Skura, B.J.; Ikonomou, M.; Higgs, D.A. Evaluation of poultry fat and blends of poultry fat with cold-pressed flaxseed oil as supplemental dietary lipid sources for juvenile sablefish (*Anoplopoma fimbria*). *Aquac. Res.* **2013**, *44*, 300–316. [[CrossRef](#)]
64. Dawson, M.R.; Alam, M.S.; Watanabe, W.O.; Carroll, P.M.; Seaton, P.J. Evaluation of poultry by-product meal as an alternative to fish meal in the diet of juvenile black sea bass reared in a recirculating aquaculture system. *N. Am. J. Aquac.* **2018**, *80*, 74–87. [[CrossRef](#)]
65. Onomu, A.J.; Okuthe, G.E. The role of functional feed additives in enhancing aquaculture sustainability. *Fishes* **2024**, *9*, 167. [[CrossRef](#)]
66. Ringø, E.; Harikrishnan, R.; Soltani, M.; Ghosh, K. The effect of gut microbiota and probiotics on metabolism in fish and shrimp. *Animals* **2022**, *12*, 3016. [[CrossRef](#)]
67. Panteli, N.; Mastoraki, M.; Nikouli, E.; Lazarina, M.; Antonopoulou, E.; Kormas, K.A. Imprinting statistically sound conclusions for gut microbiota in comparative animal studies: A case study with diet and teleost fishes. *Comp. Biochem. Physiol. D Genomics Proteomics* **2020**, *36*, 100738. [[CrossRef](#)]
68. Carda-Diéguéz, M.; Mira, A.; Fouz, B. Pyrosequencing survey of intestinal microbiota diversity in cultured sea bass (*Dicentrarchus labrax*) fed functional diets. *FEMS Microbiol. Ecol.* **2014**, *87*, 451–459. [[CrossRef](#)]
69. Larsen, A.; Tao, Z.; Bullard, S.A.; Arias, C.R. Diversity of the skin microbiota of fishes: Evidence for host species specificity. *FEMS Microbiol. Ecol.* **2013**, *85*, 483–494. [[CrossRef](#)]
70. Liu, Y.; Wu, H.; Chen, W.; Liu, C.; Meng, Q.; Zhou, Z. Rumen microbiome and metabolome of high and low residual feed intake angus heifers. *Front. Vet. Sci.* **2022**, *9*, 812861. [[CrossRef](#)] [[PubMed](#)]
71. Wang, Y.; Zhang, H.; Zhu, L.; Xu, Y.; Liu, N.; Sun, X.; Hu, L.; Huang, H.; Wei, K.; Zhu, R. Dynamic distribution of gut microbiota in goats at different ages and health states. *Front. Microbiol.* **2018**, *9*, 2509. [[CrossRef](#)]
72. Wang, L.; Jin, L.; Xue, B.; Wang, Z.; Peng, Q. Characterizing the bacterial community across the gastrointestinal tract of goats: Composition and potential function. *Microbiologyopen* **2019**, *8*, e00820. [[CrossRef](#)]
73. David, L.A.; Maurice, C.F.; Carmody, R.N.; Gootenberg, D.B.; Button, J.E.; Wolfe, B.E.; Ling, A.V.; Devlin, A.S.; Varma, Y.; Fischbach, M.A.; et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014**, *505*, 559–563. [[CrossRef](#)]
74. Zhou, Y.; Xu, Z.Z.; He, Y.; Yang, Y.; Liu, L.; Lin, Q.; Nie, Y.; Li, M.; Zhi, F.; Liu, S.; et al. Gut microbiota offers universal biomarkers across ethnicity in inflammatory bowel disease diagnosis and infliximab response prediction. *mSystems* **2018**, *3*, 10–1128. [[CrossRef](#)]
75. Namkung, J. Machine learning methods for microbiome studies. *J. Microbiol.* **2020**, *58*, 206–216. [[CrossRef](#)]
76. Hernández Medina, R.; Kutuzova, S.; Nielsen, K.N.; Johansen, J.; Hansen, L.H.; Nielsen, M.; Rasmussen, S. Machine learning and deep learning applications in microbiome research. *ISME Commun.* **2022**, *2*, 98. [[CrossRef](#)] [[PubMed](#)]

77. Marcos-Zambrano, L.J.; Karaduzovic-Hadziabdic, K.; Loncar Turukalo, T.; Przymus, P.; Trajkovik, V.; Aasmets, O.; Berland, M.; Gruca, A.; Hasic, J.; Hron, K.; et al. Applications of machine learning in human microbiome studies: A review on feature selection, biomarker identification, disease prediction and treatment. *Front. Microbiol.* **2021**, *12*, 634511. [[CrossRef](#)]
78. Shi, Z.; Guo, R.; Yao, F.; Liu, Z.; Zhang, J. Exploiting the gut microbiota of aquatic animals as indicators of microplastic pollution using interpretable machine learning models. *J. Hazard. Mater.* **2025**, *496*, 139178. [[CrossRef](#)]
79. Zhou, T.; Zhao, F. AI-empowered human microbiome research. *Gut* **2025**, 1–15. [[CrossRef](#)] [[PubMed](#)]
80. Cheng, R.; Ying, Z.; Yang, Y.; Zhang, C.; Zhou, W.; Zhang, Z.; Ding, H.; Zhou, Y.; Zhang, C. Changes of intestinal microbiota and liver metabolomics in yellow catfish (*Pelteobagrus fulvidraco*) before and after rice flowering in rice-fish symbiosis farmed mode. *Front. Microbiol.* **2025**, *16*, 1617168. [[CrossRef](#)] [[PubMed](#)]
81. Zhang, L.; Zhou, J.; Huang, Z.; Zhao, H.; Zhao, Z.; Mou, C.; Feng, Y.; Li, H.; Li, Q.; Duan, Y. *Lactobacillus acidophilus* in aquaculture: A review. *Microbiol. Res.* **2025**, *16*, 174. [[CrossRef](#)]
82. Al-Hisnawi, A.; Rodiles, A.; Rawling, M.D.; Castex, M.; Waines, P.; Gioacchini, G.; Carnevali, O.; Merrifield, D.L. Dietary probiotic *Pediococcus acidilactici* MA18/5M modulates the intestinal microbiota and stimulates intestinal immunity in rainbow trout (*Oncorhynchus mykiss*). *J. World Aquac. Soc.* **2019**, *50*, 1133–1151. [[CrossRef](#)]
83. Dvergedal, H.; Sandve, S.R.; Angell, I.L.; Klemetsdal, G.; Rudi, K. Association of gut microbiota with metabolism in juvenile Atlantic salmon. *Microbiome* **2020**, *8*, 160. [[CrossRef](#)] [[PubMed](#)]
84. Ramadan, H.A.I.; Jamal, M.T.; El-Wahsh, H.M.; El-Regal, M.A. Molecular analysis of bacterial communities in marbled spinefoot (*Siganus rivulatus*) and squaretail coral grouper (*Plectropomus areolatus*), in Jeddah, Saudi Arabia. *Egypt. J. Aquat. Res.* **2025**, *51*, 198–206. [[CrossRef](#)]
85. Singh, B.K.; Thakur, K.; Kumari, H.; Mahajan, D.; Sharma, D.; Sharma, A.K.; Kumar, S.; Singh, B.; Pankaj, P.P.; Kumar, R. A review on comparative analysis of marine and freshwater fish gut microbiomes: Insights into environmental impact on gut microbiota. *FEMS Microbiol. Ecol.* **2025**, *101*, fiae169. [[CrossRef](#)]
86. Huang, Q.; Sham, R.C.; Deng, Y.; Mao, Y.; Wang, C.; Zhang, T.; Leung, K.M.Y. Diversity of gut microbiomes in marine fishes is shaped by host-related factors. *Mol. Ecol.* **2020**, *29*, 5019–5034. [[CrossRef](#)]
87. Jose, J.A.; Alex, A.; Philip, S. Gut microbiome analysis reveals core microbiota variation among allopatric populations of the commercially important euryhaline cichlid *Etroplus suratensis*. *Microbiol. Res.* **2025**, *16*, 210. [[CrossRef](#)]
88. Soh, M.; Er, S.; Low, A.; Jaafar, Z.; de Boucher, R.; Seedorf, H. Spatial and temporal changes in gut microbiota composition of farmed Asian seabass (*Lates calcarifer*) in different aquaculture settings. *Microbiol. Spectr.* **2025**, *13*, e01989-24. [[CrossRef](#)]
89. Morshed, S.M.; Chen, Y.Y.; Lin, C.H.; Chen, Y.P.; Lee, T.H. Freshwater transfer affected intestinal microbiota with correlation to cytokine gene expression in Asian sea bass. *Front. Microbiol.* **2023**, *14*, 1097954. [[CrossRef](#)]
90. Chakraborty, D.; Jousset, A.; Wei, Z.; Banerjee, S. Rare taxa in the core microbiome. *Trends Microbiol.* **2025**, *33*, 727–737. [[CrossRef](#)] [[PubMed](#)]
91. Neu, A.T.; Allen, E.E.; Roy, K. Defining and quantifying the core microbiome: Challenges and prospects. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104429118. [[CrossRef](#)] [[PubMed](#)]
92. Ruiz, A.; Torrecillas, S.; Kashinskaya, E.; Andree, K.B.; Solovyev, M.; Gisbert, E. Comparative study of the gut microbial communities collected by scraping and swabbing in a fish model: A comprehensive guide to promote non-lethal procedures for gut microbial studies. *Front. Vet. Sci.* **2024**, *11*, 1374803. [[CrossRef](#)] [[PubMed](#)]
93. Chen, X.; Sun, C.; Dong, J.; Li, W.; Tian, Y.; Hu, J.; Ye, X. Comparative analysis of the gut microbiota of mandarin fish (*Siniperca chuatsi*) feeding on compound diets and live baits. *Front. Genet.* **2022**, *13*, 797420. [[CrossRef](#)]
94. Jarett, J.K.; Kingsbury, D.D.; Dahlhausen, K.E.; Ganz, H.H. Best practices for microbiome study design in companion animal research. *Front. Vet. Sci.* **2021**, *8*, 644836. [[CrossRef](#)] [[PubMed](#)]
95. Johnson, A.J.; Zheng, J.J.; Kang, J.W.; Saboe, A.; Knights, D.; Zivkovic, A.M. A guide to diet-microbiome study design. *Front. Nutr.* **2020**, *7*, 79. [[CrossRef](#)]
96. Marcos-Zambrano, L.J.; López-Molina, V.M.; Bakir-Gungor, B.; Frohme, M.; Karaduzovic-Hadziabdic, K.; Klammsteiner, T.; Ibrahim, E.; Lahti, L.; Loncar-Turukalo, T.; Dharmo, X.; et al. A toolbox of machine learning software to support microbiome analysis. *Front. Microbiol.* **2023**, *14*, 1250806. [[CrossRef](#)]
97. Kirk, D.; Kok, E.; Tufano, M.; Tekinerdogan, B.; Feskens, E.J.M.; Camps, G. Machine learning in nutrition research. *Adv. Nutr.* **2023**, *13*, 2573–2589. [[CrossRef](#)]
98. Gordon-Rodriguez, E.; Quinn, T.P.; Cunningham, J.P. Data augmentation for compositional data: Advancing predictive models of the microbiome. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 20551–20565. [[CrossRef](#)]
99. Wen, L.Y.; Zhang, X.M.; Li, Q.F.; Min, F. KGA: Integrating KPCA and GAN for microbial data augmentation. *Int. J. Mach. Learn. Cyber.* **2023**, *14*, 1427–1444. [[CrossRef](#)]
100. Wen, L.Y.; Chen, Z.; Xie, X.N.; Min, F. Microbial data augmentation combining feature extraction and transformer network. *Int. J. Mach. Learn. Cyber.* **2024**, *15*, 2539–2550. [[CrossRef](#)]

101. Karwowska, Z.; Aasmets, O.; Estonian Biobank Research Team; Kosciolk, T.; Org, E. Effects of data transformation and model selection on feature importance in microbiome classification data. *Microbiome* **2025**, *13*, 2. [[CrossRef](#)] [[PubMed](#)]
102. Oudah, M.; Henschel, A. Taxonomy-aware feature engineering for microbiome classification. *BMC Bioinform.* **2018**, *19*, 227. [[CrossRef](#)]
103. Han, H.; Li, Y.; Qi, Y.; Mangiola, S.; Ling, W. Deciphering gut microbiome in colorectal cancer via robust learning methods. *Genes* **2025**, *16*, 452. [[CrossRef](#)]
104. Han, J.; Zhang, H.; Ning, K. Techniques for learning and transferring knowledge for microbiome-based classification and prediction: Review and assessment. *Brief. Bioinform.* **2024**, *26*, bbaf015. [[CrossRef](#)]
105. Zakaria, M.; Francisco, M.E.; Sanyal, S.K.; Hossain, A.; Mandal, S.C.; Haque, M.I.M. A review on modulation of gut microbiome interaction for the management of shrimp aquaculture and proposal of the introduction of deep learning-based approach for shrimp disease detection. *Microbe* **2025**, *7*, 100299. [[CrossRef](#)]
106. Diéguez, A.L.; Balboa, S.; Magnesen, T.; Jacobsen, A.; Lema, A.; Romalde, J.L. Comparative study of the culturable microbiota present in two different rearing systems, flow-through system (FTS) and recirculation system (RAS), in a great scallop hatchery. *Aquac. Res.* **2020**, *51*, 542–556. [[CrossRef](#)]
107. Pollock, J.; Glendinning, L.; Wisedchanwet, T.; Watson, M. The madness of microbiome: Attempting to find consensus “best practice” for 16S microbiome studies. *Appl. Environ. Microbiol.* **2018**, *84*, e02627-17. [[CrossRef](#)]
108. Janakiev, T.; Dimkić, I.; Aleksić, J.; Grbić, M.L.; Knežević, A.; Kosel, J.; Tavzes, Č.; Unković, N. Beneficial bacteria-based bioformulations as potential biocontrol and biocleaning solutions for stone heritage conservation. *World J. Microbiol. Biotechnol.* **2025**, *41*, 200. [[CrossRef](#)] [[PubMed](#)]
109. Hug, L.A.; Hatzenpichler, R.; Moraru, C.; Soares, A.R.; Meyer, F.; Heyder, A.; Probst, A.J. A roadmap for equitable reuse of public microbiome data. *Nat. Microbiol.* **2025**, *10*, 2384–2395. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.